Optics Communications 452 (2019) 203–210

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Modulation format identification based on constellation diagrams in adaptive optical OFDM systems

Yuanyuan Ma [a], Mingyi Gao [a,*], Junfeng Zhang [a], Yang Ye [a], Wei Chen [b], Hongliang Ren [c], Yonghu Yan [b]

[a] School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
[b] Key Lab. of New Fiber Tech. of Suzhou City, Jiangsu Heng tong Fiber Science and Technology Corporation, Suzhou 215200, China
[c] College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

ARTICLE INFO

Keywords:
Fiber optics communications
Modulation format identification
Constellation diagrams

ABSTRACT

Adaptive orthogonal frequency division multiplexing (OFDM) systems are promising for high-bit-rate short-reach communication by optimizing the allocation of modulation format and power to each subcarrier. Because adaptive OFDM systems involve multiple modulation formats, automatic modulation format identification (AMFI) is significant for subsequent signal processing and symbol decision of OFDM receivers. In this work, we proposed and experimentally demonstrated a blind modulation format identification technique based on constellation diagrams achieved from channel estimation of OFDM system. In the proposed AMFI scheme, the identification feature is the number of signals' constellation clusters. First, we use the peak-density clustering algorithm to track the centers of signals' constellation clusters by plotting the density–distance graph, where the clusters' centers have much larger density and distance values than that of the other clusters' points. Then, we use the K nearest neighbor regression algorithm to automatically calculate the number of signals' constellation clusters. Finally, we evaluated and measured the identification accuracies of QPSK, 8-QAM, 8-PSK, 16-QAM, 64-QAM and 128 QAM signals in simulation and experiment.

1. Introduction

Intensity modulation/direct detection (IM/DD) orthogonal frequency division multiplexing (OFDM) based on quadrature amplitude modulation (QAM) is one of good candidates for next-generation passive optical network because of the several advantages, such as strong anti-dispersion ability, high spectral efficiency, flexible bandwidth allocation and strong scalability [1]. However, the data rate and the cost are still limited by the bandwidths of the various electro-optical components. Fortunately, the adaptive OFDM systems can utilize the narrow-bandwidth components with lower cost to implement high-bit-rate operation by optimizing the allocation of modulation format and power to each subcarrier [2,3]. Nevertheless, symbol decision always depends on modulation format and the receiver must have prior knowledge of modulation format. It will be feasible to develop a universal receiver for adaptive OFDM systems if the modulation format can be automatically identified.

Modulation format identification (MFI) of QAM signals has been investigated based on the amplitude features and the constellation features [4–9]. In the OFDM systems, the constellation diagrams of QAM signals can be recovered by the modulation-format independent channel estimation prior to the symbol decision. Thus, the constellation diagrams with various templates are ready for MFI. Convolutional neural network (CNN) with amounts of training dataset has been proposed to identify the modulation formats by recognizing the images of signals' constellation diagrams [8,9]. However, the CNN-based system performance mainly relies on the training dataset, which has to been trained by amounts of data to optimize its accuracy. Besides, the overfitting is also a critical issue if the parameters are adjusted to fit the training data too well.

Furthermore, blind modulation format identification is especially attractive because additional training information is not required and the identification accuracy is independent on training dataset. Density-peak clustering algorithm has been proposed to track the centers of clusters and calculate the number of clusters for predicting the modulation format [5,6,10,11]. Such techniques utilize the clustering algorithm to classify the clusters without need of extra training information, where amplitude and Stokes features of QAM signals are used. However, the MFI accuracy based on the density-peak clustering method is limited by accurately distinguishing the clusters' centers from other clusters' halos and for the signals with more clusters' centers, it is becoming challenging to achieve the accurate number of constellation clusters.

* Corresponding author.
E-mail address: mygao@suda.edu.cn (M. Gao).

https://doi.org/10.1016/j.optcom.2019.07.039
Received 23 March 2019; Received in revised form 25 June 2019; Accepted 18 July 2019
Available online 21 July 2019

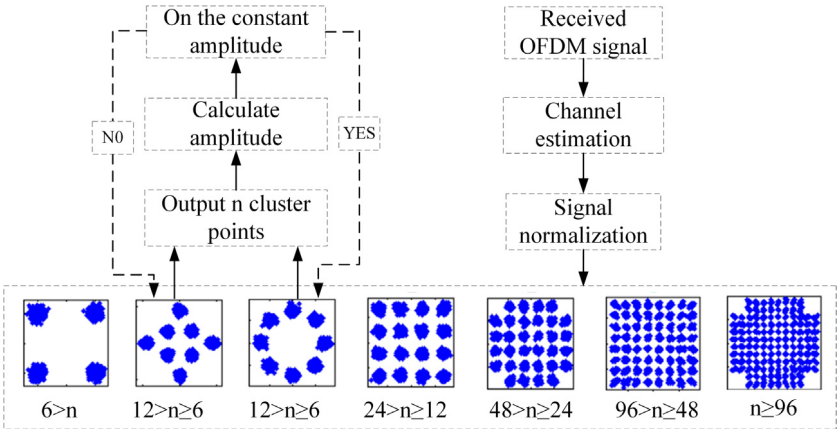

Fig. 1. Procedure diagram of the proposed MFI.

Moreover, the key parameter of the cutoff distance has not been well considered in most literatures.

In this work, we propose a blind MFI scheme for adaptive multi-modulation-format OFDM systems based on the constellation features extracted prior to symbol decision. We utilize the density-peak clustering algorithm to search the centers of QAM signals' constellation clusters. In order to accurately calculate the number of the QAM signals' constellation clusters, we apply the K nearest neighbor (KNN) regression algorithm to predict the density–distance product values. These product values are used to automatically calculate the number of constellation clusters and identify the QAM signals' modulation format. We experimentally demonstrated the proposed technique in an IM-DD OFDM optical communication system and analyzed the influence of the key parameters on MFI accuracy.

2. Principle

In the proposed MFI technique, first, the received OFDM signals with channel estimation are normalized. And then, the normalized data is processed to estimate the number of the constellation clusters n as the identification feature, as shown in Fig. 1. According to the estimated n, the modulation format can be identified, where we use QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM, 64-QAM and 128 QAM signals as examples. Note that n is in the same range for 8-PSK and 8-QAM signals and therefore further amplitude judgment is necessary, where 8-PSK signal has the constant amplitude for all constellation clusters instead of 8-QAM signal.

From Fig. 1, it is obvious that the key of the proposed MFI technique is to yield n from the received data. We use the density-peak clustering algorithm to search the centers of the constellation clusters and automatically calculate the number of constellation clusters n. For the signals with channel estimation, the constellation diagrams can be extracted. The centers of the different clusters in the constellation diagrams always have local maxima in the density of data points surrounded by lower-local-density neighbors [12]. These points with the highest local density have larger distance from higher-local-density points. The detailed procedure is followed.

First, we calculate the local density ρ of data, expressed as,

$$\rho_i = \sum_{j \neq i} e^{-(dist_{i,j}/D_c)^2} \tag{1}$$

where Gaussian kernel function is used. Assume that the dataset is $X = [X_1, X_2 \ldots X_N]$, N is the number of the data and i, $j \in N$. $dist_{i,j}$ is the distance between point X_i and point X_j and D_c is the cutoff distance. If the distance $dist_{i,j}$ between the points is less than D_c and the point X_j will be considered within the cluster of X_i. D_c plays a significant role in the clustering algorithm and has an important influence on the identification accuracy. The choice D_c of will be discussed in detail in Section 5.

For each data, there are N−1 distances from other N−1 data. Thus, there are N (N−1) distance values of N data points and half of them are reduplicate. These distances can be sorted in ascending order as $D_1 \leq D_2 \leq D_3 \ldots \leq D_{N(N-1)/2}$. If D_c is denoted as D_Z and there are Z distances with the values less than D_Z. Generally, Z is approximately 1%~2% of the total distance [12]. Once Z is determined, the local density ρ of each data can be calculated according to Eq. (1).

Second, we calculate the minimum distance δ of each point from the higher-local-density points. Assuming that $\rho_1 \geq \rho_2 \geq \rho_3 \ldots \geq \rho_N$ is the local density of all points sorted in descending order, then the data point X_i $(i \in N)$ with the local density ρ_i will have $i-1$ higher-local-density neighbors. The minimum distance δ_i of point X_i from the higher-local-density points can be expressed as,

$$\delta_i = \min_{j=1:\,i-1} (dist_{i,j}) \tag{2}$$

For the highest-local-density point X_j with the local density ρ_1 in the local-density array of descending order, δ will be set to the maximum distance between the point X_j and other points of the dataset.

Generally, the decision graph of the density ρ versus distance δ and the product γ_i bar graph of the density ρ_i and the distance δ_i can be used to calculate the number of constellation clusters n by plotting a decision line to distinguish the centers of constellation clusters and the other points [5,12].

The product γ_i of the density ρ_i and the distance δ_i is defined as,

$$\gamma_i = \rho_i \times \delta_i \tag{3}$$

Fig. 2 shows the decision graphs of the QPSK, 16-QAM and 64-QAM signals with different optical signal-to-noise ratios (OSNRs). The centers of different constellation clusters are those distinct density maxima in the decision graphs, as shown by the red dots in Fig. 2. As shown in Fig. 2, for the degraded signals with lower OSNRs and the higher-order signals, the centers of constellation clusters are always mixed with the other points, it becomes difficult to distinguish them. Usually, the number of the constellation clusters is achieved by plotting the decision line in the product bar graph with the descending order as shown in Fig. 3(a), (d) and (g), where the decision line is chosen at the position where the product bar graph starts anomalously, i.e., falls sharply [5]. As shown in Fig. 3(a), (d) and (g), with increase of the modulation order, the product bar graph falls sharply for more times. It is also more challenging to plot accurate decision lines in the product bar graphs of the 64-QAM signal than that of the QPSK and 16-QAM signals, because for higher-order signals, the difference between the centers of constellation clusters and the other points becomes trivial. Meanwhile, for the systems involved multiple modulation formats, the

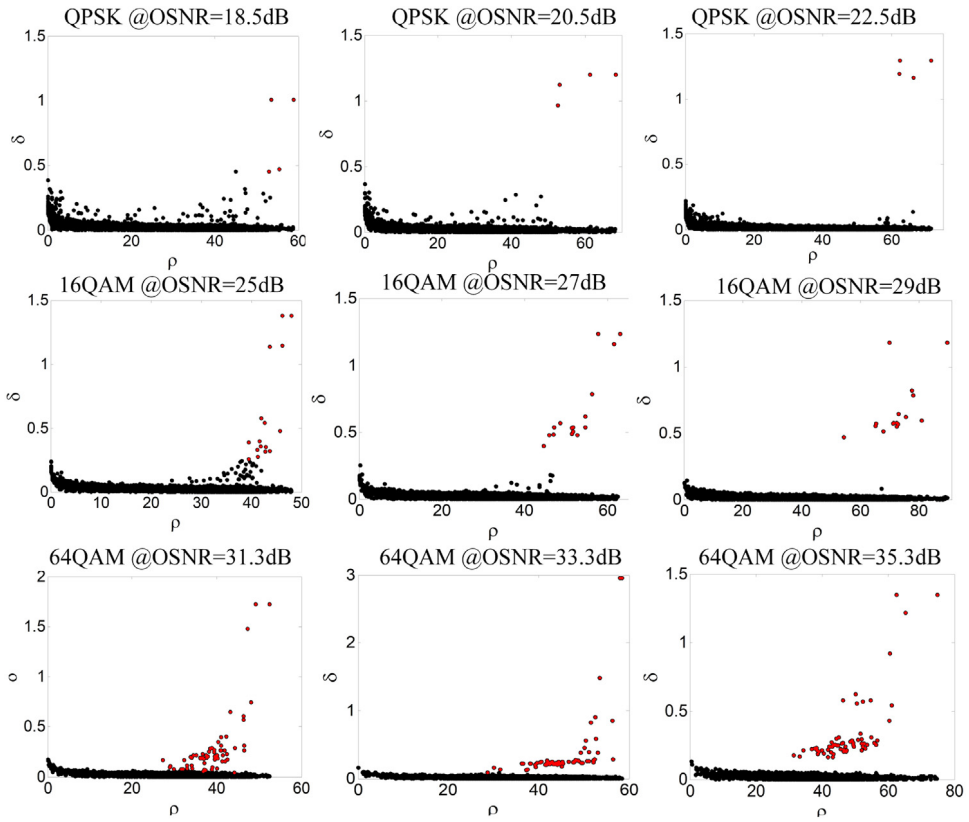

Fig. 2. Decision graphs of QPSK, 16-QAM and 64-QAM signals with various OSNRs.

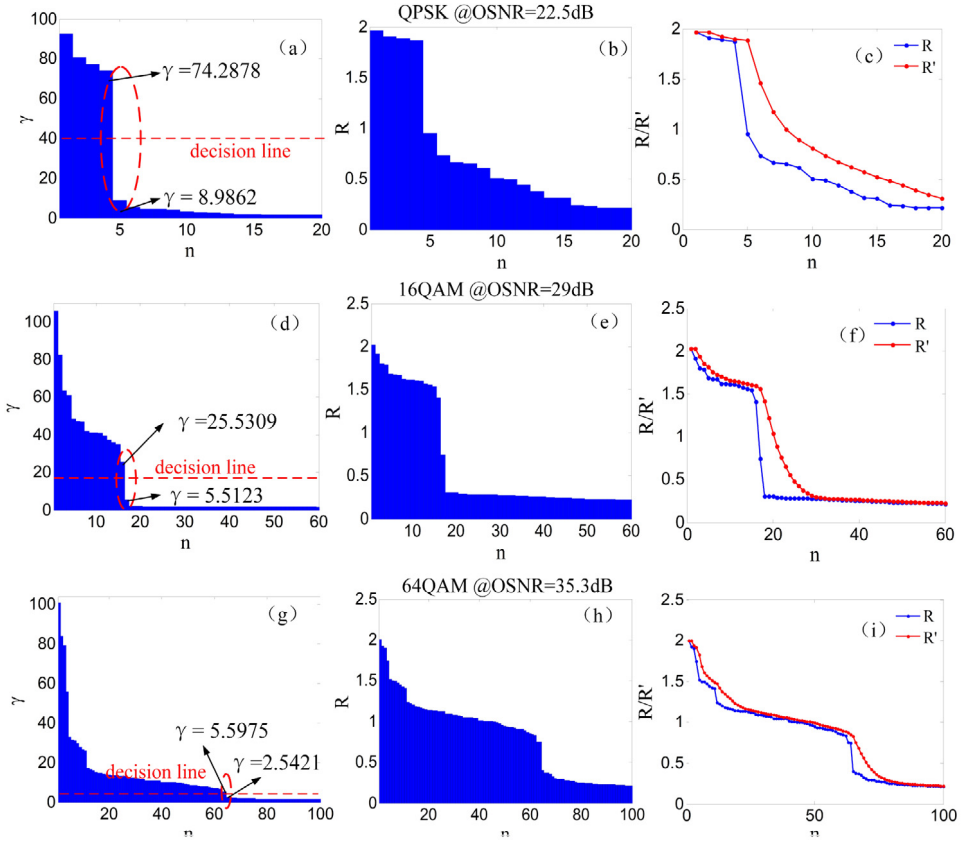

Fig. 3. Product γ bar graphs, logarithmic product R = log(γ) bar graphs and the measured R and KNN predicted R′ values . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

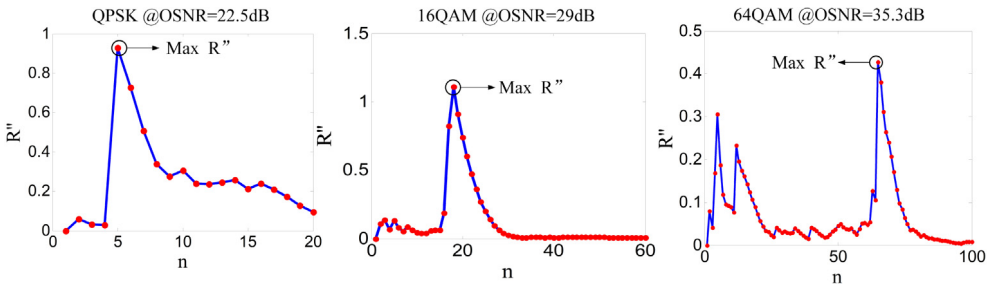

Fig. 4. Difference graphs between the measured R vale and the predict R' value.

issue will be how to find an optimum universal decision line for all signals with the diverse distributions of the product γ value, as shown in Fig. 3(a), (d) and (g).

In order to automatically yield the accurate number of the constellation clusters from the product bar graph, we first plot the logarithm of the product bar graph to enlarge its transilience in Fig. 3(a), (d) and (g), i.e. R = log(γ), as shown in Fig. 3(b), (e) and (h). Then, we utilize the KNN regression method to find the transilience position in the product bar graph instead of roughly plotting a decision line. The KNN regression algorithm searches for the nearest K points for the new prediction points, and then averages the values of the K points as the predicted values of the new prediction points. The choice of the appropriate K value will be discussed in Section 5. Therefore, the product value can be estimated by the R values of the surrounding K samples. In this work, we adopt the weighting KNN regression algorithm and the predicted R′ value can be expressed as [13],

$$R(m)' = \begin{cases} \dfrac{\sum_{i=1}^{K}(k-i+1)^2 \times R(m-i)}{\sum_{i=1}^{K} i^2} & m > K \\ \dfrac{\sum_{i=1}^{m-1} i^2 \times R(i)}{\sum_{i=1}^{m-1} i^2} & m \le K \end{cases} \tag{4}$$

where X_m is the mth data of the dataset X = [X_1, X_2…X_N], R (m) is the product value of X_m, R (m)′ is the predicted product value of X_m by KNN regression algorithm, and K is the K value of the KNN regression algorithm. The product R values are sorted in descending order and the differences of R values between neighbors' points are much smaller than that of the points far away from each other. The weighting factor is inversely proportional to the number of points. The weighting coefficient of the farthest point is square of 1 and that of next point is square of 2, and so on. The measured product value R and the KNN predicted value R' are plotted in Fig. 3(c, f, i) by the blue curve and the red curve, respectively.

Finally, we calculate the difference R'' between the estimated R' value and the measured R value, given by

$$R'' = R' - R \tag{5}$$

The maximum R'' difference value corresponds to the transilience position of the product bar graph, as shown in Fig. 4. The number of the constellation clusters can be automatically achieved,

$$n = n_{\max(R'')} - 1 \tag{6}$$

3. Simulations

Simulations of QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals have been conducted to verify the feasibility of the proposed MFI technique. The modulated optical OFDM signal is first loaded onto the optical carrier by the Mach–Zehnder modulator (MZM). Then, an Erbium doped fiber amplifier (EDFA) and a variable optical attenuator (VOA) are used to vary OSNRs of the signal. Finally, the optical signal is received by the photodetector. The identification accuracy is evaluated by the measured recognition rates with various OSNR values according to the estimated number of constellation clusters.

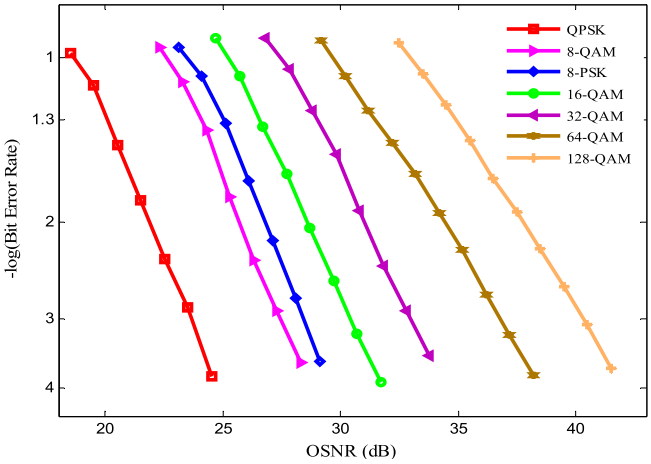

Fig. 5. BER vs. OSNR (dB) curves of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals.

Generally, different modulation signals have different OSNR sensitivity and higher-order modulation signals require higher OSNR to yield desirable BER performance. For worse signals, it is difficult for signal recovery and it is not meaningful to identify their modulation formats. Therefore, we plot the BER curves against OSNR of QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals in Fig. 5 and set the BER of 1×10^{-1} to 1×10^{-4} as the references to measure the OSNR ranges of various signals for the MFI. The OSNRs ranges of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals are 18.5 dB to 24.5 dB, 22 dB to 28 dB, 23 dB to 29 dB, 25 dB to 32 dB, 26.5 dB to 33.5 dB, 29.5 dB to 38.5 dB, and 32.5 dB to 41.5 dB, respectively. The following simulations and experiments will focus on the above approximate OSNR ranges.

Figs. 6 and 7 are the calculated identified numbers of constellation clusters and the corresponding recognition rates of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals as functions of OSNR. Dashed curves denote the decision line-based results and the solid curves represent the KNN-based results. Here, we use 7 modulation formats and each signal has different OSNR variation range to yield desirable BER performance. For different signals with various OSNRs, the product bar graphs in Fig. 3 will have diverse distributions.

For the system with multiple modulation formats, we first calculate the variation range of decision lines for each modulation format with higher OSNRs. According to these variation ranges, we will choose a universal decision line suitable for all modulation signals. However, it is challenging to yield a universal decision line for all signals, because higher-order signals, such as 64-QAM and 128-QAM signals, have no larger jumps between the clusters' centers and the other points in the product bar graphs, as shown in Fig. 3. If the decision line is chosen for most modulation signals, i.e., the line with γ = 10.5, the decision line-based MFI method can achieve better recognition rates on the

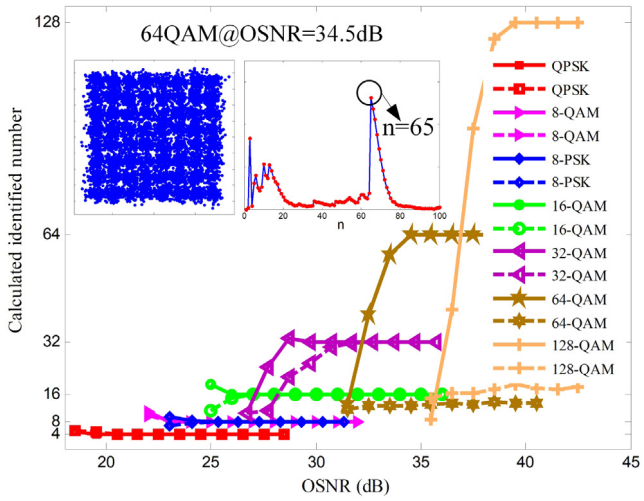

Fig. 6. Calculated identified number of constellation clusters of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals vs. OSNR (dB).

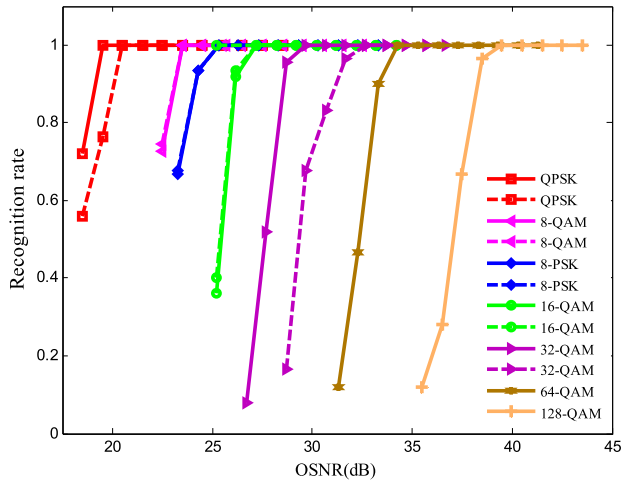

Fig. 7. Recognition rates of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals vs. OSNR (dB).

lower-order signals, such as QPSK, 8-PSK, 8-QAM, 16-QAM and 32-QAM signals. Unfortunately, the designated decision line will not work on the higher-order signals, such as 64-QAM and 128-QAM signals.

From Figs. 6 and 7, it can be seen that lower-OSNR signals suffer from severe noise and it becomes challenging to accurately identify the number of constellation clusters. Therefore, the identified number of constellation clusters becomes inaccurate with the decrease of the OSNR values, so does the recognition rate. It is notable that the identified number of constellation clusters is in the same range for 8-QAM and 8-PSK signals and the further amplitude judgment is required. Here, we utilize 50 testing data for each modulation signal at each OSNR and the length of the testing samples is 8192. The proposed KNN-based method can achieve 100% recognition rate for QPSK with OSNR of over 19.5 dB, 8-QAM with OSNR of over 23 dB and 8-PSK with OSNR of over 25 dB. With the increase of the modulation order, the required OSNR values to achieve 100% recognition rate are also increased to 27 dB, 30 dB, 34.5 dB, and 39.5 dB for 16-QAM, 32-QAM, 64-QAM and 128-QAM signals, respectively. Meanwhile, the decision line-based method can work well only for lower-order signals, such as QPSK, 8-QAM, 8-PSK and 16-QAM signals and has worse performance on the 32-QAM signals, as shown by dashed curves in Figs. 6 and 7. The reason is that the designated universal decision line is not suitable for the 64-QAM and 128-QAM signals. Actually, for the higher-order signals, the clusters' centers and the other clusters' points are mixed up as shown in Fig. 2 and the differences between them are very trivial in the product bar graphs of Fig. 3. Therefore, the variation of decision line is larger between the lower-order signals and the higher-order signal and it is difficult to choose an optimum universal decision line for all signals. On the contrary, the proposed KNN-based method can automatically track the transilience position in the product bar graph instead of roughly plotting a decision line, which enables it to yield better recognition performance. Therefore, the proposed AMFI method is especially suitable for the system with multiple modulation formats involved higher-order modulations.

4. Experimental setup and results

In order to evaluate the proposed MFI technique, we set OFDM experiments of various modulation signals, such as QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM, and 64-QAM, as shown in Fig. 8. The 12.5-Gbuad OFDM signals with various QAM/PSK mapping are generated offline and loaded into a 50-GSamples/s arbitrary waveform generator (AWG). Then we use an external cavity laser (ECL) and an intensity modulator to convert the electrical OFDM signal into the optical OFDM signal. The launched power of optical OFDM signal into the 30-km single mode fiber (SMF) is adjusted to 0 dBm by a variable optical attenuator (VOA). The fiber attenuation is 0.2 dB/km, the fiber dispersion at 1550 nm is 18 ps/nm km, the fiber nonlinear coefficient is 1.2 W^{-1} km^{-1} and the fiber polarization mode dispersion is 0.2 ps/$\sqrt{}$km. We compare the cases with and without 30-km SMF transmission. After that, the output optical signal from the SMF is launched into a variable optical attenuator (VOA) followed by an Erbium doped fiber amplifier (EDFA) to vary the noise level and adjust the signal' OSNR. Before the photodetector, another VOA is used to yield −3-dBm input power and a 50-GSamples/s real-time oscilloscope is utilized to acquire data. Finally, modulation formation identification and OFDM demodulation are conducted offline.

We measure the identified numbers of constellation clusters and the recognition rates of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM and 64-QAM signals, as shown in Figs. 9 and 10. The solid lines indicate the experimental results without fiber transmission and the dotted lines represent that with 30-km fiber transmission. Here, we utilize 50 testing data for each modulation of each OSNR and the length of the testing samples is 8192.

For the deteriorated signals with lower OSNR, the constellation diagrams are very ambiguous and the identified numbers of constellation clusters are not accurate. With the increase of OSNR, the constellation diagrams become distinct and it is easier to identify the accurate numbers of constellation clusters, as shown in Fig. 9. For QPSK, 8-QAM and 8-PSK signals, the required OSNR values to achieve 100% recognition rate are 21 dB, 23.5 dB and 26.5 dB. For 16-QAM, 32-QAM, and 64-QAM signals, there are more constellation clusters and higher OSNR values are required to yield 100% recognition rate, such as 27.5 dB for 16-QAM, 30 dB for 32-QAM and 37 dB for 64-QAM, as shown in Fig. 10. As seen in Figs. 9 and 10, the fiber transmission has less influence on the identification accuracy. The measured results in Figs. 9 and 10 have the similar tendencies as the simulation results in Figs. 6 and 7.

5. Discussions

In this section, we will discuss the effect of two key parameters on the recognition rate, i.e., the cutoff distance D_c of the peak-density clustering algorithm and the K value of KNN regression algorithm.

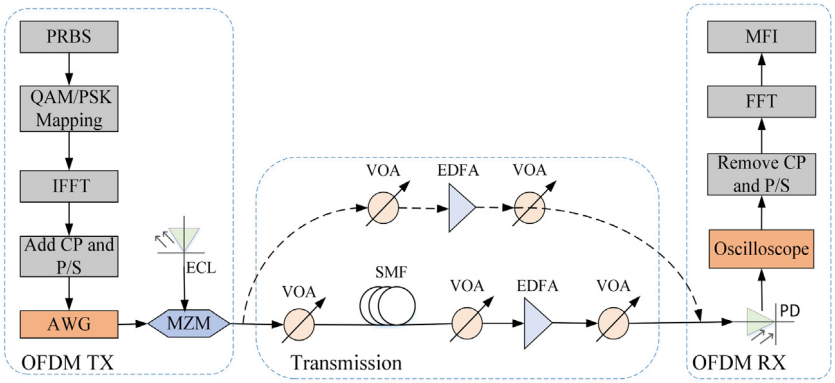

Fig. 8. Experimental setup for QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM and 64-QAM optical OFDM communication system.

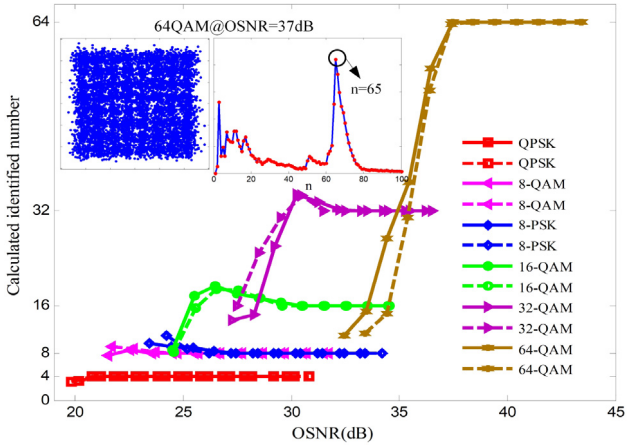

Fig. 9. Measured identified number of constellation clusters of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM and 64-QAM signals vs. OSNR (dB).

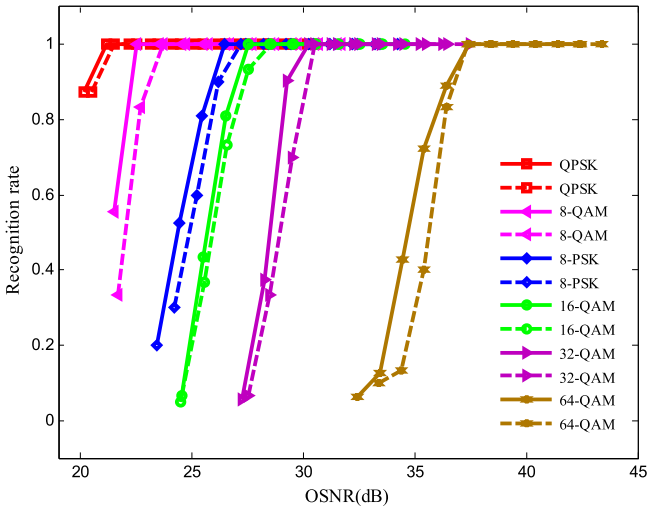

Fig. 10. Measured recognition rates of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM and 64-QAM signals vs. OSNR (dB).

5.1. Cutoff distance D_c

D_c is the cut-off distance and used to calculate the local density ρ of data in Eq. (1). More data with distance less than D_c will give a higher local density. To achieve a high recognition rate, it is necessary to optimize D_c. On the one hand, too larger D_c yields higher local density but simultaneously multiple clusters will be classified into one cluster, which deceases the recognition rate. On the other hand, a cluster may be divided into multi-subcategories because of too small D_c. In this work, we use QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals to investigate the influence of cutoff distance on the recognition rate.

For a universal modulation-format-agnostic receiver, the cutoff distance D_c has to set prior to modulation format identification. Therefore, an accurate cutoff distance is much helpful to increase recognition rate of multi-modulation systems. In Ref. [12], D_c is set so that the average number of localities is about 1% to 2% of the total. In this work, 2% is chosen and then the obtained cutoff distance D_c is weighted with the weighting factor from 0.05 to 1. The step interval of weighting factor is set to 0.05.

In order to track the optimal weighting factor of D_c for each modulation format, we plot the recognition rate curves versus the weighting factor in Fig. 11. We measure experimentally the modulation format recognition rate of QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM and 64-QAM signals as the function of D_c weighting factor, where the solid and dashed curves denote the results without/with fiber transmission, respectively. Limited by the receiver sensitivity of direct detection, 128-QAM signal cannot be recovered in experiment. As shown in Fig. 11, the recognition rates of lower-order modulation signals are not sensitive to the D_c weighting factor, such as QPSK, 8-PSK and 8-QAM signals. The recognition rates of higher-order modulation signals highly rely on the D_c weighting factor, as shown by the curves of 16-QAM, 32-QAM and 64-QAM signals. For the higher-order modulation signals, it is significant to choose an optimal D_c value. As shown in Fig. 11 that for 16-QAM, 32-QAM and 64-QAM signals, the optimum weighting factors of D_c are 0.5, 0.4 and 0.3, respectively.

In order to verify the reliability of the above results, we measure experimentally the modulation format recognition rate of 16-QAM and 64-QAM signal with various OSNR values versus D_c weighting factor in Fig. 12. The measured corresponding constellation diagrams are inserted in Fig. 12. As shown in Fig. 12, the deteriorated signals with lower OSNR values have the reduced variation ranges of optimum D_c. Therefore, the D_c optimization is very important for deteriorated signals. For the deteriorated signals with lower OSNRs, the recognition rates are lower because the deteriorated signals have very messy constellation diagrams, as shown in Fig. 12. If extra equalization algorithms are utilized to improve the constellation diagrams of the deteriorated signals, the proposed AMFI can work to yield high recognition rate. In addition, the signal with the same modulation format and different OSNRs has same optimal D_c coefficient, i.e., the optimal D_c coefficient heavily depends on the modulation format. In the adaptive multi-modulation OFDM system, the optimal D_c coefficient should be determined according to the highest-order modulation format involved, as shown in Fig. 12. In this work, we use 64-QAM as the highest-order modulation and therefore the weighting factor of D_c is set to 0.3.

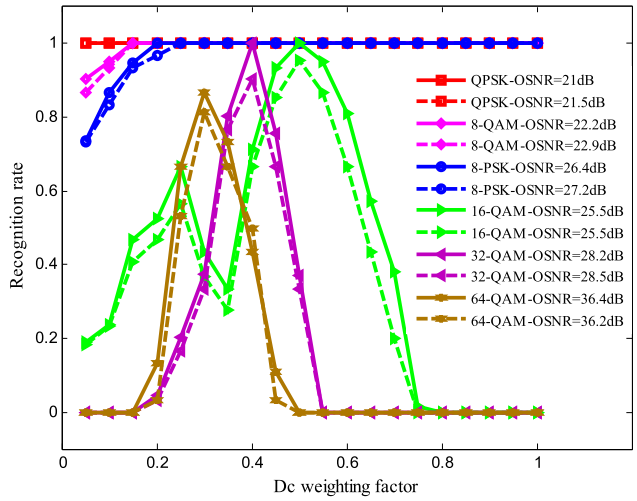

Fig. 11. Measured recognition rate of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM and 64-QAM versus D_c weighting factor with various OSNR.

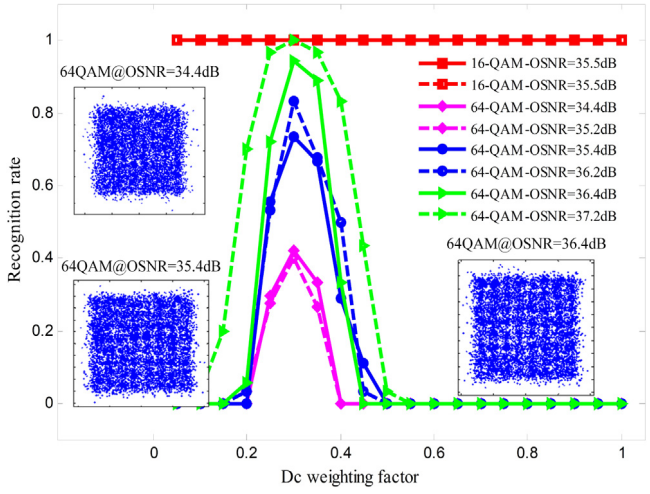

Fig. 12. Measured recognition rate of 16-QAM and 64-QAM with various OSNRs versus D_c weighting factor.

5.2. K value

Next, we will discuss the effect of the K value of K nearest neighbor regression algorithm. As discussed in Fig. 3, when the log(R) value is sorted in descending order, there is a transilience to distinguish the constellation clusters' centers. KNN regression algorithm is used to locate the transilience of log(R) so as to identify the modulation format of the signal by calculating the number of the constellation clusters. In KNN regression algorithm, the K previous values are used to predict the current values and the issue will be how many previous values are necessary to accurately predict the current values. Smaller K values will cause more prediction errors and contrarily larger K values will cost more computational resource.

In the KNN regression algorithm, K value plays an important role in recognition rate. Generally, larger K values can give more accurate predicted results to improve the recognition rate of the proposed method. Meanwhile, the computational consumption of KNN regression algorithm also depends on the K value with the complexity of O(K). Therefore, it is significant to optimize the K value in the system. Fig. 13 shows the recognition rate as a function of K value of KNN regression algorithm for various signals, where K value is varied from 1 to 20. As shown in Fig. 13, K value has no influence on the recognition rate of

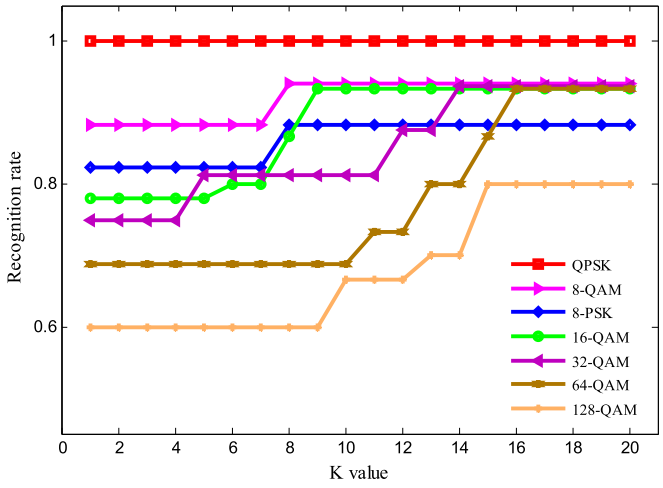

Fig. 13. Calculated recognition rate of QPSK, 8-QAM, 8-PSK, 16-QAM, 32-QAM,64-QAM and 128-QAM vs. K value.

QPSK signal because of the sparse distribution of its constellation clusters. For other modulational signals with dense constellation clusters, as expected, larger K values improve the recognition rate and higher-order signal requires larger K value. However, when the algorithm attains the convergence, the recognition rate cannot be improved furthermore by increasing K values. For the system with multiple modulation formats, the optimum K value of the highest-order signal should be chosen. In this experiment, the 64-QAM is used as the highest-order modulation and thus K value is set to 16. Meanwhile, compared with the O(K) complexity of the KNN regression, the computation cost of the proposed MFI method mainly relies on the calculation of the local density and the distance. When there are N constellation points, the complexity is $O(N^2)$.

6. Conclusion

We have proposed a blind modulation format identification based on the peak-density clustering algorithm and KNN regression algorithm for adaptive OFDM communication systems. We utilized the peak-density clustering algorithm to search the centers of signals' constellation clusters and plot the product bars of the density and the distance. Then, we applied the KNN regression algorithm to predict the transilience's location of the product bars so as to calculate the number of the constellation clusters' centers for modulation format identification. KNN regression method can achieve accurate prediction of the constellation clusters' centers and therefore improve the identification accuracy of modulation format. Simulations and experiments have been implemented to verify the feasibility of the proposed method for QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM, 64-QAM and 128-QAM signals. We also discussed the influence of key parameters on the identification accuracy. The proposed blind modulation format identification will be helpful to construct the universal modulation-format-agnostic OFDM receiver.

Funding

This work was supported in part by Open Fund of IPOC 2017B001 (BUPT), China, Suzhou Industry Technological Innovation Projects, China (SYG201809), the National Natural Science Foundation of China Project (61307082) and Natural Science Foundation of Zhejiang Province, China (LY16F050009).

References

[1] J. Zhang, Y. Zheng, X. Hong, C. Guo, Increase in capacity of an IM/DD OFDM-PON using super-nyquist image-induced aliasing and simplified nonlinear equalization, J. Lightwave Technol. 35 (19) (2017) 4105–4113.

[2] B. Wang, P. Ho, Energy-efficient routing and bandwidth allocation in OFDM-based optical networks, IEEE/OSA J. Opt. Commun. Networking 8 (2) (2016) 71–84.

[3] M. Hadi, M.R. Pakravan, Energy-efficient fast configuration of flexible transponders and grooming switches in OFDM-based elastic optical networks, IEEE/OSA J. Opt. Commun. Networking 10 (2) (2018) 90–103.

[4] F.N. Khan, et al., Modulation format identification in coherent receivers using deep machine learning, IEEE Photon. Technol. Lett. 28 (17) (2016) 1886–1889.

[5] L. Jiang, et al., Blind density-peak-based modulation format identification for elastic optical networks, J. Lightwave Technol. 36 (14) (2018) 2850–2858.

[6] J. Zhang, et al., Blind and noise-tolerant modulation format identification, IEEE Photonics Technol. Lett. (2018) 1–4.

[7] Ahmadi Negar, Using fuzzy clustering and TTSAS algorithm for modulation classification based on constellation diagram, Eng. Appl. Artif. Intell. 23 (3) (2010) 357–370.

[8] D. Wang, et al., Intelligent constellation diagram analyzer using convolutional neural network-based deep learning, Opt. Express 25 (15) (2017) 17150–17166.

[9] S. Peng, et al., Modulation classification based on signal constellation diagrams and deep learning, IEEE Trans. Neural Netw. Learn. Syst. (2018) 1–10.

[10] P. Chen, J. Liu, X. Wu, K. Zhong, X. Mai, Subtraction-clustering-based modulation format identification in stokes space, IEEE Photonics Technol. Lett. 29 (17) (2017) 1439–1442.

[11] X. Mai, et al., Stokes space modulation format classification based on non-iterative clustering algorithm for coherent optical receiver, Opt. Express 25 (3) (2017) 2038–2050.

[12] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, Science 344 (6191) (2014) 1492.

[13] J. Song, J. Zhao, F. Dong, J. Zhao, Z. Qian, Q. Zhang, A novel regression modeling method for PMSLM structural design optimization using a distance-weighted KNN algorithm, IEEE Trans. Ind. Appl. 54 (5) (2018) 4198–4206.