This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Hierarchical Dirichlet Process Based Gamma Mixture Modelling for Terahertz Band Wireless Communication Channels

ERHAN KARAKOCA[1,2], (Student Member, IEEE),
GÜNEŞ KARABULUT KURT[3],(Senior Member, IEEE)
ALİ GÖRÇİN[1,4], (Senior Member, IEEE)
[1]Informatics and Information Security Research Center (BİLGEM), TÜBİTAK, 41470 Kocaeli, Turkey
[2]Department of Electronics and Communications Engineering, İstanbul Technical University, 34469 İstanbul, Turkey
[3]Poly-Grames Research Center, Department of Electrical Engineering Polytechnique Montréal, 4551 Montréal, Canada
[4]Department of Electronics and Communications Engineering, Yıldız Technical University, 34220 İstanbul, Turkey

Corresponding author: Erhan Karakoca (erhankrkca@gmail.com).

ABSTRACT Due to the unique channel characteristics of Terahertz (THz), comprehensive propagation channel modeling is essential to understand the spectrum and develop reliable communication systems in these bands. In this work, we propose the utilization of the hierarchical Dirichlet Process Gamma Mixture Model (DPGMM) to characterize THz channels statistically in the absence of any prior knowledge. DPGMM provides mixture component parameters and the required number of components. A revised expectation-maximization (EM) algorithm is also proposed as a pre-step for DPGMM. Kullback-Leibler Divergence (KL-divergence) is utilized as an error metric to examine the amount of inaccuracy of the EM algorithm and DPGMM when modeling the experimental probability density functions (PDFs). DPGMM and EM algorithm are implemented over the measurements taken at frequencies between 240 GHz and 300 GHz. By comparing the results of the DPGMM and EM algorithms for the measurement datasets, we demonstrate how well the DPGMM fits the target distribution. It is shown that the proposed DPGMM can accurately describe the various THz channels as well as the EM algorithm, and its flexibility allows it to represent more complex distributions better than the EM algorithm. We also demonstrated that DPGMM can be used to model any wireless channel due to its versatility.

INDEX TERMS Terahertz communications, statistical channel modeling, expectation maximization, Dirichlet process, Gamma mixture model.

I. INTRODUCTION

THE demand for high-bandwidth instant online connectivity grows every day. That is prompting the emergence of new data-hungry technological instruments, resulting in an everlasting increase in wireless data traffic load [1], [2]. As a result, both academia and the industry increased their interest in higher frequency bands featuring wider bandwidths to comply with the demand. The Terahertz (THz) band between 0.1 THz and 10 THz is one of the spectra that has been examined from various perspectives recently [3], [4] and appears to have a promising future. Since the THz band is not standardized and allows bandwidths in the orders of 100 GHz, researchers push for the design of THz wireless systems that will enable communication [5]. Owing to its appealing features, the THz band is expected to play a key role in sixth-generation (6G) communication systems by providing data rates up to Terabit per second levels [6].

Besides these, THz band comes at the cost of severe losses such as propagation losses and molecular absorption. Because of these issues, THz band has distinct channel characteristics than other commonly utilized frequency bands. Even various THz sub-bands have different channel characteristics, demonstrating the importance of channel modeling in THz band and the requirement for a range of flexible channel modeling approaches.

A. RELATED WORKS

Several works employing various approaches have been conducted to achieve the goal of accurate channel modeling in THz bands. Generally, channel models are divided into two

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

types *i.e.*, deterministic and statistical.

In terms of deterministic channel modeling, ray-tracing is frequently utilized. In [7], the frequency domain ray-tracing approach is employed to estimate the THz indoor propagation channel based on the measurements conducted at 275 to 325 GHz. In [8], the smart rail mobility channel is characterized by ray-tracing at the 300 GHz band. Ray-tracing based on a stochastic channel model of high data rate data download at 220 GHz and 340 GHz is proposed in [9]. Deterministic modeling methods provide accurate channel modeling results; however, the detailed geometric structure of the propagation environment, as well as the transmitter and receiver locations must be known beforehand, which becomes a significant problem, especially in case of mobility. Furthermore, deterministic approaches suffer from high computational complexity.

On the other hand, statistical modeling approaches can be employed to estimate the THz propagation characteristics. A statistical path loss model for 240 to 300 GHz band is proposed in [10]. Furthermore, a two-slope statistical path loss model for short-range THz communication links between 275 and 325 GHz based on real measurement results is introduced in [11]. The suitability of the $\alpha-\mu$ distribution for measurements taken at the different transmitter and receiver configurations in a shopping mall, an airport check-in area, and a university entrance hall was investigated in [12] to model THz small scale fading accurately. In addition to these works, the ergodic capacity of THz wireless systems is evaluated in [13] by defining THz wireless channels with a $\alpha-\mu$ distribution. Statistical modeling techniques are able to characterize THz channels in different environments rather than assuming a specific place. Beyond this, the fundamental advantage of statistical channel modeling is the lower complexity when compared to deterministic models, which enables quick channel model generation based on essential characteristics [14].

Contemporary THz band statistical channel modeling approaches mostly assume that the channel can be characterized based on a single distributional representation. However, taking into account the variable channel characteristics of THz sub-bands, mixed models should be considered. In fact, although the use of mixture models in density estimation is common [15]–[19], it is rarely used in studies to model wireless communication channels. For example, [20] employed a mixture of Gamma distributions to model the signal-to-noise ratio (SNR) of various wireless channels by matching probability density functions (PDFs) and moment generation functions (MGFs), and it enables the evaluation of channel capacity, outage, and error rate owing to its mathematically tractable form and high accuracy representation. Also, in [20], the required number of mixture components is determined by checking values of mean squared error and Kullback-Leibler Divergence (KL-divergence) with the increasing number of mixture components. In [21], shadow fading of the empirical results collected in real scenario modeled with lognormal mixtures. In [22], the Gaussian mixture model is used to describe wireless channels, where the component parameters and the number of required components are determined by EM and Bayesian information criteria (BIC), respectively. When THz bands are investigated, [23] proposed Gamma mixture model (GMM) for 240 to 300 GHz band characteristics and represented the bands accurately. In this work, expectation-maximization (EM) algorithm is utilized to infer the parameters of Gamma distribution components.

B. MOTIVATION

GMM provided a significant step forward for THz channel characterization, however there are major improvement points in terms of applications since matching PDFs and MGFs requires special parametrization for every distribution. Moreover, the performance of the EM algorithm is highly dependent on initialization parameters [24], [25] thus, it has to be tuned every time with the required number of mixture components to be provided a priori. To that end, BIC can be utilized to determine the number of required mixture components. However, in such a scenario, BIC should scan all potential mixture component numbers to one with the minimum BIC. This additional step should be repeated for each different distribution. To overcome the impracticality of this process, there a more generalized method is strived for. Motivated by this requirement, in this work, the utilization of a Dirichlet Process Mixture Model (DPMM) for THz band small scale fading modeling is proposed.

DPMM is widely used in many fields involving nonparametric Bayesian estimation problems such as classification of brain tissues [26], protein fold [27], hyperspectral images [28] and clustering of genes [29], texts [30], medical images [31] and density estimation [32] of the relevant data in many fields. DPMM has the ability to grow its representation as more data are observed assuming that data come from distinct sub-clusters, with each sub-cluster's data described by a separate probability density function. Without any prior knowledge, it allows inferring the required number of clusters and statistical properties of each corresponding clusters as data observed. Because of these properties, DPMMs are accurate and adaptable models, especially when the underlying distribution of the data is either known, or can be closely approximated by the assumed model. As a result of this, DPMM has gained considerable popularity in the field of machine learning because of its flexibility, especially in unsupervised learning, due to its clustering and latent feature extraction property.

The one important thing that has to be considered is clusters in actual data do not always possess symmetric distribution like the Gaussian distribution. The Gamma distribution is a versatile alternative to the Gaussian distribution, and because of its skewness, it can describe both long-tailed and asymmetric distributions. As a result, any arbitrary PDF can be modeled by utilizing a Gamma mixture [32]. Because of this, Gamma kernels are preferred in DPMM studies such as [31]–[34], to model the density of the real

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

TABLE 1. Distinguishing features and contributions of the paper.

Reference	Subject	Approach	Mixtures	Prior knowledge	Band of interest
[20]	Approximation of SNR distribution for any fading channel	PDF, MGF and moment matching, Bayesian information criteria	Gamma	required	no bandwidth restriction
[21]	Modeling of shadow fading	EM and DPMM	Lognormal and Gaussian	not required	< 6GHz
[23]	Modeling of short distance THz channel	Maximum likelihood and EM	Gamma	required	200 - 300 GHz
Proposed	Accurate modeling of THz band channel and giving an approach capable of model any of wireless communication channel	Hierarchical DPMM	Gamma	not required	no bandwidth restriction

data and their belonging clusters accurately. As previously stated, due to its variety, also THz sub-bands do not always have the same form and rarely follow a smooth distribution. For these reasons, Gamma distribution is preferred as the mixture kernel in our proposed model, and we denote it as the Dirichlet Process Gamma Mixture Model (DPGMM). Unlike the previous mixture-based channel modeling works, distribution parameters and the number of components required for different channel types can be inferred using a single DPGMM without any further processing and prior knowledge. This allows a single DPGMM to be used to describe various distributions from simple to complex.

C. CONTRIBUTIONS

- Motivated by the requirements described, EM algorithm is revised in this work for GMM and as a follow up to the works of [20], [21], [23] we proposed a flexible hierarchical DPGMM for the THz band, which determines the required number of mixture components and corresponding mixture component parameters according to the variable structure of THz channels. We clarified our contributions on Table 1.
- To demonstrate the validity of the models, the proposed DPGMM and EM algorithm are applied to several measurements conducted between 240 and 300 GHz.
- The ability of the DPGMM and EM algorithm to estimate experimental distributions is assessed by the KL-divergence error metric, and results are illustrated. Furthermore, an extensive comparison of the DPGMM and EM algorithm for the GMM is carried out.
- The flexibility of DPGMM is validated, and its ability to span a broad range on the positive axis is proven by applying it to a very large measurement result. Therefore, it is shown that DPGMM is not only applicable to THz channels, but it can also be used to characterize a wide variety of wireless communications channels, including those with single distribution characteristics. Finally, the python code for the developed model is also made available[1] for further research activities in this domain.

[1]DPGMM source codes available to reproduce the results and further research activities: https://github.com/erhankarakoca/DPGMM-Channel-Modelling

D. ORGANIZATION OF THE PAPER

This paper is organized as follows. Section II describes the mathematical foundations of the signal model and the GMM, as well as the expectation maximization algorithm for the GMM. Section III firstly explains the Dirichlet process (DP) and its construction with the stick-breaking process. Then introduces the DPGMM by giving detailed information about its hierarchical structure. Section IV provides detailed information about the measurement dataset and simulation settings. Following that, the performance of the DPGMM and EM algorithms on the measurement dataset is shown, and both models are contrasted. Finally, Section V concludes the study.

II. PRELIMINARIES

A. SIGNAL MODEL

The signal collected at the receiver can be expressed in time domain as

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

with f_c being the carrier frequency of the transmitted signal, j representing imaginary number $\sqrt{-1}$. $x_I(t)$ and $x_Q(t)$ define the in-phase and quadrature parts of the received signal, respectively. $\mathfrak{Re}\{\cdot\}$ represents the real-valued terms of the complex baseband signal $r(t)$.

The impulse response of the multipath channel at the passband can be given as

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

where L is the total number of multipath sources and δ is the Dirac delta function. a_l and t_l denote attenuation and delay coefficients for the given l^{th} multipath source. The corresponding complex baseband representation of the multipath channel impulse response is

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

In the case of only the line-of-sight (LOS) component, $L = 1$ in Eq. (3) and LOS channel can be given as

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

where a_0 is amplitude and $2\pi f_c t$ phase of the channel. Propagation delay; $t_0 = d/c$, where d is the spacing between

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

transmitter and receiver and c is the speed of the light constant.

Measurement dataset [35] used in this study is obtained in an anechoic chamber that only allows LOS propagation. Even though this setting moves the measurements away from rich scattering scenarios, the loss factors become a factor that can be defined as misalignment between antennas, hardware impairments, and path loss. Therefore, the received signal representation can be simplified to a combination of distance dependent path loss and misalignment between antennas. The effect of these losses on channel amplitude a_0 can be expressed as

$$P_{rx} = P_{tx} + 10n\log_{10}(d) + M, \tag{5}$$

where P_{rx} is the received power calculated as the change in the power of the transmitted signal P_{tx} due to path loss. Path loss exponent is denoted as n and M is the random antenna gain due to the effect of misalignment between antennas. Eq. (5) expresses the received signal power in the LOS condition at the receiver that we are trying to model. Since the signals under consideration are wideband and the channel is varied at THz frequencies, P_{rx} also fluctuates with frequency. For this reason, in the next step, the GMM is explained to better capture and model the power clusters and differences in P_{rx}.

B. GAMMA MIXTURE MODEL

Let us define two parameter $\alpha > 0$ and $\beta > 0$. The Gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx. \tag{6}$$

If both sides of Eq. (6) are divided by $\Gamma(\alpha)$ and by changing of variables as $x = \beta y$ follows

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}dx = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}dy. \tag{7}$$

Then probability density function $f(x|\alpha,\beta)$ of Gamma distribution can be defined as

$$f(x \mid \alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x},\ x \geq 0,\ \alpha > 0,\ \beta > 0, \tag{8}$$

where parameters shape α and rate β for all positive values of x and it sums to one.

The finite GMM with K components can be written as

$$p\left(\mathbf{x}|\alpha_1,\beta_1,\pi_1,\ldots,\alpha_K,\beta_K,\pi_K\right) = \sum_{k=1}^{K}\pi_k\mathcal{G}\left(\alpha_k,\beta_k\right), \tag{9}$$

where $\mathbf{x} = \{x_1,\ldots,x_n\}$ is the positive vector of observations, π_k mixing proportions or weights of the k^{th} mixture component that sum to one $\sum_{k=1}^{K}\pi_k = 1$ and $\mathcal{G}$ denotes Gamma distribution defined in Eq. (8) with parameters α_k and β_k which are the shape and rate parameters of the k^{th} mixture component, respectively. The Gamma mixture was chosen for this study because it has traceable cumulative distribution function (CDF) and MGF, thus it can provide an approximation for small-scale fading channels [20]. Furthermore, by adjusting its parameters, a wide range of distributions can be represented with high accuracy and as showed in [32], arbitrary PDFs on $(0,\infty)$ can be modeled using Gamma mixture. As a result, $P_{rx} = \{x_1,\ldots,x_n\}$ can be modeled as a Gamma mixture, so we will need to find the appropriate number of Gamma clusters and their weights in addition to the parameters associated with each Gamma cluster. For ease of expression, the Gamma distribution parameters will be expressed together as $\theta_k = \{\alpha_k,\beta_k\}$ and throughout the remainder of this study, methods are described to find the Gamma mixture parameters in a way that models the empirical distribution of P_{rx} observation vector.

C. EXPECTATION MAXIMIZATION

When some data is absent or latent variables are present, an iterative process called EM which is commonly used for density estimation such as clustering for mixture models is used to obtain the maximum likelihood estimates of the parameters. In this context, EM algorithm is herein utilized to fit the empirical PDFs and find Gamma mixture parameters of the P_{rx} measurements.

When the implementation of the EM algorithm is considered, it is observed that the number of Gamma components to be used for modeling P_{rx} measurements is a priori information for its two-step process; expectation (E-step) and maximization (M-step) [36]. EM iterates until it reaches the desired convergence point, which may not be optimal [37], updating the randomly assigned values for the parameters of the mixture model $\theta_{1:M} = (\theta_1,\ldots,\theta_M)$ at each iteration. The membership coefficients for the entire measurement set $(i = 1,\ldots,N)$ and all mixture components $(k = 1,\ldots,M)$ are computed in the E-step using the current parameters $\theta_{1:M}$ as follows

$$\phi_{ik} = \frac{\pi_k p_k\left(x_i \mid \theta_k\right)}{\sum_{k=1}^{M}\pi_k p_k\left(x_i \mid \theta_k\right)}, \tag{10}$$

where $\sum_{k=1}^{M}\phi_{ik} = 1$. Then, using the measurement data and membership coefficients derived in the E-step, new parameter values α, β, and π in the mixture model can be obtained for each Gamma component by updating the equations in the M-step as follows [38]

$$\pi_k^{new} = \frac{\sum_{i=1}^{N}\phi_{ik}}{N}, \tag{11}$$

$$\mathbb{E}\left[X_k\right]^{\text{new}} = \frac{\sum_{i=1}^{N}\phi_{ik}x_i}{\sum_{i=1}^{N}\phi_{ik}} = \alpha\beta, \tag{12}$$

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

Using the mixture parameters identified by Eq. (11), Eq. (12) and Eq. (13) we can represent the P_{rx} observation vector in terms of a mixture of Gamma distribution.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

III. DIRICHLET PROCESS GAMMA MIXTURE MODEL

A. DIRICHLET PROCESS

Dirichlet distribution is convenient in describing random probability mass functions for finite categorical sets, and it can be thought as a generalization of Beta distribution for multivariate data sets. For K categorical event, Dirichlet distribution denoted as $Dir(a_1, \ldots, a_K)$ can be given as

$$f\left(x_1, \ldots, x_K; a_1, \ldots, a_K\right) = \frac{\Gamma\left(\sum_{i=1}^{K} a_i\right)}{\prod_{i=1}^{K} \Gamma\left(a_i\right)} \prod_{i=1}^{K} x_i^{a_i - 1}, \tag{14}$$

where x_i represents a specific category, $\sum_{i=1}^{K} x_i = 1$ and $x_i \geq 0$ for all $i \in \{1, \ldots, K\}$ also a_i denotes the intensity of the specified category.

The infinite dimensional extension of Dirichlet distribution is a stochastic process called as DP, whose realizations are probability distributions over some measurable set S. Therefore, each draws from the DP, itself a distribution. This process is denoted as $DP(a, H)$, where H is the expected value of the process called as base distribution and where a is positive real number again describing the intensity of mass around the mean called as concentration parameter. If G is a random variable drawn from $DP(a, H)$, then it can be shown that $G\left(A_1\right), \ldots, G\left(A_n\right) \sim Dir\left(aH\left(A_1\right), \ldots, aH\left(A_n\right)\right)$, where $\{A_i\}_{i=1}^{n}$ denotes any measurable finite partition of measurable set S [39].

The distinctive characteristic of the DP is that the required number of clusters is obtained from the process due to non-parametric nature which makes it an ideal candidate for clustering problems where the number of clusters in unknown a priori. The DP is commonly used in non-parametric Bayesian models as a prior on the distribution space, especially in DPMM which is known as infinite mixture models. In our analysis, we can use DP as a prior in distribution space to obtain the underlying distributions of the P_{rx} values and their corresponding parameters.

B. STICK-BREAKING CONSTRUCTION OF DIRICHLET PROCESS

DP can be built in a variety of methods such as the Blackwell–MacQueen urn scheme, the Chinese restaurant process, or the stick-breaking construction, and each of them emphasizes a distinct aspect of the DP. In this study, we employed the stick-breaking construction, which allows expressing a sample from a DP ; $G \sim DP(a, H)$ as [40]

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}(\cdot), \quad \theta_k \sim H, \tag{15}$$

where θ_k represents atoms drawn independently from the base distribution H, δ_θ point mass at θ and π_k is the probability mass at atom θ_k. The probability masses also known as weights and can be constructed as follows

$$\pi_k = V_k \prod_{j=1}^{k-1} \left(1 - V_j\right), \quad V_k \sim \mathcal{B}(1, a), \tag{16}$$

where V_i denotes a broken piece, a is the concentration parameter and $\mathcal{B}$ denotes Beta distribution. This entire process may be illustrated by first breaking a unit-length stick randomly drawn by Beta distribution as, V_1 and then continuing to break the remaining portion of the stick $1 - V_i$ randomly drawn by Beta distribution as $V_2 \ldots V_k$. The weights are indicated by the length of each broken piece as $\pi_1 = V_1, \pi_2 = V_2(1 - V_1), \ldots, \pi_k = V_k(1 - V_{k-1})$. Then atoms θ_k are drawn from the base distribution H to associate these weights. Upon breaking the stick, it becomes shorter, and the length of the higher indexed atoms decreases stochastically, whose rate of decrease depends on the DP concentration parameter a because it plays a key role in determining the weights in each iteration. This procedure assures that $\sum_{k=1}^{\infty} \pi_k = 1$. Note that for G to be true from a DP, an infinite number of weights and atoms must be drawn. However, in practice, it is possible to truncate summation on Eq. (15) with only a finite number of K draws while still providing a very good approximation. In Section IV-D, we will go through how to choose the truncation number.

C. DIRICHLET PROCESS GAMMA MIXTURE MODEL

Stick breaking construction of the DP shows that samples from the process are discrete distribution. Thus, actually, DP is not a proper prior for continuous distributions. Therefore, in non-parametric density estimation DP is used indirectly by supporting kernel function $\mathcal{K}(\cdot)$ as described by

$$f(\mathbf{x}) = \int \mathcal{K}\left(x \mid \theta\right) G(\theta) \mathrm{d}\theta. \tag{17}$$

Choosing a DP prior on G as Eq. (15) and using it in Eq. (17) can be turned into a sum of the infinite mixture of kernels [41]

$$f(\mathbf{x}) = \sum_{k=1}^{\infty} \pi_k \mathcal{K}\left(x \mid \theta_k\right), \tag{18}$$

which can be denoted as DPMM. In Eq. (17) the summation is set to infinity. This allows the model to describe new mixture components that may occur as new classes are added corresponding to the sampled atoms. However, it doesn't imply that infinitely many components are occupied.

Instead of simply sampling the data from the DP, sampling the mixture parameters θ_i from the DP and then utilizing these values as input in the continuous kernel function $\mathcal{K}(\cdot)$ enables the construction of DPMM for non-parametric density estimation as hierarchically defined below

$$\begin{aligned} x_i &\sim \mathcal{K}\left(\theta_k\right), \\ \theta_k &\sim G, \\ G &\sim \mathrm{DP}\left(a, H\right). \end{aligned} \tag{19}$$

Thus, DPGMM can be defined by setting the kernel function $\mathcal{K}(\cdot)$ as Gamma distributed

$$\mathcal{K}(x \mid \theta_k) \equiv \mathcal{G}(x \mid \alpha_k, \beta_k), \tag{20}$$

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

then it gets into the form of

$$f(\mathbf{x}) = \sum_{k=1}^{\infty} \pi_k \mathcal{G}\left(x \mid \alpha_k, \beta_k\right). \tag{21}$$

In order to find Gamma mixture parameters θ_k, base distribution H and its hyperparameters needs to be defined for DPGMM. To that end a modified hierarchical model out of [31] is utilized to estimate the mixture component parameters with fundamental modifications in terms of construction of DP, defining priors and hyperpriors, and computing algorithm.

Assuming that α and β follow prior distributions; $p(\alpha)$ and $p(\beta)$, respectively depending on hyperparameters; $\lambda, \kappa, \nu, \upsilon$. In order to add more flexibility to the model, we assume that also these hyperparameters follow some distributions; $p(\lambda), p(\kappa), p(\nu)$ and $p(\upsilon)$ depending on hyper-priors; $\vartheta, \varpi, \varsigma, \varepsilon, \mu, \vartheta, \varrho$.

The prior distributions for the mixture component parameters α and β are all assumed independent of each other. The Inverse-Gamma distribution ($\mathcal{IG}$) is used as a prior for the shape parameter α, and hyperparameters with shape parameter λ and ratio parameter κ are utilized for this prior

$$p\left(\alpha_k \mid \lambda, \kappa\right) \sim \frac{\kappa^{\lambda} \alpha_k^{-\lambda-1} \mathrm{e}^{-\kappa/\alpha_k}}{\Gamma(\lambda)}. \tag{22}$$

An $\mathcal{IG}$ prior with hyperpriors ϑ and ϖ are employed for λ and an exponential ($\mathcal{E}xp$) prior with hyperprior ς is used for κ

$$p(\lambda \mid \vartheta, \varpi) \sim \frac{\varpi^{\vartheta} \mathrm{e}^{-\varpi/\lambda}}{\Gamma(\vartheta)\lambda^{\vartheta+1}}, \tag{23}$$

$$p(\kappa \mid \varsigma) \sim \zeta \mathrm{e}^{-\varsigma\kappa}. \tag{24}$$

With these above specifications for the hyperparameter α, its posterior can cover a very large range in positive axis with given vague hyperpriors on ϑ, ϖ, and ς.

Also, Gamma distribution is used as a prior for the rate parameter β, and hyperparameters with shape parameter ν and ratio parameter υ are utilized for this prior

$$p\left(\beta_k \mid \nu, \upsilon\right) \sim \frac{\upsilon^{\nu}}{\Gamma(\nu)} \beta_k^{\nu-1} \mathrm{e}^{-\upsilon\beta_k}. \tag{25}$$

An $\mathcal{IG}$ prior with hyperpriors ϵ and μ is used for υ and a $\mathcal{G}$ prior with hyperpriors φ and ϱ is used for ν

$$p(\nu \mid \epsilon, \mu) \sim \frac{\mu^{\epsilon} \mathrm{e}^{-\mu/\nu}}{\Gamma(\epsilon)\nu^{\epsilon+1}}, \tag{26}$$

$$p(\upsilon \mid \varphi, \varrho) \sim \frac{\upsilon^{\varphi-1} \varrho^{\varphi} \mathrm{e}^{-\varrho\upsilon}}{\Gamma(\varphi)}. \tag{27}$$

Based on these specifications for the hyperparameter β, its posterior can cover especially small values between 0 and 1 with given vague hyperpriors on ϵ, μ and φ, ϱ, which are specified in Section IV-B.

In addition to base distribution also concentration parameter a which also governs the distribution over the number of components K, has to be defined for the DPGMM. In [31], Inverse-Gamma distribution was chosen as a prior to a. However, utilization of $\mathcal{IG}$ leads to the inclusion of an excessive number of mixture components in the estimation process thus estimation flexibility becomes limited. Thus, in this study, conjugacy of the Gamma distribution to the $\mathcal{B}(1, a)$ is exploited to provide flexibility to the cluster sizes and mixture weights. So, $\mathcal{G}(1, 1)$ is used as a prior for concentration parameter a, therefore it became possible to represent distributions using fewer components while maintaining estimation accuracy. Please note that the proposed model's hierarchical structure allows flexibility and adaptability to a wide range of data with various distributional characteristics. Furthermore, finding robust parameters for models may be challenging, and misspecified parameters may diminish model performance; consequently, using a hierarchical model alleviates this difficulty while also providing flexibility and robustness to the model framework [42].

The structure of our model summarized in Fig. 1 and DPGMM can be described all together hierarchically using the previously mentioned structure of the weights, base distribution, and the concentration parameter as described below:

$$a \sim \mathcal{G}(1, 1), \tag{28a}$$

$$V_1, \ldots, V_K \sim \mathcal{B}(1, a), \tag{28b}$$

$$\pi_k = V_k \prod_{j=1}^{k-1} \left(1 - V_j\right), \tag{28c}$$

$$\lambda_k \sim \mathcal{IG}(\vartheta, \varpi), \tag{28d}$$

$$\kappa_k \sim \mathcal{E}xp(\varsigma), \tag{28e}$$

$$\nu_k \sim \mathcal{G}(\epsilon, \mu), \tag{28f}$$

$$\upsilon_k \sim \mathcal{IG}(\varphi, \varrho), \tag{28g}$$

$$\alpha_k \sim \mathcal{IG}(\lambda_k, \kappa_k \mid \vartheta, \varpi, \varsigma) \tag{28h}$$

$$\beta_k \sim \mathcal{G}(\nu_k, \upsilon_k \mid \epsilon, \mu, \varphi, \varrho), \tag{28i}$$

$$x_i \sim \sum_{k=1}^{K} \pi_k \mathcal{G}\left(x \mid \alpha_k, \beta_k\right). \tag{28j}$$

Since our prior assumptions for mixture parameters independent of each other, a draw from the base distribution $(\alpha_k, \beta_k) \sim H(\theta_k)$ can be defined as

$$\alpha_k, \beta_k \sim \mathcal{IG}(\lambda_k, \kappa_k \mid \vartheta, \varpi, \varsigma)\mathcal{G}(\nu_k, \upsilon_k \mid \epsilon, \mu, \varphi, \varrho). \tag{29}$$

We aim to find posterior distribution for the DPGMM. A mixture model which gives the best evidence of assigning each data point $\mathbf{x} = (x_1, \ldots, x_i)$ to only one cluster, can be found by selecting the model that maximizes the integrated likelihood, which is [43]

$$\mathbf{f}(\mathbf{x}, \mathbf{z} \mid \psi) = \int_{\Theta} \mathbf{f}(\mathbf{x}, \mathbf{z} \mid \theta, \psi) p(\theta \mid \psi) d\theta, \tag{30}$$

with

$$\mathbf{f}(\mathbf{x}, \mathbf{z} \mid \theta, \psi) = \prod_{i=1}^{n} f\left(x_i, z_i \mid \theta, \psi\right), \tag{31}$$

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

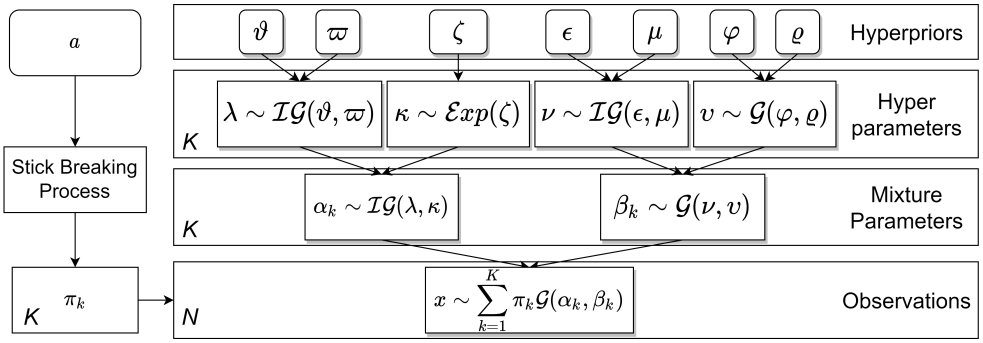


FIGURE 1. Hierarchical Dirichlet process Gamma mixture model structure.

and

$$f\left(x_i, z_i \mid \theta, \psi\right) = \prod_{k=1}^{K} \left(\pi_k p\left(x_i \mid \theta_k, \psi_k\right)\right)^{\mathbb{I}(z_i=k)}, \quad (32)$$

where $\Theta = \{\theta_{1:K}, \pi_{1:K}, \psi_{1:K}\}$ denotes parameter space of the model, $\theta_k = \{\alpha_k, \beta_k\}$ and $\psi_k = \{\lambda_k, \kappa_k, \mu_k, \nu_k\}$ denotes hyperparameters. In addition, z_i is an indicator variable that describes the assignment of the observation x_i to a specific mixture component between 1 and K. It is also known as a latent variable or missing allocation variable introduced by the nature of DP clustering property and its incidence represents formation of a new component on the mixture, which is controlled by a. The indicator function for z_i is denoted by $\mathbb{I}$. By following Eq. (31) and Eq. (32), the likelihood function for the model can be given as

$$p(\mathbf{x}, \mathbf{z} \mid \Theta) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\pi_k \, p\left(x_i \mid \theta_k, \psi_k\right)\right)^{\mathbb{I}(z_i=k)}, \quad (33)$$

Hence by using the Bayes rule, the posterior probability is equal to the likelihood times the prior divided by the evidence

$$p(\Theta \mid \mathbf{x}, \mathbf{z}) = \frac{p(\Theta)p(\mathbf{x}, \mathbf{z} \mid \Theta)}{\int p(\Theta)p(\mathbf{x}, \mathbf{z} \mid \Theta)d\Theta} \propto p(\Theta)p(\mathbf{x}, \mathbf{z} \mid \Theta). \quad (34)$$

Also, posterior density proportional to joint probability, $p(\Theta \mid \mathbf{x}, \mathbf{z}) \propto p(\mathbf{x}, \mathbf{z}, \Theta)$ where the marginal likelihood in the denominator in Eq. (34), can be disregarded as it does not depend on the model parameters. In a sense, determining the joint probability function provides information on the posterior density. We can define the joint probability of the model taking into account hyperparameters and hyperpriors as [44]

$$\begin{aligned} &p\left(\mathbf{x}, z_{1:K}, \theta_{1:K}, \pi_{1:K}, \psi_{1:K}\right) = \\ &\prod_{i=1}^{n} \prod_{k=1}^{K} \left[\pi_k \, p\left(x_i \mid \theta_k, \psi_k\right)\right]^{\mathbb{I}(z_i=k)} \times \\ &\prod_{k=1}^{K} p\left(\alpha_k, \beta_k \mid \psi_k\right) \prod_{k=1}^{K-1} \mathcal{B}\left(1, a\right). \end{aligned} \quad (35)$$

When dealing with DPGMMs, the posterior distributions and posterior of the model parameters are usually analytically intractable when a non-conjugate base distribution is chosen for the mixture kernel due to Eq. (34) involves integration. However, by having $p(\mathbf{x}, \mathbf{z}, \Theta)$ as Eq. (35) we can simulate our posterior distribution and model parameters Θ, rather than computing them. Hence, the inference is done through computational Markov Chain Monte Carlo (MCMC) simulations. To generate samples from DPGMM posterior distributions, several MCMC simulation algorithms based on Gibbs sampling, Metropolis-Hastings and Hamiltonian Monte Carlo (HMC) have been developed, which can subsequently be utilized for inference of all parameters and posterior distribution.

Furthermore, the No-U-Turn Sampler (NUTS) [45] a variant of the HMC that is a newer and more efficient algorithm than the others, can also be used for posterior sampling. By leveraging first-order gradient information, NUTS and HMC provide faster convergence than other sampling methods, especially in complex and high-dimensional datasets. However, unlike HMC, NUTS contains many self-tuning procedures for adaptively adjusting the tunable parameters of HMC like step size and desired number of steps. We also should mention that, the NUTS algorithm is available in the PyMC3 [46] which is a python based probabilistic programming package, and it allows fitting Bayesian models with a variety of MCMC simulation algorithms. Moreover, it provides extra advantages in gradient computation. For these particular and distinctive reasons, the PyMC3 package with the NUTS algorithm is applied to posterior sampling in this study.

IV. GAMMA MIXTURE MODEL FOR TERAHERTZ BAND WIRELESS CHANNELS

In this section, first the measurement setup and the datasets will be briefly described in Section IV-A. After explaining pre-processing of the dataset and post-processing settings of the DPGMM and EM algorithm in Section IV-B, we are presented the error metric KL-divergence in Section IV-C to quantify performance with respect to the experimental PDFs with the proposed approaches driven by the simulation settings. Finally, in Section IV-D, we will discuss and compare the results of the DPGMM and EM algorithm.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

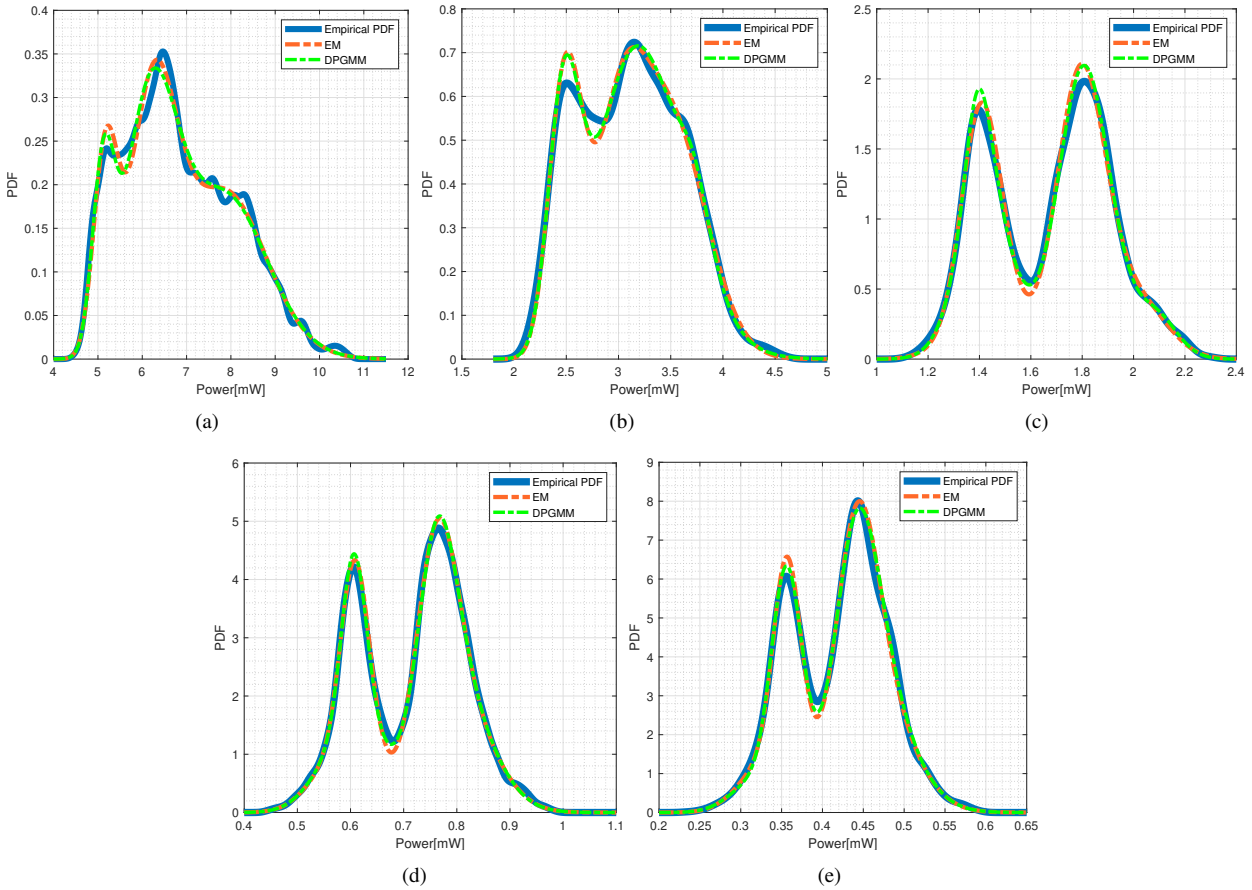


FIGURE 2. **DPGMM and EM Gamma mixture model for empirical PDFs (a) 20 cm, (b) 30 cm, (c) 40 cm, (d) 60 cm, (e) 80 cm.**

A. MEASUREMENT DATASET

The measurements utilized in this paper were obtained at the Turkish Science Foundation's (TÜBİTAK) anechoic chamber [47], which are also available at [35]. The data consist of complex S_{21} parameters taken at five different points away from the transmitter *i.e.*, at 20 cm, 30 cm, 40 cm, 60 cm, and 80 cm respectively. A laser-based system is used for measurements to precisely align the transmitter and receiver, ensuring LOS condition and measurement reliability. Measurements cover the 60 GHz band between 240-300 GHz with 4096 points of S_{21} measurements. The spectrum resolution becomes 14.468 MHz with this configuration. Also, IF bandwidth is set to 100 Hz for the measurements, which allows for an increase in the observed dynamic range and a decrease in the noise floor. Detailed information about measurement setup can be found in [23].

One important thing to mention, due to the short wavelengths of THz frequencies, smaller antennas will be utilized, which will also result in an expansion of communication applications in next-generation communication systems. One such application area is the microscale communications [9]. Microscale communication includes high-speed wireless connections between personal mobile terminals, PC/kiosk/cloud servers, and wireless nanosensor communications, which link tiny devices across considerably shorter distances [48]. The measurements we employed in this study are consistent with microscale THz communication scenarios.

B. DATA PROCESSING AND SIMULATION SETTINGS

First, by using S_{21} parameters obtained from measurements, the received power P_{rx} is calculated by

$$P_{rx} = |S_{21}|^2 P_{tx}, \tag{36}$$

where P_{tx} power of the transmitted signal and $|S_{21}|$ is amplitude response of transmission channel. Second, the proposed models are used to estimate the underlying distributions of empirical measurement PDFs, which are constructed from the histogram of P_{rx} values. In EM, mixture component number must be given a priori. However, because the required number of mixture components is unknown beforehand, we used the number of mixture components obtained from DPGMM in EM as input to perform the EM algorithm. Note that also, instead of starting from random points to run the EM algorithm, the first estimate is obtained via k-means algorithm, which is then used by the EM algorithm to run its iterations [38]. Then EM Gamma mixture parameters can be found by using Eqs. (11), (12), and (13) and with DPGMM, parameters can be inferred using the hierarchical structure of Eqs. (28) by sampling from posterior by NUTS.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

TABLE 2. Mixture parameters and error metrics for PDF estimations at distinct distances.

Distance	K	Method	Mixture Parameters			KL-Divergence
			π	α	β	
20 cm	3	EM	0.41942	141.51136	0.04461	0.03343
			0.41571	82.46476	0.09704	
			0.16485	345.43516	0.01502	
		DPGMM	0.47049	108.87149	0.05768	0.03603
			0.39966	83.45134	0.09634	
			0.12972	424.98806	0.01208	
30 cm	3	EM	0.39304	140.15848	0.02194	0.00581
			0.34110	145.96878	0.02469	
			0.26585	238.59034	0.01050	
		DPGMM	0.72306	63.55631	0.05088	0.00378
			0.21586	270.00446	0.00921	
			0.06098	657.81747	0.00568	
40 cm	4	EM	0.49563	353.30886	0.00510	0.00928
			0.28849	336.61776	0.00419	
			0.11952	105.15786	0.01376	
			0.09634	370.67005	0.00549	
		DPGMM	0.52684	323.77472	0.00560	0.00860
			0.28225	159.00929	0.00901	
			0.12817	894.49836	0.00155	
			0.06264	658.57817	0.00317	
60 cm	4	EM	0.43426	366.50425	0.00208	0.07901
			0.21705	522.76836	0.00116	
			0.17665	218.71135	0.00380	
			0.17202	93.00082	0.00648	
		DPGMM	0.31911	480.53104	0.00159	0.07699
			0.27953	194.58752	0.00419	
			0.21470	91.27561	0.00668	
			0.18655	671.84839	0.00090	
80 cm	4	EM	0.42664	287.17883	0.00154	0.12921
			0.21421	454.25301	0.00078	
			0.20415	162.81782	0.00295	
			0.15497	73.67085	0.00474	
		DPGMM	0.32959	130.13821	0.00358	0.12347
			0.32336	287.00556	0.00154	
			0.25716	368.78112	0.00096	
			0.08985	79.30695	0.00421	

Also, vague values given to the hyperpriors for DPGMM as $\vartheta = 1, \varpi = 1, \varsigma = 0.001, \epsilon = 1, \mu = 1, \varphi = 1, \varrho = 1$.

C. ERROR METRIC

In this paper, KL-divergence is performed to compare the experimental distributions with predicted distributions using DPGMM and EM. The KL-divergence is a metric that is used to assess the difference between two probability distributions over the same probability space $\mathcal{X}$. Let $Q(x)$ be the distribution whose distance from the reference distribution $P(x)$ is to be measured, then KL-divergence can be given as

$$\mathcal{D}_{KL}(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right), \qquad (37)$$

where P_x represent experimental distribution and Q_x represent predicted distributions via EM algorithm and DPGMM.

D. RESULTS AND COMPARISON

The results of the EM algorithm and DPGMM obtained by using the P_{rx} histograms are presented in this section. In order to plot measurements and compare results, we transformed P_{rx} histograms to the empirical PDFs.

In Fig. 2 we showed estimated PDFs utilizing DPGMM and EM algorithm to the P_{rx} values for five distinct distances. Different histograms are used to illustrate the flexibility and accuracy of the proposed model, with a variety of distributions with different forms. As can be seen from Fig. 2, the DPGMM and EM algorithm can describe all empirical PDFs very well. Although the hyperprior values are not reliant on the measurement data, they do incorporate the distributions acquired from the sub-THz band observations. It can also cover a large distribution space specified on the positive axis.

Table 2 contains estimated parameter values of mixture components and KL-divergence values for the DPGMM and

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

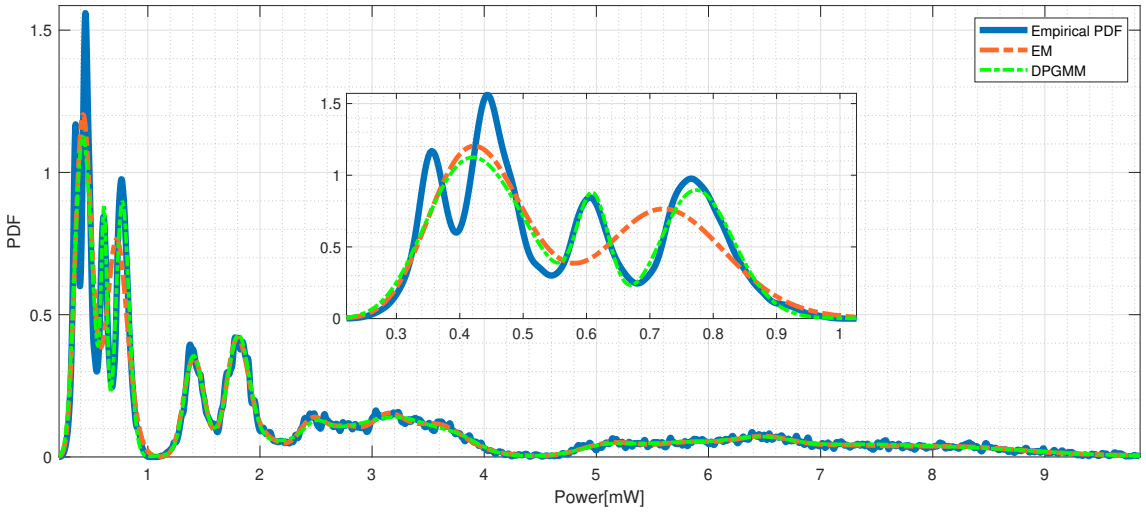


FIGURE 3. **Empirical PDF for combination of all measurement data and Gamma mixture models.**

the EM algorithm. As shown in Table 2, although there is a difference in the KL-divergence values for DPGMM and EM algorithm, it is not significant because the values are small and very close to each other. Furthermore, as seen in Fig. 1, we may neglect these discrepancies because the DPGMM and EM algorithm both match the empirical PDFs fairly well. The important thing to be considered here is the ability of the DPGMM to fit the empirical PDFs as much as the EM algorithm, even if the required number of mixture components is not known a priori.

Furthermore, we combined the P_{rx} values of all measurements collected at different distances and modeled the obtained PDF using the proposed methods as shown in Fig. 3. In our case, the performance of the EM algorithm decreases when the data size and data dimensionality increase, as given in Table 3. EM algorithm was not able to model $0.5 - 0.7$ mW P_{rx} region accurately, as can be seen in Fig. 3. One reason is that the EM algorithm converges to the local maxima rather than the global maxima. Also, the EM algorithm does not always guarantee convergence to the local maxima, it only guarantees convergence to a point with zero gradients according to the parameters and depending on the initialization step [25]. As a result, the EM algorithm might occasionally become stuck at the saddle points [36], [37]. Besides, DPGMM is able to model the combination of all P_{rx} values very well.

TABLE 3. **Error metrics for all measurement data.**

Distance	*K*	Method	KL-Divergence
All	17	EM	0.07697
		DPGMM	0.03759

In addition to these, we must address the truncation number K. Normally DP consists of an infinite number of distributions however, for computational convenience, K should be chosen in a way that can accurately describe the DP. It has to be large enough to represent the empirical PDF of all measurement data while being cost-effective for posterior sampling. For our DPGMM, we choose $K = 30$, which is sufficient for all measurement histograms. This does not imply that the clusters occupy K components in the data samples; instead, the model allows flexibility by introducing new mixture components up to 30, if necessary, as samples are added. As an illustration, for the empirical PDF of all P_{rx} data in Fig. 3, it is sufficient to use a maximum of 17 mixture components with DPGMM as shown in Fig. 4.

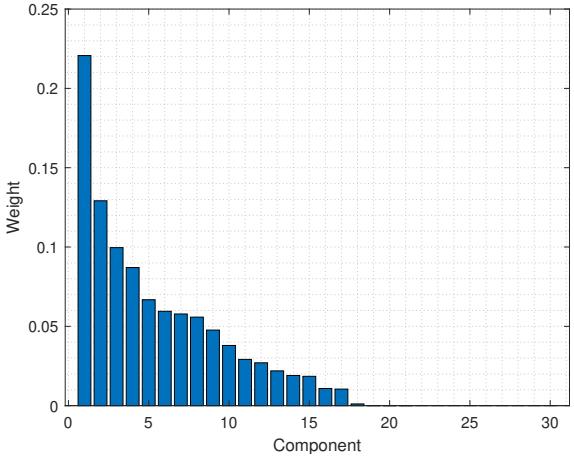


FIGURE 4. **Number of required components inferred with DPGMM for empirical PDF of all measurement data combination.**

Finally, one additional mixture component is added to the number of mixture components extracted using DPGMM for each distance, and the EM algorithm is applied to the P_{rx} measurements with this mixture component number. The KL-divergence values for EM obtained with these settings are given in Table 4, along with the KL-divergence values of DPGMM obtained previously. Although the number of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

components increased by one, the KL-divergence values of the EM algorithm vary very little, as shown in Table 4. This demonstrates that the number of mixture components provided by DPGMM is sufficient for modeling empirical PDFs, and adding more components will not significantly improve overall performance.

TABLE 4. The comparison of error metrics of the DPGMM and EM algorithm which is initialized with one additional number of mixture component.

Distance	K	Method	KL-Divergence
20 cm	4	EM	0.03298
	3	DPGMM	0.03603
30 cm	4	EM	0.00402
	3	DPGMM	0.00581
40 cm	5	EM	0.00937
	4	DPGMM	0.00860
60 cm	5	EM	0.07843
	4	DPGMM	0.07699
80 cm	5	EM	0.12915
	4	DPGMM	0.12347

E. PROCESSING TIMES

Since DPGMM uses a MCMC based sampling algorithm, as a result, we use this algorithm for a situation where we have no prior knowledge, so we start solving the problem by considering certain points as the starting point. This initial state may be close or far from the result, so we move posterior by taking each sample and updating our estimates according to the Bayesian probability with the observations. The processing time of MCMC algorithms increases as the model parameters increase, this is firstly because of the dimensionality, where the volume of the sample space increases with the number of parameters [49]. As a result, the MCMC algorithm is computationally heavier than the EM algorithm [50]. The processing times are given in the Table 5. Note that, the dataset contains all measurements consist of 10240 data points others are 2048 points.

TABLE 5. Processing times of EM algorithm and DPGMM (with MCMC) for different dataset

	20 cm	30 cm	40 cm	60 cm	80 cm	All
EM	8.98s	9.45s	9.31s	6.60s	6.13s	22.97s
DPGMM	32m25.4s	25m15.4s	21m29.4s	9m13.4s	22m25.4s	179m53.4s

However, it should also be taken into account that, in contrast to MCMC, for the EM algorithm we give the starting points by using the k-means algorithm instead of giving the starting points randomly so that it can converge to the target distribution faster with the number of mixture components that should be given as a priori in the EM algorithm. This allows the EM algorithm to run faster and guarantees to convergence the result. Otherwise, the EM algorithm cannot converge, or the density estimation performance is insufficient.

To minimize calculation time, the variational inference approach can be applied to the same model. Variational inference delivers solutions with equivalent accuracy to MCMC sampling at a faster rate in many situations [51].

V. CONCLUDING REMARKS AND FUTURE WORKS

In this work, we proposed utilizing a flexible hierarchical DPGMM for the sub-THz wide-band channel modeling. The main reasons that Gamma distribution is chosen as a kernel for the mixture are because of its flexibility, skewness, ability the model tailed distributions and moreover its CDF, MGF and moments are tractable. The proposed model is applied to the measurement dataset, which is between 240 GHz and 300 GHz. Also EM algorithm is utilized for the same dataset in order to compare the parameter estimation performance of DPGMM and EM algorithm. In the EM algorithm the required number of mixture components must be known a priori, however the proposed DPGMM is able to infer the required number of mixture components and corresponding component parameters. The simulation results reveal that the proposed DPGMM can accurately describe the various sub-THz channels that occur as a result of different scenarios, as much as the EM algorithm. However, for the higher dimensional histogram, DPGMM is able to describe underneath distributions better than the EM algorithm. Due to the given hyperprior specifications and hierarchical structure of the model, it is shown that DPGMM can be used for any type of wireless channel, not just for the THz band.

For Gamma mixture wireless channels, average channel capacity, the outage probability, and the symbol error rate were derived earlier. By making use of this provided knowledge, analytical analyzes can be made as future work using the mixture parameters given in this study. In addition to future works, measurements taken in different scenarios such as misalignment fading can also be modeled with given DPGMM. We believe that this work paves the way for modeling and performance analysis of not only the THz band also any kind of wireless communication channel in 5G and 6G networks.

ACKNOWLEDGMENTS

The authors would like to thank Ali Rıza Ekti and Kürşat Tekbıyık for their assistance.

REFERENCES

[1] S. Cherry, "Edholm's law of bandwidth," *IEEE Spectrum*, vol. 41, no. 7, pp. 58–60, 2004.
[2] Cisco, "Cisco annual internet report (2018–2023) white paper," *Cisco: San Jose, CA, USA*, 2020.
[3] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, 2014.
[4] K. Tekbıyık, A. R. Ekti, G. Karabulut Kurt, and A. Görçin, "Terahertz band communication systems: Challenges, novelties and standardization efforts," *Physical Communication*, vol. 35, p. 100700, 2019.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603

[5] K. Tekbiyik, A. R. Ekti, G. Karabulut Kurt, A. Gorcin, and H. Yanikomeroglu, "A holistic investigation of terahertz propagation and channel modeling toward vertical heterogeneous networks," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 14–20, 2020.

[6] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Network*, vol. 33, no. 4, pp. 70–75, 2019.

[7] S. Priebe, M. Kannicht, M. Jacob, and T. Kürner, "Ultra broadband indoor channel measurements and calibrated ray tracing propagation modeling at THz frequencies," *Journal of Communications and Networks*, vol. 15, no. 6, pp. 547–558, 2013.

[8] K. Guan, D. He, B. Ai, Y. Chen, C. Han, B. Peng, Z. Zhong, and T. Kuerner, "Channel characterization and capacity analysis for THz communication enabled smart rail mobility," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4065–4080, 2021.

[9] 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 Transactions on Terahertz Science and Technology*, vol. 7, no. 5, pp. 502–513, 2017.

[10] A. R. Ekti, A. Boyaci, A. Alparslan, İ. Ünal, S. Yarkan, A. Görçin, H. Arslan, and M. Uysal, "Statistical modeling of propagation channels for terahertz band," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, 2017, pp. 275–280.

[11] K. Tekbıyık, E. Ulusoy, A. R. Ekti, S. Yarkan, T. Baykaş, A. Görçin, and G. Karabulut Kurt, "Statistical channel modeling for short range line–of–sight terahertz communication," in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–5.

[12] E. N. Papasotiriou, A.-A. A. Boulogeorgos, K. Haneda, M. F. de Guzman, and A. Alexiou, "An experimentally validated fading model for THz wireless systems," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[13] E. N. Papasotiriou, A.-A. A. Boulogeorgos, M. F. De Guzman, K. Haneda, and A. Alexiou, "A new look to THz wireless links: Fading modeling and capacity assessment," in *IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021, pp. 1–5.

[14] C. Han and Y. Chen, "Propagation modeling for wireless communications in the terahertz band," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 96–101, 2018.

[15] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 577–588, 1995.

[16] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual Review of Statistics and its Application*, vol. 6, pp. 355–378, 2019.

[17] C. Rasmussen, "The infinite Gaussian mixture model," *Advances in Neural Information Processing Systems*, vol. 12, 1999.

[18] K. Roeder and L. Wasserman, "Practical Bayesian density estimation using mixtures of normals," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 894–902, 1997.

[19] T. S. Ferguson, "Bayesian density estimation by mixtures of normal distributions," in *Recent Advances in Statistics*. Elsevier, 1983, pp. 287–302.

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

[21] S. Büyükçorak, M. Vural, and G. Karabulut Kurt, "Lognormal mixture shadowing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4386–4398, 2014.

[22] 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 Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8309–8321, 2015.

[23] K. Tekbıyık, A. R. Ekti, G. Karabulut Kurt, A. Görçin, and S. Yarkan, "Modeling and analysis of short distance sub-terahertz communication channel via mixture of Gamma distribution," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 2945–2954, 2021.

[24] M. Meila and D. Heckerman, "An experimental comparison of several clustering and initialization methods," *arXiv preprint arXiv:1301.7401*, 2013.

[25] D. Karlis and E. Xekalaki, "Choosing initial values for the EM algorithm for finite mixtures," *Computational Statistics & Data Analysis*, vol. 41, no. 3-4, pp. 577–590, 2003.

[26] A. R. F. da Silva, "A Dirichlet process mixture model for brain mri tissue classification," *Medical image analysis*, vol. 11, no. 2, pp. 169–182, 2007.

[27] B. Shahbaba and R. Neal, "Nonlinear models using Dirichlet process mixtures." *Journal of Machine Learning Research*, vol. 10, no. 8, 2009.

[28] H. Wu and S. Prasad, "Dirichlet process based active learning and discovery of unknown classes for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4882–4895, 2016.

[29] D. B. Dahl, "Model-based clustering for expression data via a Dirichlet process mixture model," *Bayesian inference for gene expression and proteomics*, vol. 4, pp. 201–218, 2006.

[30] J. Yin and J. Wang, "A Dirichlet multinomial mixture model-based approach for short text clustering," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 233–242.

[31] T. Elguebaly and N. Bouguila, "A hierarchical nonparametric Bayesian approach for medical images and gene expressions classification," *Soft Computing*, vol. 19, no. 1, pp. 189–204, 2015.

[32] 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.

[33] K. Copsey and A. Webb, "Bayesian Gamma mixture model approach to radar target recognition," *IEEE Transactions on Aerospace and Electronic systems*, vol. 39, no. 4, pp. 1201–1217, 2003.

[34] T. E. Hanson, "Modeling censored lifetime data using a mixture of Gammas baseline," *Bayesian Analysis*, vol. 1, no. 3, pp. 575–594, 2006.

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

[36] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley & Sons, 2007, vol. 382.

[37] S. K. Ng, T. Krishnan, and G. J. McLachlan, "The em algorithm," in *Handbook of computational statistics*. Springer, 2012, pp. 139–172.

[38] G. Vegas-Sanchez-Ferrero, J. Seabra, O. Rodriguez-Leor, A. Serrano-Vida, S. Aja-Fernandez, C. Palencia, M. Martin-Fernandez, and J. Sanches, "Gamma mixture classifier for plaque detection in intravascular ultrasonic images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 1, pp. 44–61, 2014.

[39] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, pp. 209–230, 1973.

[40] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica sinica*, pp. 639–650, 1994.

[41] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. CRC Press, 2013.

[42] D. Görür and C. E. Rasmussen, "Dirichlet process Gaussian mixture models: Choice of the base distribution," *Journal of Computer Science and Technology*, vol. 25, no. 4, pp. 653–664, 2010.

[43] C. Biernacki, G. Celeux, and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 719–725, 2000.

[44] O. Zobay, "Mean field inference for the Dirichlet process mixture model," *Electronic Journal of Statistics*, vol. 3, pp. 507–545, 2009.

[45] M. D. Hoffman and A. Gelman, "The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1593–1623, 2014.

[46] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," *PeerJ Computer Science*, vol. 2, p. e55, 2016.

[47] "Tubitak MAM, miltal millimeter wave and terahertz technology laboratory," https://me.mam.tubitak.gov.tr/en/research-areas/millimeter-wave-and-terahertz-technology, (Accessed on 03/13/2022).

[48] I. F. Akyildiz and J. M. Jornet, "Electromagnetic wireless nanosensor networks," *Nano Communication Networks*, vol. 1, no. 1, pp. 3–19, 2010.

[49] S. Brooks, "Markov chain Monte Carlo method and its application," *Journal of the Royal Statistical Society: series D (the Statistician)*, vol. 47, no. 1, pp. 69–100, 1998.

[50] T. Rydén, "Em versus Markov chain Monte Carlo for estimation of hidden markov models: A computational perspective," *Bayesian Analysis*, vol. 3, no. 4, pp. 659–688, 2008.

[51] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3197603



ERHAN KARAKOCA (Student Member, IEEE) completed his B.Sc. (with high Hons.) in electronics and communication engineering at Yıldız Technical University in Istanbul, Turkey. He is currently pursuing an M.Sc. in telecommunications engineering at Istanbul Technical University in Istanbul, Turkey. Also, he is a researcher in the Hisar Lab of TÜBİTAK BİLGEM. Next-generation wireless communication systems, terahertz wireless communications, Bayesian statistics, and machine learning are among his research interests.



GÜNEŞ KARABULUT KURT (Senior Member, IEEE) is currently an Associate Professor of Electrical Engineering at Polytechnique Montréal, Montréal, QC, Canada. She received the B.S. degree with high honors in electronics and electrical engineering from the Bogazici University, Istanbul, Turkey, in 2000 and the M.A.Sc. and the Ph.D. degrees in electrical engineering from the University of Ottawa, ON, Canada, in 2002 and 2006, respectively. From 2000 to 2005, she was a Research Assistant at the University of Ottawa. Between 2005 and 2006, Gunes was with TenXc Wireless, Canada. From 2006 to 2008, she was with Edgewater Computer Systems Inc., Canada. From 2008 to 2010, she was with Turkcell Research and Development Applied Research and Technology, Istanbul. Gunes was with Istanbul Technical University between 2010 and 2021. She is a Marie Curie Fellow and has received the Turkish Academy of Sciences Outstanding Young Scientist (TÜBA-GEBIP) Award in 2019. She is an Adjunct Research Professor at Carleton University. She is also currently serving as an Associate Technical Editor (ATE) of the IEEE Communications Magazine and a member of the IEEE WCNC Steering Board.

PLACE PHOTO HERE

ALİ GÖRÇİN (Senior Member, IEEE) received the B.Sc. degree in electronics and telecommunications engineering and the master's degree in defense technologies from Istanbul Technical University, and the Ph.D. degree in wireless communications from the University of South Florida (USF). After working at Turkish Science Foundation (TÜBİTAK) on avionics projects for more than six years, he moved to the U.S. to pursue Ph.D. degree. He worked with Anritsu Company during his tenure with USF and worked with Reverb Networks and Viavi Solutions after his graduation. He is currently an Assistant Professor at Yıldız Technical University, Istanbul. He is also the President of TÜBİTAK BİLGEM.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/