Expert Systems With Applications 90 (2017) 13–22

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Blind signal modulation recognition through clustering analysis of constellation signature

Gaurav Jajoo[a,*], Yogesh Kumar[a], Sandeep Kumar Yadav[b], Bibhas Adhikari[c], Ashok Kumar[d]

[a] Department of EE, Indian Institute of Technology Jodhpur, Residency Road, Ratanada, Jodhpur 342001, Rajasthan, India
[b] Electrical Engineering, Indian Institute of Technology Jodhpur, Residency Road, Ratanada, Jodhpur 342001, Rajasthan, India
[c] Department of Mathematics, Indian Institute of Technology Kharagpur, India
[d] DRDO Delhi, India

ARTICLE INFO

Article history:
Received 28 December 2016
Revised 19 June 2017
Accepted 29 July 2017
Available online 5 August 2017

Keywords:
K-means clustering
Linear regression
OPTICS
Blind signal

ABSTRACT

Automatic recognition of digital modulation schemes is becoming an active research area in many covert operations. It has many military applications where surveillance and electronic warfare requires a type of modulation in intercepted signal to prepare jamming signals. Most of the approaches are based on modulated signal's component, but the modulation type can be best identified with the use of constellation diagram. The proposed technique is able to recognize M-QAM, M-ASK, and M-PSK modulation scheme in Additive White Gaussian Noise (AWGN) environment. As the constellation points form clusters in the I-Q plane, the order of the modulation can be obtained by estimating the correct number of clusters, which is calculated by OPTICS algorithm. The least square error has been calculated using linear regression from the obtained constellation points, to identify either ASK or PSK and QAM. The error is least for ASK which differentiates ASK from PSK and QAM. To identify between the PSK and QAM, k-means clustering is employed to find the number of centroids equal to order of modulation estimated by OPTICS. With the difference in maximum and minimum absolute value of the centroids, PSK or QAM is recognized. The proposed method shows an improvement in the classification accuracy which reaches 100% using 1024 symbols at 20 dB compared to 98.89%, 98.05%, and 98% when using more complex classifiers like Support Vector Machine, Naive Bayes Classifier, KNN respectively. The method used is unsupervised whereas most of the methods in the literature require training phase to set the thresholds or weights for final model to detect modulation type. This algorithm is also implemented in LabVIEW, and tested on real-time signals. An intelligent system is made which does not require any knowledge of symbol rate, carrier frequency, and any training phase to set thresholds, and detects the type of modulation blindly in real time. Modulated RF signals are generated by NI PXIe-5673 (RF transmitter). NI PXI 5600 is used to downconvert RF signal and NI PXI-5142 (100 MS/s OSP digitizer) is used to sample the downverted signal.

1. Introduction

Automatic modulation classification is the process to classify the modulation format of a signal corrupted by noise and fading channel. It is an intermediate step between signal detection and data demodulation and has significant roles in many fields of military, civil and security applications where friendly signals should be securely transmitted and received, whereas hostile signals must be identified and jammed (Dobre, Abdi, Bar-Ness, & Su, 2007). It is used in link adaptation where modulation used at transmitter adapts according to channel in order to increase transmission reliability. If AMC system is deployed at receiver, it can significantly increase the data throughput and spectrum efficiency by saving bits in each frame containing information of transmitted modulation type. Modulation Identification carries paramount importance in universal Modems and is a precursor to FEC decoding etc. Therefore robust blind modulation identification acts as a foundation of any Communications Intelligence (COMINT) system. AMC is

* Corresponding author. URL: http://home.iitj.ac.in/~sy (Sandeep Kumar Yadav), http://www.facweb.iitkgp.ernet.in/~bibhas/ (Bibhas Adhikari).

E-mail addresses: jajoo.1@iitj.ac.in (G. Jajoo), kumar.33@iitj.ac.in (Y. Kumar), sy@iitj.ac.in (S.K. Yadav), bibhas@maths.iitkgp.ernet.in (B. Adhikari), ak.drdo@gmail.com (A. Kumar).

http://dx.doi.org/10.1016/j.eswa.2017.07.053
0957-4174/

currently an important research area in the design of intelligent radio systems, including Cognitive Radio and Software Defined Radio (SDR).

AMC involves two steps: First, preprocessing of the received signal in which different parameters are estimated like carrier frequency, symbol rate, channel state information, timing and waveform recovery, etc. and the second step is the algorithm to classify the modulation format (Dobre et al., 2007; Hameed, Dobre, & Popescu, 2009). For the latter step, algorithm is broadly grouped into two categories: Likelihood-Based Classifiers and Feature Based methods. Likelihood-Based (LB) approach requires perfect knowledge of channel to give the optimal performance and has high computational complexity compared to feature based approach. In LB method, the likelihood function is calculated for every modulation scheme under consideration and compared to conclude the type of modulation. Maximum Likelihood Estimation (MLE) calculates the likelihood values for modulation hypotheses with all the parameters of the signal known but the modulation format, and classifies the signal to the modulation hypothesis with the maximum value of likelihood function. But how-ever, MLE cannot handle any unknown parameter and to overcome this limitation, Average Likelihood Ratio Test (ALRT) was first used by Polydoros and Kim, 1990 in which the unknown parameter is replaced by its integral in the likelihood function for all of its values. ALRT gets complex when unknown parameters are introduced and also if the model considered is not accurate, the classification becomes suboptimal. Hence Generalized Likelihood Ratio Test (GLRT) is used in which unknown parameters are replaced by Maximum Likelihood Estimates i.e. likelihood function is maximized with respect to unknown parameters, which is computationally less complex. In general, GLRT is a biased classifier in case of nested modulation type like QPSK, 8-PSK, 16-QAM, 64-QAM, thus, Hybrid Likelihood Ratio Test is introduced in which some parameters are handled by ALRT and remaining by GLRT (Panagiotou, Anastasopoulos, & Polydoros, 2000).

Feature Based Methods provide near optimal performance with reduced computational complexity (Dobre et al., 2007). Feature Based method is divided into two steps; first is the feature extraction, in which key features are extracted and the second is pattern recognizer, which identifies modulation type based on the features extracted. Various spectral based features are used by Nandi and Azzouz to classify digital modulation schemes. The features exploit the unique characteristics of different signal modulations in three key signal aspects, namely the amplitude, phase, and frequency. Wavelet Transform based features identify the transients at the phase change during symbol change and also give the time and frequency information, based on which modulation scheme is identified (Ho, Prokopiw, & Chan, 2000; Hong & Ho, 1999). Higher order statistics (cumulants and moments) based features are efficient tools for detection of spectrally equivalent modulations like MPSK and MQAM whose mean and second order statistics are same (Dobre, Bar-Ness, & Su, 2003; Marchand, Lacoume, & Le Martret, 1998; Wu, Saquib, & Yun, 2008). Fourth order cumulants has been used by Swami and Saddler to classify MPSK, MPAM, and MQAM while higher order (up to eighth order) is used in Swami and Sadler, (2000); Wu et al. (2008).

In the feature-based modulation detection, decision making is based on a decision tree. At each level, different features are used and threshold values are optimized which is not very efficient. To overcome such problems, Machine Learning techniques have been employed to identify the modulation type. These are easier to implement and computationally efficient. The feature extraction subsystem calculates the prominent characteristics from the raw data. These features are given to pattern recognizer system which uses machine learning tools viz. KNN classifier, Support Vector Machine (SVM), Neural Network, etc. to classify the signal. Nandi and Azzouz have used Artificial Neural Network along with spectral based features and further Wong and Nandi used ANN with Genetic Programming for modulation classification (Wong & Nandi, 2001). Neural Network does not use the decision tree, instead, it uses nonlinear mapping of features to modulation type. Genetic Algorithm (GA) has been used along with ANN to reduce the dimensionality of feature space (Aslam, Zhu, & Nandi, 2012). Cyclostationary property of linear modulated signal is another feature to detect the modulation type (Dobre & Hameed, 2006). These features along with ANN, LDA, SVM, and KNN are used to classify different modulation schemes (Aslam, Zhu, & Nandi, 2010) and their performances are compared in Satija, Manikandan, and Ramkumar (2014). Overview of feature based classification methods have been given in Hazza, Shoaib, Alshebeili, and Fahad (2013). Other classical approaches, new trends, and comparison of all techniques are given in Dobre et al. (2007).

AMC becomes a challenging task particularly in a noncooperative environment when no prior information of the signal is available. Most of the approaches require channel information or SNR information like in Abdelmutalab, Assaleh, and El-Tarhuni (2016) to set thresholds else their probability of correct classification drops. Also, most of the approaches are not tested practically on real time signal and consider specific order modulation schemes. The proposed approach solves all these problems and has the following advantages:

- It provides higher accuracy of classification compared to most of the approaches in the literature.
- It classifies the modulation scheme without any prior information of SNR. It does not need any training phase to set thresholds.
- The proposed algorithm is also implemented on hardware and tested on real-time signals.
- Method is generalized for any order of modulation in the pool of PSK, QAM, and ASK.

In this paper, Order of modulation has been calculated by Ordering Points To Investigate the Clustering Structure (OPTICS) clustering algorithm, and results show that the proposed algorithm works with 100% accuracy above 10 dB SNR for lower order modulation schemes considered when there is no symbol timing error in the signal received through Additive White Gaussian Noise (AWGN) channel.

The paper is organized in 6 sections. In Section 2, Signal Model and System Model considered in the paper have been explained. In Section 3, the algorithm proposed for the modulation recognition using cluster analyses is explained. Implementation of the algorithm on real time signal is explained in Section 4 followed by the results and discussion in Section 5. We conclude the paper in Section 6.

2. Signal model and system model

It is assumed that the signal is sampled without timing error from rectangular pulses (in simulation). Modulation recognition is a step between the detection of low-level energy signal and full demodulation. In a noisy channel, even if there is a phase error, constellation shape remains unchanged and algorithm is able to classify the modulation type on this basis. Fig. 1 shows the system model to extract the constellation signature from a received signal (Haykin, 2008; Mobasseri, 2000).

Constellation is deformed due to channel noise, carrier frequency, and symbol time estimation error. Channel considered in the present paper is AWGN. Probability distribution function of the Gaussian noise is given by

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{1}$$

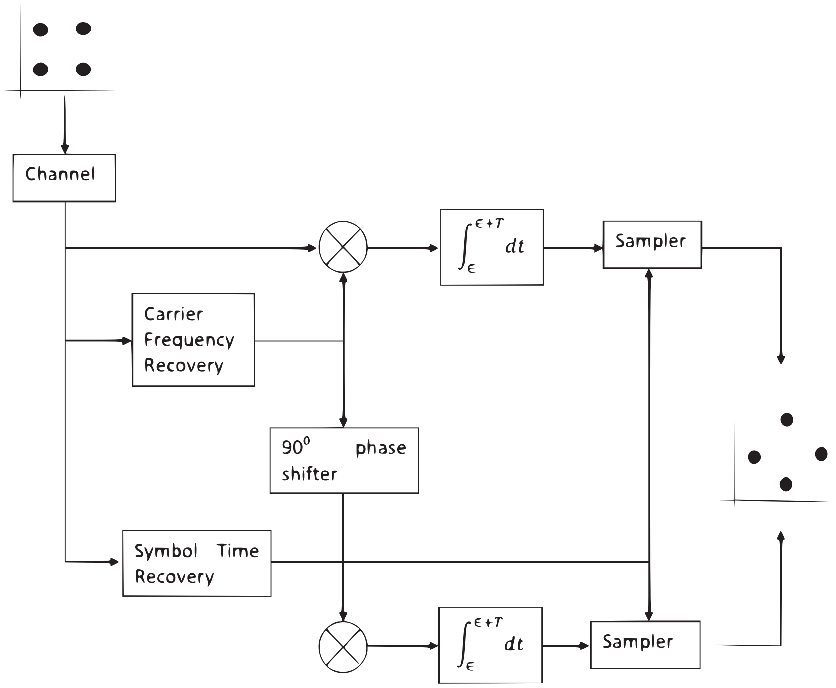

Fig. 1. Method to obtain constellation signature.

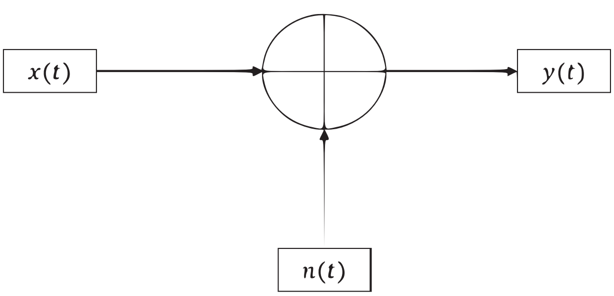

Fig. 2. AWGN noise model.

Where μ is mean and σ is the standard deviation associated with the distribution. Fig. 2 shows the model of AWGN. Thus

$$y(t) = x(t) + n(t); \quad 0 \le t \le NT_S \tag{2}$$

Where y(t) is the received noisy signal, x(t) is the modulated signal, and n(t) is Additive White Gaussian Noise, N is number of symbols transmitted and T_s is symbol time. Eq. (1) and Eq. (2) have been taken from Azzouz and Nandi (1995). Constellation diagrams for various modulation schemes are given in Haykin (2008).

For recovery of the constellation, the system depends on many algorithms which include baud rate estimation, carrier frequency and phase recovery and many other necessary channel equalization (Mobasseri, 2000). Cyclostationary based features gives symbol rate estimate (Phukan & Bora, 2014). Many algorithms for blind carrier estimation are also available in Yu, Shi, and Su (2004); Zhang and Wang (2010).

3. Methodology

3.1. Order of modulation calculation by ordering points to investigate the clustering structure (OPTICS)

To calculate modulation order, OPTICS clustering has been used. It returns three values of all the objects i.e. Core Distance, Reachability Distance (RD) and the Order in which data points are processed. By the graph obtained between RD and Order of processing of data points, the order of modulation is estimated.

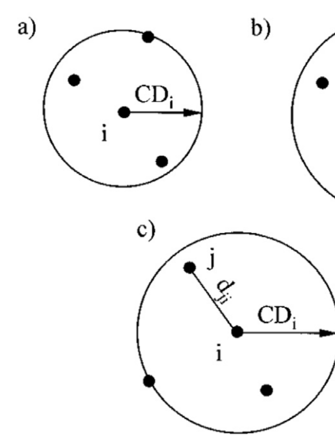

Fig. 3. For minpoints = 3,(a) core distance of the i_{th} object (CD_i) (b) Core distance of p_{th} object (c) Reachability distance of j_{th} object (RD_j), RD_j = max (d_{ji}, CD_i).

To clearly understand OPTICS, following definitions are needed:

Core Distance (CD): Core distance of any i^{th} point is the minimum distance of the radius from the data point i such that, k (minpoints) points come under the circle drawn from data point i.

Reachability Distance (RD): Reachability Distance of any j^{th} object is maximum of the distance between j^{th} object and nearest object and the core distance of the nearest object.

Fig. 3 shows CD and RD for minpoints equal to 3. OPTICS is a density based clustering algorithm and can be considered as an extension of the Density-based spatial clustering of applications with noise (DBSCAN) algorithm. Clusters with different densities which are not well separated with each other will not be identi-

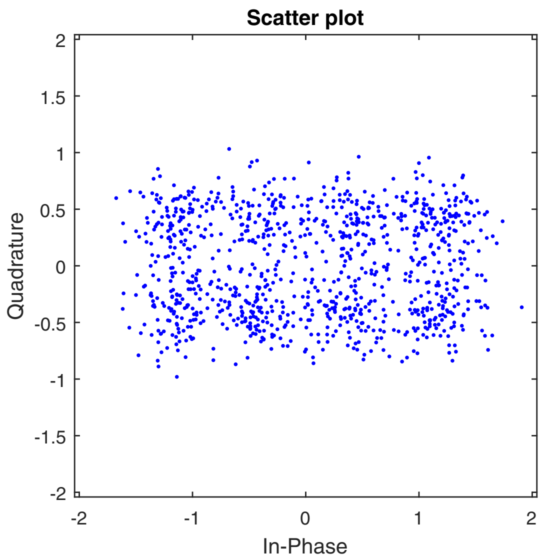

Fig. 4. 8 QAM at 10 dB SNR with 1000 constellation points.

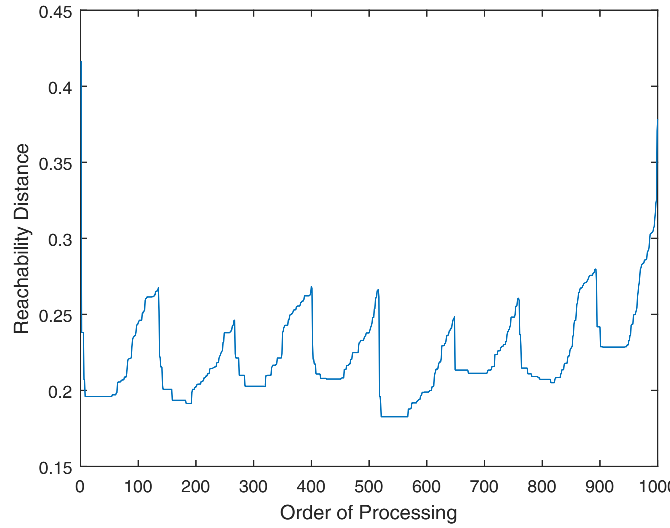

Fig. 5. Reachability distance versus order of processing of constellation points for Fig. 4.

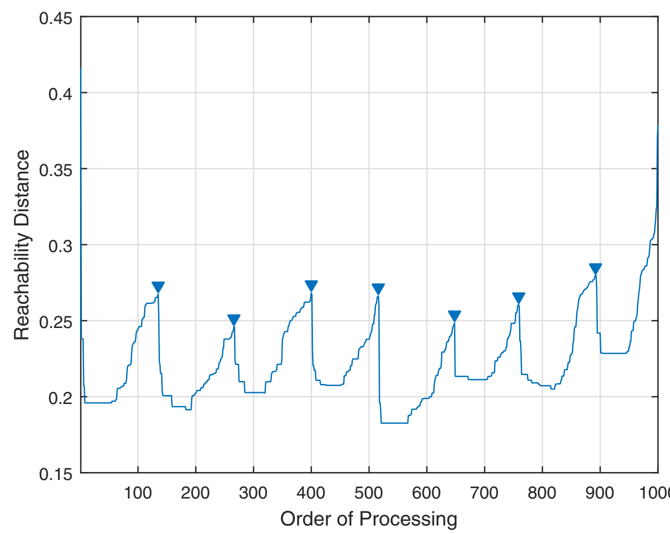

Fig. 6. Peak detection of RD versus order of processing obtained in Fig. 5.

fied as a separate cluster with DBSCAN. This problem is solved in OPTICS algorithm explained in Daszykowski, Walczak, and Massart (2002).

Steps to calculate CD, RD, and Order of processing:

Step1 Randomly first object is selected called current object. This object is marked as processed and its Reachability distance is taken as a large value.

Step2 RD for all objects is found out with respect to the current object.

Step3 In this step, data object whose RD is minimum is marked processed and is now considered as the current object. Again RD of the remaining data points is calculated with respect to this current object. If present RD of any object which is not processed is less than previous RD, then RD is updated with the new lower RD. Order of the entire processed object is saved.

Step4 Go to step 3 till all the objects are processed.

Once we get RD and Order of processing for all the data points, peaks from the graph between RD and Order of processing are calculated. The input to the OPTICS is the constellation points we have shown in Fig. 4. There are 1000 data points of 8 QAM at 10 dB SNR in AWGN channel. Fig. 5 shows the plot between RD and Order of processing with minpoints 50. Optimum value of the number of clusters is obtained by the number of peaks in the plot between RD and order of processing of objects shown in Fig. 6. The number of peaks extracted is 7 for order 8 because RD of the initial point selected is made large hence actual number of clusters will be 1 greater than the estimated number of peaks.

Sometimes in peak detection, the lower variations are also estimated as a peak. To avoid the detection of false peaks, a threshold value (threshold value is taken as 60% of the mean of two highest values) is set. Peaks greater than the threshold value are taken into consideration. As the order of modulation cannot be any arbitrary number, it can be only in the power of 2, the number of clusters (number of peaks + 1) nearest to 2^N (N is the integer which makes the value of 2^N nearest to number of peaks + 1) is considered as an actual order of modulation. All the graphs for lower modulation schemes (BPSK, QPSK, 8PSK, 8QAM, and 4ASK) are obtained with 1000 data points and minpoints (input parameter for OPTICS clustering) is taken to be 50. For higher modulation schemes (16QAM, 32QAM, 16PSK, and 8ASK), results are simulated with 2000 data points and minpoints 30.

Fig. 7 shows the method used for recognition of modulation format. The order of Modulation is calculated using Density based OPTICS clustering algorithm. After estimating the order, domain of modulation is calculated i.e. ASK, PSK or QAM.

3.2. Identification of modulation domain

Identification of the modulation type of the signal from the pool of PSK, QAM, and ASK is done using linear regression and k-means clustering. Fig. 7 shows the block diagram for the modulation domain classification and Algorithm 1 explains the process for the same. To decide between ASK or PSK and QAM, a straight line with slope b_2 and intercept b_1 is regressed on the given data points (in the present case data extracted is constellation points let's say (x_k, y_k), k = 1, 2, 3…N. Thus the regression line is given by

$$Y = b_1 + b_2 X$$

Algorithm 1
Modulation domain classification.

Calculate error ‘e’ through linear regression and normalize it with respect to length of the data points.
if $e <= e_{th}$ then
Modulation is ASK else
Calculate M number of centroids using k-means clustering algorithm where M is modulation order calculated using OPTICS.
Centroids = C_i i = 1; 2; M.
Calculate $d = \max(|C_i|) - min(|C_i|)$ i = 1, 2...M.
if $d >= d_{th}$ then
Modulation domain is QAM else
Modulation is PSK end if
end if

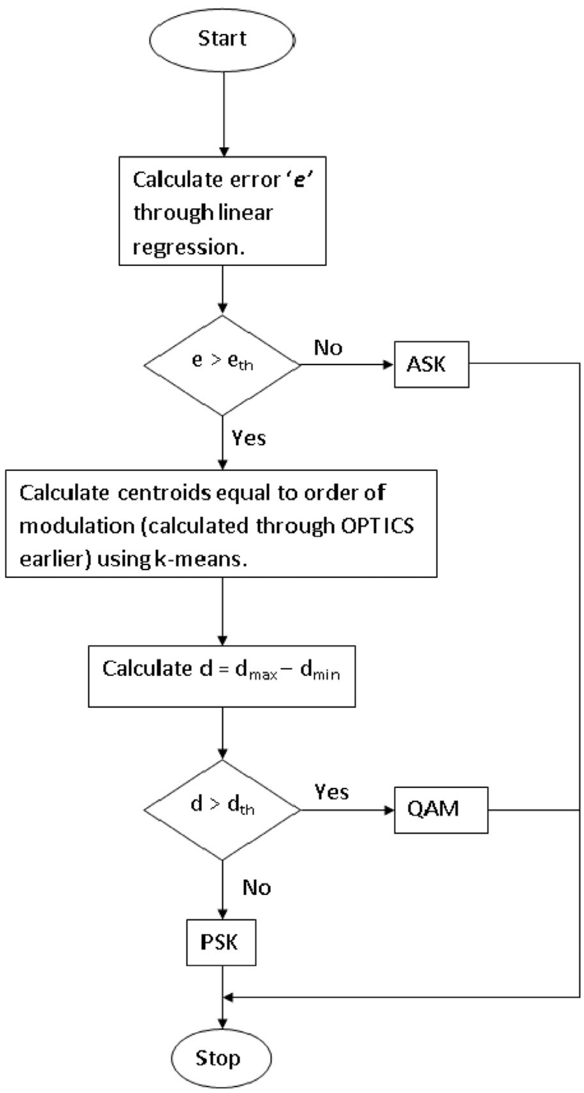

Fig. 7. Proposed algorithm for modulation domain classification.

Where, b_1 and b_2 are constants. We try to minimize the squared error given by

$$e(b_1, b_2) = \sum_{k=1}^{N} \varepsilon_k{}^2 = \sum_{k=1}^{N} (y_k - (b_1 + b_2 x_k))^2$$

Minimizing ‘e’ with respect to b_1 and b_2, the following values of slope and intercept of the regressed line are obtained.

$$b_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$
$$b_1 = \bar{y} - b_2 \bar{x}$$

Where $\bar{x}$ and $\bar{y}$ are the mean of (x_k, y_k), for k = 1, 2, 3...N respectively. b_1 and b_2 are obtained directly from the constellation points extracted and square error is calculated and divided by the number of constellation points. Fig. 8(a) and 8(b) show a line regressed on 8ASK constellation and 8QAM constellation respectively. From the figures, it is clear that error in case of ASK is less compared to QAM and the same follows for PSK as well. On this basis, ASK is differentiated from PSK and QAM.

If the value of ‘e’ falls under PSK and QAM class then, k centroids is estimated using k-means clustering algorithm where k is given by the modulation order estimated in Section 3.1. To calculate centroids using k-means, the following steps are applied:

Step1 k initial centroids i.e. seed points are randomly considered from the data points. Here k is given by Order of modulation.
Step2 Assign each data point to nearest centroid. After this, means of the assigned points in all k clusters are calculated and centroids are updated for each cluster with the calculated mean.
Step3 Repeat Step2 until no more reassignment is possible.

Once we get the final k centroids, their absolute values are calculated. Difference in Maximum absolute value (d_{max}) and Minimum absolute value (d_{min}) is calculated. Fig. 8(c) shows d_{max} and d_{min} value for PSK constellation. This difference for PSK will be very less as compared to QAM constellation shown in Fig. 8(d). From this, PSK and QAM are separated.

4. Implementation on real time signal

The methodology described in Section 3 is implemented on labVIEW and tested on hardware. For the generation of RF signal, NI PXIe-5450 (400 MS/s I/Q Signal Generator), NI PXIe-5611 (I/Q vector modulator) and NI PXIe-5652 (RF Signal generator) have been used. Onboard clock and signal with the center frequency ranging from 550 MHz to 650 MHz have been used for the data. The signal is received using NI PXI 5600 (RF downconverter) and NI PXI-5142 (100MS/s OSP Digitizer). In the first step of implementation, Carrier Frequency Off set (CFO) is estimated and corrected. Further to extract the constellation points, Symbol Rate is estimated and signal is downsampled explained in Section 4.1, 4.2 and 4.3.

4.1. Carrier frequency offset estimation and correction

If local carrier generated at receiver to downconvert the signal is not equal to the carrier frequency of the received signal, constellation points rotate from their actual positions. Hence carrier frequency offset needs to be corrected. CFO is estimated and corrected using Sethi and Ray (2013). Fig. 9(a), 9(b), and 9(c) show the power spectrum for three signals with carrier frequencies 608 MHz, 615 MHz, and 594 MHz respectively.

If $\Delta\omega$ is the CFO in the downconverted signal, n^{th} constellation point rotates with the value of $n\Delta\omega T$, where T is the sampling

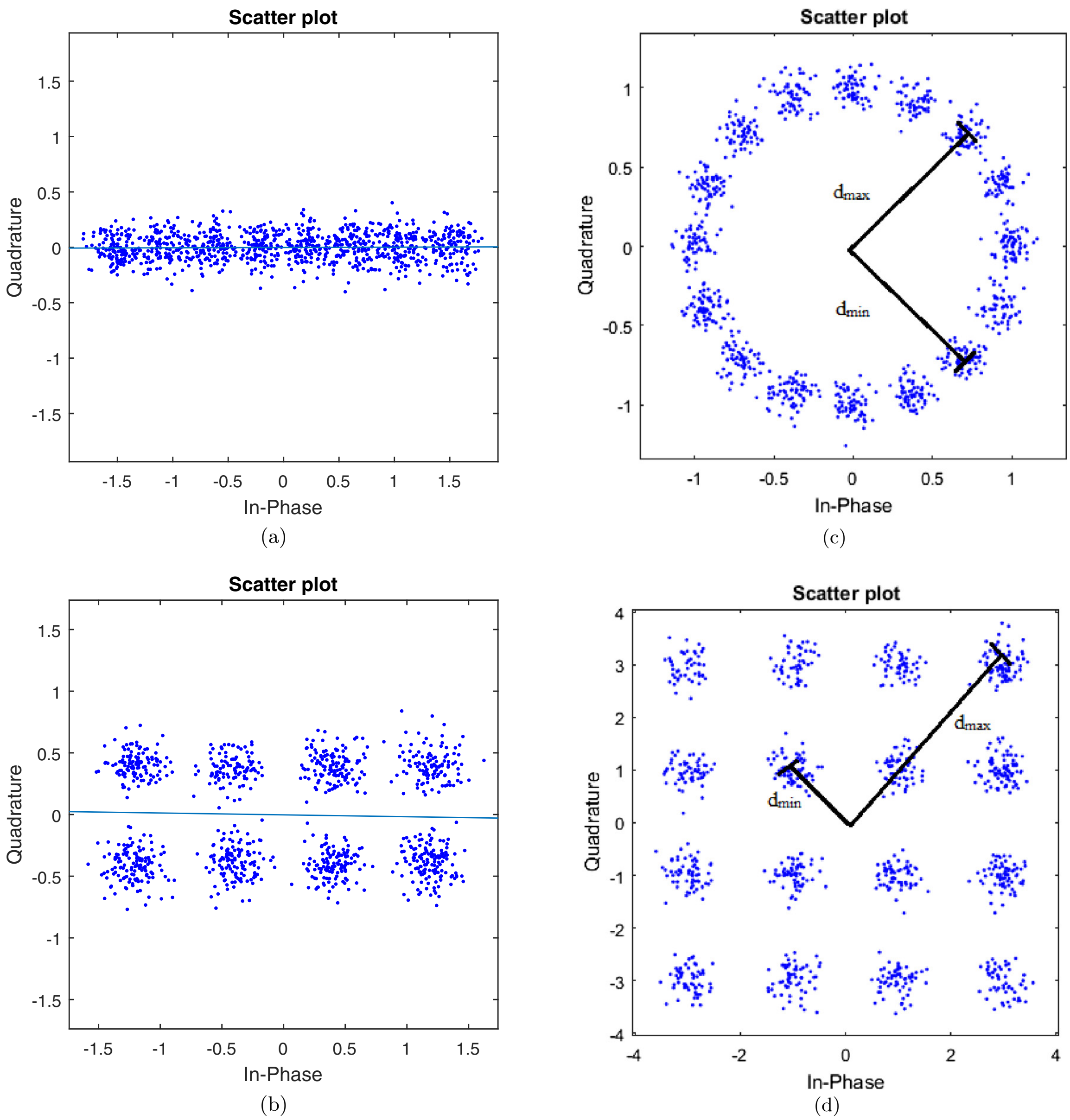

Fig. 8. (a), (b) Line regressed for 8ASK and 8QAM respectively. (c), (d) d_{max} and d_{min} for PSK and QAM respectively.

time. For calculating CFO, the downverted and sampled baseband signal is passed through eighth order nonlinearity.

$$R(\mathrm{f}) = \left|FFT(\mathrm{r}^8(\mathrm{n}))\right|$$

After calculating R(f), the frequency corresponding to the peak in the R(f) is calculated. This value of frequency is divided by 8 to get the actual CFO.

$$\Delta f = \frac{peak frequency}{8}$$
$$\Delta\omega = 2\pi\,\Delta f$$

The estimated CFO is used to correct the sampled baseband signal.

$$\begin{aligned}\mathrm{u(n)} &= \mathrm{r(n)e}^{\mathrm{j}n(-\Delta\omega)}\mathrm{u(n)} = (\mathrm{I(n)} + \mathrm{jQ(n)})(\cos(-\mathrm{n}\Delta\omega)\\ &+\mathrm{j}\sin(-\mathrm{n}\Delta\omega))\mathrm{u(n)} = \mathrm{I(n)}\cos(\mathrm{n}\Delta\omega) + \mathrm{jQ(n)}\cos(\mathrm{n}\Delta\omega)\\ &-\mathrm{jI(n)}\sin(\mathrm{n}\Delta\omega) + \mathrm{Q(n)}\sin(\mathrm{n}\Delta\omega)\mathrm{u(n)} = [\mathrm{I(n)}\cos(\mathrm{n}\Delta\omega)\\ &+\mathrm{Q(n)}\sin(\mathrm{n}\Delta\omega)] + \mathrm{j}[\mathrm{Q(n)}\cos(\mathrm{n}\Delta\omega) - \mathrm{I(n)}\sin(\mathrm{n}\Delta\omega)]\end{aligned}$$

Finally u(n) is the corrected baseband signal. In the next section, symbol rate estimation is described.

4.2. Symbol rate estimation

After estimation and correction of CFO, symbol rate is estimated so that the received signal can be downsampled to the symbol rate for extraction of the constellation points. Digitally modulated signals have an underlying symbol rate. Estimation of symbol rate is based on the received complex baseband signal. Symbol Rate estimation is done using Sethi and Ray (2013). Discrete Fourier Transform of the CFO corrected baseband signal's absolute square value is calculated and frequency corresponding to the peak in search range is set as an estimate of the symbol rate.

$$D(\mathrm{f}) = \mathrm{DFT}\left[{u_i}^2(\mathrm{n}) + {u_q}^2(\mathrm{n})\right] \tag{3}$$

Where u(n) is complex baseband signal.

$$u(\mathrm{n}) = \mathrm{u}_i(\mathrm{n}) + \mathrm{u}_q(\mathrm{n}) \tag{4}$$

Where $\mathrm{u_i(n)}$ and $\mathrm{u_q(n)}$ are real and imaginary part of u(n).

Root Raised cosine pulse shaping with roll off factor equals 0.5 is used at the transmitting end. Fig. 10(a) and 10(b) show the com-

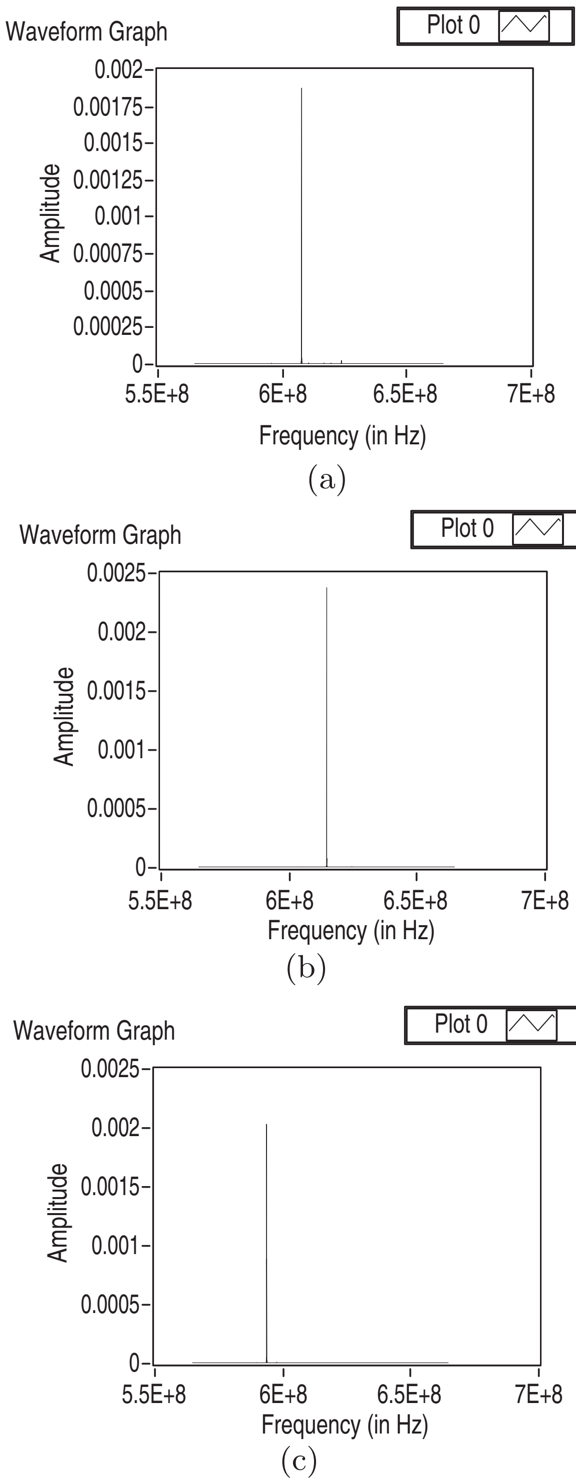

Fig. 9. Power spectrum of the received signal with carrier frequency (a) 608 MHz (b) 615 MHz (c) 594 MHz.

plex baseband In-phase component and Quadrature-phase component for QPSK with root raised cosine pulse shaping and roll off factor equal to 0.5.

Fig. 11(a) shows the graph for D(f), and 11(b) shows the same graph in frequency range $f_s/2$ and f_s obtained from Eq. (3) to estimate the symbol rate. Fig. 11(a) and 11(b) are obtained for data sampled at 1600 K, and the length L of u(n) taken is 131,056. X-axis is from 0 to f_s i.e. 0 to 1600 KHz. Last point of FFT which is 131056th point corresponds to f_s i.e. 1600 KHz. Peak next to $f_s/2$ is

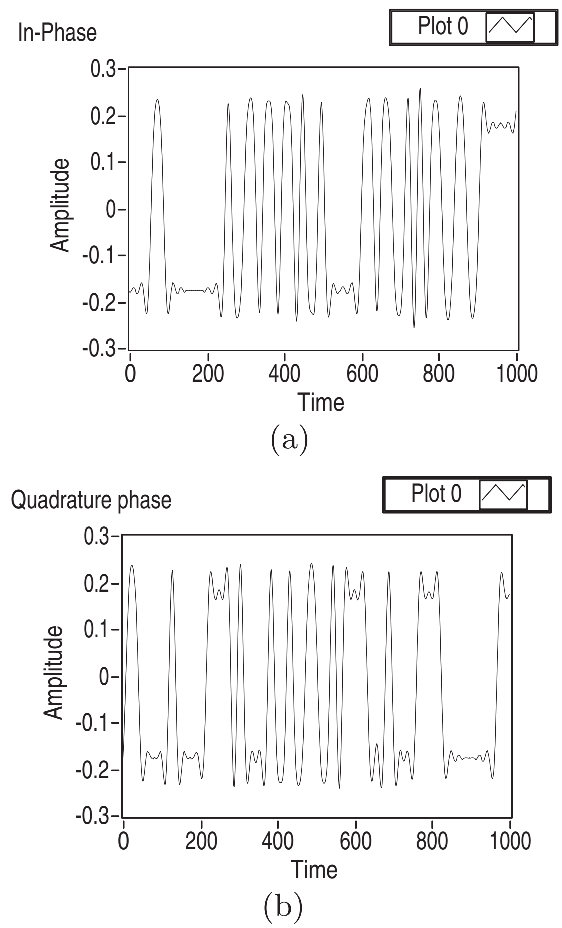

Fig. 10. Complex baseband pulse with root raised cosine pulse shaping and roll off factor equals 0.5 for QPSK (a) In-phase component (b) Quadrature phase component.

extracted which is obtained at 73719th index. After shifting D(f) by $f_s/2$ i.e. 65,528 points, new index value of peak is 8191 corresponds to frequency 100KHz which is the estimated symbol rate. After estimation of the symbol rate, constellation points are extracted as explained in next section.

$$SR_{est} = \frac{kf_s}{N}$$

Where SR_{est} is estimated symbol rate, k is 8191, f_s is 1600KHz and N is length of FFT which is 131,056.

4.3. Constellation extraction

After the CFO estimation and correction, symbol rate is estimated. The signal is down-sampled by Oversampling Factor (given in Eq. (5)) which gives one sample per symbol and constellation points for the linear modulation schemes are obtained.

$$\text{OversamplingFactor} = (\text{SamplingRate})/(\text{SymbolRate}) \tag{5}$$

Eq. (5) has been taken from Schmogrow (2014). Sampling rate used at reciever is 1600 KHz. When the estimated symbol rate is 100 KHz, oversampling factor comes out to be 16. So digitized baseband signal is downsampled by 16. After extracting the constellation points, the algorithm proposed in Section 3 has been implemented. Fig. 12 shows the constellation points obtained for different modulation schemes.

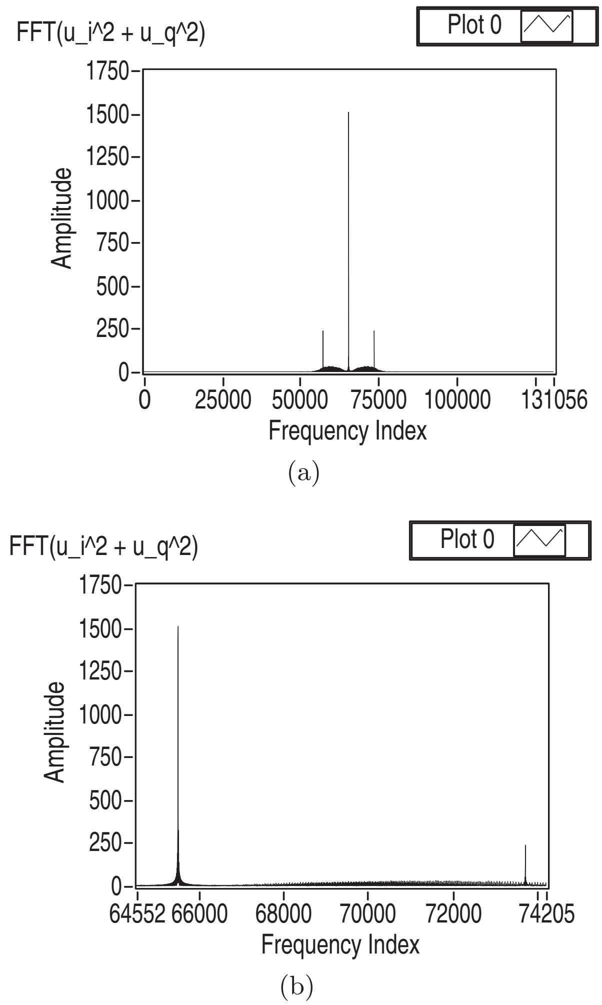

Fig. 11. (a) D(f) (b) D(f) between fs/2 and fs.

Table 1
Confusion matrix for different modulation schemes at 10 dB SNR.

	BPSK	QPSK	8-PSK	8-QAM	4-ASK
BPSK	100	0	0	0	0
QPSK	0	100	0	0	0
8-PSK	0	0	100	0	0
8-QAM	0	0	0	100	0
4-ASK	0	0	0	0	100

Table 2
Comparison of Naïve classifier and SVM with our proposed method at different values of SNR.

N	SNR (dB)	Probability of correct classification (%)		
		Naïve	SVM	Proposed
1024	14	96.66	97.89	99.5
	16	97.29	98.65	100
	18	97.76	98.93	100
	20	98.05	98.89	100
2046	14	99.43	99.04	100
	16	99.42	99.26	100
	18	99.4	99.36	100
	20	99.66	99.4	100

Table 3
Comparison of KNN, GP-KNN, GP-tree and proposed classifier.

N	SNR (dB)	Probability of correct classification (%)			
		KNN	GP-KNN	GP-Tree	Proposed
1024	15	97	99	99.31	99.6
	20	98	100	99.55	100
2048	15	100	100	-	100
	20	100	100	-	100

5. Results and discussions

In this section, the performance of the proposed method for AMC has been analyzed in AWGN channel and is compared with the existing methods in literature. Almost all the methods in the literature requires to train the classifier for setting the thresholds and weights for training models but in our proposed approach, no weights and thresholds have to be set and it works without any information of the SNR.

5.1. Proposed classifier

In this section, results of the proposed method for classifying various modulation schemes are explained. All the simulations are performed in MATLAB. In Fig. 13, a classification problem among BPSK, QPSK, 8PSK, 8QAM, and 4ASK is considered under AWGN channel from 5 dB to 10 dB. It is simulated for 1000 symbol points with the value of minpoints (input parameter of the OPTICS clustering algorithm) as 50. It is observed that BPSK and QPSK are classified with 100% accuracy above 6 dB SNR. 8PSK, 8QAM, and 4ASK are classified reliably above 9 dB SNR. Similarly, Fig. 14 is showing the accuracy of 4 modulation scheme i.e. 8ASK, 16PSK, 16QAM, and 32QAM. Simulation results of these schemes are found out with 2000 symbols and minpoints value 30, as 32 order modulation scheme is considered with all the symbols being equiprobable. It is observed that 16QAM, and 32QAM are accurately classified above 14 dB SNR while 16PSK is classified accurately above 16 dB. ASK uses only one dimension for symbols hence its accuracy is less compared to other schemes and reaches 100% above 17 dB SNR. Table 1 shows the confusion matrix of BPSK, QPSK, 8PSK, 8QAM, and 4ASK. It shows that at 10 dB SNR, all the modulation schemes under consideration are recognized correctly.

5.2. Comparison to related work

Table 2 compares the classification probability of Naive Bayes Classifier (Wong, Ting, & Nandi, 2008), Support Vector Machine (Wong et al., 2008) and the proposed method. Naive Bayes classifier and SVM are given for 4 classes i.e. BPSK, QPSK, 16QAM, and 64QAM. Hence for fair comparison, the results for the same configuration are calculated. For 1024 symbols, the value of minpoints taken is 9 as 64 order modulation is involved. For 2046 symbols, value of minpoints taken is 17. Table 2 shows that Naive Bayes and SVM classifies with accuracy of 96.66% and 97.89% respectively while the proposed method classifies with an accuracy of 99.5% at 14 dB SNR. On increasing the number of symbols to 2046 at 14 dB, Naive Bayes classifier classifies with 99.43% while the proposed method achieved the classification accuracy of 100%. This shows that the new method has outperformed the Naive Bayes classifier and SVM classifier. It is also shown in Aslam et al. (2012) that cumulants based classification for these four classes achieve 98.2% accuracy at 20 dB (compared to 100% using the proposed method), 87.1% at 15 dB (compared with 100% in our approach). These results clearly show the improvement in the classification accuracy.

Table 3 compares the classification accuracy of KNN, GP-KNN, GP-Tree, taken from Aslam et al. (2012) and our method. Proposed

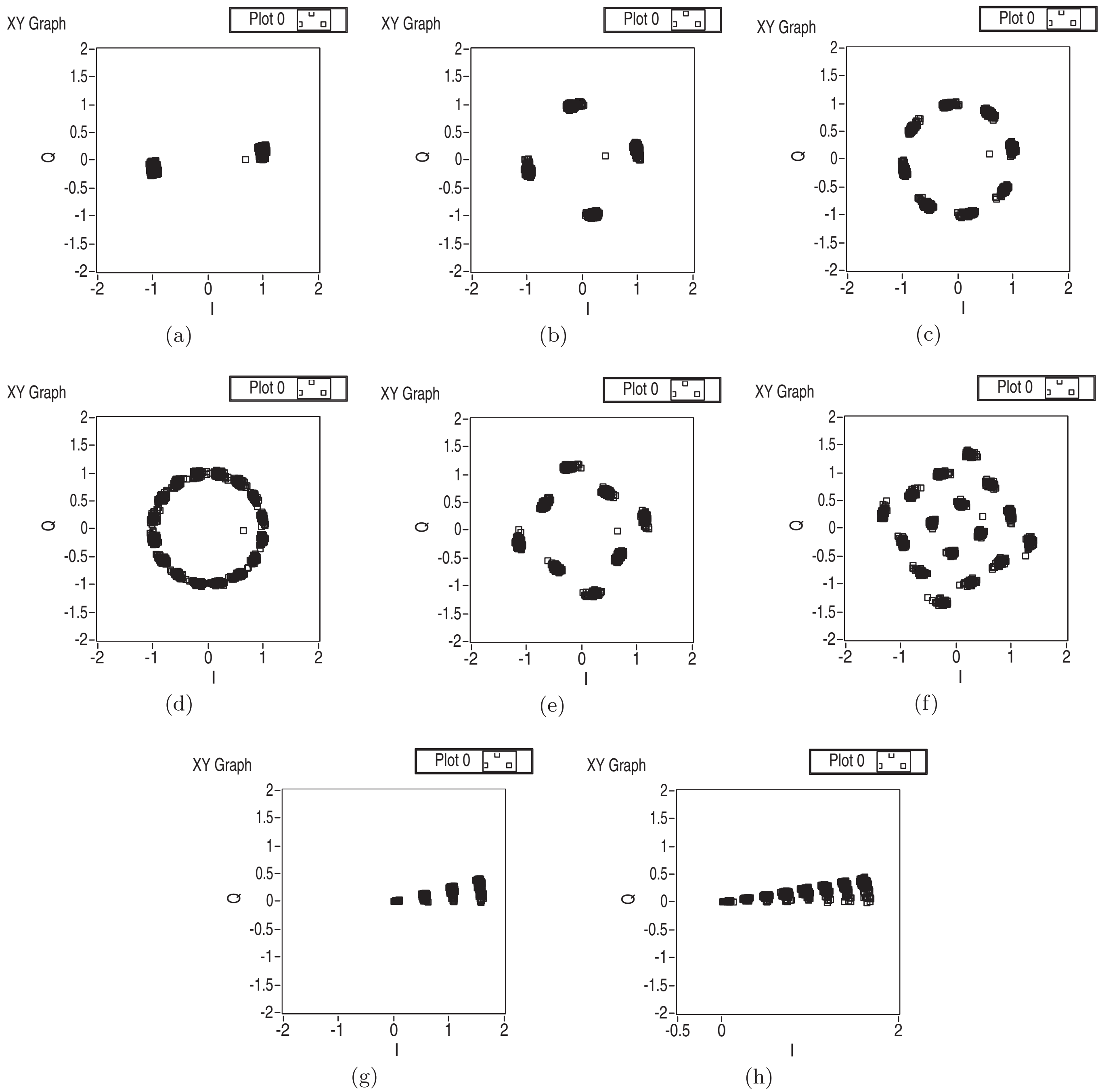

Fig. 12. Acquired Data points for (a) BPSK (b) QPSK (c) 8 PSK (d) 16 PSK (e) 8QAM (f) 16 QAM (g) 4ASK (h) 8 ASK.

classifier is simulated for the same configuration used in Aslam et al. (2012). When the number of symbols is 1024, KNN, GP-KNN and GP-Tree achieve 97%, 99% and 99.31% accuracy respectively while proposed method achieves 99.6% accuracy at 15 dB SNR. From Table 3, it is clear that our method achieves better accuracy compared to KNN, GP-KNN, and GP-Tree. Other then the accuracy, to provide a final training model, GP-KNN and SVM require iterative approaches which consume significant processing time as compared to our approach which works without any training. Mirarab and Sobhani (2007) achieved 95% accuracy at 15 dB SNR with 2000 number of symbols while we achieved 100% accuracy.

Sometimes using cumulants and higher order statistics, separation of modulation scheme is not possible as there is no optimum threshold defined like for 16QAM and 64QAM explained in Abdelmutalab et al. (2016). In the proposed algorithm, the modulation order calculated through OPTICS returns the number of clusters nearest to 2^N. Therefore, it can classify any higher order modulation scheme like 64QAM, 128QAM, 256QAM, etc. For example if the transmitted modulation is 256QAM, and the estimated number of peaks is in between 193 and 383, it will give order as 256. So the margin of error is very less as far as constellation extracted is good. In practical implementation of algorithm, if value of OversamplingFactor is non integer, the signal is resampled with the integer multiple of symbol rate and then downsampled.

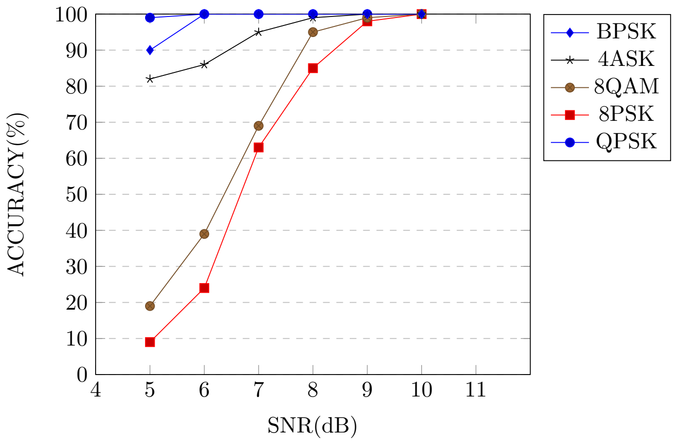

Fig. 13. Accuracy versus SNR for different modulation schemes.

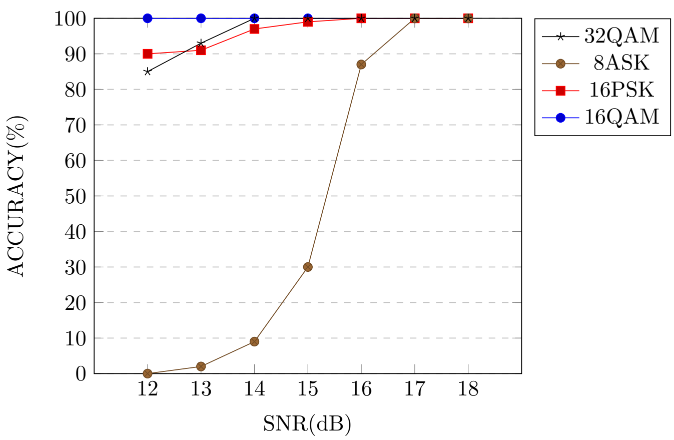

Fig. 14. Accuracy versus SNR for higher order modulation schemes.

6. Conclusion

In this paper, an autonomous method is proposed to classify MQAM, MASK, and MPSK. The method has low computational complexity. By virtue of the results obtained, BPSK and QPSK are accurately estimated above 6 dB SNR while 4ASK, 8QAM, and 8PSK are estimated reliably above 9 dB SNR. This makes the algorithm efficient in practical communication systems. The results show that higher order modulation schemes (16QAM, 32QAM, 16PSK, and 8ASK) are classified correctly above 17 dB SNR in AWGN channel. The method can recognize modulation schemes of ASK, PSK, and QAM for any order.

Acknowledgment

This project is funded by Defense Research and Development Organization under project number S/DRDO/SKY/20150004.

References

Abdelmutalab, A., Assaleh, K., & El-Tarhuni, M. (2016). Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers. *Physical Communication, 21*, 10–18.

Aslam, M. W., Zhu, Z., & Nandi, A. K. (2010). Automatic digital modulation classification using genetic programming with k-nearest neighbor. In *Military communications conference, 2010-MILCOM 2010* (pp. 1731–1736). IEEE.

Aslam, M. W., Zhu, Z., & Nandi, A. K. (2012). Automatic modulation classification using combination of genetic programming and knn. *IEEE Transactions on Wireless Communications, 11*, 2742–2750.

Azzouz, E. E., & Nandi, A. K. (1995). Automatic identification of digital modulation types. *Signal Processing, 47*, 55–69.

Daszykowski, M., Walczak, B., & Massart, D. L. (2002). Looking for natural patterns in analytical data. 2. tracing local density with optics. *Journal of Chemical Information and Computer Sciences, 42*, 500–507.

Dobre, O. A., Abdi, A., Bar-Ness, Y., & Su, W. (2007). Survey of automatic modulation classification techniques: Classical approaches and new trends. *IET Communications, 1*, 137–156.

Dobre, O. A., Bar-Ness, Y., & Su, W. (2003). Higher-order cyclic cumulants for high order modulation classification. In *Military communications conference, 2003. MILCOM'03. 2003 IEEE: 1* (pp. 112–117). IEEE.

Dobre, O. A., & Hameed, F. (2006). Likelihood-based algorithms for linear digital modulafition classification in fading channels. In *2006 Canadian conference on electrical and computer engineering* (pp. 1347–1350). IEEE.

Hameed, F., Dobre, O. A., & Popescu, D. C. (2009). On the likelihood-based approach to modulation classification. *IEEE Transactions on Wireless Communications, 8*, 5884–5892.

Haykin, S. (2008). *Communication systems*. John Wiley & Sons.

Hazza, A., Shoaib, M., Alshebeili, S. A., & Fahad, A. (2013). An overview of feature-based methods for digital modulation classification. In *Communications, signal pro-cessing, and their applications (ICCSPA), 2013 1st international conference on* (pp. 1–6). IEEE.

Ho, K., Prokopiw, W., & Chan, Y. (2000). Modulation identification of digital signals by the wavelet transform. *IEEE Proceedings-Radar, Sonar and Navigation, 147*, 169–176.

Hong, L., & Ho, K. (1999). Identification of digital modulation types using the wavelet transform. In *Military communications conference proceedings, 1999. MILCOM 1999. IEEE: 1* (pp. 427–431). IEEE.

Marchand, P., Lacoume, J.-L., & Le Martret, C. (1998). Multiple hypothesis modulation classification based on cyclic cumulants of different orders. In *Acoustics, speech and signal processing, 1998. Proceedings of the 1998 IEEE international conference on: 4* (pp. 2157–2160). IEEE.

Mirarab, M., & Sobhani, M. (2007). Robust modulation classification for psk/qam/ask using higher-order cumulants. In *Information, communications & signal processing, 2007 6th international conference on* (pp. 1–4). IEEE.

Mobasseri, B. G. (2000). Digital modulation classification using constellation shape. *Signal Processing, 80*, 251–277.

Panagiotou, P., Anastasopoulos, A., & Polydoros, A. (2000). Likelihood ratio tests for modulation classification. *MILCOM 2000. 21st century military communications conference proceedings*: 2. IEEE 670{674).

Phukan, G. J., & Bora, P. K. (2014). An algorithm for blind symbol rate estimation using second order cyclostationarity. In *2014 international conference on signal processing and communications (SPCOM)* (pp. 1–6). IEEE.

Satija, U., Manikandan, M., & Ramkumar, B. (2014). Performance study of cyclostationary based digital modulation classification schemes. In *2014 9th international conference on industrial and information systems (ICIIS)* (pp. 1–5). IEEE.

Schmogrow, R. M. (2014). *Real-time digital signal processing for Software-defined optical transmitters and receivers*: vol. 13. KIT Scientific Publishing.

Sethi, A., & Ray, B. (2013). Blind carrier/timing recovery and detection of modulation scheme. US Patent 8,605,830.

Swami, A., & Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications, 48*, 416–429.

Wong, M. D., & Nandi, A. K. (2001). Automatic digital modulation recognition using spectral and statistical features with multi-layer perceptrons. In *Signal processing and its applications, sixth international, symposium on. 2001: 2* (pp. 390–393). IEEE.

Wong, M. D., Ting, S. K., & Nandi, A. K. (2008). Naive bayes classification of adaptive broadband wireless modulation schemes with higher order cumulants. In *Signal processing and communication systems, 2008. ICSPCS 2008. 2nd international conference on* (pp. 1–5). IEEE.

Wu, H.-C., Saquib, M., & Yun, Z. (2008). Novel automatic modulation classification using cumulant features for communications via multipath channels. *IEEE Transactions on Wireless Communications, 7*, 3098 {3105.

Yu, Z., Shi, Y. Q., & Su, W. (2004). A blind carrier frequency estimation algorithm for digitally modulated signals. In *Military communications conference, 2004. MILCOM 2004. 2004 IEEE: 1* (pp. 48–53). IEEE.

Zhang, L., & Wang, J. (2010). A fast blind carrier frequency estimation algorithm for digitally modulated signals. In *2010 6th international conference on wireless communications networking and mobile computing (WiCOM)* (pp. 1–3). IEEE.