Specific Emitter Identification via Hilbert–Huang Transform in Single-Hop and Relaying Scenarios

Jingwen Zhang, *Student Member, IEEE*, Fanggang Wang, *Member, IEEE*, Octavia A. Dobre, *Senior Member, IEEE*, and Zhangdui Zhong

***Abstract*— In this paper, we investigate the specific emitter identification (SEI) problem, which distinguishes different emitters using features generated by the nonlinearity of the power amplifiers of emitters. SEI is performed by measuring the features representing the individual specifications of emitters and making a decision based on their differences. In this paper, the SEI problem is considered in both single-hop and relaying scenarios, and three algorithms based on the Hilbert spectrum are proposed. The first employs the entropy and the first- and second-order moments as identification features, which describe the uniformity of the Hilbert spectrum. The second uses the correlation coefficient as an identification feature, by evaluating the similarity between different Hilbert spectra. The third exploits Fisher's discriminant ratio to obtain the identification features by selecting the Hilbert spectrum elements with strong class separability. When compared with the existing literature, we further consider the identification problem in a relaying scenario, in which the fingerprint of different emitters is contaminated by the relay's fingerprints. Moreover, we explore the identification performance under various channel conditions, such as additive white Gaussian noise, non-Gaussian noise, and fading. Extensive simulation experiments are performed to evaluate the identification performance of the proposed algorithms, and results show their effectiveness in both single-hop and relaying scenarios, as well as under different channel conditions.**

***Index Terms*— Feature extraction, Hilbert spectrum, relay, specific emitter identification (SEI).**

I. Introduction

SPECIFIC emitter identification (SEI) is the process of discriminating individual emitters by comparing the features carried by the received signal with a categorized feature set, and choosing the class that best matches these features [2]. SEI has been intensively studied for military communications [3]; recently, it has become increasingly important with the advent of new technologies, such as cognitive radio [4] and self-organized networking [5].

Based on the operation mode of the emitter, SEI is applicable either to the transient signal or to the steady-state signal. The transient signal, commonly known as the turn-on signal, provides unique and distinguishable characteristics suited for feature extraction and emitter identification [6], [7]. To measure the transient features, a main approach is to extract the transient signal from noise by detecting its starting and ending points [8]. Basically, the transient features are emitter-specific and consistent, which is advantageous for identification; nevertheless, they are difficult to be captured since the duration of the transient signal is extremely short. Furthermore, the transient features are easily hampered by non-ideal and complicated channel conditions, which may negatively affect the identification results [9].

The steady-state signal is transmitted by emitters operating under stable conditions. Although the steady-state features tend to be corrupted by the transmitted information, leading to difficulties in extracting them, the investigation of the steady-state signal has considerable practical implications, as this is easily detected and captured. A number of feature extraction schemes have been studied in such a case, such as bispectrum [10] and bio-inspired algorithm [11], etc. Among those schemes, the most widely used approaches rely on the time-frequency representation-based feature. A time-frequency representation maps the signal onto a two-dimensional plane of time and frequency, which provides the temporal and spectral information simultaneously. In [12], a signal detection and identification system is proposed based on the short-time Fourier transform (STFT). However, STFT is a linear transform which cannot be adopted to analyze a non-linear signal. A similar radar waveform identification algorithm is presented in [13] utilizing the Wigner and Choi-Williams distributions. Another class-dependent scheme is presented in [14], which employs smoothing regular quadratic time-frequency representations to extract features for radar emitter identification. In [15], an SEI scheme employing the quadratic time-frequency representation together with a sequential classifier is adopted. However, the difficulty with the quadratic time-frequency representations is the inevitable cross-terms problem. A new method, namely the Hilbert-Huang transform (HHT), is developed for analyzing non-linear and non-stationary signals [16]. Note that since HHT offers sharper results than the traditional methods, in this paper

Manuscript received July 12, 2015; revised October 27, 2015 and December 21, 2015; accepted January 6, 2016. Date of publication January 22, 2016; date of current version March 16, 2016. This work was supported in part by the National Natural Science Foundation under Grant 61571034, Grant U1334202, and Grant 61501390, in part by the Fundamental Research Funds for the Central Universities under Grant 2015JBM112, in part by the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2016ZT013 and Grant RCS2014ZT09, and in part by Shenzhen Science and Technology Innovation Commission (SZSTI), Basic Research Project No. JCYJ20150507154758604. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athanasios Vasilakos. *(Corresponding Author: Fanggang Wang.)*

J. Zhang, F. Wang, and Z. Zhong are with the Sate Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: 12111009@bjtu.edu.cn; wangfg@bjtu.edu.cn; zhdzhong@bjtu.edu.cn).

O. A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3R5, Canada (e-mail: odobre@mun.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2016.2520908

we employ the time-frequency distribution provided by the HHT.

Furthermore, the majority of the current SEI literature concentrates on single-hop communication scenarios; however, relays are widely deployed in communication systems [17]. For instance, in satellite communications, the satellites that relay and amplify the signals between the emitters and receivers are adopted for a variety of applications, such as television, telephone, internet, wireless and miliary communications. In such a case, the features carried by the received signal are the mixture of the fingerprints of the emitter and relay; thus, the latter contaminates the features of the emitters, which leads to a negative effect on SEI. Recently, we have proposed the entropy, and first- and second-order moments (EM^2) algorithm to address the SEI problem in single-hop and relaying scenarios [1]. Despite its advantage of effectively identifying emitters in the relaying scenario, the algorithm is developed for the additive white Gaussian noise (AWGN) channel, while SEI under non-ideal conditions, such as non-Gaussian noise and fading, remains a challenge.

In this paper, we propose three approaches to SEI and study them in various channels under the single-hop and relaying scenarios. The main contributions are summarized as follows:

- We propose an algorithm, referred to as the EM^2 algorithm, which relies on the energy entropy, and first- and second-order moments of the Hilbert spectrum as identification features, to evaluate the uniformity of the Hilbert spectrum. This algorithm is also proposed in our previous conference paper [1].
- A correlation-based (CB) algorithm is additionally proposed, which exploits the correlation coefficient to measure the similarity between different Hilbert spectra.
- An algorithm employing the Fisher's discriminant ratio (FDR) is also proposed, to select elements of the Hilbert spectrum with strong class separability as identification features.
- We further extend the identification problem to the relaying scenario, in which the fingerprint of emitters is contaminated by the relay's fingerprint. To the best of our knowledge, all the existing literature focuses on the SEI problem in typical single-hop communication systems. We are the first to propose SEI algorithms applicable to relaying scenarios. In addition, we also consider SEI under non-ideal conditions, such as non-Gaussian noise and fading.
- Extensive simulation results are performed to verify the identification performance of the proposed algorithms under various channel conditions in both single-hop and relaying scenarios. Furthermore, we validate the identification performance of the EM^2 algorithm when different emitters employ diverse modulation formats and transmit powers.

The rest of this paper is organized as follows. Section II briefly introduces the HHT and support vector machine (SVM) concepts. The system model is presented in Section III, and the proposed algorithms are introduced in Section IV. Section V shows simulation results, and conclusions are drawn in Section VI.

II. Preview of HHT and SVM

A. Hilbert-Huang Transform

In contrast to the traditional transforms, such as Fourier and wavelet transforms, HHT is a more adaptive tool which can be applied to non-linear and non-stationary signals [16]. The fundamental of HHT consists of the empirical mode decomposition (EMD) and Hilbert spectrum analysis. The former is a sifting process to decompose any signal into a infinite set of intrinsic mode function (IMF), while the latter offers the time-frequency distribution, referred to as the Hilbert spectrum, by performing the Hilbert transform on each IMF.

1) Empirical Mode Decomposition: The EMD aims at representing a given signal by a collection of IMFs, which is the foundation of obtaining physically meaningful instantaneous frequency [16]. The definition of the IMF satisfies the following assumptions: a) The number of extrema either equals the number of zero-crossings, or the difference is one at most; b) At any point, the sum of the upper and lower envelopes is zero.

Let $z(t)$ be the original signal; the sifting process of decomposition is described as follows [16]:

i) Identify all local maxima and minima to obtain the upper and lower envelopes by cubic spline fitting, where the signal is covered between the upper and lower envelopes;

ii) Calculate the mean of the upper and lower envelopes, denoted by $\mu_{10}(t)$. The first component $z_{10}(t)$ is defined as the difference between $z(t)$ and $\mu_{10}(t)$, i.e., $z_{10}(t) = z(t) - \mu_{10}(t)$;

iii) Repeat steps i) and ii) p times until $z_{1p}(t)$ becomes an IMF

$$z_{1p}(t) = z_{1(p-1)}(t) - \mu_{1p}(t), \quad p = 1, 2, \ldots, \tag{1}$$

where $\mu_{1p}(t)$ is the mean of the upper and lower envelopes of $z_{1(p-1)}(t)$. The stopping criterion is when

$$\xi = \sum_{t=0}^{T_s} \frac{|z_{1(p-1)}(t) - z_{1p}(t)|^2}{z_{1(p-1)}^2(t)} \tag{2}$$

is below a pre-defined value ε, with T_s as the length of the signal[1];

iv) Let $c_1(t) = z_{1p}(t)$ be the first IMF; this is subtracted from $z(t)$ to obtain the residual

$$d_1(t) = z(t) - c_1(t). \tag{3}$$

The residual, containing the IMF, is considered as a new signal to implement the sifting process;

v) Repeat steps i) to iv) on all residuals $d_q(t)$, $q = 1, \ldots, Q$, with Q as the number of IMFs, and

$$\begin{aligned} d_2(t) &= d_1(t) - c_2(t), \\ &\cdots \\ d_Q(t) &= d_{Q-1}(t) - c_Q(t). \end{aligned} \tag{4}$$

The sifting process is stopped when $d_Q(t)$ either is less than the pre-defined value of ξ or becomes a monotonic function, which contains no oscillation.

[1] An empirical value of ξ is usually set between 0.2 and 0.3 [16].

From (3) and (4), $z(t)$ can be expressed as

$$z(t) = \sum_{q=1}^{Q} c_q(t) + d_Q(t). \tag{5}$$

2) Hilbert Spectrum Analysis: By applying the Hilbert transform to each IMF, the original signal in (5) that omits the residual $d_Q(t)$ can be expressed as [16]

$$z(t) = \Re\left[\sum_{q=1}^{Q} a_q(t) \exp\left(\jmath \int \omega_q(t) dt\right)\right], \tag{6}$$

where $\Re[\cdot]$ returns the real component of a complex variable, $\jmath = \sqrt{-1}$, $a_q(t) = \sqrt{c_q^2(t) + \hat{c}_q^2(t)}$ and $\omega_q(t) = \frac{d\theta_q(t)}{dt}$ are the instantaneous amplitude and frequency of each IMF component $c_q(t), q = 1, \ldots, Q$, respectively, with $\theta_q(t) = \arctan \frac{\hat{c}_q(t)}{c_q(t)}$ as the phase function and $\hat{c}_q(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_q(\tau)}{t-\tau} d\tau$ as the Hilbert transform of each IMF.

Based on the instantaneous amplitude and frequency of each IMF component, the Hilbert spectrum $\mathcal{H}(\omega, t)$ is obtained [16]. We use the squared value of the Hilbert spectrum, which provides information on the time-frequency energy distribution.

B. Support Vector Machine

The SVM is a supervised learning classifier for a two-class classification problem, i.e., $K = 2$ [18]. Suppose that $\{\mathbf{v}_i, \imath_i\}$ is the training set, where $\mathbf{v}_i$ is the training vector and $\imath_i \in \{1, -1\}$ is the class label.

1) Linear SVM: We first consider the simplest case when the positive and negative data can be linearly separated by a hyperplane $\chi(\mathbf{v}) = 0$. The decision function $\chi(\mathbf{v})$ is given by [18]

$$\chi(\mathbf{v}) = \mathbf{w}^{\mathrm{T}}\mathbf{v} + b, \tag{7}$$

where $\mathbf{w}$ is the normal vector to the hyperplane, $b/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin, and $\|\cdot\|$ denotes the Euclidean norm.

The goal of the SVM is to optimize the hyperplane to make the margin $m_d = m_{d,+} + m_{d,-}$ the largest, where $m_{d,+}$ and $m_{d,-}$ are the closest distances from the positive and negative points to the hyperplane, respectively. To simplify the derivation, two hyperplanes that bound the margin are defined as

$$\begin{aligned} \mathbf{w}^{\mathrm{T}}\mathbf{v} + b &= 1, \\ \mathbf{w}^{\mathrm{T}}\mathbf{v} + b &= -1. \end{aligned} \tag{8}$$

With the geometrical theory, it is clear that the distance between the two hyperplanes is $m_d = \frac{2}{\|\mathbf{w}\|}$. Since the task is to optimize $\mathbf{w}$ and b to maximize m_d, the classification problem is converted into a constrained optimization function[2] [18], [19]

$$\begin{aligned} &\min \ \frac{1}{2}\|\mathbf{w}\|^2 \\ &s.t. \ \imath_i \left(\mathbf{w}^{\mathrm{T}}\mathbf{v}_i + b\right) \geq 1, \quad i = 1, \ldots, \bar{N}, \end{aligned} \tag{9}$$

where $\bar{N}$ is the number of training data. By introducing non-negative Lagrange multipliers $\lambda_i \geq 0$, (9) can be converted into a dual quadratic programming optimization problem, given by [18], [19]

$$\begin{aligned} &\max_{\boldsymbol{\lambda}} \ \sum_{i=1}^{\bar{N}} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{\bar{N}} \lambda_i \lambda_j \imath_i \imath_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle \\ &s.t. \ \lambda_i \geq 0, \quad i = 1, \ldots, \bar{N}, \\ &\quad \sum_{i=1}^{\bar{N}} \lambda_i \imath_i = 0, \end{aligned} \tag{10}$$

where $\mathbf{w}$ and b are replaced by λ_i, with $\mathbf{w} = \sum_{i=1}^{\bar{N}} \lambda_i \imath_i \mathbf{v}_i$, and $b = -\frac{1}{2}(\max_{i:\imath_1=-1} \mathbf{w}^{\mathrm{T}}\mathbf{v}_i + \min_{i:\imath_1=1} \mathbf{w}^{\mathrm{T}}\mathbf{v}_i)$.

By solving the optimization problem in (10), the decision function can be re-written as [18], [19]

$$\chi(\mathbf{v}) = \sum_{i=1}^{\bar{N}} \lambda_i \imath_i \langle \mathbf{v}_i, \mathbf{v} \rangle + b, \tag{11}$$

where $\langle \cdot, \cdot \rangle$ represents the inner product. The decision is determined by the sign of (11), where the test data $\mathbf{u}_l$ satisfying $\chi(\mathbf{u}_l) > 0$ is labeled as 1; otherwise, it is labeled as -1, i.e., the decision rule of the correct classification is $\imath_l \chi(\mathbf{u}_l) > 0$.[3]

2) Non-Linear SVM: Generally, $\mathbf{v}_i$ cannot be simply distinguished by the linear classifier; in such a case, by applying a non-linear mapping function ϕ, $\mathbf{v}_i$ is mapped to a high-dimensional space $\mathfrak{F}$, where classes can be separated by a hyperplane. Similarly, the decision function is expressed as [18]

$$\chi(\mathbf{v}) = \sum_{i=1}^{\bar{N}} \lambda_i \imath_i \langle \phi(\mathbf{v}_i), \phi(\mathbf{v}) \rangle + b. \tag{12}$$

Note that computing $\phi(\cdot)$ in $\mathfrak{F}$ may be intractable; thus, the kernel function $\kappa(\mathbf{v}_i, \mathbf{v})$ is introduced to replace the inner product $\langle \phi(\mathbf{v}_i), \phi(\mathbf{v}) \rangle$ [20].[4] Using $\mathbf{w} = \sum_{i=1}^{\bar{N}} \lambda_i \imath_i \mathbf{v}_i$ and the kernel function, (12) can be re-written as [18]

$$\chi(\mathbf{v}) = \sum_{i=1}^{\bar{N}} \lambda_i \imath_i \kappa(\mathbf{v}_i, \mathbf{v}) + b. \tag{13}$$

The decision rule is the same as for the linear classifier.

3) Multi-Class SVM: Furthermore, we consider the case when $K > 2$. Generally, the multi-class classification problem is solved by reducing it to several binary classification problems. Common approaches include one-versus-one [21], one-versus-all [21] and binary tree architecture [22] techniques. Since the one-versus-one technique has better performance and is more suitable for practical use [21], we adopt

[2]It is noted that the optimization problem of minimizing $\|\mathbf{w}\|$ is difficult to solve since $\|\mathbf{w}\|$ involves the square root. By substituting $\|\mathbf{w}\|$ with $\|\mathbf{w}\|^2$, this problem can be solved easily without changing the solution [19].

[3]For the ease of calculation, we define the class label to be ± 1. In fact, the value is arbitrary. For instance, if $\imath_l \in \{1, 2\}$, the test data $\mathbf{u}_l$ satisfying $\chi(\mathbf{u}_l) > 0$ is labeled as 2; otherwise, it is labeled as 1, and the decision rule of the correct classification becomes $(\imath_l - 1.5)\chi(\mathbf{u}_l) > 0$ [19].

[4]Conventional kernel functions are the Gaussian radial base function (RBF), $\kappa(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|^2/2\gamma^2}$, and the polynomial kernel, $\kappa(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})^d$, where d is the sum of the exponents in each term, i.e., all the non-zero terms in the polynomial have the same degree d [20].

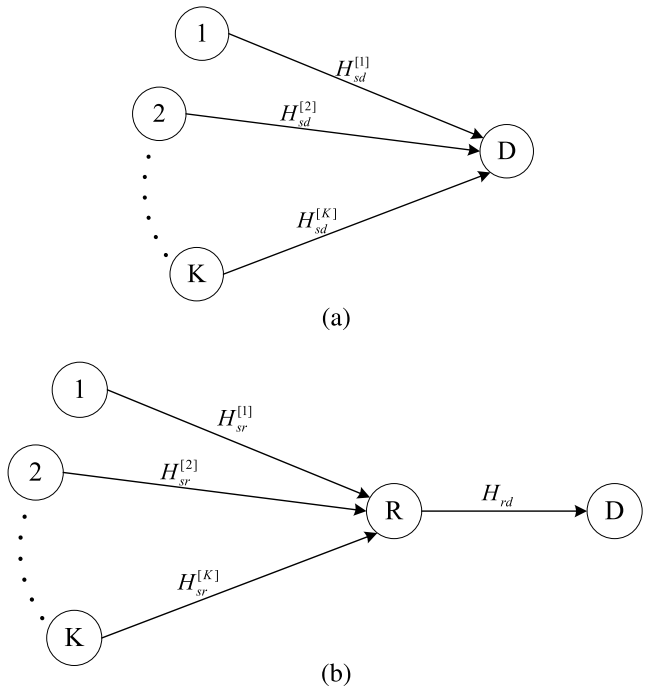

Fig. 1. (a) The system model of the single-hop scenario; (b) The system model of the relaying scenario.

it as the multi-class SVM classifier in this paper. In such a case, the classification is implemented by a max-win voting mechanism, i.e., the class with the highest number of votes is chosen as the classification result [21].

III. System Model

A. Single-Hop Scenario

Consider a typical time division multiple access communication system with K emitters,[5] as shown in Fig. 1 (a). SEI aims at distinguishing the unique emitter by extracting the specific information carried by the transmitted signal; therefore, a model should be introduced to define the difference between emitters. Note that a power amplifier is a major component of an emitter, whose non-linear system response characteristic is the main cause of the specific feature of the emitter, also referred to as the fingerprint. In general, the Taylor polynomial model is employed to describe the non-linear system [26]–[28]. Let L_s be the order of the Taylor polynomial; then, for emitter k, the system response function of the power amplifier is expressed as

$$\Phi^{[k]}(x_n) = \sum_{l=1}^{L_s} \alpha_l^{[k]} (x_n)^l, \tag{14}$$

where

$$x_n = s_n e^{j2\pi nfT}, \tag{15}$$

is the input signal at the power amplifier, with s_n as the baseband modulated signal at time n, f as the carrier frequency, and T as the sampling period, and $\{\alpha_l^{[k]}\}$ are the coefficients of the Taylor polynomial. It can be seen that with the same order L_s for diverse emitters, the different coefficients represent the fingerprints. $\Phi^{[k]}(x_n)$ denotes the output signal at the power amplifier of emitter k, i.e., the transmitted signal of emitter k, through which the emitter-specific information is carried.

At the receiver, the discrete-time received signal at time n can be expressed as[6]

$$r_n = H_{sd}^{[k]} \Phi^{[k]}(x_n) + w_n, \quad k = 1, \ldots, K, \tag{16}$$

where $H_{sd}^{[k]}$ is the fading coefficient of the channel from emitter k to the receiver, which is unknown at the receiver, and w_n is the additive noise. At the receiver, the emitter-specific feature is extracted from r_n to identify different emitters.

Hence, by substituting (14) into (16), the received signal can be written as

$$r_n = H_{sd}^{[k]} \sum_{l=1}^{L_s} \alpha_l^{[k]} (x_n)^l + w_n. \tag{17}$$

B. Relaying Scenario

We next consider a two-hop communication system in the presence of an amplify-and-forward (AF) relay, as shown in Fig. 1 (b). For emitter k, the signal received at the relay is given by

$$y_n = H_{sr}^{[k]} \Phi^{[k]}(x_n) + \eta_n, \quad k = 1, \ldots, K, \tag{18}$$

where $H_{sr}^{[k]}$ is the fading coefficient of the channel from emitter k to the relay, and η_n is the additive noise.

Then, the signal at the receiver, forwarded by the relay, can be written as

$$\begin{aligned} r_n &= H_{rd} \Psi(y_n) + v_n \\ &= H_{rd} \Psi\left(H_{sr}^{[k]} \Phi^{[k]}(x_n) + \eta_n \right) + v_n, \end{aligned} \tag{19}$$

where $\Psi(\cdot)$ denotes the system response characteristic of the power amplifier of the relay, H_{rd} is the fading coefficient of the channel from the relay to receiver, and v_n is the additive noise. Note that H_{sr} and H_{rd} are unknown at the receiver. Similarly, $\Psi(\cdot)$ can be expressed through the Taylor polynomial model as

$$\Psi(y_n) = \sum_{m=1}^{L_r} \beta_m (y_n)^m, \tag{20}$$

where L_r is the order of the power amplifier Taylor polynomial of the relay, and $\{\beta_m\}$ denote the fingerprint of the relay. Therefore, the signal at the receiver can be further written as

$$\begin{aligned} r_n &= H_{rd} \sum_{m=1}^{L_r} \beta_m (y_n)^m + v_n \\ &= H_{rd} \sum_{m=1}^{L_r} \beta_m \left(H_{sr}^{[k]} \sum_{l=1}^{L_s} \alpha_l^{[k]} (x_n)^l + \eta_n \right)^m + v_n. \end{aligned} \tag{21}$$

Hence, the features carried by the received signal are the mixture of the fingerprint of the emitter and relay, and the latter may cause a negative effect on SEI.

[5] We assume that the number of the emitters is *a priori* known at the receiver; this can be determined by an algorithm for detecting the number of sources, e.g., those proposed in [23]–[25].

[6] Note that we do not use a notation to show the dependency of r_n on the emitter k, as the receiver does not know from which emitter the information originates.

IV. Hilbert Spectrum-Based SEI Algorithms

In this section, we propose three SEI algorithms as follows: the EM^2 algorithm is introduced in Section IV-A, and measures the uniformity of the Hilbert spectrum; the CB algorithm is proposed in Section IV-B, and extracts features by calculating the correlation coefficient between Hilbert spectra; the FDR algorithm is presented in Section IV-C, and selects the Hilbert spectrum elements with strong class separability. The computational complexity is analyzed in Section IV-D, and a comparison between the three proposed algorithms is provided in Section IV-E.

A. Entropy and First- and Second-Order Moments (EM^2)-Based Algorithm

The EM^2 algorithm extracts features by measuring the uniformity of the Hilbert spectrum. The energy entropy is calculated to extract the emitter-specific feature displayed in the Hilbert spectrum, as shown in the first subsection below. The Hilbert spectrum can be considered as a two-dimensional greyscale digital image, and the first- and second-order moments are employed to analyze this image, as presented in the second sub-section below.

1) Energy Entropy: In order to measure the uniformity of the time-frequency distribution of the Hilbert spectrum, the definition of information entropy is introduced to calculate the energy entropy. First, the Hilbert spectrum is divided into several time-frequency slots. Suppose that $\mathcal{H}_{ij}(\omega, t)$, $i = 1, \ldots, G_t, j = 1, \ldots, G_\omega$ denotes the (i, j)th time-frequency slot, where G_t is the number of time slots with resolution Δt and G_ω is the number of frequency slots with resolution $\Delta\omega$. Then, the energy of each time-frequency slot is described as

$$E_{ij} = \int_{(i-1)\Delta t}^{i\Delta t} \int_{(j-1)\Delta\omega}^{j\Delta\omega} \mathcal{H}_{ij}(\omega, t) d\omega dt, \quad i = 1, \ldots, G_t, \quad j = 1, \ldots, G_\omega. \tag{22}$$

By using the expression of the information entropy [29], the energy entropy of the Hilbert spectrum is defined as

$$I = -\sum_{i=1}^{G_t} \sum_{j=1}^{G_\omega} p_{ij} \log p_{ij}, \tag{23}$$

where $p_{ij} = E_{ij}/E$ is the proportion of the energy of each time-frequency slot, with $E = \sum_i \sum_j E_{ij}$ as the total energy of the Hilbert spectrum.

2) First- and Second-Order Moments: In image processing, a greyscale image is composed of shades of grey with intensity information. The value of each pixel in the greyscale image, taking ζ-bit greyscale as an example, is a scalar ranging from 0 to $2^\zeta - 1$, i.e., varying from black at the weakest intensity (0) to white at the strongest intensity $(2^\zeta - 1)$. To calculate the first- and second-order moments, we first convert the Hilbert spectrum to a ζ-bit greyscale image as [9]

$$B_{m,n} = \left\lfloor (2^\zeta - 1) \times \frac{\mathcal{H}_{m,n}}{\max_{\{m,n\}} \mathcal{H}_{m,n}} \right\rfloor, \tag{24}$$

where $B_{m,n}$ is the (m, n)th value of the greyscale image matrix $\mathbf{B}$, $\mathcal{H}_{m,n}$ is the (m, n)th value of the Hilbert spectrum matrix $\boldsymbol{\mathcal{H}}$,[7] and $\lfloor \cdot \rfloor$ is the floor function, which equals the nearest lower integer value. Thus, the largest value of the Hilbert spectrum is converted to the greyscale $(2^\zeta - 1)$, while other values are linearly scaled.

We use the first- and second-order moments to evaluate how similar two greyscale images are based on the intensity information distribution [30]. The first-order moment measures the average intensity of the greyscale image, which is defined as [31]

$$\mu = \frac{1}{N_H} \sum_{m=1}^{M} \sum_{n=1}^{N} B_{m,n}, \tag{25}$$

where M and N are the number of rows and columns of the greyscale image matrix, respectively, and $N_H \triangleq M \times N$ is the size of the matrix. The second-order moment interprets the standard deviation of the shades of grey, and is given by [31]

$$\varsigma = \left[\frac{1}{N_H} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(B_{m,n} - \mu \right)^2 \right]^{\frac{1}{2}}. \tag{26}$$

The training and identification procedures of the EM^2 algorithm are summarized in Algorithm 1.

B. Correlation-Based (CB) Algorithm

The CB algorithm employs the correlation coefficient between Hilbert spectra as the identification feature to measure their similarities. Note that the shape of the time-frequency distribution of signals from the same emitter are similar, while that from different emitters are significantly different; hence, the correlation coefficient can be exploited for identification, as follows.

The correlation coefficient is defined to measure the similarity between different Hilbert spectra. Let $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{H}}_j$ be the Hilbert spectrum matrices of the ith and jth training sequence, respectively. The correlation coefficient between $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{H}}_j$ is defined as (27), as shown at the bottom of the next page, where $\mathcal{H}_{i,m,n}$ $(\mathcal{H}_{j,m,n})$ denotes the (m, n)th element of the Hilbert spectrum matrix $\boldsymbol{\mathcal{H}}_i$ $(\boldsymbol{\mathcal{H}}_j)$, and $\mathbb{E}[\cdot]$ is the mean of values of the Hilbert spectrum matrix. Equation (27) measures the linear dependence between $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{H}}_j$, which indicates that $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{H}}_j$ from the same emitter have a larger $\rho^{(i,j)}$; otherwise, $\rho^{(i,j)}$ is close to 0.

The training and identification procedures of the CB algorithm are summarized in Algorithm 2.

C. Fisher's Discriminant Ratio (FDR) Algorithm

The FDR algorithm selects the elements of the Hilbert spectrum that provide well separation between two classes. Note that a number of elements in the Hilbert spectrum are

[7] As the Hilbert spectrum $\mathcal{H}(\omega, t)$ is a two-dimensional continuous spectrum, it is represented through a matrix, referred to as the Hilbert spectrum matrix. The indices of the columns and rows of this matrix correspond to the sampling point and instantaneous frequency, respectively, and the elements of the matrix are combination of the instantaneous amplitudes of the IMFs [16].

Algorithm 1 Training and Identification Procedures of the EM^2 Algorithm

Training Procedure: Let $\boldsymbol{\mathcal{H}}_i, i = 1, \ldots, \bar{N}$, be the Hilbert spectrum matrix of the training sequence i, with $\bar{N}$ as the total number of training sequences over all K classes.[8]

1: Calculate the energy entropy of $\boldsymbol{\mathcal{H}}_i$ using (23), denoted as $\bar{I}_i$[9];

2: Map $\boldsymbol{\mathcal{H}}_i$ to $\mathbf{B}_i$ employing (24), where $\mathbf{B}_i$ is the Hilbert greyscale image matrix of the training sequence i. Then, compute its first- and second-order moments using (25) and (26), denoted as $(\bar{\mu}_i, \bar{\varsigma}_i)^{\mathrm{T}}$;

3: Form the training vector of training sequence i as $\mathbf{v}_i = \left(\bar{I}_i, \bar{\mu}_i, \bar{\varsigma}_i\right)^{\mathrm{T}}$;

4: Let $\{\mathbf{v}_i, \imath_i\}$ be the training set, where $\imath_i \in \{1, \ldots, K\}$ is the label of each class. Use the training set to train the SVM classifier, i.e., to obtain the optimal $\mathbf{w}$ and b of the decision hyperplane $\chi(\mathbf{v})$.

Identification Procedure: Let $\boldsymbol{\mathcal{H}}_l, l = 1, \ldots, N$, be the Hilbert spectrum matrix of the lth test sequence of an unknown class, where N is the number of test sequences.

5: Use (23) to (26) to compute the energy entropy I_l, the first- and second-order moments $(\mu_l, \varsigma_l)^{\mathrm{T}}$ of $\boldsymbol{\mathcal{H}}_l$;

6: Obtain the test vector of test sequence l, denoted as $\mathbf{u}_l = (I_l, \mu_l, \varsigma_l)^{\mathrm{T}}$;

7: The identification task is implemented by employing the SVM classifier defined in the training procedure. For $K = 2$, $\mathbf{u}_l$ which satisfies $\chi(\mathbf{u}_l) > 0$ is labeled as class 2; otherwise, it is labeled as class 1. For $K > 2$, the one-versus-one technique is applied, where the decision depends on the max-win voting mechanism, i.e., the class with the highest number of votes is considered as the identification result.

Algorithm 2 Training and Identification Procedures of the CB Algorithm

Training Procedure: Let $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{H}}_j$ be the Hilbert spectrum matrices of the training sequences i and j, respectively, where $i, j = 1, \ldots, \bar{N}$, with $\bar{N}$ as the total number of the training sequences over all K classes.

1: Calculate the correlation coefficient $\bar{\rho}^{(i,j)}$ between $\boldsymbol{\mathcal{H}}_i$ and $\boldsymbol{\mathcal{H}}_j$ using (27). For the training sequence i, the training vector is $\bar{N}$-dim, denoted as $\bar{\boldsymbol{\rho}}_i = \left[\bar{\rho}^{(i,1)}, \ldots, \bar{\rho}^{(i,\bar{N})}\right]^{\mathrm{T}}$;

2: Let $\left\{\bar{\boldsymbol{\rho}}_i, \imath_i\right\}$ be the set of training data with $\imath_i \in \{1, \ldots, K\}$ as the label of each class. Then, the set is input into the SVM classifier for training, i.e., to obtain the optimal $\mathbf{w}$ and b of the decision hyperplane $\chi(\bar{\boldsymbol{\rho}})$.

Identification Procedure: Let $\boldsymbol{\mathcal{H}}_l, l = 1, \ldots, N$, be the Hilbert spectrum matrix of the test sequence of an unknown class, where N is the number of test sequences.

3: For the test sequence l, the correlation coefficient $\rho^{(l,i)}$ between $\boldsymbol{\mathcal{H}}_l$ and $\boldsymbol{\mathcal{H}}_i$ is calculated using (27), and the $\bar{N}$-dim test vector $\boldsymbol{\rho}_l = \left[\rho^{(l,1)}, \ldots, \rho^{(l,\bar{N})}\right]^{\mathrm{T}}$ is obtained;

4: Classify the test sequence by employing the SVM classifier. For $K = 2$, $\boldsymbol{\rho}_l$ which satisfies $\chi(\boldsymbol{\rho}_l) > 0$ is labeled as class 2; otherwise, it is labeled as class 1. For $K > 2$, the one-versus-one technique is applied, where the decision depends on the max-win voting mechanism, i.e., the class with the highest number of votes is considered as the identification result.

featureless, having no effect on or even being counterproductive to the correct identification performance, and thus, they need to be screened out. Hence, the Fisher's discriminant ratio is defined for this problem, as follows.

Let $\mathcal{H}_i^{[k]}(\omega, t), i = 1, \ldots, \bar{N}_0$, be the Hilbert spectrum of the ith training sequence of the kth class at the time-frequency spot (ω, t), where $\bar{N}_0$ is the number of training sequences for each class. Suppose that (k_1, k_2) is the combination of two classes arbitrarily chosen from all K classes, where $k_1 \neq k_2$; hence, there are $K(K-1)/2$ possible combinations (k_1, k_2) for all K classes. Then, the FDR at (ω, t) between class k_1 and k_2 is defined as

$$\mathcal{F}^{(k_1,k_2)}(\omega, t) = \frac{\left(\mathbb{E}_i\left[\mathcal{H}_i^{[k_1]}(\omega, t)\right] - \mathbb{E}_i\left[\mathcal{H}_i^{[k_2]}(\omega, t)\right]\right)^2}{\sum_{k=k_1,k_2} \mathbb{D}_i\left[\mathcal{H}_i^{[k]}(\omega, t)\right]}, \tag{28}$$

where $\mathbb{E}_i\left[\mathcal{H}_i^{[k]}(\omega, t)\right]$ and $\mathbb{D}_i\left[\mathcal{H}_i^{[k]}(\omega, t)\right]$ represent the mean and variance of the training sequences of class k at (ω, t), respectively. Equation (28) indicates that $\mathcal{F}^{(k_1,k_2)}(\omega, t)$ measures the separability at (ω, t) between classes k_1 and k_2, i.e., the element with larger $\mathcal{F}^{(k_1,k_2)}(\omega, t)$ provides a large separation between the mean of two classes and a small within-class variance, and has stronger separability.

For each combination (k_1, k_2), we define $\boldsymbol{\Omega} = \{\mathcal{F}_1^{(k_1,k_2)}(\omega, t), \ldots, \mathcal{F}_{N_H}^{(k_1,k_2)}(\omega, t)\}$ as the original FDR sequence, with N_H as the length of the FDR sequence, i.e., the number of the time-frequency spots. Then, $\boldsymbol{\Omega}$ is sorted in descending order and the re-arranged FDR sequence is expressed as $\tilde{\boldsymbol{\Omega}} = \{\tilde{\mathcal{F}}_1^{(k_1,k_2)}(\omega, t), \ldots, \tilde{\mathcal{F}}_{N_H}^{(k_1,k_2)}(\omega, t)\}$, which satisfies $\tilde{\mathcal{F}}_1^{(k_1,k_2)}(\omega, t) \geq \ldots \geq \tilde{\mathcal{F}}_{N_H}^{(k_1,k_2)}(\omega, t)$.

$$\rho^{(i,j)} = \frac{\sum_m \sum_n \left(\mathcal{H}_{i,m,n} - \mathbb{E}[\boldsymbol{\mathcal{H}}_i]\right)\left(\mathcal{H}_{j,m,n} - \mathbb{E}[\boldsymbol{\mathcal{H}}_j]\right)}{\sqrt{\left(\sum_m \sum_n \left(\mathcal{H}_{i,m,n} - \mathbb{E}[\boldsymbol{\mathcal{H}}_i]\right)^2\right)\left(\sum_m \sum_n \left(\mathcal{H}_{j,m,n} - \mathbb{E}[\boldsymbol{\mathcal{H}}_j]\right)^2\right)}}, \quad i, j = 1, \ldots, N_R, \tag{27}$$

[8]In this paper, a class is a set of received signals coming from one emitter, or a cluster of features represents one emitter. Therefore, the label of classes is expressed in the same notation as emitters.

[9]In order to distinguish variables between the training and test procedures, we use $\bar{I}$ to represent the energy entropy of the training sequence. The subscript i of $\bar{I}_i$ denotes the energy entropy of the training sequence i. This also holds for the notations of the moments.

Let $\{(\tilde{\omega}_1, \tilde{t}_1), \ldots, (\tilde{\omega}_{N_H}, \tilde{t}_{N_H})\}$ be the time-frequency slots which correspond to the FDR sequence $\tilde{\Omega}$; then, the time-frequency spots corresponding to the S largest FDR $\tilde{\mathcal{F}}_1^{(k_1,k_2)}(\omega, t), \ldots, \tilde{\mathcal{F}}_S^{(k_1,k_2)}(\omega, t)$ are chosen as optimal time-frequency spots, denoted as $\mathcal{Z}^{(c)} = \{(\tilde{\omega}_s^{(c)}, \tilde{t}_s^{(c)}), s = 1, \ldots, S, c = 1, \ldots, C\}$, where $C = K(K-1)/2$ represents the number of all possible combinations (k_1, k_2). The total set of optimal time-frequency spots is defined as the union of $\mathcal{Z}^{(c)}$, i.e., $\mathcal{Z} = \bigcup_{c=1}^{C} \mathcal{Z}^{(c)}$. For the same $(\tilde{\omega}_s, \tilde{t}_s)$ between different combinations (k_1, k_2), only one is retained in order to avoid duplication, i.e., $\mathcal{Z} = \{(\tilde{\omega}_1, \tilde{t}_1), \ldots, (\tilde{\omega}_D, \tilde{t}_D)\}$, where D is the number of optimal time-frequency spots without duplication, with $D \leq S \times \frac{K(K-1)}{2}$.

The training and identification procedures of the FDR algorithm are summarized in Algorithm 3.

Algorithm 3 Training and Identification Procedures of the FDR Algorithm

Training Procedure: Let $\mathcal{H}_i(\omega, t)$ be the Hilbert spectrum of the training sequence i at time-frequency spot (ω, t), where $i = 1, \ldots, \bar{N}$, with $\bar{N}$ as the total number of the training sequences over all K classes.

1: For the combination (k_1, k_2), use (28) to compute the original FDR sequence $\Omega = \{\mathcal{F}_1^{(k_1,k_2)}(\omega, t), \ldots, \mathcal{F}_{N_H}^{(k_1,k_2)}(\omega, t)\}$;

2: Obtain the FDR sequence (descending order of elements) $\tilde{\Omega} = \{\tilde{\mathcal{F}}_1^{(k_1,k_2)}(\omega, t), \ldots, \tilde{\mathcal{F}}_{N_H}^{(k_1,k_2)}(\omega, t)\}$;

3: Select the time-frequency spots that correspond with the S largest FDR in $\tilde{\Omega}$ to form the optimal time-frequency spots set $\mathcal{Z}^{(c)} = \{(\tilde{\omega}_s^{(c)}, \tilde{t}_s^{(c)}), s = 1, \ldots, S, c = 1, \ldots, C\}$;

4: Repeat steps 1 to 3 over all $C = K(K-1)/2$ possible combinations (k_1, k_2) to obtain the total set of optimal time-frequency spots, $\mathcal{Z} = \bigcup_{c=1}^{C} \mathcal{Z}^{(c)}$. For the same $(\tilde{\omega}_s, \tilde{t}_s)$ between different combinations (k_1, k_2), only one of them is retained in order to avoid duplication;

5: For the training sequence i, the Hilbert spectrum elements corresponding to $(\tilde{\omega}_1, \tilde{t}_1), \ldots, (\tilde{\omega}_D, \tilde{t}_D)$ are extracted to form a D-dim training vector, expressed as $\mathbf{v}_i = \left[\mathcal{H}_i(\tilde{\omega}_1, \tilde{t}_1), \ldots, \mathcal{H}_i(\tilde{\omega}_D, \tilde{t}_D)\right]^{\mathrm{T}}$;

6: Let $\{\mathbf{v}_i, \iota_i\}$ be the set of training data with $\iota_i \in \{1, \ldots, K\}$ as the label of each class. Then, the data is input into the SVM classifier for training, i.e., to obtain the optimal $\mathbf{w}$ and b of the decision hyperplane $\chi(\mathbf{v})$.

Identification Procedure: Let $\mathcal{H}_l(\omega, t)$, $l = 1, \ldots, N$, be the Hilbert spectrum of test sequence l at time-frequency spot (ω, t) of a unknown class, where N is the number of test sequences.

7: For the test sequence l, extract the elements corresponding to the D optimal time-frequency spots as the test vector, i.e., $\mathbf{u}_l = \left[\mathcal{H}_l(\tilde{\omega}_1, \tilde{t}_1), \ldots, \mathcal{H}_l(\tilde{\omega}_D, \tilde{t}_D)\right]^{\mathrm{T}}$;

8: Utilize the SVM classifier to identify the test sequence. For $K = 2$, $\mathbf{u}_l$ which satisfies $\chi(\mathbf{u}_l) > 0$ is labeled as class 2; otherwise, it is labeled as class 1. For $K > 2$, one-versus-one technique is applied, where the decision depends on the max-win voting mechanism, i.e., the class with the highest number of votes is considered as the identification results.

D. Computational Complexity

We consider the computational complexity of the proposed algorithms, which involves performing the HHT, extracting the identification features, and applying the SVM. Since the HHT is empirical and has no well-established formula, a rough complexity analysis is provided. For each sequence, the complexity of performing the HHT is of order $\mathcal{O}(pQN_l + pQ\log N_l + QN_l \log N_l)$, where p is the number of iterations to obtain the IMF, Q is the number of IMFs and N_l is the length of the sequence. The process is performed N_S times to obtain the Hilbert spectrum of all sequences, where $N_S \triangleq \bar{N} + N$ is the total number of training and test sequences over all K classes. Furthermore, the complexity of training an SVM classifier is of order $\mathcal{O}(R^3 + \bar{N}N_{SV})$, where R is the number of free support vectors and N_{SV} is the number of support vectors [33]. At the test stage, the complexity is of order $\mathcal{O}(N_{SV})$ [33].[10]

The computational complexity of the proposed feature extraction schemes is analyzed as follows. For the EM^2 algorithm, based on (22) to (26), the computational complexity is linear, i.e., of order $\mathcal{O}(KN_s(N_H + G))$, where $N_s = \bar{N}_0 + N_0$ is the total number of training and test sequences for each class, N_H is the size of the Hilbert spectrum matrix, and $G = G_t \times G_\omega$ is the number of the time-frequency slots with G_t as the number of time slots with resolution Δt, and G_ω as the number of frequency slots with resolution $\Delta\omega$. For the CB algorithm, the complexity of calculating the correlation coefficient using (27) is of order $\mathcal{O}(N_H)$, and the complexity of obtaining the training and test vectors is of order $\mathcal{O}\left(K^2 N_s \bar{N}_0 N_H\right)$. The FDR algorithm involves calculating the FDR with (28) and sorting, whose complexity is of order $\mathcal{O}\left((K^2 - K) \times \left(N_H^2 + N_H N_s\right)\right)$.

Furthermore, we compare the proposed algorithms with the conventional approach in [9] in terms of complexity; a performance comparison will be also provided later on in the paper. The approach proposed in [9] employs the principal component analysis (PCA) to extract features from the Hilbert spectrum. PCA is a statistical technique to project data to a set of orthogonal variables—called principal components—without much loss of information. Note that the number of principal components is less than the dimensions of the original data, and that PCA is widely used in dimension reduction [35]. The algorithm proposed in [9] can be simply summarized as follows. The Hilbert spectrum of the training sequence i is first considered as a matrix of size $M \times N$, and vectorized into an N_H-dimensional vector $\mathcal{H}_i'$, where $N_H = M \times N$. Then, for all $\bar{N}_0$ training

[10] In practice, when the training samples contain outliers, the soft margins are applied to the the optimization problem in (9), where an additional parameter C is added to control the compromise between large and small margin violations [34], i.e., the Lagrange multipliers λ_i have the upper bound C. Then, the samples satisfying $\lambda_i = C$ are defined as the free support vectors R, and those satisfying $0 < \lambda_i < C$ are referred to as the bounded support vectors. Both the free and bounded support vectors are called the support vectors, N_{SV} [33].

TABLE I
COMPARISON OF THE COMPLEXITY FOR THE PROPOSED ALGORITHMS

Algorithm	Computational complexity
EM^2	$\mathcal{O}\left(KN_s(N_H+G)\right)$
CB	$\mathcal{O}\left(K^2N_s\bar{N}_0N_H\right)$
FDR	$\mathcal{O}\left((K^2-K)\times\left(N_H^2+N_HN_s\right)\right)$
[9]	$\mathcal{O}\left(K^2N_H^2\left(N_s^2-\bar{N}_0N_0\right)\right)$

sequences, the vectors obtained by the vectorization are denoted as $\tilde{\boldsymbol{\mathcal{H}}}=[\boldsymbol{\mathcal{H}}'_1,\cdots,\boldsymbol{\mathcal{H}}'_{\bar{N}_0}]$. For PCA to work properly, $\tilde{\boldsymbol{\mathcal{H}}}$ is standardized by subtracting the mean from each dimension. Next, the eigenvector $\boldsymbol{e}_i$ is computed by solving $\zeta_i\boldsymbol{e}_i=\boldsymbol{\Sigma}\boldsymbol{e}_i$, where ζ_i is the eigenvalue corresponding to the eigenvector $\boldsymbol{e}_i$ and $\boldsymbol{\Sigma}=\tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{\mathcal{H}}}^{\mathrm{T}}$ is the covariance matrix. Then, the eigenvalues are sorted in descending order, and the N_D eigenvectors corresponding to the N_D largest eigenvalues are selected to construct the projection matrix $\boldsymbol{W}\in\mathbb{R}^{N_H\times N_D}$. The N_D-dimensional feature is finally obtained as $\hat{\boldsymbol{\mathcal{H}}}_i=\boldsymbol{W}^{\mathrm{T}}\times\boldsymbol{\mathcal{H}}'_i$, i.e., the original N_H-dimensional data is mapped to an N_D-dimensional feature ($N_D\ll N_H$). The algorithm proposed in [9] chooses N_D such that the feature contains 90% information of the original data. Note that PCA is a linear technique, and the original algorithm in [9] employs the kernel Fisher discriminant analysis (KFDA) to deal with the non-linear data, and then uses the K-nearest neighbors (K-NN) classifier to distinguish different classes.[11] The computational complexity of PCA is quadratic, and hence, the complexity of the feature extraction process of the algorithm in [9] is of order $\mathcal{O}\left(K^2N_H^2\left(N_s^2-\bar{N}_0N_0\right)\right)$, which is computationally more complex than for the proposed algorithms.

E. Comparison of the Algorithms

The comparison of the proposed EM^2, CB, and FDR algorithms and the algorithm in [9] is provided in Table I and Table II: Table I presents the comparison in terms of computational complexity, while Table II shows the SNR required to achieve a probability of correct identification $P_c=0.8$ under various channel conditions (see Section V for the simulation setup). It can be seen that the FDR algorithm achieves the best identification performance, followed by the CB algorithm and finally the EM^2 algorithm. On the other hand, the FDR algorithm has the highest computational complexity and the EM^2 algorithm has the lowest complexity, which means that the improvement in the classification performance is at the cost of an increase in the complexity. Furthermore, it is noted that all three proposed algorithms are computationally more efficient than the algorithm in [9]. The identification performance of the proposed algorithms outperforms that of the algorithms in [9], which basically fails to identify the fingerprint in a reasonable SNR in non-Gaussian and fading channels. Therefore, the advantage of the proposed algorithms can be clearly seen.

[11] In order to make a fair comparison with the three proposed algorithms, we used the algorithm proposed in [9] with SVM as classifier.

V. NUMERICAL RESULTS

In this section, we evaluate the identification performance of the proposed algorithms through various numerical experiments, and compare their performance with that of the conventional algorithm in [9]. The probability of correct identification P_c is used as a performance measure, and is calculated based on 2000 Monte Carlo trials.

The modulation format of the emitter is quadrature amplitude modulation of order 4 (4-QAM), unless otherwise mentioned. The number of training and test sequences is $\bar{N}_0=N_0=50$ for each class, the duration of each sequence is 50 μs and the sampling frequency is 10 GHz, i.e., each observed sequence contains 500 sampling points. The carrier frequency is set to 2 GHz. The order of the Taylor polynomial in (17) and (21) is $L_s=L_r=3$. The time resolution of the Hilbert spectrum equals the time sampling interval, and the number of the frequency bins of the Hilbert spectrum is set to 400, i.e., the size of the Hilbert spectrum is 500×400. Unless otherwise mentioned, to compute the energy entropy for the EM^2 algorithm, the time and frequency resolutions Δt and $\Delta\omega$ that divide the Hilbert spectrum into several time-frequency slots are set to 10 times the time and frequency resolutions of the Hilbert spectrum, respectively.

The coefficients of the Taylor polynomial, denoted here as $\boldsymbol{\alpha}^{[k]}=\left(\alpha_1^{[k]},\ldots,\alpha_{L_s}^{[k]}\right)$, for emitter $k, k=1,2,\ldots,5$, are set to $\boldsymbol{\alpha}^{[1]}=(1,0.5,0.3)^{\mathrm{T}}$, $\boldsymbol{\alpha}^{[2]}=(1,0.08,0.6)^{\mathrm{T}}$, $\boldsymbol{\alpha}^{[3]}=(1,0.01,0.01)^{\mathrm{T}}$, $\boldsymbol{\alpha}^{[4]}=(1,0.01,0.4)^{\mathrm{T}}$ and $\boldsymbol{\alpha}^{[5]}=(1,0.6,0.08)^{\mathrm{T}}$; for $K=2$, the coefficient matrix of the power amplifier Taylor polynomial model of the emitter is set to $\mathcal{A}_2^{\mathrm{T}}=\left(\boldsymbol{\alpha}^{[1]};\boldsymbol{\alpha}^{[2]}\right)$, for $K=3$, it is $\mathcal{A}_3=\left(\boldsymbol{\alpha}^{[1]};\boldsymbol{\alpha}^{[2]};\boldsymbol{\alpha}^{[3]}\right)$, and for $K=5$, it is $\mathcal{A}_5=\left(\boldsymbol{\alpha}^{[1]};\boldsymbol{\alpha}^{[2]};\boldsymbol{\alpha}^{[3]};\boldsymbol{\alpha}^{[4]};\boldsymbol{\alpha}^{[5]}\right)$ [36]. The coefficient matrix of the power amplifier Taylor polynomial model of the relay is $\mathcal{B}^{\mathrm{T}}=(1,0.1,0.1)$. The SVM classifier is implemented by utilizing the LIBSVM toolbox [37], where the Gaussian RBF, $\kappa(\mathbf{x},\mathbf{y})=e^{-\|\mathbf{x}-\mathbf{y}\|^2/2\gamma^2}$, is selected as the kernel function with $\gamma=0.1$, and the one-against-one technique is adopted to solve the multi-class problem.

A. Algorithms Performance in AWGN Channel

For the AWGN channel, the channel fading coefficient in (16), (18) and (19) is set to 1, and $\upsilon_n,\eta_n,\upsilon_n$ represent the additive white Gaussian noise, with $\upsilon_n,\eta_n,\upsilon_n\sim\mathcal{N}_c(0,\sigma^2)$. The SNR is defined as $1/\sigma^2$.

Fig. 2 shows the identification performance of the EM^2 algorithm for $K=2$ and $K=3$ in the single-hop (dotted line) and relaying (solid line) scenarios. When $K=2$, the EM^2 algorithm achieves an acceptable identification performance ($P_c\geq0.8$) for SNR $\geq$ 0 dB and SNR $>$ 2 dB under the single-hop and relaying scenarios, respectively. When $K=3$, P_c reaches 0.8 for SNR $=$ 6 dB in the single-hop scenario and SNR $=$ 8 dB in the relaying scenario. The identification performance of the algorithm in [9] is also illustrated in Fig. 2 for comparison. The advantage of the EM^2 algorithm is apparent, especially at low SNR. The identification accuracy of the EM^2 algorithm achieves more than 30% improvement at SNR $=$ 8 dB for both $K=2$ and $K=3$ in both single-hop and relaying scenarios. Furthermore, it is noticed

TABLE II

COMPARISON OF THE SNR WHEN $P_c = 0.8$ FOR THE PROPOSED ALGORITHMS

EM^2	CB	FDR	[9]	AWGN				Non-Gaussian				Fading			
$K = 2$	Single-hop	-1 dB	-1 dB	-3 dB	15 dB	1 dB	0 dB	-1 dB	18 dB	4 dB	0 dB	0 dB	25 dB
	Relaying	2 dB	0 dB	1 dB	18 dB	4 dB	4 dB	3 dB	21 dB	9 dB	7 dB	4 dB	N/A
$K = 3$	Single-hop	5 dB	4 dB	-2 dB	20 dB	6 dB	6 dB	0 dB	22 dB	9 dB	8 dB	0 dB	N/A
	Relaying	7 dB	7 dB	2 dB	22 dB	9 dB	9 dB	4 dB	N/A	12 dB	13 dB	5 dB	N/A

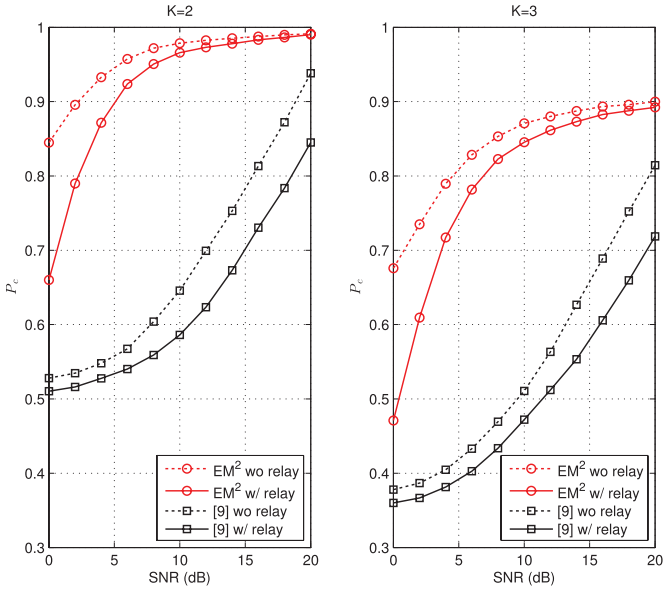

Fig. 2. The identification performance of the EM^2 algorithm in AWGN channel: P_c versus SNR for $K = 2$ and $K = 3$, with $\bar{N}_0 = N_0 = 50$, w/wo relay.

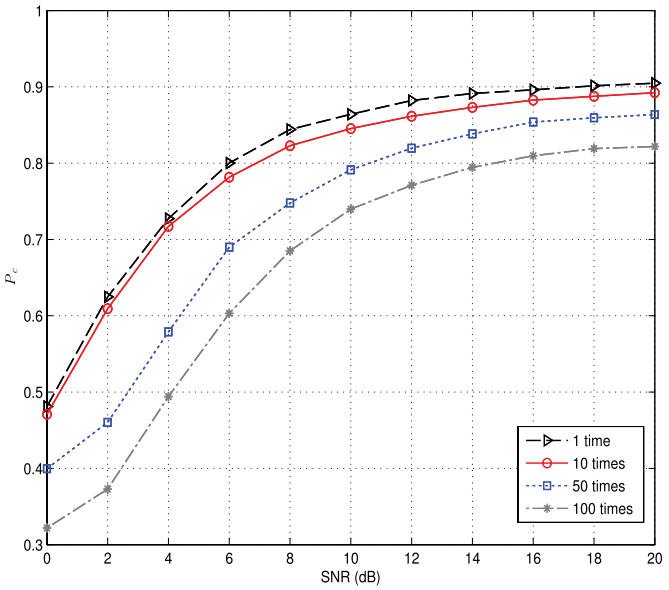

Fig. 3. The identification performance of the EM^2 algorithm in AWGN channels: P_c versus SNR for $K = 3$, with $\bar{N}_0 = N_0 = 50$, w relay. The time and frequency resolutions Δt and $\Delta\omega$ is $1, 10, 50, 100$ times the time and frequency resolutions of the Hilbert spectrum.

that the identification performance of the EM^2 algorithm in the single-hop and relaying scenarios is similar at high SNR, while the algorithm in [9] presents a gap of 10% between these cases; this indicates that the EM^2 algorithm can combat the negative effect of the relay on the fingerprint.

Fig. 3 illustrates how the time and frequency resolutions Δt and $\Delta\omega$, used to calculate the energy entropy, affect the identification performance of the EM^2 algorithm; P_c is plotted as a function of SNR, with curves parameterized by Δt and $\Delta\omega$. The case when $K = 3$ in the relaying scenario is investigated. It can be seen that with the increase of the time and frequency resolutions,[12] the identification performance enhances. It is worth mentioning that the high resolution corresponds to high computational complexity, which means that the selection of the time and frequency resolutions is a trade-off between identification performance and computational complexity. Additionally, when compared with the result when the time and frequency resolutions Δt and $\Delta\omega$ equal those of the Hilbert spectrum, the performance when Δt and $\Delta\omega$ is 10 times those of the Hilbert spectrum barely degrades, while the complexity is lower; this indicates that 10 times is a reasonable value to be chosen.

In Fig. 4, P_c is plotted as a function of the number of training sequences for each class $\bar{N}_0$, with curves parameterized by the number of test sequences for each class N_0. The case of $K = 2$ and $K = 3$ are investigated in the relaying scenario, at SNR $= 14$ dB. It can be seen that the identification performance is reasonably robust against the training and test sequences. When N_0 is fixed, P_c only slightly increases with $\bar{N}_0$, especially when $\bar{N}_0 \geq 70$, where P_c has a ceiling for $K = 2$ and $K = 3$, respectively. This implies that the identification performance is essentially independent of the number of training sequences. In addition, when $\bar{N}_0$ is fixed, the identification performance degrades no more than 2% for $K = 2$ and decreases less than 3% for $K = 3$, which indicates that the number of test sequences barely influences the identification accuracy.

For the case when different emitters employ different modulation formats, the identification performance is illustrated in Fig. 5. For $K = 2$, the modulation types of the two emitters are 4-QAM and 16-QAM, respectively, and for $K = 3$, the emitters utilize 4-QAM, 16-QAM and 64-QAM, respectively. Note that the receiver intends to classify the emitters by extracting features that represent the fingerprints of the emitters, and does not apply any modulation classification technique to obtain the modulation format, such as in [38]–[42]; the effect of the modulation type, which is unknown at the receiver, is shown here. From Fig. 5, we can see that the SEI performance with different modulation formats outperforms that with the same modulation type in both single-hop and relaying scenarios. This can be easily explained, as the difference between different emitters

[12]Note that the time and frequency resolutions, Δt and $\Delta\omega$, are multiple integers of those of the Hilbert spectrum. The lower these integers are, the higher the resolutions are.

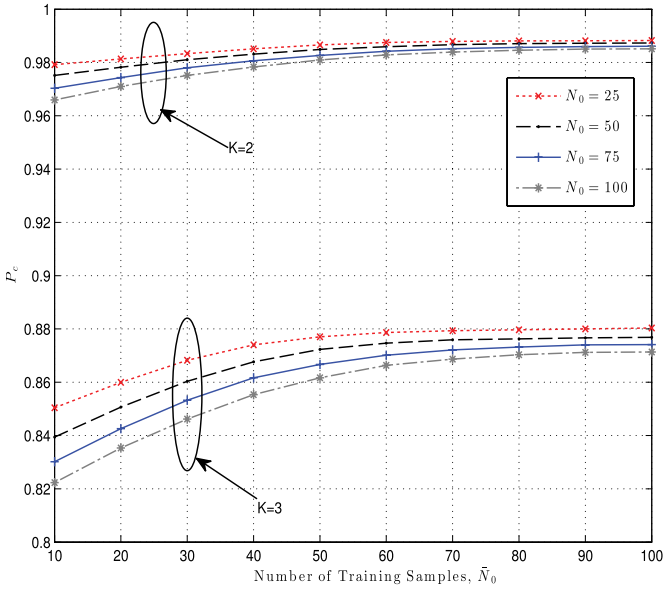

Fig. 4. The identification performance of the EM^2 algorithm in AWGN channel: P_c versus the number of training sequences for each class $\bar{N}_0$ for $K = 2$ and $K = 3$ in the relaying scenario. The number of test sequences for each class is $N_0 = 25, 50, 75, 100$ and SNR $= 14$ dB.

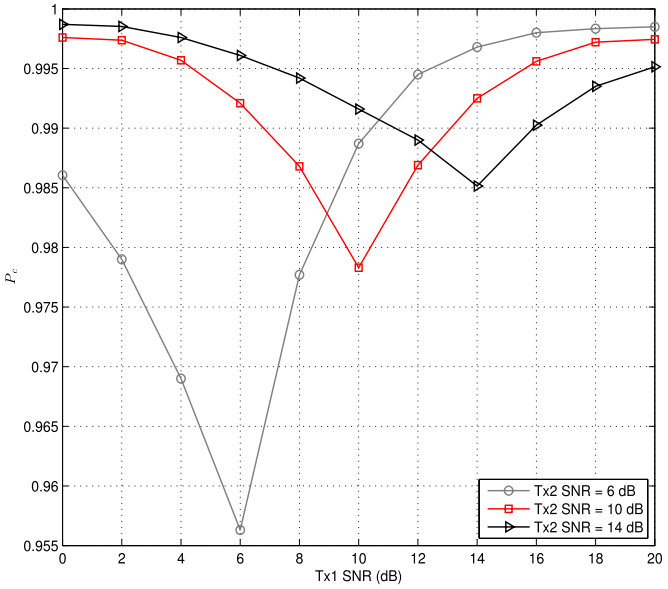

Fig. 6. The identification performance of the EM^2 algorithm when emitters use different powers: P_c versus SNR for $K = 2$, with $\bar{N}_0 = N_0 = 50$, wo relay.

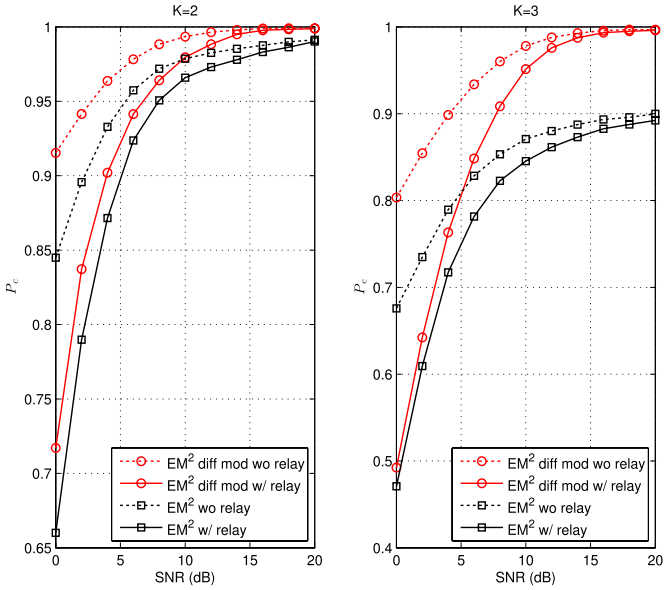

Fig. 5. The identification performance of the EM^2 algorithm when emitters use different modulation formats: P_c versus SNR for $K = 2$ and $K = 3$, with $\bar{N}_0 = N_0 = 50$, w/wo relay. For $K = 2$, emitters use 4-QAM and 16-QAM. For $K = 3$, emitters use 4-QAM, 16-QAM, and 64-QAM.

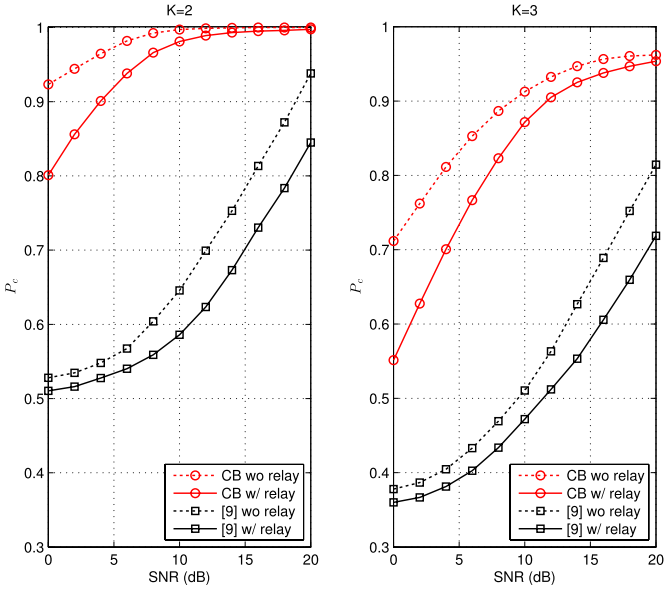

Fig. 7. The identification performance of the CB algorithm in AWGN channel: P_c versus SNR for $K = 2$ and $K = 3$, with $\bar{N}_0 = N_0 = 50$, w/wo relay.

becomes more significant when emitters employ different modulation formats, which makes the identification task easier.

The identification performance when different emitters use different powers is evaluated in Fig. 6. By taking $K = 2$ under the single-hop scenario as example, P_c is plotted versus the SNR of the signals from emitter 1, with curves parameterized by the SNR of the signals from emitter 2. As it can be seen, the worst identification performance is achieved when the powers of the two emitters are the same. As the difference in the transmit powers becomes larger, the identification performance improves significantly. This can be easily explained as the transmit power translates into the SNR at the receiver, which represents an effective feature for SEI; the difference between the transmit powers enhances the separability of transmitted signals, which is beneficial to identification.

Fig. 7 shows the identification performance of the CB algorithm in AWGN channels for $K = 2$ and $K = 3$ in the single-hop and relaying scenarios. When $K = 2$, the CB algorithm provides a good identification performance ($P_c \geq 0.8$) for SNR ≥ 0 dB in both scenarios. When $K = 3$, the identification performance attains $P_c = 0.8$ for SNR $= 4$ dB in the single-hop scenario and for SNR $= 7$ dB in the relaying scenario. Furthermore, when compared with the identification performance of the EM^2 algorithm shown in Fig. 2, the CB algorithm achieves a gain of 2 dB for

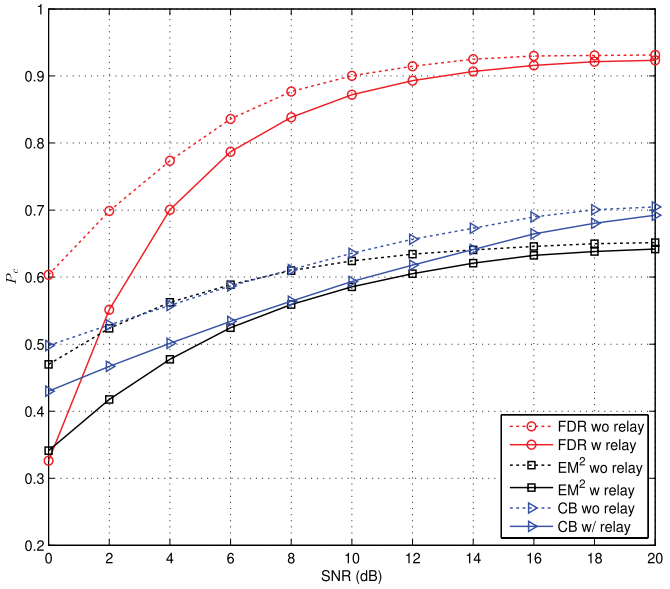

Fig. 8. The identification performance of the FDR algorithm in AWGN channel: P_c versus SNR for $K = 5$, with $\bar{N}_0 = N_0 = 50$, w/wo relay, $S = 50$.

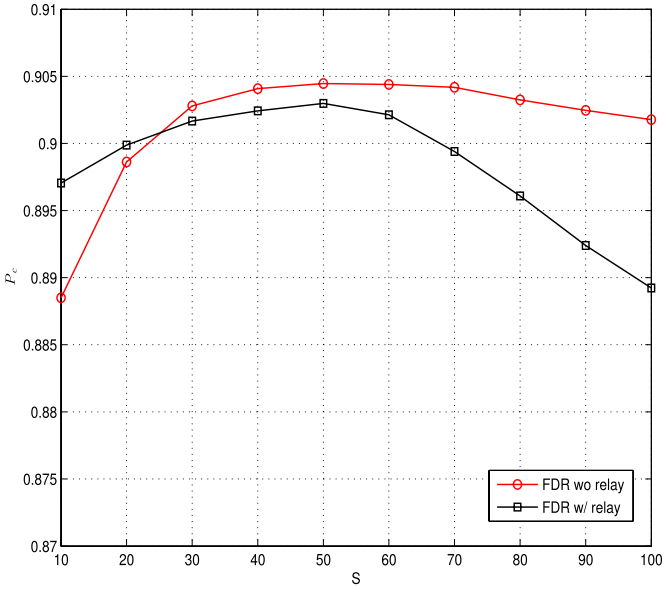

Fig. 9. The identification performance of the FDR algorithm in AWGN channel: P_c versus the number of time-frequency spots chosen for each combination S for $K = 5$, w/wo relay. SNR = 14 dB.

$P_c \geq 0.8$ when $K = 2$; when $K = 3$, P_c of the CB algorithm has a 5% improvement over that of the EM^2 algorithm.

Fig. 8 illustrates the identification performance of the FDR algorithm, along with that of the EM^2 and CB algorithms in AWGN channels when $K = 5$, both in the single-hop and relaying scenarios. The number of optimal time-frequency spots for each combination is set to $S = 50$. It is clear that since the FDR algorithm extracts elements with strong separability for identification, a reliable performance is achieved even for $K = 5$. In such a case, $P_c \geq 0.8$ is obtained for SNR = 5 dB in the single-hop scenario and for SNR = 7 dB in the relaying scenario. On the other hand, the identification performance of the EM^2 and CB algorithms degrades dramatically, as the features tend to be similar to each other when the number of emitters increases.

In Fig. 9, P_c is plotted as a function of the number of time-frequency spots chosen for each combination, S. The case of $K = 5$ under both single-hop and relaying scenarios at SNR = 14 dB is investigated. It can be seen that the identification performance of the FDR algorithm is not significantly sensitive to the change of S, especially in the single-hop scenario. When S is small, P_c first increases as S increases, then when S is large enough to effectively identify different emitters, it achieves a ceiling, after which it decreases slightly, because more featureless Hilbert spectrum elements are included as S increases. It is noticed that P_c is better when $30 \leq S \leq 60$ under both the single-hop and relaying scenarios; a reasonable value of S can be selected within range.

Table III summarizes the identification performance of the three proposed algorithms, along with that of the algorithm in [9], for SNR $= 4, 12$ and 20 dB in AWGN channels. It can be seen that when $K = 2$ and $K = 3$, all the proposed algorithms attain good identification performance in both single-hop and relaying scenarios. As expected, the FDR algorithm obtains the best identification performance, then followed by the CB algorithm and finally the EM^2 algorithm. The FDR algorithm can effectively identify emitters when $K = 5$ since it extracts features with strong separability. In addition, the identification performance of the proposed algorithms outperforms that of the algorithm in [9], especially in the relaying scenario.

B. Algorithms Performance Comparison in Non-Gaussian Noise Channel

Here we compare the identification performance of the proposed algorithms, along with the one in [9], in the presence of non-Gaussian noise. The received signal is given by (16), (18) and (19), with the distribution of ν_n, η_n and υ_n following the Middleton Class A model, i.e., with the probability distribution function (pdf) given by [43], [44]

$$f_{ClassA}(x) = e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!\sqrt{(2\pi\sigma_m^2)}} e^{-\frac{x^2}{2\sigma_m^2}}, \qquad (29)$$

where A is the impulse index, and $\sigma_m^2 = \frac{\frac{m}{A}+\Gamma}{1+\Gamma}$ is the noise variance with Γ as the ratio of of the intensity of the independent Gaussian component and the intensity of the impulsive non-Gaussian component. In simulations, A and Γ are set to 0.1 and 0.05, respectively. Furthermore, the number of terms in the Class A pdf is finite and set to $M = 500$, i.e., $m \in [0, M-1]$ [43]. No fading is considered.

The identification performance of the three proposed algorithms and the algorithm in [9], for SNR $= 4$, 12 and 20 dB in non-Gaussian noise channels is summarized in Table IV. When compared with results for $K = 2$ and $K = 3$ shown in Table III (AWGN case), it can be seen that the proposed algorithms have little loss in the identification performance in the single-hop scenario. It is also noticed that in the relaying scenario, the effect of the non-Gaussian noise leads to more reduction in the identification performance of the proposed algorithms. However, with the increase of the SNR, the negative effect of the non-Gaussian noise is overcome. The identification performance degradation may virtually be ignored

TABLE III
IDENTIFICATION PERFORMANCE P_c FOR SNR = 4, 12, 20 dB IN AWGN CHANNELS

EM^2	CB	FDR	[9]	4 dB				12 dB				20 dB			
$K=2$	Single-hop			0.933	0.965	0.992	0.548	0.983	0.999	1.000	0.700	0.991	1.000	1.000	0.938
	Relaying			0.872	0.901	0.935	0.528	0.973	0.989	0.997	0.623	0.990	0.998	0.999	0.845
$K=3$	Single-hop			0.789	0.811	0.972	0.405	0.880	0.933	0.994	0.563	0.899	0.962	0.998	0.814
	Relaying			0.717	0.701	0.889	0.381	0.862	0.905	0.981	0.512	0.892	0.954	0.994	0.719
$K=5$	Single-hop			0.562	0.558	0.774	0.254	0.634	0.656	0.915	0.275	0.652	0.705	0.931	0.327
	Relaying			0.477	0.501	0.700	0.247	0.605	0.618	0.893	0.260	0.642	0.692	0.923	0.297

TABLE IV
IDENTIFICATION PERFORMANCE P_c FOR SNR = 4, 12, 20 dB IN NON-GAUSSIAN NOISE CHANNELS

EM^2	CB	FDR	[9]	4 dB				12 dB				20 dB			
$K=2$	Single-hop			0.903	0.914	0.975	0.548	0.978	0.987	0.997	0.699	0.989	0.996	0.999	0.938
	Relaying			0.822	0.825	0.857	0.528	0.964	0.967	0.991	0.623	0.980	0.990	0.996	0.845
$K=3$	Single-hop			0.736	0.753	0.941	0.405	0.861	0.913	0.989	0.563	0.891	0.953	0.995	0.814
	Relaying			0.607	0.624	0.808	0.381	0.845	0.869	0.966	0.512	0.881	0.942	0.986	0.719
$K=5$	Single-hop			0.527	0.503	0.729	0.254	0.609	0.617	0.896	0.275	0.635	0.689	0.925	0.327
	Relaying			0.407	0.431	0.628	0.247	0.567	0.574	0.875	0.260	0.619	0.672	0.912	0.297

TABLE V
IDENTIFICATION PERFORMANCE P_c FOR SNR = 4, 12, 20 dB IN FADING CHANNELS

EM^2	CB	FDR	[9]	4 dB				12 dB				20 dB			
$K=2$	Single-hop			0.811	0.899	0.966	0.514	0.957	0.978	0.995	0.556	0.990	0.998	0.999	0.681
	Relaying			0.628	0.723	0.818	0.501	0.904	0.904	0.986	0.525	0.982	0.984	0.993	0.591
$K=3$	Single-hop			0.650	0.728	0.905	0.363	0.855	0.885	0.984	0.401	0.898	0.933	0.995	0.544
	Relaying			0.628	0.500	0.770	0.348	0.804	0.764	0.955	0.385	0.982	0.917	0.983	0.451
$K=5$	Single-hop			0.502	0.488	0.712	0.208	0.598	0.611	0.891	0.229	0.619	0.662	0.918	0.285
	Relaying			0.353	0.396	0.579	0.204	0.548	0.549	0.860	0.214	0.603	0.639	0.906	0.248

for SNR $\geq$ 12 dB, which indicates that the proposed algorithms are applicable under non-Gaussian noise conditions. Furthermore, it can be seen that the FDR algorithm is robust against the non-Gaussian noise channel when $K = 5$ under both scenarios, and outperforms the EM^2 and CB algorithms. All proposed algorithms achieve an improved identification performance when compared with the algorithm in [9], which performs poorly especially in the relaying scenario.

C. Algorithms Performance Comparison in Flat-Fading Channel

The identification performance in flat-fading channel is additionally examined. The received signal is given by (16), (18) and (19), with $\nu_n, \eta_n, \upsilon_n \sim \mathcal{N}_c(0, \sigma^2)$, and the channel fading coefficients H_{sd}, H_{sr} and H_{rd} as zero-mean independent Gaussian random variables with variance σ_H^2. The SNR is defined as σ_H^2/σ^2. It is worth noting that the channel fading coefficients are unknown at the receiver.

Table V lists the identification performance of the three proposed algorithms and the algorithm in [9], for SNR = 4, 12 and 20 dB in fading channels for both single-hop and relaying scenarios. When compared to results in Table III (AWGN case), it can be seen that the fading degrades the performance. A more severe degradation is noticed for the CB algorithm in the relaying scenario. The primary reason is that the similarity between Hilbert spectra is corrupted, leading to less accurate identification performance. On the other hand, for the FDR algorithm, a relatively reduced performance degradation is noticed in both scenarios when compared with the AWGN case, which indicates that this is applicable in fading channels. Furthermore, the proposed algorithms significantly outperform the algorithm in [9], which basically fails under fading conditions.

D. Complexity Analysis

The computational complexity of the three proposed algorithms and the algorithm in [9] is shown in Fig. 10. From Section IV-D, it can be seen that the computational complexity is related to the number of emitters K, the total number of training and test sequences from each emitters N_s, i.e., $N_s = \bar{N}_0 + N_0$, and the size of the Hilbert spectrum matrix N_H. Here we omit G for the EM^2 algorithm, as G is either of lower or same order as N_H, i.e., the complexity order of the EM^2 algorithm is $\mathcal{O}(KN_sN_H)$. Moreover, as N_s is of higher order than $\bar{N}_0$ and N_0, the complexity orders of the CB algorithm and the algorithm in [9] are expressed as $\mathcal{O}\left(K^2N_s^2N_H\right)$ and $\mathcal{O}\left(K^2N_s^2N_H^2\right)$, respectively. Fig. 10 (a) illustrates the complexity versus the number of emitters K. Parameters N_s and $N_{\mathcal{H}}$ are set to $N_s = 100$ and $N_H = 2 \times 10^5$ in simulation, respectively. It is noticed that the FDR algorithm has the highest complexity among the proposed algorithms, followed by CB and EM^2, i.e., the improvement in the identification performance is at the cost of an increase in the computational complexity. Furthermore, it

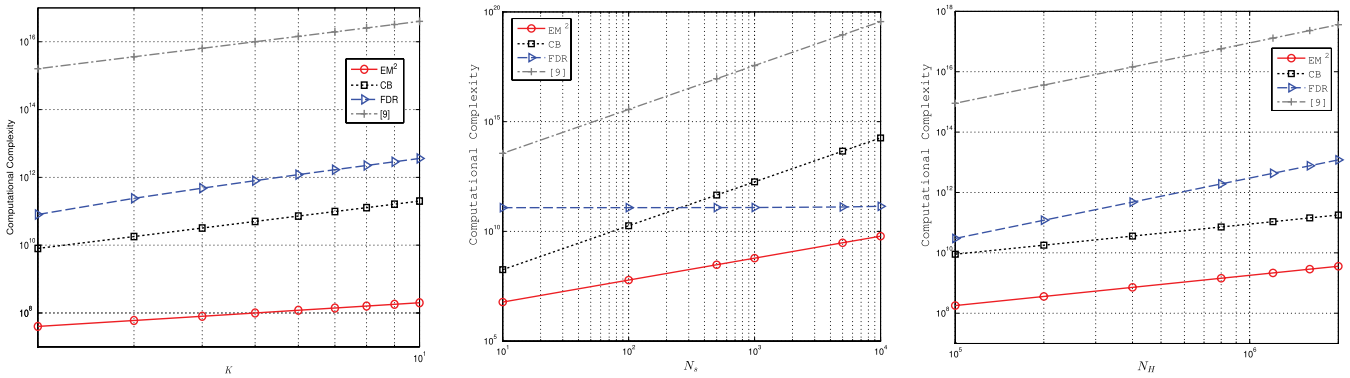

Fig. 10. The computational complexity of the proposed algorithms and the algorithm in [9]: (a) versus K, with $N_s = 100$ and $N_H = 2 \times 10^5$; (b) versus N_s, with $K = 3$ and $N_H = 2 \times 10^5$; and (c) versus N_H, with $K = 3$ and $N_s = 100$.

can be seen from Fig. 10 (b) and Fig. 10 (c) that the complexity of the EM^2 and CB algorithms is more affected by N_s, whereas that of the FDR algorithm is highly related to N_H. Given $N_H = 2 \times 10^5$, when $N_s < 4 \times 10^2$, the complexity of the CB algorithm is lower than that of the FDR algorithm. When N_s increases, the complexity of the EM^2 algorithm is approaching that of the FDR algorithm. Furthermore, the complexity of all the proposed algorithms is lower than that of the algorithm in [9] as expected, since the algorithm in [9] involves performing high computational PCA.

VI. Conclusion

In this paper, we propose three algorithms for SEI, i.e., EM^2, CB and FDR, which extract identification features from the Hilbert spectrum. When compared with the conventional method, the proposed algorithms provide an improved performance in AWGN, non-Gaussian and fading channels, for both single-hop and relaying scenarios; the conventional method fails under fading conditions. Out of the three proposed algorithms, FDR provides the best performance, with increased, yet acceptable complexity. Due to the good performance of the proposed algorithms, we consider to further investigate them in our future works, e.g., by studying the SEI problem under receiver imperfections. Moreover, in the relaying scenarios, we will further consider the SEI problem with the relay following the decode-and-forward strategy, and investigate the system model that the emitters and receiver have direct transmission links.

Acknowledgement

The authors are grateful to the anonymous reviewers and the Editor, Dr. Athanasios Vasilakos, for their constructive comments.

References

[1] J. Zhang, F. Wang, Z. Zhong, and O. Dobre, "Novel Hilbert spectrum-based specific emitter identification for single-hop and relaying scenarios," in *Proc. IEEE Globecom*, 2015, pp. 1–6.

[2] K. I. Talbot, P. R. Duley, and M. H. Hyatt, "Specific emitter identification and verification," *Technol. Rev. J.*, pp. 113–133, Jan. 2003.

[3] R. G. Wiley, *ELINT: The Interception and Analysis of Radar Signals*. London, U.K.: Artech House, 2006.

[4] K. Kim, C. M. Spooner, I. Akbar, and J. H. Reed, "Specific emitter identification for cognitive radio with application to IEEE 802.11," in *Proc. IEEE GLOBECOM*, Nov./Dec. 2008, pp. 1–5.

[5] Z. Zhang, K. Long, and J. Wang, "Self-organization paradigms and optimization approaches for cognitive radio technologies: A survey," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 36–42, Apr. 2013.

[6] H. C. Choe *et al.*, "Novel identification of intercepted signals from unknown radio transmitters," in *Proc. SPIE Wavelet Appl. II*, vol. 2491. Apr. 1995, pp. 504–516.

[7] O. Ureten and N. Serinken, " Bayesian detection of Wi-Fi transmitter RF fingerprints," *IET Elect. Lett.*, vol. 41, no. 6, pp. 373–374, Mar. 2005.

[8] Y. Huang and P. M. Djuric, "Bayesian detection of transient signals in colored noise," in *Proc. ICASSP*, 2000, pp. 725–728.

[9] S. Xu, "On the identification technique of individual transmitter based on signal prints," Ph.D. dissertation, Dept. Elect. Eng., Huazhong Univ. Sci. Technol., Wuhan, China, 2007.

[10] X.-D. Zhang, Y. Shi, and Z. Bao, "A new feature vector using selected bispectra for signal classification with application in radar target recognition," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1875–1885, Sep. 2001.

[11] Z. Zhang, K. Long, J. Wang, and F. Dressler, "On swarm intelligence inspired self-organized networking: Its bionic mechanisms, designing principles and optimization approaches," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 513–537, Feb. 2014.

[12] G. López-Risueño, J. Grajal, and A. Sanz-Osorio, "Digital channelized receiver based on time-frequency analysis for signal interception," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 879–898, Jul. 2005.

[13] J. Lunden and V. Koivunen, "Automatic radar waveform recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 124–136, Jun. 2007.

[14] B. W. Gillespie and L. E. Atlas, "Optimizing time-frequency kernels for classification," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 485–496, Mar. 2001.

[15] L. Li, H.-B. Ji, and L. Liang, "Quadratic time-frequency analysis and sequential recognition for specific emitter identification," *IET Signal Process.*, vol. 5, no. 6, pp. 568–574, Sep. 2011.

[16] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A*, vol. 454, pp. 903–995, Mar. 1998.

[17] J. Sydir and R. Taori, "An evolved cellular system architecture incorporating relay stations," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 115–121, Jun. 2009.

[18] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[19] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[20] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[21] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[22] G. Madzarov and D. Gjorgjevikj, "Multi-class classification using support vector machines in decision tree architecture," in *Proc. IEEE EUROCON*, May 2009, pp. 288–295.

[23] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, May 2004.
[24] B. Nadler, "Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2746–2756, May 2010.
[25] F. Haddadi, M. Malek-Mohammadi, M. M. Nayebi, and M. R. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452–457, Jan. 2010.
[26] T. L. Carrol, "A nonlinear dynamics method for signal identification," *Chaos*, vol. 17, no. 2, p. 023109, Jun. 2007.
[27] C. Liu, H. Xiao, Q. Wu, and F. Li, "Linear RF power amplifier design for wireless signals: A spectrum analysis approach," in *Proc. IEEE ICASSP*, Apr. 2003, pp. 568–571.
[28] L. W. Couch, II, *Digital and Analog Communication Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
[29] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[30] H. Yu, M. Li, H.-J. Zhang, and J. Feng, "Color texture moments for content-based image retrieval," in *Proc. Int. Conf. Image Process.*, Jun. 2002, pp. 929–932.
[31] M. A. Stricker and M. Orengo, "Similarity of color images," in *Proc. SPIE Storage Retr. Image Video Database III*, vol. 2420. Mar. 1995, pp. 381–392.
[32] J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-98-14, Apr. 1998.
[33] L. Bottou and C. Lin, "Support vector machine solvers," *Large Scale Kernel Mach.*, pp. 301–320, 2007.
[34] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[35] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002.
[36] D. Xu, "Research on mechanism and methodology of specific emitter identification," Ph.D. dissertation, Dept. Elect. Eng., Nat. Univ. Defense Technol., Changsha, China, 2008.
[37] C. Chang and C. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin
[38] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov–Smirnov test," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2324–2332, Aug. 2010.
[39] O. A. Dobre, "Signal identification for emerging intelligent radios: Classical problems and new challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 2, pp. 11–18, Apr. 2015.
[40] F. Hameed, O. A. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.
[41] O. A. Dobre, M. Oner, S. Rajan, and R. Inkol, "Cyclostationarity-based robust algorithms for QAM signal identification," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 12–15, Jan. 2012.
[42] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.
[43] *RFI/Impulsive Noise Toolbox 1.2 for MATLAB*. [Online]. Available: http://user.ece.utexas.edu/bevans/projects/rfi/software/index.html, accessed Apr. 7, 2009.
[44] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods an results for class A and class B noise models," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999.

Jingwen Zhang received the B.E. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2012, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Rail Traffic Control and Safety.

She has been a Visiting Scholar with the Electrical Engineering Department, University of California at Los Angeles, Los Angeles, CA, USA, since 2015. Her current research interests are in the field of cognitive radio and signal processing.

Fanggang Wang received the B.Eng. and Ph.D. degrees from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, in 2005 and 2010, respectively.

He was a Visiting Scholar with the Electrical Engineering Department, Columbia University, New York City, New York, USA, from 2008 to 2010. He was a Postdoctoral Fellow with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong, from 2010 to 2012. He joined the State Key Laboratory of Rail Traffic Control and Safety, School of Electronic and Information Engineering, Beijing Jiaotong University, in 2010, where he is currently an Associate Professor. His research interests are in wireless communications, signal processing, and information theory. He chaired two workshops on wireless network coding (NRN 2011 and NRN 2012) and served as an Editor for several journals and the Technical Program Committee Member for several conferences.

Octavia A. Dobre (M'05–SM'07) was with the New Jersey Institute of Technology, USA, and the Polytechnic Institute of Bucharest, Romania. She was a Visiting Professor with the Université de Bretagne Occidentale, France, and the Massachusetts Institute of Technology, USA, in 2013. She is currently an Associate Professor with the Faculty of Engineering and Applied Science, Memorial University, Canada. She has coauthored over 170 journal and conference papers in these areas and gave over 40 invited and keynote talks to the industry and academia. Her research has been supported by the Natural Sciences and Engineering Research Council of Canada, the Mathematics of Information Technology and Complex Systems, the Canada Foundation for Innovation, Research and Development Corporation, the Atlantic Canada Opportunities Agency, Defence and Research Development Canada, the Communications Research Centre Canada, Altera Corporation, DTA Systems, ThinkRF, and Agilent Technologies. Her research interests include blind signal identification and parameter estimation techniques, cognitive radio systems, spectrum sensing techniques, transceiver optimization algorithms, dynamic spectrum access, and cooperative wireless communications, as well as underwater and optical communications. She was a recipient of a Royal Society Scholarship at Westminster University, U.K. (2000), and a Fulbright Fellowship at the Stevens Institute of Technology, Hoboken, USA (2001). She serves as the Editor-in-Chief of the IEEE COMMUNICATIONS LETTERS, and an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. She was a Senior Editor and an Editor of the IEEE COMMUNICATIONS LETTERS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and a Guest Editor of the *IEEE Communications Magazine* and the IEEE JOURNAL OF SELECTED TOPICS ON SIGNAL PROCESSING. She served as the General Chair of CWIT, and the Cochair of Technical Symposia at numerous conferences, such as the IEEE VTC, the IEEE GLOBECOM, CrownCom, the IEEE WCNC, and the IEEE ICC. She is the Chair of the Women in Communications Engineering ComSoc Standing Committee, and a registered Professional Engineer in the province of Newfoundland and Labrador, Canada.

Zhangdui Zhong received the B.E. and M.S. degrees from Beijing Jiaotong University, Beijing, China, in 1983 and 1988, respectively. He is currently a Professor, the Director of the School of Computer and Information Technology, and a Chief Scientist of the State Key Laboratory of Rail Traffic Control and Safety with Beijing Jiaotong University. He is also the Director of the Innovative Research Team with the Ministry of Education, Beijing, and a Chief Scientist of the Ministry of Railways, Beijing. He has authored or coauthored seven books, five invention patents, and over 200 scientific research papers in his research area. His interests include wireless communications for railways, control theory and techniques for railways, and GSM-R systems. He was a recipient of the Mao Yisheng Scientific Award of China, the Zhan TianYou Railway Honorary Award of China, and the Top 10 Science/Technology Achievements Award of Chinese Universities.