Artificial Intelligence-Based Cyber-Physical Events Classification for Islanding Detection in Power Inverters

Mohammad Babakmehr, Member, IEEE, Farnaz Hariri, Member, IEEE, Payman Dehghanian, Member, IEEE, and Johan Enslin, Fellow, IEEE

1Clemson University, Restoration Institute, Charleston, SC, 2George Washington University, Washington DC

Abstract—Along with the rapid integration of distributed generation units (DGU) into the power grids is the rise in unconventional and unpredictable patterns of the undesirable cyber-physical intrusions and faults; this dramatically increases the risk of islanding possibilities and threatens the sustainability of the energy delivery infrastructure. Classification of cyber-physical events and developing solutions to mitigate their impacts before rising to an islanding situation is a critical monitoring task in DGUs. Passive islanding detection has been widely applied to studying the behavior of voltage signals at the point of common coupling, which is a sophisticated challenge due to cross-similarity among fault (event) patterns and their fast dynamics. In this paper, a novel quadratic Time-frequency decomposition, namely HSS-Transform, is applied over an alternative complex representation of 3-phase signal defined by the synchronous reference frame transformation. We further exploit the principles of Informative Sparse representation-based Classification (TISC) to develop a comprehensive artificial intelligence framework for fast and reliable classification of DGU islanding and non-islanding events with the focus on practical limitations and requirements of a smart power electronics inverter as the desirable observational site. Different from the state-of-the-art techniques, TISC does not need any training procedure, while due to its linear mathematical formulation acts inherently fast with low computational burden on the inverter processing unit. Moreover, the simultaneous 3-phase feature extraction strategy ensures preservation of the between-phase information.

Index Terms—Smart grids, Cyber-physical events, Power electronic inverters, Islanding detection, Time-frequency analysis, Pattern recognition, Sparse classification, Artificial intelligence.

I. INTRODUCTION

SMART grid technologies have revolutionized the traditional architecture of the power grids from a unidirectional single-layer physical system into a multilayer interconnected network system with a real-time two-way communications between generation, transmission, distribution systems, and loads [1]. The rise in the (mostly renewable-based) distributed generation resources (such as wind and solar) in addition to smart loads (aka prosumers) at the distribution level, traces a decentralized architecture for the grid, where we face not only a two-way flow of information but also a two-way flow of energy, aka Active Distribution Networks (Fig. 1.a) [2]. In modern active distribution networks of the future, advanced power electronics and Artificial Intelligence-based (AI) technologies such as the Internet of Things (IoT) will play a major role [3].

Power generation decentralization on the one hand, and integration of cyber-based internet-accessible platforms on the other hand, lights up a variety of unconventional cyber-physical monitoring, control, and security challenges in DGU-enabled architectures [4]-[5]. In such a highly distributed cyber-energy environment, the real-time fault or in general, “event” analysis plays a critical role to better improve both cyber and physical situational awareness and correspondingly enhance the security and sustainability of the grid. Some such events can not only result in catastrophic damages to the transmission and distribution equipment, but also significantly increase the risk of unintentional islanding situations in DGUs which may correspondingly cause wildfires and large-scale blackouts.

Bridge between DGUs and the main grid [6], power electronic inverters are most desirable monitoring sites for both cyber and physical event recognition. Furthermore, interfacing with energy storage units and other advanced technologies such as electric vehicles and IoT services opens new pathways into the capabilities of inverters as ideal observation points for distributed control across the grid [6]-[7]. In this regard, there are two major challenges towards a sustainable active distribution network: 1) Data integrity and interoperability, 2) Cybersecurity [8]. Both concepts are tightly related to the reliability of data collection and transfer. However, considering the enormous amount of information that may arrive in control rooms from hundreds of inverters every second, any central monitoring and control framework would have to be equipped with big data analytics with various data processing and analysis challenges [9]. Nowadays, commercial inverters are taking both single and 3-phase measurements from electrical currents and voltage waveforms for local control [10]. As such, on top of their daily control functionalities, these measurements can be used to perform cyber-physical event detection, classification, and system protection in a decentralized manner while only the high-level reports can be communicated to the utility (Fig. 1). This has led to PMU-embedded inverter architectures [10]. The passive islanding detection through cyber-physical events classification is among the emerging trends in power electronics applications, well-suited for smart inverters within the futuristic power systems [11].

A. Literature Review and the State of the Art

Recently, advanced signal processing and artificial intelligence approaches have been widely studied for event detection and classification in power systems within different levels and a variety of applications such as islanding detection and power quality event classification, from generation sites [12] to transmission lines [13]-[14] all the way down to the small-scale PV-based microgrids [15]. In the literature, power system events classification is typically considered as a general pattern recognition problem, which can be divided into the following standard 5-step procedure [6]: (1) phenomena recording and raw data preprocessing; (2) potential patterns detection/data segmentation; (3) feature extraction (FE), where Time-Frequency analysis is the most popular such as Short Time Fourier Transform (STFT) [16], Wavelets [17], S-transform [18], Time-Time Transform [19], Mathematical Morphology [20], etc.; (4) feature selection (FS) or dimensionality reduction; and (5) categorization, or classification. A variety of mathematical methods has been
developed to implement each of these steps in the literature. From simple linear to highly complex and nonlinear FE, FS and classification approaches are available, and one may select the optimal approach based on the data characteristics [21]-[22].

B. Challenges and Motivations

For the sake of islanding detection, there are two major sets of unresolved challenges when implementing the state-of-the-art algorithms in power electronic inverters (or any other candidate monitoring device with limited computational resources such as phasor measurement units): (1) Analytical problems, (2) Practical problems.

Analytical problems are general algorithmic issues associated with any pattern recognition problem, including (1.1) Optimal feature extraction that was a major topic of research in islanding detection [21]-[22], (1.2) Feature selection or dimensionality reduction for computational efficiency, (1.3) Optimal classifier training. Practical issues include (2.1) Computational resource limitations, (2.2) Ignorance of the collaborative information along with 3-phase systems, (2.3) Generalizability, and (2.4) Sensitivity to noise and uncertainties.

C. Contributions

In this paper and in response to the above challenges, we first introduce a novel instantaneous complex signal quadratic time-frequency decomposition to extract 3-phase features from a voltage or current signal in 3-phase inverters. We particularly develop an S-transform-based realization for the S-Method [30] to generate time-frequency images of the 3-phase cyber-physical events (CPE). We verify our approach over a comprehensive set of 20 critical (islanding and non-islanding) events including but not limited to Line to Line faults, Line to ground fault, Frequency jump, Phase ramp, Voltage sag caused by generation interruption, Flickers, Arc furnaces, Harmonics, and so on (Please refer to Section V for a full list). These events can be either purely cyber-driven events caused by cyber-attacks or be physical fault-driven events. Inspired by the theory of sparse-based classification [23], we formulate the CPE detection and classification problem in terms of a sparse representation classification problem (SRC) with lower time and implementation complexities compared to the state-of-the-art classification approaches in the literature. SRC has been used in power quality events classification and partial discharge classification in [24]-[25].

The main privileges of the proposed SRP framework, referred to as Time-frequency Informative Sparse representation-based CPEs Classification (TISC-CPE), compared to the state-of-the-art techniques are:

- **Simultaneous 3-Phase Feature extraction:** Exploiting an alternative 2-dimensional orthogonal representation for the 3-phase signals based on Park Transform [26] that contains the instantaneous 3-phase information (2.2).

---

2 In this work we would focus on challenges in 3-phase systems and particularly on power electronic inverters, however, a very similar but simpler approach can be used for single phase systems while our mathematical frameworks can in general be adapted for similar applications along with other devices and observational points in a power network.
IEEE Journal of Emerging and Selected Topics in Power Electronics

- A modified Time-frequency representation that can result in an enhanced set of features called TF images to address the feature extraction limitations in state-of-the-art in covering the wide range of CPE dynamics and behaviors.
- Feature selection independency. Due to the blessing of dimensionality (Section IV), feature space can be reduced by selecting random features without decreasing the classifier accuracy (refer to Section V)(1.2).
- Training-free property. Training data samples are only stacked into a matrix among a linear optimization formulation, where no training procedure is needed (1.3, 2.3).
- Linearity. The presented method results in an inherently fast linear classifier with much lower computational cost (2.1).
- Informative training sample selection. A high dimensional convex hull approximation technique is used to find the prominent training samples (referred to as informative samples) and discard the rest, which correspondingly decreases the data space dimension without significant changes in the classification accuracy (Section IV) (2.1, 1.3).
- Performance enhancement for higher classes. Unlike the conventional classifiers, due to the collaborative formulation of the TISC, increasing event classes results in higher sparsity in the proposed framework and better performance in addition to robustness against noise (2.4, 2.3).

We will demonstrate that, despite its simplicity, TISC has a competitive performance in CPE classification compared to the frontier ANN [27] and SVM [28], while it is easily adaptable to uncertainties that may result in unexpected changes in the data characteristics.

II. NOTATIONS AND PROBLEM DEFINITIONS

Assume for a 110V-60Hz 3-phase inverter a set of $N = \sum_{j=1}^{J} n_j$ labeled 3-phase sinusoidal electrical voltage segments (each of length 15-cycles) has been detected and recorded by a continuous windowing process from $J$ classes of CPEs with a 9.6 kHz sampling frequency (160 sample/cycle) and is available as the training set\(^2\). We consider each of these 3-phase known-class vectors, $y_i \in \mathbb{R}^{3 \times M}, i = 1,2,\ldots,N,$ and $M = 2400,$ to contain a CPE pattern.

The final goal is to develop a framework that takes a feature vector, $d_{\text{test}}^\text{ext}$ extracted from a new detected cyber-physical data point, $y_{\text{test}}$, and assigns or maps it into an individual class of CP events $c_j, j = 1,2,\ldots,J$. We illustrate this combined mapping as a mathematical function $c_i = A(FG_{\text{test}})$, $c_i \in C = \{c_j | j = 1,2,\ldots,J\}$. Roughly speaking, operator $D$ stands for the data-feature mapping resulting from FE and FS procedure, while the feature-class (classification) mapping is represented by the (in general nonlinear) operator $A$ (such a mapping is usually done by training a classifier). The optimal framework for selection of these mappings is unknown in advance and may change based on the system and associated data specifications or applications.

In this paper, first, we develop a new time-frequency decomposition that exploits the flexibilities of the S-transform in time-frequency plane tiling [29] with high resolution of a modified version of Winer-Ville Distribution [30] named S-Method. We aim to develop an optimal feature extraction method that is adapted to address all technical requirements for classifying complicated, yet within a wide range of dynamics and occurrence time range, cyber-physical events and scenarios. This includes short term transient arcs to resistive long-term harmonics. Next, we exploit a modified sparse representation-based classification approach to perform the operator $A$ that decreases the size of the feature vectors in the mapping $F$. Furthermore, we utilize an optimal training set selection approach to find the most informative feature vectors for each class of CPEs. We also discuss that, due to the sparse recovery-based formulation of the proposed method, the classification mapping $A$ is a fast and reliable linear operator compared to the complex and nonlinear methods proposed in the literature, well suited for limited computational resources in power inverters.

We split our methodology into two main Sections as follows: I: Instantaneous 3-phase feature extraction using HSS-Transform, II: Feature selection and classification using informative sparse classifier.

III. INSTANTANEOUS 3-PHASE FEATURE EXTRACTION

In this section, we propose a novel simultaneous 3-Phase Time-Frequency Representation (TFR) to extract distinguished features from different patterns of CPEs that are unique enough for event classification. The major advantages of our proposed approach are: (1) exploiting simultaneous 3-Phase information; (2) using enhanced Wigner-Ville distribution, we aim to preserve the best possible time-frequency resolution; (3) utilizing the Hyperbolic S-transform, we combine the advantageous of time-frequency resolution of STFT and multiscal e resolution of Wavelets altogether.

A. Instantaneous 3-Phase Signal Processing Tools

In a nutshell, time-domain-based instantaneous power theories are 3D mathematical signal mappings (comparable to 1D Fourier/wavelet transform in 1-Phase system analysis) that transforms any 3-phase signal into an alternative coupled (usually) orthogonal feature space at each sample of time. Each coupled component in this new feature space is called an instantaneous power component. Here we use a famous such transform called Synchronous reference frame ($dq$) method to convert a 3-phase signal to an alternative orthogonal mathematical space. We investigate that tracking the associated trajectories of the TFR of 3-phase signals in these mathematical domains can be utilized as a new, fast, and real-time CPE indicator without imposing any power quality issue.

B. Direct-Quadrature Transformation ($dq$)

The Synchronous Reference Frame Method or in short $dq$ transforms maps a 3-phase electrical signal from the abc frame to the synchronous reference frame ($dq$) as follows [16]:

\[
\begin{bmatrix}
    v_d \\
    v_q
\end{bmatrix} = \sqrt{3} \begin{bmatrix}
    \cos(\theta) & \cos(\theta - \frac{2\pi}{3}) & \cos(\theta + \frac{2\pi}{3}) \\
    -\sin(\theta) & -\sin(\theta - \frac{2\pi}{3}) & -\sin(\theta + \frac{2\pi}{3})
\end{bmatrix} \begin{bmatrix}
    v_a \\
    v_b \\
    v_c
\end{bmatrix}
\]

\(2\) According to [3]-[4], a data point constructed from sampling 10-20 cycles of a signal with frequency of 10 kHz is an appropriate segment to represent a cyber-physical event for most classes of cyber-physical events.
where $\theta$ is the synchronization angle, which is time-invariant and represents the angular position of the $dq$ frame, and it is detected by a Phase Locked Loop (PLL). Thanks to the fact of orthogonality of $d$ and $q$ components, we define the following orthogonal complex signal:

$$x = v_d + jv_q$$  \hspace{1cm} (2)

If the system is voltage-balanced, $x$ carries the entire 3-phase information in an alternative mathematical representation. It is worth noting that, by projecting an $N$-dimensional dataset into an $M$-dimension ($N>M$), we will always loose a portion of the information available in the original data space. In this particular problem (3-phase event detection), it is a bit tricky. In general, the $dq$ transform has an extra component named zero component $dq0$. The zero component reflects the unbalanced term of the voltage or current signals in 3-phase vector space. The type of the unbalances can be either homopolar or heteropolar. In fact, homopolar unbalance system is the one where the vectorized sum of the voltages $V_a$, $V_b$ and $V_c$ is not zero. While in the heteropolar condition, the vectorized sum is still equal to zero but the vectors do not satisfy the standard definition of the balance condition where the length of all vectors (namely 110 volt) should be equal and each should be out-phase from the others on the order of 120 degrees.

Our approach can handle heteropolar unbalance since the effect of such an unbalanced condition can be reflected in the $d$ and $q$ components. Even in case of homopolar unbalanced situations, some sources of unbalance terms are appeared in these terms, but we will lose the extra information captured by zero component and this will be the compromise we should take. In bulk level systems and high power (industrial) inverters, unbalance condition is usually not a big concern; however, as we go through the finer levels of power grid structures (residential level), we ignore some sources of extra information if the systems considerably unbalanced.

Now we use the time-frequency representation of $x$ (for any cyber-physical event recorded on a 3-phase voltage) to extract useful features from this complex signal for the sake of classification.

C. Time-Frequency Representation

In the last several decades, Time-Frequency Representation (TFR) has been the most popular approach in studying the non-stationary behavior of signals such as faults. Unlike the Fourier transform, it provides information regarding the distribution of the signals’ energy and power components in both time and frequency domains simultaneously. However, the time-frequency uncertainty principle implies that due to the strong correlation between time and frequency resolution, it is impossible to reach an ideal resolution along both dimensions of the time-frequency plane simultaneously [26]. A compromise is, hence, required which was the motivation for many time-frequency analyses. As such, in general, the optimal TFR approach should be selected concerning the application specifications.

D. Bilinear vs Linear Time-Frequency Decomposition

There are two major classes of TFR widely used in different applications, namely linear and bilinear (aka Cohen class) transforms [30]. Without loss of generality, in a linear TFR, the TF output of the transform is a function of a linear integral of the input signal (3), while in the bilinear or quadratic form, there is a quadrature dependency to the input signal (4):

$$L_{TFR} = \int x(t) \omega_0(t-\tau,f) dt$$  \hspace{1cm} (3)

$$Q_{TFR} = \int x(t+\frac{r}{2})\bar{x}(t+\frac{r}{2}) \omega_0(t-\tau,f) dt$$  \hspace{1cm} (4)

where $x$ is, in general, a complex signal with $\bar{*}$ denotes its conjugate and $\omega_1$ and $\omega_2$ some predesigned functions of time and frequency. In extreme cases $\omega_1$ is simplified to a complex exponential ($e^{-i2\pi ft}$) which will result in the ordinary Fourier transform(5) and $\omega_2$ would be equal to one, so $Q_{TFR}$ will be the autocorrelation function of the signal $x$. During the last two decades, linear TFRs have been vastly deployed in a variety of scientific fields, including Short Time Fourier Transform (STFT), (6), Wavelet Transforms (WT), (7), and, more recently Stockwell or in short S-Transform (9). Although linearity is always a desirable property in signal and systems, the quadratic structure of a TFR is an intuitively reasonable assumption when one is going to interpret a TFR as time-frequency energy distribution. To maintain the quadratic dependence on the signal, we often employ the squared amplitude modulus of the STFT or WT called spectrogram (SP=|STFT|) and scalogram (SCA=|WT|), respectively. Mathematically speaking, for a joint TF decomposition, Quadratic or Bilinear TFRs provide better resolution with most desirable mathematical properties (especially preserving the time and frequency shift). Among which the Wigner-Ville distribution (WVD) results in the best time-frequency resolution concerning the uncertainty principle. However, quadratic methods, in general, suffer from the cross-interference components in addition to high computational complexity. As such their applications were mostly limited to certain case studies. Here are the common mathematical formulation of FT, STFT, WT, and WVD, respectively:

**FT:**

$$X(f) = \int x(t') e^{-i2\pi ft'} dt'$$  \hspace{1cm} (5)

**STFT:**

$$X_g(t,f) = \int x(t') g(t'-t) e^{-i2\pi ft'} dt'$$  \hspace{1cm} (6)

**CWT:**

$$X_q(t,a) = \int x(t') \sqrt{|a|} \psi^*(a(t'-t)) dt'$$  \hspace{1cm} (7)

**WVD:**

$$X(t,f) = \int x(t+\frac{t'}{2})\bar{x}(t+\frac{t'}{2}) e^{-i2\pi ft'} dt'$$  \hspace{1cm} (8)

where $g$ is a predesigned window such as Gaussian, $\psi$ is a zero-mean mother wavelet function, and $a = \frac{t}{b}$ is the scaling factor of the mother wavelet that generates the daughter.

In this work, we are going to combine two modified versions of both linear and quadratic transforms and introduce an alternative TFR that can surpass the limitations of both approaches. The first one is the S-Transform, which is a modified version of STFT, and the other one is named S-Method that is a both computationally and mathematically enhanced version of the Wigner-Ville transform.
E. S-Transform

The S-Transform (ST) is a linear TFR which has been recently introduced by Stockwell in [29]-[31]. S-Transform combines the local Fourier analysis of the STFT with the multi-scale feature of the wavelet transform. In effect, it can be considered as a multi-scaled localized Fourier transform. The STFT captures dynamic frequency changes over time by exploiting a window function that provides time localization. However, the choice of window function represents a compromise. Wavelet transforms (WT) were introduced to improve the STFT performance by implementing the idea of resilient windowing or progressive resolution which enforces a finer time resolution at high frequencies and finer frequency resolution at low frequencies. Therefore, the WT does not directly measure frequency, but a similar quantity called scale. Additionally, the WT provides neither phase information, nor scale measurements, which are all relative to different local references. The ST exhibits globally referenced phase and frequency measurements like those of the DFT and STFT, as well as the progressive resolution of the WT using the following definition:

\[ ST: \quad x(t,f) = \int x(t') \frac{|f|}{2\pi} e^{-\frac{(t-t')^2}{2}} e^{-i2\pi ft'} \, dt' \quad (9) \]

In comparison with the STFT, the constant width of the localizing time window becomes \( \frac{1}{|f|} \) in the ST, i.e., the window width is scaled according to the inverse of the temporal frequency. As a result, narrower time windows are used at higher frequencies and wider time windows at lower frequencies, similar to the WT while we replaced the scale interpretation with the pure frequency.

F. S-Method

It has been claimed in the signal processing literature that the Wigner-Ville technique not only results in the best TF resolution among all TFRs, but also satisfies an exponentially large number of desirable mathematical properties. An alternative discrete form of the WVD is defined by:

\[ WVD \quad (n,k) = \sum_{m=\frac{1}{2}}^{\frac{N}{2}} x(n+m)x^*(n-m)e^{-\frac{j2\pi mk}{N+1}} \quad (10) \]

where \( x(n) \) is time-limited signal with \( n \) ranges in \( |n| < N/2 \) while a constant multiplication factor of 2 is omitted. Janovic et.al [32] showed that we can alternatively represent the Wigner-Ville decomposition in terms of (the simplest form of the windowing function) STFT if defined as follows:

\[ STFT \quad (n,k) = \sum_{m=\frac{1}{2}}^{\frac{N}{2}} x(n+m)e^{-\frac{j2\pi mk}{N+1}} \quad (11) \]

This relation has led to the following alternative TFR definition named S-method:

\[ SM \quad (n,k) = \frac{1}{N+1} \sum_{l=-L}^{L} STFT(n,k+l) \times STFT^*(n,k-l) \quad (12) \]

The S-method can represent a multicomponent signal such that the distribution of each component is its Wigner-Ville distribution, but it can avoid cross-interference components [3].

G. S-Method through S-Transform: SS Transform

Similar to other linear TFR, the ST suffers from the low TF resolution challenge in addition to the redundancy and high computational complexity [33]. To surpass the redundancy and computational complexity issues, a modified one-to-one fast discrete version of the ST has been introduced in [34]. To avoid the localization disturbance is sue in the time domain associated with Gaussian window in (9), we use a Hyperbolic ST (HST) that exploits a pseudo-Gaussian hyperbolic window [35], where the Hyperbolic format provides a frequency dependence shape along with its width and height that provides a significantly better time and frequency resolutions at low and high frequencies. Inspired by the one-to-one ST formulation, we can define the discrete version of the HST as follows:

\[ HST \quad (n,k) = \sum_{m=1}^{N} x(m,n)G(m,n)e^{-\frac{2\pi}{N+1} mk} \quad (13) \]

where \( x(m,n) \) is the frequency-shifted version of the discrete Fourier transform (DFT) of discrete signal \( x(n) \): \( x(m) = \frac{1}{N} \sum_{n=0}^{N-1} x(k)e^{-\frac{j2\pi mk}{N+1}} \), and \( G(m,n) \) is the DFT of a hyperbolic window \( h_w(t) \) which is defined as follows:

\[ h_w(t) = \frac{2fs}{\sqrt{2\pi(\alpha t + \beta)}} e^{-\frac{(t^2)(\gamma^2)}{2}} \quad (14) \]

\[ G(m,n) = \frac{2fs}{\sqrt{2\pi(\alpha t + \beta)}} e^{-\frac{(t^2)(\gamma^2)}{2}} \quad (15) \]

where \( \varphi(t) \) and \( \rho(t) \) are following the general representation of a hyperbolic function as follows: \( \frac{\alpha + \beta}{2\beta} ((t - \xi) + \sqrt{(t - \xi)^2 + y^2}) \). An introduction to select the optimized values for \( \alpha, \beta, \gamma, \) and \( \xi \) is presented in [36] and is out of the scope of this paper. To reach the best TF resolution, we use the following modification in (12) replacing for the STFT with HST. We term SS-Transform:

\[ SS T \quad (n,k) = \frac{1}{N+1} \sum_{l=-L}^{L} W(l)HST(n,k+l)HST^*(n,k-l) \quad (16) \]

with the window function

\[ W(l) = \begin{cases} \frac{1}{N+1} & \text{for } |l| \leq L \\ 0 & \text{o.w} \end{cases} \quad (17) \]

We may then use the fast S-Method approach introduced in [34] to calculate the \( L \)-th order HSS-Transform of the signal \( x \) as follows:

\[ HSS_{l} \quad (n,k) = SM_{l-1} \quad (n,k) + 2R[HST(n,k+l) \times HST^*(n,k-l)] \]

\[ HSS_{0} \quad (n,k) = H_{spectrogram} \quad (n,k) = |HST(n,k)|^2 \quad (18) \]
where $H_{\text{spectrogram}}$ indicates the spectrogram of the signal $x$ calculated by Hyperbolic S-Transform window, and symbol $\Re[.]$ stands for the real value. Figure 2 (a)-(g) present a comparative case study for the TFR mentioned above over a chirp-shaped signal with two slight frequency jump events at 0.5 and 1 seconds.

**Figure 2**

- **(a)** A Sinusoidal signal with a Chirp frequency distortion and two slight frequency jumps in 500 and 1000 milliseconds. In the rest of the panels, the power of signal has been calculated in dB and mapped accordingly with either parula or jet color maps. **(b)** The associated spectrogram resulted by the STFT with a 5th order Kaiser window of length 20% of the original signal and 70% overlap. **(c)** The associated scalogram resulted by the Morlet WT over the non-logscale frequency axis. **(d)** The associated time-frequency plane from Pseudo Wigner-Ville Transforms. **(e)** The associated time-frequency plane from original S-method. **(f)** The associated spectrogram resulted by the Hyperbolic S-transform. **(g)** The associated time-frequency plane representation from S-Transform-based S-method (SST) proposed in this work. While these different TFRs have slightly different mathematical properties, each will need specific requirements for clear visualization. For example, the Scalogram of wavelet transform usually uses a logarithmic partitioning along the frequency axis and suffers from interference in lower ends of the time-frequency plane; as a result, a cover filter is usually plotted over those time-frequency ranges (b). Moreover, the transforms such as WT or ST and SST are illustrating a different frequency thickness along the frequency axis due to their multi-scale feature and the difference between their kernels. One may clearly observe that the SST can have a clearer TFR of the signal with high resolution and minimum interference. We have changed the coloring threshold over each plot to make sure the best presentation is illustrated for each individual TFR based on its own specifications.
As we consider these images to be our features, we call them as TFR feature images (we ignored the TF axis notations).

IV. INFORMATIVE SPARSE CLASSIFICATION

This section presents an overview of the mathematical formulation and concepts of the informative sparse CP events classification framework introduced in [23] and [25]. Also, a brief review is given on the relevant sparse recovery theorem concepts and methods.

A. Sparse Representation-based CP Events Classification

Consider a set of N 3-phase CPE patterns from J CPE classes are recorded and each is available in terms of a 3-D tensor $x_j$. As instructed in Section III, first, the $d$q-Transform is applied and an orthogonal complex $x_j$ is generated (2). Next, the $L^q$ order HSS - Transform (18) of the signal $x_j$ is calculated and captured as an image called feature-image $d_j$. Finally, a training feature tensor $D = [d_1, d_2, ..., d_N] \in \mathbb{R}^{(T\times M) \times N}$ is formed such that the samples from different classes are sorted in order, i.e.,

$$D = [D_1|D_2|...|D_J].$$

From the concatenation of all associated feature images generated from training samples of the $j^{th}$ event class a feature sub-dictionary is formed and is termed as $D_j$. For the feature image $d_j$, $T$ stands for the number of pixels along the time axis while $M$ represents the number of pixels along the frequency axis. These two are usually specified by TFR specifications as well as the length of the signal, analysis window and maximum frequency component in the signal. According to the theory of SRC, if the training data points of the $j^{th}$ class are informative enough, a new feature image $d^{test}$ from the same class could be approximately linearly spanned by the elements of sub-dictionary $D_j$, means that, for a real-valued vector $s_j \in \mathbb{R}^{D_j}$:

$$d^{test} = D_j s_j.$$  

(20)

Alternatively, one may represent $d^{test}$ in terms of the whole training feature tensor,

$$d^{test} = D s.$$  

(21)

Regarding the problem statement given in Section II, $d^{test} = F y^{test}$ and $D = FY \in \mathbb{R}^{(T \times M) \times N}$ are the projection of the test sample and training dictionary to the feature space, respectively. If (20) holds, the obvious solution of (21) takes the following sparse format $s^* = [0, ..., 0, s_j, 0, ..., 0]^T$. Sparse vector $s \in \mathbb{R}^N$ is called the sparse indicator vector, with all entries equal to zero except those associated with the $j^{th}$ CPE class training sub-dictionary. Fig. 4 gives a visual demonstration of such a mathematical formulation. If the training dictionary $D$ is formed out of an underdetermined system of equations, i.e., $L < N$ (with $L = T \times M$) the theory of sparse recovery guarantees a desirable sparse format for $s^*$ as the solution of (21) by using the following formulation:

$$P_0: \quad \hat{s}_0 = \text{argmin}_{s} \|s\|_0 \quad \text{subject to} \quad f^{test} = Fs.$$  

(22)

here the $l_0$-norm stands for the number of non-zero elements or cardinality of vectors $s$. In terms of CPE classification problem, since the number of CPE classes $J$ is reasonably large and we have enough number of training data points from each CP event class $c_j (j = 1:J)$ we can ensure the training dictionary $D$ satisfies the underdetermined condition $L < N$. Although, $P_0$ is known to be NP-hard, under creation mathematical conditions on $D$, we can use the following relaxed $l_1$-norm format instead:

$$P_{11}: \quad \hat{s}_1 = \text{argmin}_{s} \|s\|_1 \quad \text{subject to} \quad ||f^{test} - Fs||_2 < \eta.$$  

(23)

This is widely known as Basic Pursuit Denoising regularization. Section IV-B introduces a well-known method for solving $P_{11}$ and the condition of the exact recovery of $s$ in CPE classification (refer to [47] for details of sparse recovery).

B. Greedy Sparse Solvers and Sparse CPC Solution

The orthogonal Matching Pursuit algorithm (Algorithm 1) is a popular alternative greedy method to be used instead of direct optimization-based solvers. In this work, we use this solver with respect to its low time complexity and tractability for large-scale problems. It has been shown in the literature that if the pairwise correlation among all columns of the training dictionary $D$ is lower than a certain threshold ($\mu_F = \max_{1 \leq l \leq J, 1 \leq s \leq L} |\langle D_l, D_s \rangle| < \frac{1}{2L-1}$), then the OMP can recover original $\hat{s} = s$, the unique solution to $d = D \hat{s}$, having sparsity $K$ or less using several measurements that scales like $K \log(L/K)$ [47].
Without loss of generality, let $D = FY$ be the training tensor/dictionary created using the data points of $J$ CP classes and for a given $d_{test}$, let $\hat{s}_1$ be the optimal solution of $NP_1$. The selected class can be obtained as the one which its corresponding subsequence in vector $\hat{s}_1$ has the minimum reconstruction residual value. Figure 4 is a visualization of a typical SRC procedure for a CP event.

Algorithm 1. Orthogonal Matching Pursuit (OMP)

```plaintext
require: matrix A, measurements $d = Ds + n$, stopping criterion $\rho_0$
initialize: $r^0 = d, s^0 = 0, l = 0, SUP = \emptyset$
repeat 
1. match: $h^l = D^l r^l$
2. identify support indicator:
   $\text{sup}^l = \{ \text{argmax}_j |h^l(j)|\}$
3. update the support:
   $SU_P^{l+1} = SU_P^l \cup \text{sup}^l$
4. update signal estimate:
   $s^{l+1} = \text{argmin}_x \{ D x : sup(z) \} ||d - Dz||_2$
   $r^{l+1} = d - As^{l+1}$
   $l = l + 1$
Until stopping criterion met
Output: $\hat{s} = s^L$
```

V. Simulation Results and Discussion

A. Data Generation

A wide range of CPEs have been selected and simulated in MATLAB/Simulink using IEEE-34 Bus system in addition to a sample Microgrid model as directed in [39], [45] and [40] (please refer to these references for full details and publicly available models and dataset). The sampling frequency was set as 9.6 kHz to follow the practical instructions in developing PMU-embedded smart inverter design [10]. Table I summarizes the specification of each event. For each of these 17 CPE scenarios/classes, 1000 possible fault cases have been generated as directed. Next, each of these faulty signals has been polluted by Gaussian noise with a set the signal to noise ratio at 10-100 dB randomly, based on current PMU standards.

1 For the sake of generality in notation and to be consistent with the classification literature concepts, we hereafter call the data or feature matrix, $Y$, or $F$ as the training matrix or training dictionary and note it by $A$.

- Table I. List of Cyber-Physical Events and their specifications

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Specifications</th>
</tr>
</thead>
</table>
| Capacitor Switching (CS)    | Refer to ref.
| Arc Furnace (AF)            | [40]           |
| Induction Motor Start-Up (IM)| Refer to ref.
| Lightening (LT)             | [40]           |
| Transformer Energizing (TE) | Refer to ref.
| Three Phase Nonlinear Load (3L) | Refer to ref.
| Magnitude Jump (MJ)         | 0.1-2pu        |
| Harmonic Distortion (HD)    | 0.5%-10%THD    |
| Amplitude Modulation (MM)   | 0.1Hz - 5Hz, 0.005-0.1pu |
| Frequency Ramp (FR)         | ±0.01Hz/s - ±1Hz/s within ±5Hz |
| Line-to-Line (LL) fault     | 0.1-1pu Magnitude drop |
| Frequency Jump (FJ)         | -5 to 5Hz     |
| Phase Jump (PJ)             | ±π/18 (rad)   |
| Out of Band Interference (OB)| 10-120Hz, 0.01-0.1pu |
| Angle Modulation (AM)       | 0.1-5Hz, 0.005-0.1pu |
| Single Line to Ground (SLG)| 0.2-1pu Magnitude drop |
| Line to Line to Ground (LLG)| 0.1-1pu Magnitude drop |

Algorithm 2. TISC-CPE

input: training data matrix $Y \in \mathbb{R}^{M \times N}$, test sample $y_{test} \in \mathbb{R}^M$
1. Extract the feature matrix from the training data, using a random transformation matrix $F \in \mathbb{R}^{L \times M} : D = FY \in \mathbb{R}^{L \times N}$
2. Extract features of the testing data from the test sample, using the matrix $F$ used in step 1: $d_{test} = F y_{test}$.
3. Calculate the approximated training dictionary $D$ using Algorithm 3.
4. Solve $P_1$ or $NP_1$ for sparse vector $\hat{s}_1$ using OMP (Alg. 1)
5. Compute $J$ purified vectors $\tilde{s}_j$ for $j = 1, \ldots, J$ using indicator function $g(\tilde{s}_j) : \mathbb{R}^L \rightarrow \mathbb{R}^L$, such that $\tilde{s}_j = g(\tilde{s}_j)$, is a new vector whose only nonzero entries are the entries in $\hat{s}_1$ that are associated with class $c_j$.
6. Compute residual $r_j = \|d_{test} - D s_j\|^2_f$ for $j = 1, \ldots, J$.
7. $j^* = \text{argmin}_j r_j$
output: Classify ($y_{test}$) $\hat{c}_{j^*}$
Algorithm 3. Informative Data Samples Selection

\textbf{Input:} Dimensionality optimized training dictionary \( D \in \mathbb{R}^{L \times N} \)
\begin{enumerate}
  \item Initiate \( D \) with any arbitrary extreme point of \( D \).
  \item Find the best element that minimizes the Hausdorff distance.
    \[ a_j^* = \arg \min_{a_j \in \partial D} d_H(D \cup a_j, D). \]
  \item \( D \leftarrow D \cup a_j^* \).
  \item Return to step 1 if the desired \( N \) or \( \varepsilon \) is not achieved.
\end{enumerate}
\textbf{Output:} Approximated training dictionary \( D \in \mathbb{R}^{L \times N} \)

\section{Feature-wise Comparison}

In the first case study, a set of 500 CPE classes from each of the 16 CPE classes in addition to a normal 3-Phase sinusoidal signal (total 17 classes) have been selected and the corresponding feature dictionary \( D \in \mathbb{C}^{(MKT) \times N} \) has been formed by the concatenation of the associated time-frequency feature images resulted from the most popular features in the literature [33][32]. Spectrogram of STFT, Scalogram of Morlet wavelet, S-Transform, S-Method, and HSS-Transform. These 5 feature tensors have been used to solve (23) and to find the associated CPE class for a set of 500 test data samples from each of the 17 CPE classes. Table II summarizes the classification accuracy rate. In addition to the initial 500 training samples, we have used the convex hull vertices found by the Algorithm 3 as informative training TF feature images (which varies from 10\% for SLG faults to 80\% for Arc Furnace) and reevaluated the TISC algorithm (result reported within parenthesis). As one can see, the overall classification performance remains almost the same along with all time-frequency features with limited set of informative samples. One may conclude that the HSST images are wiser choices and resulted in better classification accuracy compared to the state-of-the-art TF features which have been used in literature. It is worth noting that these TFR have been mostly used to extract features from single phase events in previous works.

Table II. Feature-wise comparison of classification accuracy of TISC for 18 classes of CP events: all training samples (vs. informative samples)

<table>
<thead>
<tr>
<th>CP event (all training samples)</th>
<th>TISC Classification Rate (%)</th>
<th>SPEC</th>
<th>SCA</th>
<th>ST</th>
<th>SM</th>
<th>HSST</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>90.85</td>
<td>70.83</td>
<td>85.80</td>
<td>50.90</td>
<td>95.80</td>
<td>94.90</td>
</tr>
<tr>
<td>AF</td>
<td>84.80</td>
<td>84.81</td>
<td>90.87</td>
<td>90.87</td>
<td>90.87</td>
<td>90.87</td>
</tr>
<tr>
<td>IM</td>
<td>90.86</td>
<td>85.95</td>
<td>95.90</td>
<td>95.78</td>
<td>98.97</td>
<td></td>
</tr>
<tr>
<td>LT</td>
<td>75.72</td>
<td>73.72</td>
<td>78.75</td>
<td>81.79</td>
<td>83.81</td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>88.84</td>
<td>88.82</td>
<td>93.92</td>
<td>93.91</td>
<td>95.94</td>
<td></td>
</tr>
<tr>
<td>3L</td>
<td>92.83</td>
<td>92.85</td>
<td>95.88</td>
<td>94.75</td>
<td>96.95</td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>80.79</td>
<td>82.80</td>
<td>88.87</td>
<td>87.82</td>
<td>91.90</td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>73.71</td>
<td>70.66</td>
<td>79.76</td>
<td>80.81</td>
<td>82.80</td>
<td></td>
</tr>
<tr>
<td>HD</td>
<td>93.85</td>
<td>92.88</td>
<td>97.93</td>
<td>96.93</td>
<td>96.93</td>
<td></td>
</tr>
<tr>
<td>LLG</td>
<td>92.90</td>
<td>92.89</td>
<td>96.92</td>
<td>97.90</td>
<td>98.94</td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>91.90</td>
<td>89.89</td>
<td>97.94</td>
<td>96.89</td>
<td>97.93</td>
<td></td>
</tr>
<tr>
<td>MJ</td>
<td>98.83</td>
<td>87.84</td>
<td>94.91</td>
<td>92.82</td>
<td>93.90</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>89.88</td>
<td>92.90</td>
<td>95.90</td>
<td>94.83</td>
<td>97.94</td>
<td></td>
</tr>
<tr>
<td>OB</td>
<td>82.79</td>
<td>85.80</td>
<td>93.90</td>
<td>91.85</td>
<td>93.91</td>
<td></td>
</tr>
<tr>
<td>SLG</td>
<td>88.84</td>
<td>90.83</td>
<td>93.92</td>
<td>93.92</td>
<td>97.94</td>
<td></td>
</tr>
<tr>
<td>FR</td>
<td>86.81</td>
<td>81.80</td>
<td>91.89</td>
<td>90.87</td>
<td>93.90</td>
<td></td>
</tr>
<tr>
<td>FJ</td>
<td>65.61</td>
<td>63.60</td>
<td>78.77</td>
<td>79.75</td>
<td>82.79</td>
<td></td>
</tr>
<tr>
<td>PJ</td>
<td>82.75</td>
<td>79.74</td>
<td>93.89</td>
<td>92.87</td>
<td>96.92</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>84.94</td>
<td>84.33</td>
<td>91.2</td>
<td>90.52</td>
<td>93.16</td>
<td></td>
</tr>
</tbody>
</table>

\section{Comparison vs State-of-the-Art Techniques}

ANNs and SVMs are the most popular methods which have been vastly utilized in a variety of pattern recognition problems. ANNs, SVMs, and their derivatives are widely employed power systems. These methods have been particularly used in single-phase power quality events classification and islanding detection [21][22]. It is widely observed that in high dimensional feature spaces, the usage of radial basis functions (RBF), projects the initial feature space into an infinite-dimensional Hilbert space which can solve the non-linearity problem with a high accuracy [41]. As such we have designed and evaluated an RBF artificial neural network (ANN) with Gaussian function as the activation function in the hidden layers’ neural units and a tangent hyperbolic function in the output layer (with 120 neurons in the hidden and 40 neurons in the output layers, respectively). The basis function centers have been determined using a hybrid genetic k-means clustering procedure introduced in [41] (for details, please refer to [42]). Moreover, we implemented an RBF 1v1 support vector machine (with 150 machines). It has been shown in the literature that RBF kernel SVMs have the minimum number of support vectors, minimum value as classification error and good classification accuracy [43]. We determined the penalty factor in addition the adjustable parameter of the RBF using an improved ant colony optimization algorithm introduced in [43]. Table III is summarizing the identification accuracy rate for TISC-CPE vs RBF-NN [44] and RBF-SVM [43] for two scenarios: 1. All 500 training feature images are used and, 2. Only a selected number of informative feature images (which varies from 10\% for SLG faults to 80\% for Arc Furnace) are used to either form the feature matrix \( D \) or to train the ANN or SVM classifiers.

Table III. CP event all training samples (vs. informative samples)

<table>
<thead>
<tr>
<th>CP event (all training samples)</th>
<th>Classification Rate (%) using HSST</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF-ANN</td>
<td>92.89(91.6)</td>
</tr>
<tr>
<td>RBF-SVM</td>
<td>94.90(90.8)</td>
</tr>
<tr>
<td>TISC</td>
<td>90.54(89.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CP event (all training samples)</th>
<th>Classification Rate (%) using HSST</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF-ANN</td>
<td>92.89(91.6)</td>
</tr>
<tr>
<td>RBF-SVM</td>
<td>94.90(90.8)</td>
</tr>
<tr>
<td>TISC</td>
<td>90.54(89.8)</td>
</tr>
</tbody>
</table>

The average performance of all approaches is slightly over 90\% with TISC faintly led by 1\% margin. This is a notable achievement considering the linear formulation of TISC with no training procedure performed. In the informative scenarios, however, TISC leads to an average accuracy margin of 4\%. This
is not surprising as we already knew that using a smaller number of training samples would affect the generalizability of the sophisticated versions of ANNs and SVMs. The more sophisticated ANN and SVM architecture is the cost we should pay to arrive at an acceptable identification accuracy without collecting extra training samples extraction procedures.

**D. Discussions on Notable Properties of TISC-CPE**

As we have discussed throughout the paper and through the numerical results, TISC-CPE has the following unique features that will help reducing the computational cost and implementation complexity in an inverter processing unit with limited computational capabilities:

1. Our specific formulation exploits an alternative orthogonal complex signal which contains simultaneous 3-Phase patterns at once (2). Single-phase methods should be implemented separately on individual phases, which results in losing possible relative coupling between-phase information such as unbalance.

2. According to the sparse recovery literature [37], due to the fact of “blessing of dimensionality”, one may implement a random projection from feature space to another alternative mathematical domain with lower dimensionality without affecting the recovery performance. In TISC-CPE, this is equivalent to generate random faces from time-frequency images resulted from HSS-Transform.

3. We may use the idea of informative sample selection to optimize the number of training samples to be concatenated into matrix $F$. This idea indicates that for each class of events, there is a certain set of informative data samples that are spanning the whole cluster of the corresponding class in the designated feature space. These informative samples are in fact the vertices of the associated convex hull of the class cluster within the feature space. This can dramatically reduce the dimensions of this matrix and reduce the computational complexity in our algorithm. Algorithm 3 is a sample code for this purpose [38].

4. Back to our conversation in Section VI-A, not only the increase of the number of CPE classes does not cause any issue in TISC-CPE formulation, but it also improves the overall recovery performance by enhancing the under-determinedness of the feature tensor by adding more columns.

5. The following table [25] compares the time complexity of a 2-layer ordinary artificial neural network vs. sparse classifier, with $N$ to be the number of training samples, $M$ the number of features or variables, and $J$ the number of output classes. Finally, $I$ is the total number of iterations needed for ANN convergence and $k$ is the sparsity of the classification signals (Section IV-A).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ANN</th>
<th>Sparse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>$I \times O(M \times J \times N)$</td>
<td>$NA$</td>
</tr>
<tr>
<td>Test</td>
<td>$O\left(p\left(j + \frac{M}{2}\right)^2\right)$</td>
<td>$O(k \times (M + N))$</td>
</tr>
</tbody>
</table>

Certainly, the computational complexity for a convolutional or deep neural network would be even much higher, depending on the network architecture. A recent study in [46] revealed that in 3-phase unbalanced systems, the complexity of advanced control algorithm is considerably high, where even advanced 32-bit microcontrollers such as TMS320F28335 ACTIVE Delfino™ 32-bit MCU with 150 MIPS, FPU, 512 KB Flash, EMIF, 12b ADC cannot handle all functionalities without suffering from high computational burdens. However, without any computational limitations and up on availability of unlimited data samples from all classes of islanding events and based on the recent findings in the area of machine learning and artificial intelligence, a Deep Neural Network should outperform any known classification technique.

**VI. CONCLUSIONS**

In this study, we explored the problem of cyber-physical events classification for islanding detection in power inverters using a new modified 3-phase time-frequency representation called Hyperbolic S-Transform-based S-Method (HSST) in addition to the sparse representation-based classifier. Time frequency representations are the most popular techniques in studying non-stationary signals such as fault events in a power system. However, nowadays, the huge dynamic range of cyber-physical faults/events that may happen in DGUs is highly increased due to penetration of renewables and cyber intrusions where HSST showed promising alternative decomposition with higher resolution for TF analysis of CPEs. This is particularly important in detection of islanding situations to avoid catastrophic damages such as wildfires or human risk factors. On the other hand, it has been widely observed that there are alternative mathematical representations for almost all the industrial (and most of the natural) phenomena where the new mathematical representation lies on a low dimensional subspace, where the signal under study has a sparse representation, such as Fourier representation for electrical signals. This low dimensionality formed the basis of our justification for exploiting the idea of informative sparse representation-based classification for cyber-physical events classification.

Training-free property and feature independency eliminate all the required effort and time for FE-FS and training steps, while informative sample selection provides a degree of freedom to the user to decrease the data space dimension. We called this combined approach as TISC-CPE; we verified its performance on a big range of CPEs and numerically compared the results vs. state-of-the-art algorithms.

**REFERENCES**


IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS


38] Sartipizadeh, Hossein, and Tyronne L. Vincent. “Computing the t Identif ication of er s. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2980045, IEEE Journal of Emerging and Selected Topics in Power Electronics 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.”