Examining Nonlinear Behavior in Optical Communication Systems Using Gaussian Mixture Model

Egor Sedov[a], Felippe Alves Pereira[b], David Saad[b], and Sergei Turitsyn[a]

[a]Aston Institute of Photonic Technologies, Aston University, B4 7ET Birmingham, UK
[b]Department of Applied Mathematics and Data Science, Aston University, B4 7ET Birmingham, UK

ABSTRACT

This study investigates the nonlinear effects on signal integrity in 16-QAM optical communication systems by focusing on received signal distributions without noise interference. Utilizing GPU-based simulations and analyzing "triplets" of consecutive signal points, we uncover that nonlinear interactions generate distinctive patterns in signal behavior, challenging the adequacy of standard Gaussian models. Our analysis employs the Gaussian Mixture Model (GMM), revealing that multi-component models offer a more accurate representation of signal distributions, highlighting the complexity of nonlinear effects. This research not only enhances our understanding of signal behavior under nonlinear conditions but also paves the way for future investigations into improving optical communication system design and reliability.

Keywords: Optical Communication, Nonlinear Effects, Long-haul, Gaussian Mixture Model, 16-QAM

1. INTRODUCTION

Optical communication channels often exhibit nonlinear effects that can significantly impact the integrity and recoverability of transmitted signals. Understanding of these effects is essential for designing robust communication systems capable of maintaining high level of performance. With a deterministic model such as the Nonlinear Schrödinger Equation (NLSE) that closely corresponds with real systems, we can study this behavior in depth. If we consider that we have precise knowledge of the final distribution of constellation points at the receiver for any given input at the transmitter, we can leverage this information to our advantage. As new symbols arrive at the receiver, we might initially be uncertain about their identity. However, with comprehensive knowledge about the ultimate distribution of all constellation points, we can ascertain how likely it is that a new point corresponds to a specific symbol. This approach allows us to accurately map each new point to the correct constellation symbol, enhancing the accuracy of the signal interpretation.

In pursuit of a deeper understanding of these effects, we turn to computational simulations. Specifically, we use a GPU-based simulation package, HpCom,[1] which allows one to generate large amount of data. By analyzing the internal structure of the received constellation points, we demonstrate that the nonlinear effects induce distinctive patterns in the data. These patterns are not merely noise; they hold valuable information about the behavior of the signal within the fiber. Through detailed analysis, we can reveal these patterns and improve our ability to predict and mitigate the impact of nonlinearity on our communications, pushing the boundaries of what is possible in optical data transmission.

In this study we initially ignore Amplified Spontaneous Emission (ASE) noise to focus on how signals change due to nonlinear effects alone. This choice is deliberate, to look closely nonlinear effects without noise interference. By considering signal changes as a predictable process due to nonlinear effects, we can better understand these effects without the randomness of noise getting in the way. This makes our analysis simpler and helps us get to the basics of how nonlinear effects influence signal changes, which is crucial for coming up with new ideas and algorithms to handle these nonlinearities better. Adding noise effects to the distortion due to non-linearity is straightforward and would be important for obtaining a complete picture of optical systems; however, it is beyond the scope of the current work.

Further author information: (Send correspondence to E.S.)
E.S.: E-mail: egor.sedoff@gmail.com

2. METHODOLOGY

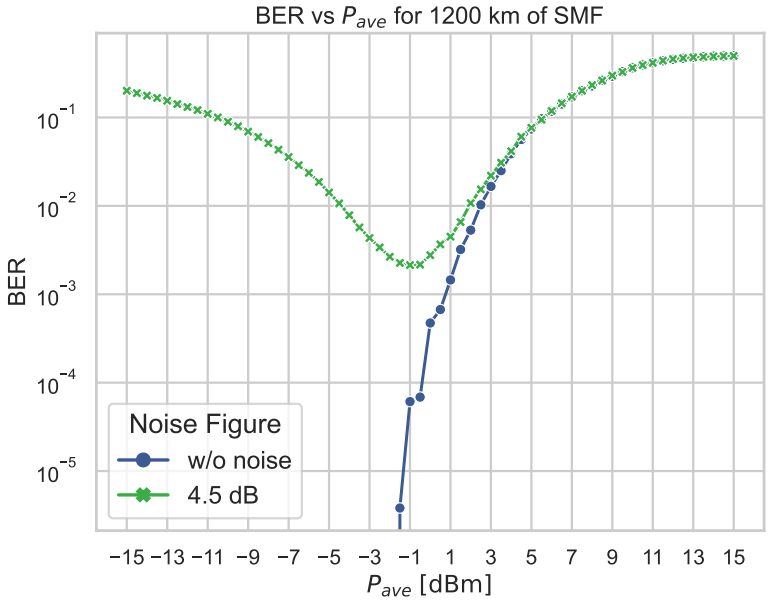


Figure 1. BER vs P_{ave} [dBm] for studied system. The green line marked with crosses represents the reference system, including the impact of additional EDFA noise characterized by a 4.5 dB Noise Figure. The blue line illustrates the ideal case, free from ASE noise, which is the focus of our study.

This study focuses on the analysis of a 16-QAM WDM signal, specifically investigating the behavior of received constellation points post-CDC and NPE. We employed an optical channel model of Standard Single-Mode Fiber (SSMF) with Erbium-Doped Fiber Amplifiers (EDFAs). The signal format under consideration is a 16-QAM WDM with single polarization and a symbol rate of 34.4 GBd. Pulse shaping was realized using a digital Root Raised Cosine (RRC) filter with a roll-off factor of 0.1. The total transmission distance spanned 15 links of 80 km each. The EDFA is assumed to be ideal, introducing no noise. The average signal power varied from -2 to 8 dBm. Signal propagation through the fiber was modeled by the Nonlinear Schrödinger Equation (NLSE), solved using the GPU-accelerated split-step Fourier method.[1] The fiber's parameters were set to a wavelength of $\lambda = 1550$ nm, a dispersion coefficient of $D = 16.8$ ps/nm $\cdot$ km, and a nonlinear coefficient of $\gamma = 1.2$ W$^{-1} \cdot$ km^{-1}. For the study, we generated 2^{24} data points for each specified average power level.

Figure 1 illustrates the dependence of the Bit Error Rate (BER) on the average signal power for the system parameters described previously. It depicts two scenarios: one with additional noise from Erbium-Doped Fiber Amplifier (EDFA) and one without. The power range under investigation is selected to align with the optimal levels for real-world transmission systems. As indicated by the graph (represented by the green line), the BER reaches its minimum around -1 dBm. It is important to note that, for the purposes of our study, we utilize data from the channel without EDFA noise. This approach allows us to specifically examine the impact of nonlinearity without the confounding effects of random noise.

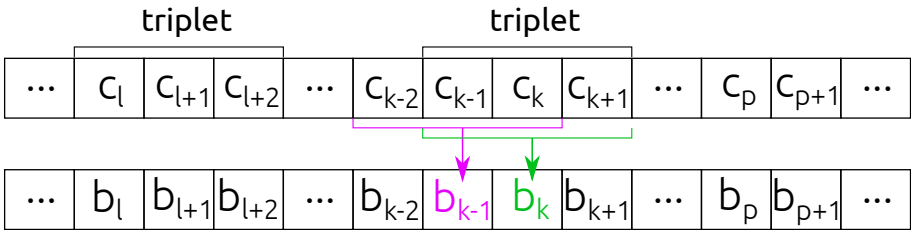


Figure 2. Schematic representation of the "triplet" concept, illustrating the extraction of data b_k corresponding to each transmitted triplet (c_{k-1}, c_k, c_{k+1}).

In this work, we introduce the concept of a "triplet", which is defined as a set of three consecutive transmitted symbols consisting of a left, central, and right points (Fig. 2). On the transmitter side, there is a sequence of transmitted symbols. Unlike real systems where the sequence is continuous and lacks a definitive beginning or end, our simulation employs a cyclic version: $\{c_0, c_1, \ldots, c_{k-1}, c_k, c_{k+1}, \ldots, c_{K-1}, c_K\}$, where K represents the total number of transmitted symbols. In the continuum of transmitted constellation symbols c_k, a triplet is any set of three points (c_{k-1}, c_k, c_{k+1}) for the transmitted sequence. Similarly, for received symbols b_k, we define a triplet as (b_{k-1}, b_k, b_{k+1}). Our focus is on the distributions of the received symbol b_k for a specific transmitted triplet (c_{k-1}, c_k, c_{k+1}). In essence, we aim to categorize our dataset and analyze the distributions for b_k based on the transmitted symbol and its adjacent neighbors. This relies on the understanding that the main impact on the nonlinearity-induced distortion comes from neighbouring signals; the received-signal variability for a given triplet results from the influence of other symbols in the sequence.

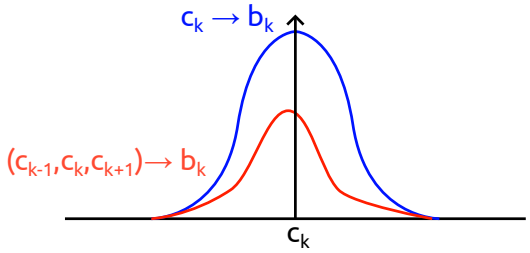


Figure 3. Schematic illustration of received symbol distributions based on transmitted triplets. The blue line represents the distribution of received symbols b_k originating from the transmission of a single symbol c_k. The red line shows the modified distribution of b_k when considering the entire transmitted triplet (c_{k-1}, c_k, c_{k+1}).

In our simulation setup, we maintain a fixed average signal power for the transmitted signal (we conduct separate analyses for each signal power level, ranging from -2 dBm up to 8 dBm). This setup provides us with complete information, including the sequence of transmitted symbols $\{c_k\}$ and the sequence of received symbols $\{b_k\}$. It is important to note that for this analysis, the received symbols $\{b_k\}$ have already been compensated for dispersion and phase shift, allowing us to concentrate solely on the deviations in the complex constellation plane caused by nonlinear interactions.

To analyze the impact of these nonlinear interactions, we examine the distributions of the received symbols $\{b_k\}$ based on their originating "triplet" (c_{k-1}, c_k, c_{k+1}). This process involves iteratively selecting indices k corresponding to each specific triplet configuration (for example, (c_{k-1}, c_k, c_{k+1}) is $(1+1\mathrm{j}, 1+3\mathrm{j}, -3-1\mathrm{j})$ or any other triplet), and gathering all corresponding b_k. For a 16-QAM system, there exist $16 \times 16 \times 16 = 2^{12}$ potential triplet combinations. With 2^{24} data points in our dataset, this results in approximately 2^{12} data points per triplet, available for further analysis. Through this method, we create distributions of received symbols $D(b_k)$ for each of the 2^{12} possible triplet combinations, thereby associating each distribution $D(b_k)$ with its respective sent triplet (c_{k-1}, c_k, c_{k+1}).

To simplify the analysis, we normalize the distributions of b_k for each triplet by subtracting the original transmission point c_k, shifting all distribution centers towards the origin $(0, 0)$. This adjustment aims to reduce the influence of the original transmission point's location on the analysis, allowing us to more clearly observe the shifts in distribution centers caused by nonlinear effects, which will be further explored in subsequent analysis steps.

Figure 3 provides a schematic representation of this process: transmitting c_k results in a distribution of received symbols b_k (depicted by the blue line). When we limit the analysis to received symbols b_k that correspond not only to the transmitted symbol c_k but to the sent triplet (c_{k-1}, c_k, c_{k+1}), we observe a modified and more concentrated distribution (indicated by the red line). The primary model utilized for distribution analysis is the Gaussian Mixture Model (GMM), which will facilitate the understanding of the underlying structure within the received signal constellations.

2.1 Gaussian Mixture Model and Likelihood

A GMM is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. The formula for a GMM can be written as:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where $\mathbf{x}$ is a data point, K is the number of Gaussian components, π_k the mixing coefficient for the k-th Gaussian component (with the constraint $\sum_{k=1}^{K} \pi_k = 1$), $\boldsymbol{\mu}_k$ the mean vector of the k-th Gaussian component, $\boldsymbol{\Sigma}_k$ is the covariance matrix of the k-th Gaussian component and $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Gaussian distribution.

In this study we use likelihood function to compare the efficacy of different models for the same data. The model with the higher likelihood (or higher log-likelihood) is generally considered to have a better fit to the data. For a given statistical model and observed data, the likelihood function $L(x|\theta)$ (where θ represents the parameters of the model and x the observed data) quantifies how well the model with specific parameter values explains the data.

Let us consider a simple example for likelihood for single one-dimensional Gaussian distribution. The likelihood function L for a Gaussian distribution gives us a measure of how likely it is to observe a given set of data points $(x_1, x_2, \ldots, x_n)$ given the parameters of the Gaussian distribution, specifically the mean μ and variance σ^2. For a set of independent and identically distributed observations from a Gaussian distribution, the likelihood function is:

$$L(x_1, x_2, \ldots, x_n|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \tag{2}$$

The log-likelihood is the natural logarithm of the likelihood function, which turns the product of probabilities into a sum of log probabilities, making it easier to work with especially for computations:

$$\log L(x_1, x_2, \ldots, x_n|\mu, \sigma^2) = \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2}\right] \tag{3}$$

If we return to two-dimensional case, for GMM (Eq. (1)) the log-likelihood is given by:

$$\log p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right), \tag{4}$$

where single two-dimensional datapoint $\mathbf{x_n} \in \mathbf{X}$ and k is the number of corresponding Gaussian component.

In our scenario, the two-dimensional data points $\mathbf{x}_n$ correspond to the received symbols $\{b_k\}$ associated with the selected triplet.

2.2 Model fitting

As previously discussed, within the dataset, we isolate all points b_k corresponding to any particular transmitted triplet (c_{k-1}, c_k, c_{k+1}). For each triplet in the 16-QAM system, this results in approximately 2^{12} data points. Given the 2^{12} combinations of triplets, we conduct a comprehensive analysis in the following manner. Initially, we assign to each triplet a unique identifier (ID) for ease of representation. Several IDs that will be referenced in this text are listed in Table 1: For these triplets, we gather all corresponding points b_k and create a centered distribution $D(b_k - c_k)_{\mathrm{ID}}$ of $b_k - c_k$. Although fitting the distribution is possible without this centering process, shifting the received points to align with the transmitted points simplifies the representation and interpretation of the data. Subsequently, we employ the BRMLtoolbox[2] to fit a GMM distribution to the centered data.

The Expectation-Maximization (EM) algorithm is a powerful iterative method used to find the maximum likelihood estimates of parameters in statistical models, especially when the model depends on unobserved latent variables. In the context of GMMs, the EM algorithm is used to estimate the parameters of the Gaussian

Table 1. Triplet unique identifiers (IDs)

Left point	Central point	Right point	identifier (ID)
$1+3\mathrm{i}$	$1+3\mathrm{i}$	$3+1\mathrm{i}$	552
$-1-1\mathrm{i}$	$1+1\mathrm{i}$	$-1-1\mathrm{i}$	1285
$3+3\mathrm{i}$	$3+3\mathrm{i}$	$3+3\mathrm{i}$	2730
$3-3\mathrm{i}$	$3-3\mathrm{i}$	$1-1\mathrm{i}$	2993

distributions (means, variances, and mixture coefficients) that best fit the data. The EM algorithm for GMMs involves two main steps repeated iteratively: the Expectation step (E-step) and the Maximization step (M-step).[3,4]

In the two-dimensional case, each data point x_i is a vector in $\mathbb{R}^2$, and the covariance matrices Σ_k of the Gaussian distributions are 2×2 matrices. The formulas for updating the means, covariance matrices, and mixture coefficients remain the same. However, the computation of the probability density $\mathcal{N}(x_i|\mu_k, \Sigma_k)$ for each data point involves the two-dimensional Gaussian density function, which takes into account the covariance between the two dimensions for each Gaussian component. This model allows for capturing more complex shapes in the data distribution by considering the correlations between dimensions, making GMMs a powerful tool for clustering and density estimation in multivariate data.

The analysis includes fitting GMM models with a number of Gaussian components ranging from one to five. For instance, a GMM with two components yields the following properties from the fitting process:

1. **Log-likelihood.** Log-likelihood of the data for the particular GMM parameters. The log-likelihood is given by Eq. (4).

2. **Mixing coefficients.** π_1 and π_2 are the mixing coefficient for the first and second Gaussian components respectively ($\pi_1 + \pi_2 = 1$).

3. **Mean Vectors.** For the first and second Gaussian component:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} \mu_{1,1} \\ \mu_{2,1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\mu}_2 = \begin{pmatrix} \mu_{1,2} \\ \mu_{2,2} \end{pmatrix}.$$

4. **Covariance Matrices.** For the first and second Gaussian component:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_{11,1} & \sigma_{12,1} \\ \sigma_{21,1} & \sigma_{22,1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} \sigma_{11,2} & \sigma_{12,2} \\ \sigma_{21,2} & \sigma_{22,2} \end{pmatrix}.$$

5. **Eigenvalues.** The eigenvalues of the covariance matrices could be obtained by solving the characteristic equation $\det(\boldsymbol{\Sigma} - \lambda \boldsymbol{I}) = 0$, where $\boldsymbol{I}$ is the identity matrix. For the first and second Gaussian components, the eigenvalues could be represented as follows:

$$\boldsymbol{\lambda}_1 = \begin{pmatrix} \lambda_{1,1} \\ \lambda_{2,1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\lambda}_2 = \begin{pmatrix} \lambda_{1,2} \\ \lambda_{2,2} \end{pmatrix}.$$

For other numbers of Gaussian components, the data is processed in a similar manner. All fitting results are stored for further analysis and are accessible from my GitHub repository or via email. Detailed information and data sets can be obtained through the following repository link or by contacting me directly. In the subsequent section, specific results and insights gleaned from the data will be presented.

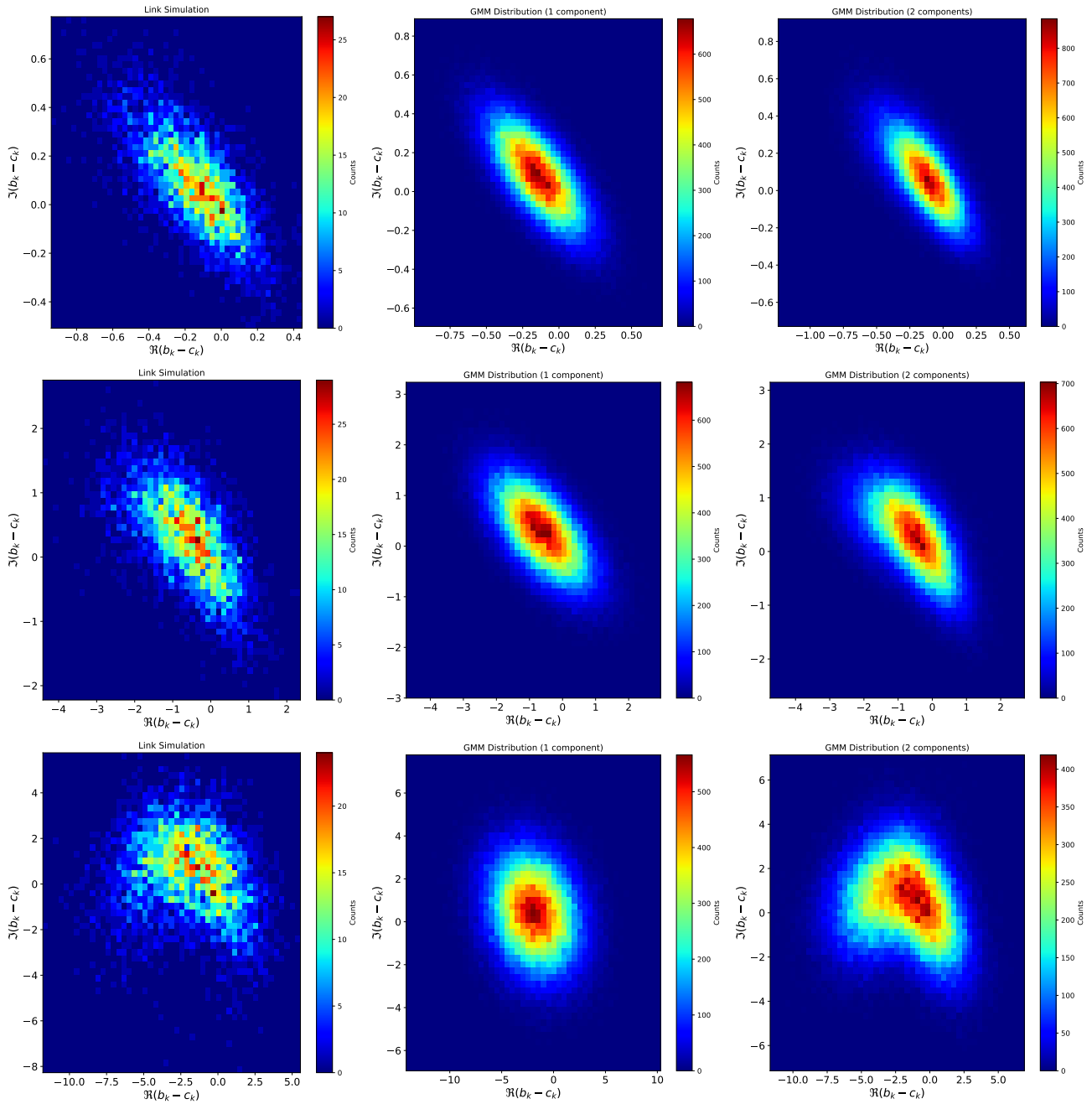


Figure 4. The distributions of the received constellation points b_k, centered with respect to the transmitted symbol c_k for triplet 2730 $[3+3\mathrm{i}, 3+3\mathrm{i}, 3+3\mathrm{i}]$. The first column illustrates the distribution of simulation data for a real communication system. The second column represents the distribution fitted with a single two-dimensional Gaussian distribution, while the third column visualizes a mixture of two Gaussians. First row corresponds to -2 dBm, second at – 0 dBm, and third – 8 dBm)

3. RESULTS

Figure 4 depicts the distributions of the received constellation points b_k, centered with respect to the transmitted symbol c_k (after Chromatic Dispersion Compensation (CDC) and Nonlinear Phase Equalisation (NPE)). Essentially, these figures represent the received symbols after subtracting the original transmitted symbol, corresponding to specific "triplets". It displays the distribution for triplet 2730 $[3+3\mathrm{i}, 3+3\mathrm{i}, 3+3\mathrm{i}]$. Notably, these distributions arise from nonlinear effects during the propagation of the WDM signal through the optical fiber, as the system excludes additional noise from EDFA amplifiers.

Figure 4 is structured as follows: rows correspond to different average signal power levels (first row at -2 dBm, second at 4 dBm, and third at 8 dBm). The first column illustrates the distribution of simulation data for a real communication system with the aforementioned parameters. These distributions are then fitted with a GMM, and the derived parameters are used to generate the visualizations in the second and third columns. The second column represents the distribution fitted with a single two-dimensional Gaussian distribution, while the third column visualizes a mixture of two Gaussians, each column containing 100000 randomly generated points. Examining Figure 4, at low average signal power, the GMMs with one and two components appear similar; however, the distribution for a single-component GMM is broader, indicating discernible differences. The disparity becomes more visible at higher power levels, where the two-component GMM notably deviates from a Gaussian shape. A striking example is at the highest power level (last row), where both the simulation distribution and the GMM with two components form a distinctive shape of a German pretzel. It is evident that a single-component GMM fails to accurately represent this distribution. This figure are merely exemplars out of 4096 different triplet distributions, yet they offer a compelling demonstration that the internal structure of distributions can be intricate and extend well beyond simple Gaussian forms.

Having observed that the shape of distributions for certain triplets can markedly differ when using single or multi-component GMM, we now proceed to compare the likelihoods of these distributions. To this end, we employ the log-likelihood measure described in the previous section (see Eq. (4)) to test the hypothesis that our simulation distribution corresponds to a specific GMM. Figure 5 presents violin plots that illustrate the log-likelihood distribution for different average signal powers and numbers of components in the GMM. A violin plot is a method of plotting numeric data and can be understood as a combination of a box plot and a kernel density plot. It includes a marker for the median of the data and a box indicating the interquartile range, as seen in standard box plots, overlaid with a kernel density estimation. The x-axis of each plot denotes the number of components in the GMM, while the y-axis shows the distribution of log-likelihood values across all possible triplets. This visualization method is particularly useful given the 2^{12} possible triplet combinations, which would be exceedingly challenging to depict using line plots (as demonstrated in Fig. 7). The average power in the graphs on Fig. 5 ranges from -2 dBm (upper left) to 8 dBm (lower right).

Regardless of the log-likelihood values, all graphs consistently reveal that the likelihood for a single-component GMM is significantly lower compared to a GMM with 2, 3, 4, or 5 components, which all exhibit approximately the same log-likelihood values. This pattern holds across all levels of average signal power. However, what varies with the power is the actual value of the log-likelihood. The principal finding, however, is as follows: the likelihood that our experimental data are drawn from a single-component GMM is exceedingly low compared to that from a multi-component GMM. This observation aligns with our earlier discussions: a standard Gaussian distribution does not adequately fit the experimental data, necessitating the use of a more complex model comprising multiple Gaussian components.

Figure 6 presents a violin plot of the log-likelihood for triplets across various power levels, contrasting a mixture with one component (left panel) against two components (right panel). In both instances, the average log-likelihood value decreases as the power increases, suggesting the potential for more intricate structural characteristics within the distribution. Notably, even at higher power levels, the two-component GMM provides a markedly superior fit to the experimental data; for instance, at 8 dBm, the average log-likelihood value for a single component is approximately -50000, whereas for two components it is around -16000. Although these values alone do not conclusively indicate that the experimental data correspond to a particular type of distribution, they do offer insights that a more suitable model might exist to describe such distributions. Furthermore, as illustrated in Fig. 5, the decline in average log-likelihood is consistent and nearly linear with increasing average power. This observation, coupled with insights from Fig. 6, indicates that adding more components to the

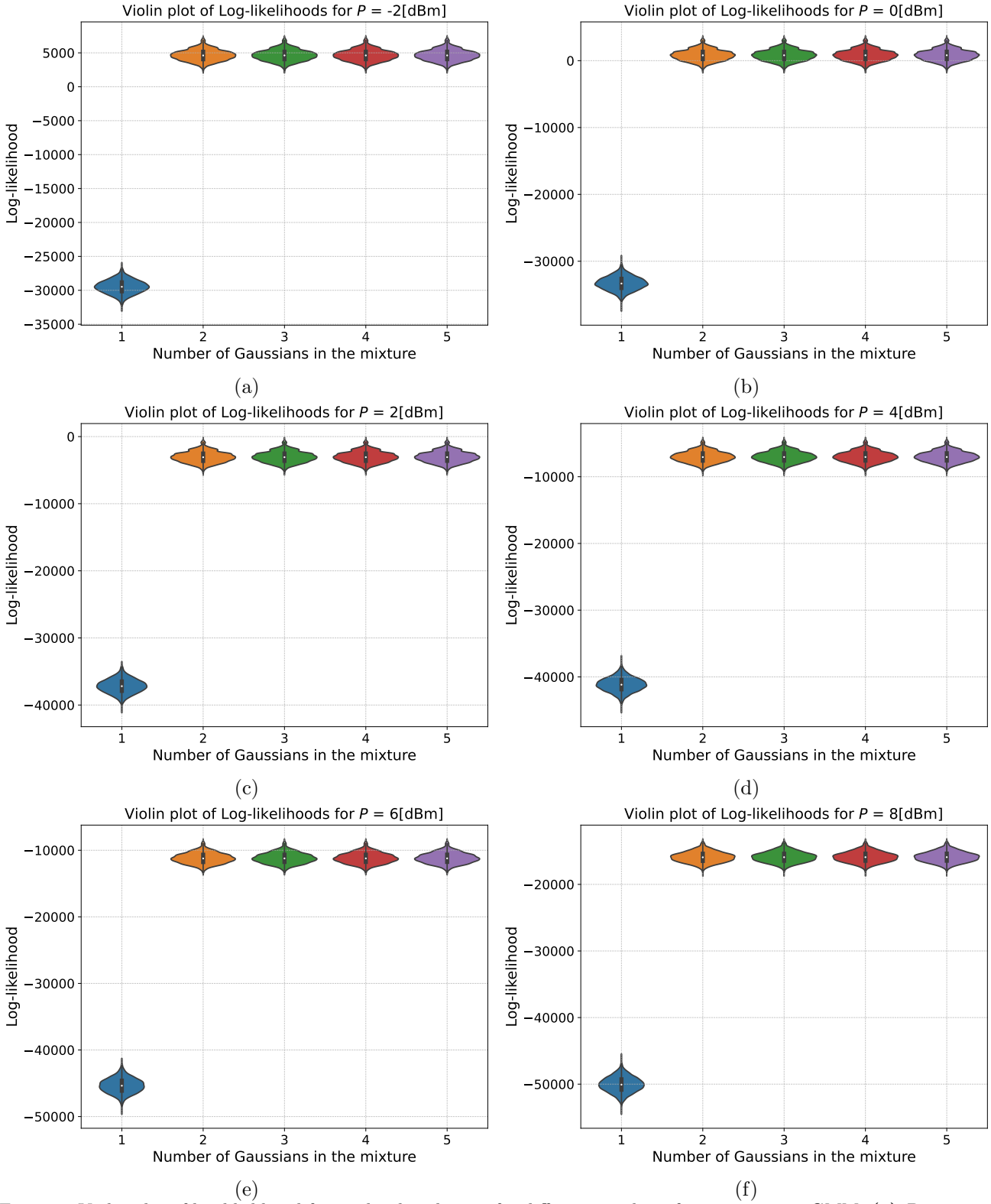
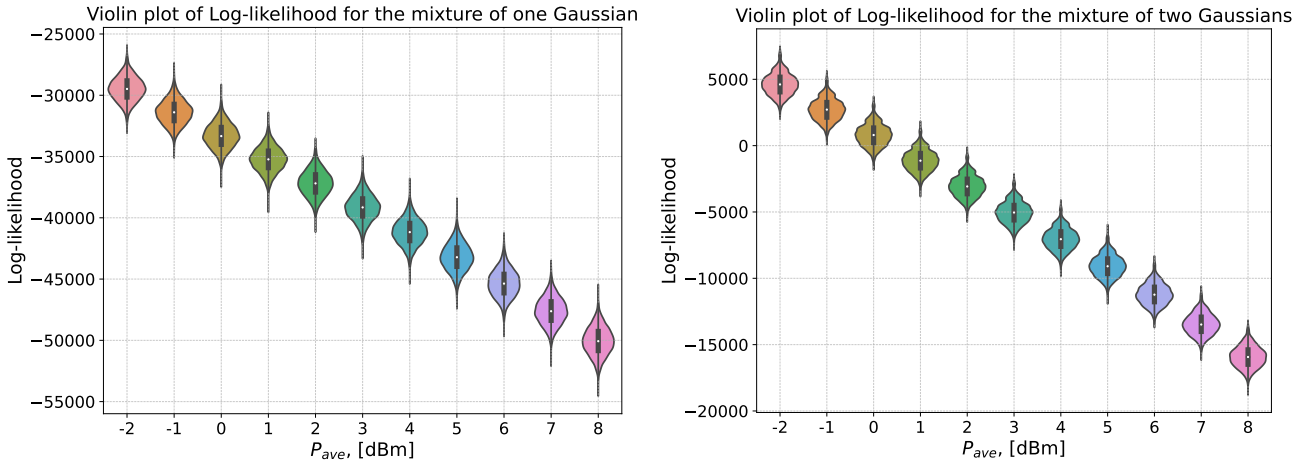
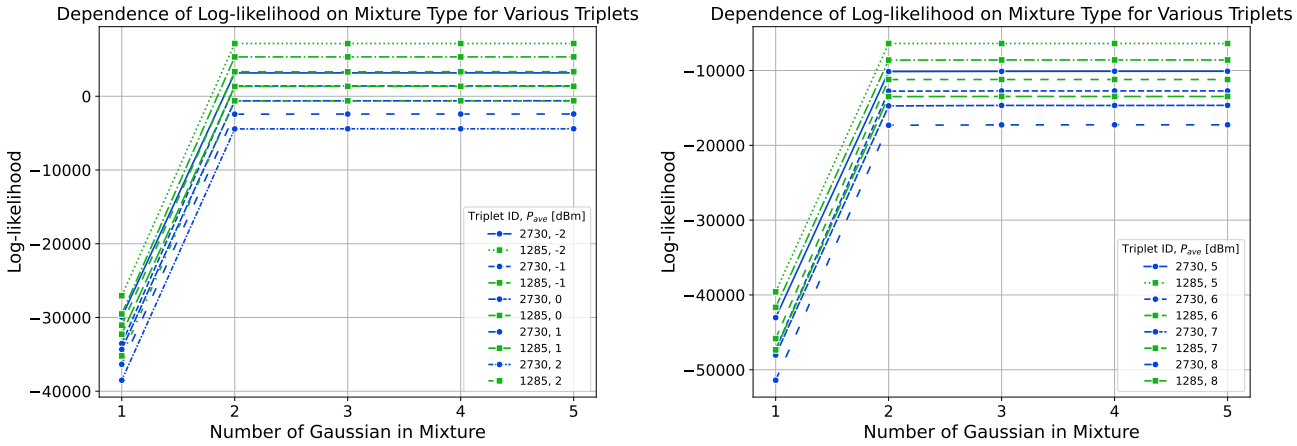


Figure 5. Violin plot of log-likelihood for triplet distribution for different number of components in GMM. **(a)** $P_{ave} = -2$ [dBm], **(b)** $P_{ave} = 0$ [dBm], **(c)** $P_{ave} = 2$ [dBm], **(d)** $P_{ave} = 4$ [dBm], **(e)** $P_{ave} = 6$ [dBm], **(f)** $P_{ave} = 8$ [dBm]



(a) (b)

Figure 6. Violin plot of log-likelihood dependence for triplet distribution on average signal power for different number of components in GMM. One Gaussian in the mixture **(left)** and mixture of two Gaussians **(right)**.



(a) (b)

Figure 7. Log-likelihood dependence for triplet distribution for different number of components in GMM. Graph shows triplets with ID 2730 $[3+3i, 3+3i, 3+3i]$ and ID 1285 $[-1-1i, 1+1i, -1-1i)]$.

mixture does not enhance the likelihood significantly. Therefore, for our simulations, a two-component mixture is deemed sufficient for an initial approximation of data representation.

Figures 7 and 8 offer a different view of the same data depicted by the violin plots in Figures 6 and 5, focusing on specific triplets. This perspective is valuable because violin plots can obscure the behavior of log-likelihood for individual triplets. These figures confirm a consistent finding: there is a substantial difference in the log-likelihood values between the single Gaussian model and a mixture of several Gaussians. The log-likelihood stabilizes with an increasing number of components in the GMM, indicating that a mixture of two Gaussians is typically adequate. The same trend is observed with the log-likelihood's response to signal power: it decreases linearly as the average power increases. Additionally, for a GMM with two components, the lines representing different triplets do not cross, affirming the consistency of this pattern.

4. CONCLUSION

In summary, this research marks an important step towards understanding how nonlinearities influence signal structures in 16-QAM systems, revealing that these structures deviate significantly from Gaussian models. Our

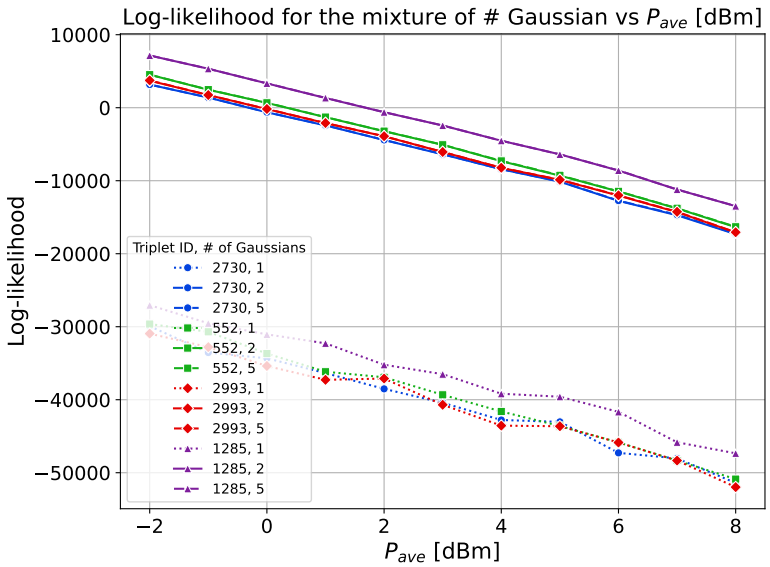


Figure 8. Dependence of log-likelihood for triplet distribution on average signal power for different number of components in GMM. Graph shows triplets with ID 2730 [3+3i, 3+3i, 3+3i], ID 1285 [−1−1i, 1+1i, −1−1i], ID 2993 [3−3i, 3−3i, 1−1i] and ID 552 [1 + 3i, 1 + 3i, 3 + 1i].

investigation underscores the complexity inherent in the distribution of received signals, where a Gaussian Mixture Model enhances the fit but reaches a saturation point in effectiveness with additional components. While the validation of these findings through error rate calculations has not been undertaken within the scope of this paper, it forms a critical component of our ongoing research agenda.

Future directions will explore mixture models beyond the GMM to discover if alternative or more intricate models could more accurately capture the nuances of signal distributions. The immediate next step involves developing an algorithm that utilizes the mixture-model framework to predict the distribution category of received signals. By calculating the likelihood across different models, we can identify the most fitting distribution for a given signal point.

Moreover, the analytical insights into signal distribution patterns hold promise for channel monitoring applications. Notably, discrepancies from anticipated distributions may signal underlying system issues, providing a novel diagnostic approach to maintain the operational integrity of optical communication networks.

Expanding this analysis to encompass various modulation formats and system configurations represents an exciting avenue for further research. The use of the HpCom[1] package will be instrumental in generating relevant statistics, facilitating a broad-based exploration of optical communication systems. Embracing these avenues for future investigation will not only deepen our understanding of the dynamics within optical communication systems but also contribute to enhancing their efficiency and reliability.

ACKNOWLEDGMENTS

All authors acknowledge funding from the EPSRC project TRANSNET (EP/R035342/1).

REFERENCES

[1] Sedov, E., “Hpcom - high performance communication library for python,” (Apr. 2023).
[2] Barber, D., [*Bayesian Reasoning and Machine Learning*], Cambridge University Press (2012).
[3] Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H., [*The elements of statistical learning: data mining, inference, and prediction*], vol. 2, Springer (2009).
[4] Murphy, K. P., [*Machine learning: a probabilistic perspective*], MIT press (2012).