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

Mohammad Babakmehr, *Member, IEEE*, Farnaz Harirchi, *Member, IEEE*, Payman Dehghanian, *Member, IEEE*, and Johan H. Enslin, *Fellow, IEEE*

***Abstract*— Along with the rapid integration of distributed generation units (DGUs) into the power grids is the rise in unconventional and unpredictable patterns of the undesirable cyber–physical intrusions and faults; this drastically 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 article, a novel quadratic time–frequency decomposition, namely HSS-transform, is applied over an alternative complex representation of three-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 nonislanding events with the focus on practical limitations and requirements of a smart power electronics inviter 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 three-phase feature extraction strategy ensures preservation of the between-phase information.**

***Index Terms*— Artificial intelligence (AI), cyber–physical events (CPEs), islanding detection, pattern recognition, power electronic inverters, smart grids, sparse classification, time–frequency analysis.**

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 (also known as 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, also known as active distribution networks [see Fig. 1(a)] [2]. In modern active distribution networks of the future, advanced power electronics and artificial intelligence (AI)-based 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 distributed generation unit (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.

Bridging 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 toward a sustainable active distribution network: 1) data integrity and interoperability and 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 three-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 (CPE) detection, classification, and system protection in a decentralized manner, while only the high-level reports

Manuscript received November 1, 2019; revised January 21, 2020; accepted February 7, 2020. Date of publication March 11, 2020; date of current version October 1, 2021. This work was partially supported by the US Department of Energy Office of Electricity Industry Technology and Practices Innovation Challenge (EITPIC) Prize. Recommended for publication by Associate Editor Yongheng Yang. *(Corresponding author: Mohammad Babakmehr.)*

Mohammad Babakmehr, Farnaz Harirchi, and Johan H. Enslin are with the Restoration Institute, Clemson University, Charleston, SC 29405 USA (e-mail: mohammad.babakmehr@gmail.com).

Payman Dehghanian is with the Department of Electrical and Computer Engineering, George Washington University, Washington, DC 20052 USA.

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

Digital Object Identifier 10.1109/JESTPE.2020.2980045

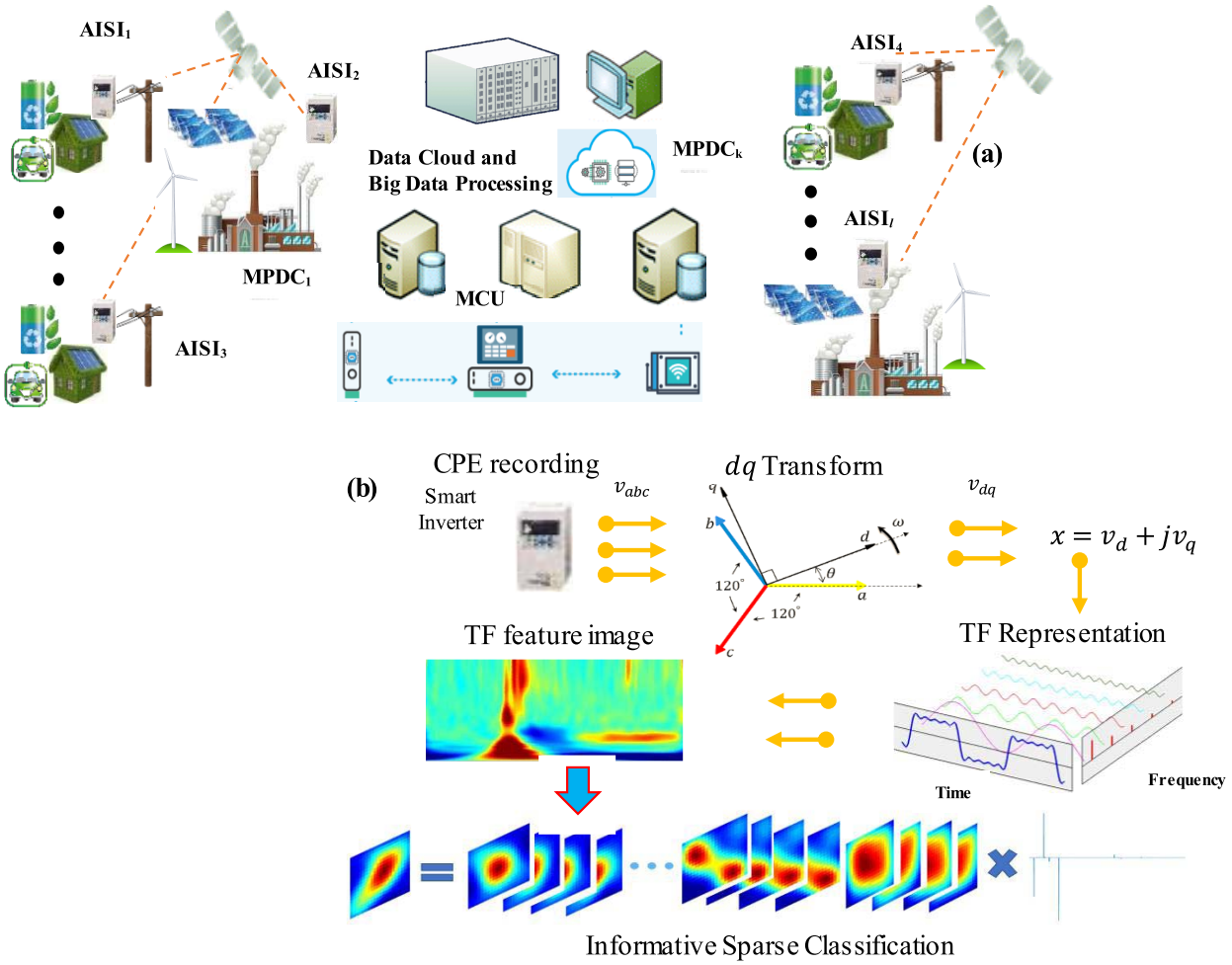

Fig. 1. (a) Sustainable active distribution network architecture with AI-enhanced smart inverters (AISI), master phasor data concentrator (MPDC), and MCU. (b) TISC-CPE.

can be communicated to the utility (see Fig. 1). This has led to phasor measurement unit (PMU)-embedded inverter architectures [10]. The passive islanding detection through CPEs 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 AI 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 photovoltaic cell (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 five-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 (ST) [18], time–time transform [19], mathematical morphology [20], and so on; 4) feature selection (FS) or dimensionality reduction; and 5) categorization or classification. A variety of mathematical methods have 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 PMUs): 1) analytical problems and 2) practical problems.

Analytical problems are general algorithmic issues associated with any pattern recognition problem, including optimal FE that was a major topic of research in islanding detection [21], [22], feature selection or dimensionality reduction for computational efficiency, and optimal classifier training. Practical issues include computational resource limitations, ignorance of the collaborative information along with three-phase systems, generalizability, and sensitivity to noise and uncertainties.

C. Contributions

In this article and in response to the abovementioned challenges, we first introduce a novel instantaneous complex signal

quadratic time–frequency decomposition to extract three-phase features from a voltage or current signal in three-phase inverters.[1] We particularly develop an ST-based realization for the S-method [30] to generate time–frequency images of the three-phase CPEs. We verify our approach over a comprehensive set of 20 critical (islanding and nonislanding) 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, and harmonics (please refer to Section V for a full list). These events can be either purely cyber-driven events caused by cyberattacks or can 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 (SRC) problem with lower time and implementation complexities compared with 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] and [25].

The main privileges of the proposed SRP framework, referred to as time–frequency decomposition-based informative sparse classification for cyber–physical events (TISC-CPE), compared with the state-of-the-art techniques are given in the following.

- *Simultaneous Three-Phase FE:* Exploiting an alternative 2-D orthogonal representation for the three-phase signals based on the Park transform [26] that contains the instantaneous three-phase information (see Section I-B).
- A modified time–frequency representation can result in an enhanced set of features called TF images to address the FE 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 (see Sections I-B and V).
- Training-Free Property: Training data samples are only stacked into a matrix among a linear optimization formulation, where no training procedure is needed (see Section I-B).
- Linearity: The presented method results in an inherently fast linear classifier with much lower computational cost (see Section I-B).
- 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 (see Sections I-B and IV).
- 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 (see Section I-B).

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

II. Notations and Problem Definitions

Assume, for a 110-V–60-Hz three-phase inverter, that a set of $N = \sum_{j=1}^{J} n_j$ labeled three-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 three-phase known-class vectors, $\boldsymbol{y}_i \in \mathbb{R}^{3\times M}$, $i = 1, 2, \ldots, N$, and $M = 2400$, to contain a CPE pattern.

The final goal is to develop a framework that takes a feature vector, $\boldsymbol{d}^{\text{test}}$ extracted from a new detected cyber–physical data point, $\boldsymbol{y}^{\text{test}}$, 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 $c_i = A(F(\boldsymbol{y}^{\text{test}})), c_i \in C = \{c_j | j = 1, 2, \ldots, J\}$. Roughly speaking, operator D stands for the data-feature mapping resulting from the 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 the selection of these mappings is unknown in advance and may change based on the system and associated data specifications or applications.

In this article, first, we develop a new time–frequency decomposition that exploits the flexibilities of the ST in time–frequency plane tiling [29] with high resolution of a modified version of Winer–Ville distribution (WVD) [30] named S-method. We aim to develop an optimal FE method that is adapted to address all technical requirements for classifying complicated, yet within a wide range of dynamics and occurrence time range, CPEs 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 with 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: 1) instantaneous three-phase FE using HSS-transform and 2) feature selection and classification using informative sparse classifier.

[1] In this article, we would focus on challenges in three-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.

[2] According to [3] and [4], a data point constructed from sampling 10–20 cycles of a signal with a frequency of 10 kHz is an appropriate segment to represent a cyber–physical event for most classes of cyber–physical events.

III. Instantaneous Three-Phase Feature Extraction

In this section, we propose a novel simultaneous three-phase time–frequency representation (TFR) to extract the distinguished features from different patterns of CPEs that are unique enough for event classification. The major advantages of our proposed approach are as follows: 1) exploiting simultaneous three-phase information; 2) using enhanced WVD, we aim to preserve the best possible time–frequency resolution; and 3) utilizing the hyperbolic ST (HST), we combine the advantageous of Fourier basis of STFT and multiscale resolution of wavelets altogether.

A. Instantaneous Three-Phase Signal Processing Tools

In a nutshell, time-domain-based instantaneous power theories are 3-D mathematical signal mappings (comparable to 1-D Fourier/wavelet transform (WT) in one-phase system analysis) that transforms any three-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 transform called synchronous reference frame (dq) method to convert a three-phase signal to an alternative orthogonal mathematical space. We investigate that tracking the associated trajectories of the TFR of three-phase signals in these new mathematical domains can be utilized as a new, fast and real-time CPE indicator without imposing any power quality issue.

B. Direct-Quadrature Transformation ($\boldsymbol{dq}$)

The synchronous reference frame method or in short dq transform maps a three-phase electrical signal from the abc frame to the synchronous reference frame (dq) as follows [16]:

$$\begin{bmatrix} v_d \\ v_q \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \cos(\theta) & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ -\sin(\theta) & -\sin\left(\theta - \frac{2\pi}{3}\right) & -\sin\left(\theta + \frac{2\pi}{3}\right) \end{bmatrix} \times \begin{bmatrix} v_a \\ v_b \\ v_c \end{bmatrix} \quad (1)$$

where θ is the synchronization angle, which is time-variant and represents the angular position of the dq frame, and it is detected by a phase-locked loop (PLL). Due to the fact of orthogonality of d and q components, we define the following orthogonal complex signal:

$$x = v_d + jv_q. \quad (2)$$

If the system is voltage-balanced, x carries the entire three-phase information in an alternative mathematical representation. It is worth noting that, by projecting an N-dimensional data set 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 (three-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 three-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 V) should be equal and each one should be out-phase from the others on the order of 120°.

Our approach can handle heteropolar unbalancity since the effect of such an unbalanced condition can be reflected in the d and q components. Even in the 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 system is considerably unbalanced.

Now, we use the TFR of $\boldsymbol{x}$ (for any CPE recorded on a three-phase voltage) to extract useful features from this complex signal for the sake of classification.

C. Time–Frequency Representation

In the last several decades, TFR has been the most popular approach in studying the nonstationary 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 resolutions, 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 Versus Linear Time–Frequency Decomposition

There are two major classes of TFR widely used in different applications, namely linear and bilinear/quadratic (also known as 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 dependence to the input signal

$$L_{\text{TFR}} = \int x(t)\omega_l(t-\tau, f)dt \quad (3)$$

$$Q_{\text{TFR}} = \int x\left(t + \frac{\tau}{2}\right)x^*\left(t + \frac{\tau}{2}\right)\omega_q(t-\tau, f)dt \quad (4)$$

where x is, in general, a complex signal with $*$ denoting its conjugate and ω_l and ω_q are some predesigned functions of time and frequency. In extreme cases, ω_l is simplified to a complex exponential ($e^{-i2\pi ft}$), which will result in the ordinary Fourier transform (5) and ω_q 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 STFT (6), WTs (7), and, more recently Stockwell or in short ST (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 $=$ $|\mathrm{STFT}|^2$) and scalogram (SCA $=$ $|\mathrm{WT}|^2$), 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 them, the 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. The common mathematical formulations of FT, STFT, WT, and WVD, respectively, are

$$\mathbf{FT}: X(f) = \int x(t')e^{-i2\pi f t'}dt' \tag{5}$$

$$\mathbf{STFT}: X_g(t,f) = \int x(t')g^*(t'-t)e^{-i2\pi f t'}dt' \tag{6}$$

$$\mathbf{CWT}: X_\varphi(t,a) = \int x(t')\sqrt{a}\psi^*(a(t'-t))dt' \tag{7}$$

$$\mathbf{WVD}: X(t,f) = \int x\left(t+\frac{t'}{2}\right)x^*\left(t+\frac{t'}{2}\right)e^{-i2\pi f t'}dt' \tag{8}$$

where g is a predesigned window such as Gaussian, ψ is a zero-mean mother wavelet function, and $a = f/f_0$ is the scaling factor of the mother wavelet that generates the daughters.

In this article, 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 ST, which is a modified version of STFT, and the other one is named ST that is a both computationally and mathematically enhanced version of the Wigner–Ville transform.

E. S-Transform

The ST is a linear TFR that has been recently introduced in [29]–[31]. ST combines the local Fourier analysis of the STFT with the multiscale feature of the WT. In effect, it can be considered as a multiscaled 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. WTs 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. In addition, the WT provides neither phase information nor phase measurements, which are all relative to different local references. The ST exhibits globally referenced phase and frequency measurements such as those of the DFT and STFT, as well as the progressive resolution of the WT using the following definition:

$$\mathbf{ST}: X(t,f) = \int x(t')\frac{|f|}{\sqrt{2\pi}}e^{-\frac{(t'-t)^2f^2}{2}}e^{-i2\pi f t'}dt'. \tag{9}$$

In comparison with the STFT, the constant width of the localizing time window becomes $1/|f|$ 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 replacing 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

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

where $x(n)$ is time-limited signal with n ranges in $|n| < N/2$, while a constant multiplication factor of 2 is omitted. Stankovic *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 it is defined as follows:

$$\mathrm{STFT}(n,k) = \sum_{m=-\frac{N}{2}}^{m=\frac{N}{2}} x(n+m)e^{-\frac{j2\pi}{N+1}2mk}. \tag{11}$$

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

$$\mathrm{SM}(n,k) = \frac{1}{N+1}\sum_{l=-L}^{l=L}\mathrm{STFT}(n,k+l)\mathrm{STFT}^*(n,k-l). \tag{12}$$

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

G. S-Method Through ST: 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 [31]. To avoid the localization disturbance issue in the time domain associated with Gaussian window in (9), we use an HST that exploits a pseudo-Gaussian hyperbolic window [34], 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:

$$\mathrm{HST}(n,k)=\sum_{m=1}^{N}X(m,n)G(m,n)e^{-j\frac{2\pi}{N+1}mk} \tag{13}$$

where $X(m,n)$ is the frequency-shifted version of the discrete Fourier transform (DFT) of discrete signal $x(n)$: $X(m)=1/N\sum_{m=1}^{N}x(k)e^{-j\frac{2\pi}{N+1}mk}$, and $G(m,n)$ is the DFT of a hyperbolic window (h_w), which is defined as follows:

$$h_w(t)=\frac{2f_s}{\sqrt{2\pi(\alpha+\beta)}}e^{\left(-\frac{f_s^2\varphi(t)^2}{2}\right)} \tag{14}$$

$$G(m,n)=\frac{2f_s}{\sqrt{2\pi(\alpha+\beta)}}e^{\left(-\frac{f_s^2\phi(f)^2}{2}\right)} \tag{15}$$

where $\varphi(t)$ and $\phi(f)$ are following the general representation of a hyperbolic function as follows: $(\alpha+\beta)/(2\alpha\beta)((\tau-t-\xi)+((\tau-t-\xi)^2+\gamma^2)^{1/2})$. An introduction to select the optimized values for α,β,γ, and ξ is presented in [35] and is out of the scope of this article. To reach the best TF resolution, we use the following modification in (12) replacing for the STFT with HST. We term as SS-transform

$$\mathrm{SST}(n,k)=\frac{1}{N+1}\sum_{l=-L}^{l=L}W(l)\mathrm{HST}(n,k+l)\mathrm{HST}^{*}(n,\text{k-l}) \tag{16}$$

with the window function

$$W(l)=\begin{cases}\dfrac{1}{N+1}, & \text{for } |l|\le L\\ 0, & o.w.\end{cases} \tag{17}$$

We may then use the fast S-method approach introduced in [31] to calculate the Lth-order HSS-transform of the signal x as follows:

$$\begin{aligned}\mathrm{HSST}_L(n,k)&=\mathrm{SM}_{L-1}(n,k)+2\Re\big[\mathrm{HST}(n,k+L)\\&\quad\times\mathrm{HST}^{*}(n,k-L)\big]\\ \mathrm{HSST}_0(n,k)&=H_{\text{spectrogram}}(n,k)=|\mathrm{HST}(n,k)|^2\end{aligned} \tag{18}$$

where $H_{\text{spectrogram}}$ indicates the spectrogram of the signal x calculated by HST window and symbol $\Re[.]$ stands for the real value. Fig. 2(a)–(g) shows a comparative case study for the TFR mentioned earlier over a chirp-shaped signal with two slight frequency jump events at 0.5 and 1 s.

Fig. 3(b)–(e) shows the TFR feature images extracted from a three-phase fault [see Fig. 3(a)], resulted by a capacitor switching event using STFT, WT, ST, and hyperbolic S-transform-based S-method (HSST), respectively. 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 on the mathematical formulation and concepts of the informative sparse CPEs classification framework using the theory of sparse representation-based classification introduced in [23] and [25]. Also, a brief review is given on the relevant sparse recovery theorem concepts and methods.

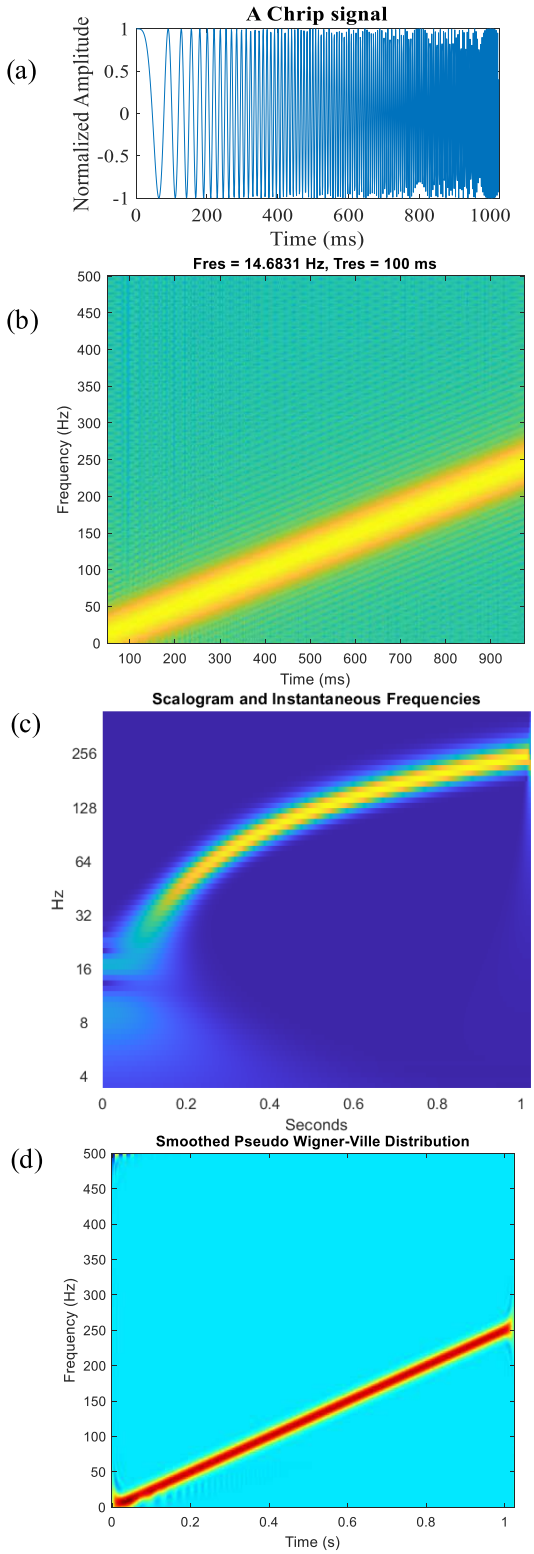

Fig. 2. (a) Sinusoidal signal with a chirp frequency distortion and two slight frequency jumps in 500 and 1000 ms. In the rest of subfigures, the power of signal has been calculated in dB and mapped accordingly with either parula or jet color maps. (b) Associated spectrogram resulted by the STFT with a fifth-order Kaiser window of length 20% of the original signal and 70% overlap. (c) Associated scalogram resulted by the Morlet WT over the nonlog-scale frequency axis. (d) Associated time–frequency plane from pseudo Wigner–Ville Transform.

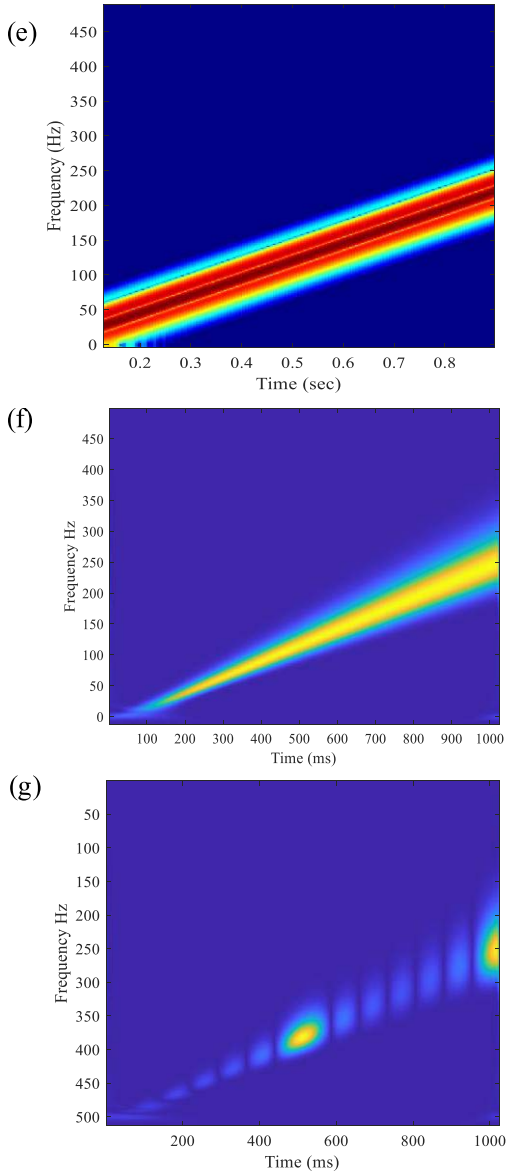

Fig. 2. *(Continued.)* (e) Associated time–frequency plane from the original S-method. (f) Associated spectrogram resulted by the HST. (g) Associated time–frequency plane representation from the ST-based S-method (SST) proposed in this article. While these different TFRs have slightly different mathematical properties, each will need specific requirements for clear visualization. For example, the scalogram of WT usually uses a logarithmic partitioning along the frequency axis and suffers from interference in lower ends of the time–frequency plan; 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 multiscale 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 that the best presentation is illustrated for each individual TFR based on its own specifications.

A. Sparse Representation-Based CPEs Classification

Consider a set of N three-phase CPE patterns from J CPE classes are recorded, and each one is available in terms of a 3-D tensor y_i. As instructed in Section III, first, the dq-transform is applied and an orthogonal complex x_i is generated (2). Next, the Lth-order HSStransform (18) of

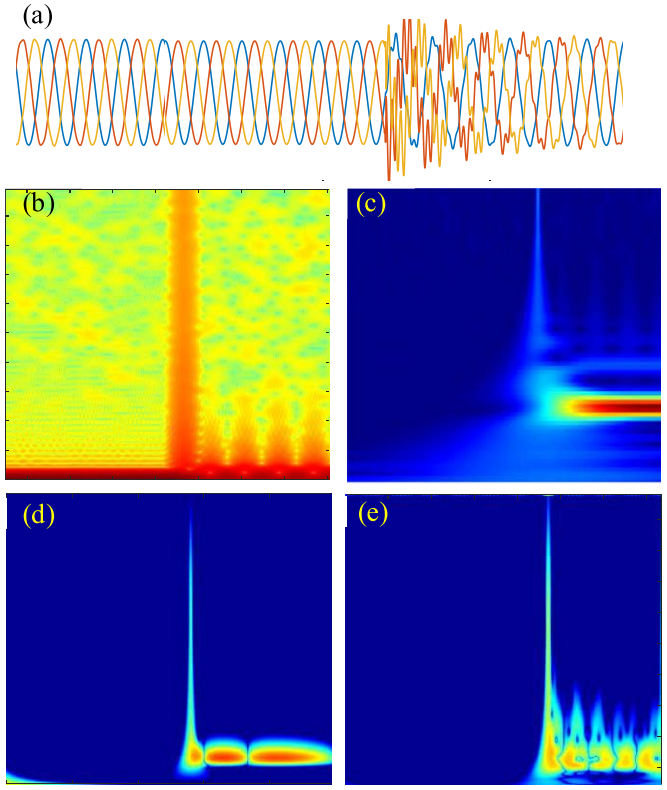

Fig. 3. Comparison between TFR feature images of (a) a three-phase fault resulted from three-phase capacitor switching event, (b) STFT, (c) WT, (d) ST, and (e) HSST.

the signal x_i is calculated and captured as an image called feature image $\boldsymbol{d}_i$. Finally, a training feature tensor $D = [\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d}_N] \in \mathbb{R}^{(T\times M)\times N}$ is formed such that the samples from different classes are sorted in order, that is

$$D = [D_1|D_2|\cdots|D_J]. \tag{19}$$

From the concatenation of all associated feature images generated from training samples of the jth event class, a feature subdictionary is formed and is termed D_j. For the feature image $\boldsymbol{d}_i$, 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 jth class are informative enough, a new feature image $\boldsymbol{d}^{\boldsymbol{test}}$ from the same class could be approximately linearly spanned by the elements of subdictionary D_j, which means that, for a real-valued vector $\boldsymbol{s}_j \in \mathbb{R}^{n_j}$

$$\boldsymbol{d}^{\text{test}} = D_j \boldsymbol{s}_j. \tag{20}$$

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

$$\boldsymbol{d}^{\text{test}} = D\boldsymbol{s}. \tag{21}$$

Regarding the problem statement given in Section II, $\boldsymbol{d}^{\boldsymbol{test}} = F\boldsymbol{y}^{\boldsymbol{test}}$ 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

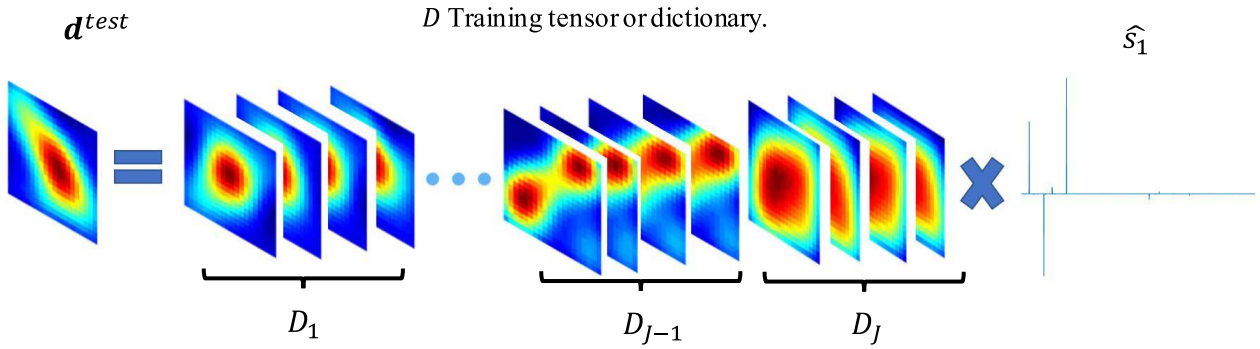

Fig. 4. Visualization of the details of the mathematical solution of the TISC algorithm for CPEs.

the following sparse format $\boldsymbol{s^*} = [\mathbf{0}, \ldots, \mathbf{0}, \boldsymbol{s}_j, \mathbf{0}, \ldots, \mathbf{0}]^T$. Sparse vector $\boldsymbol{s} \in \mathbb{R}^N$ is called the sparse indicator vector, with all entries equal to zero except those associated with the jth CPE class training subdictionary. Fig. 4 shows 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 $\boldsymbol{s^*}$ as the solution of (21) by using the following formulation:

$$P_0 : \widehat{\boldsymbol{s}}_0 = \underset{s}{\operatorname{argmin}} \|s\|_0 \quad \text{s.t. } f^{\text{test}} = Fs. \tag{22}$$

Here, the l_0-norm stands for the number of nonzero elements or cardinality of vector $\boldsymbol{s}$. 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 that 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:

$$P_{N1} : \hat{\mathbf{s}}_1 = \underset{s}{\operatorname{argmin}} \|s\|_1 \quad \text{s.t. } \|f^{\text{test}} - Fs\|_2 < \eta. \tag{23}$$

This is widely known as basic pursuit denoising regularization. Section IV-B introduces a well-known method for solving P_{N1} and the condition of the exact recovery of $\boldsymbol{s}$ in the CPE classification (refer to [46] for details of sparse recovery).

B. Greedy Sparse Solvers and Sparse CPC Solution

The orthogonal matching pursuit (OMP) algorithm (Algorithm 1) is a popular alternative greedy method to be used instead of direct optimization-based solvers. In this article, 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 ($\mu_F = \max_{1 \le i,j \le L, i \ne j} |\langle D_i, D_j \rangle| < 1/(2L-1)$), then the OMP can recover original $\hat{\boldsymbol{s}} = \boldsymbol{s}$ and the unique solution is $d = D\widehat{\boldsymbol{s}}$, having sparsity K or less using several measurements that scales such as $K\log(L/K)$ [46].

Without loss of generality, let $D = FY$ be the training tensor/dictionary created using the data points of J CP classes,[3] and for a given $\boldsymbol{d}^{\text{test}}$, let $\widehat{\boldsymbol{s}_1}$ be the optimal solution of NP_1. The selected class can be obtained as the one which its corresponding subsegment in vector $\widehat{\boldsymbol{s}_1}$ has the minimum reconstruction residual value. Fig. 4 shows a visualization of a typical SRC procedure for a CP event.

Algorithm 1 OMP

require: matrix A, measurements $\boldsymbol{d} = \boldsymbol{Ds} + \boldsymbol{n}$, stopping criterion

initialize: $\boldsymbol{r^0} = \boldsymbol{d}, \boldsymbol{s^0} = \mathbf{0}, \boldsymbol{l} = \mathbf{0}, \boldsymbol{SUP} = \emptyset$

repeat

1. match: $\boldsymbol{h^l} = \boldsymbol{D^* r^l}$
2. identify support indicator:
   $\boldsymbol{sup^l} = \{\boldsymbol{argmax}_j |\boldsymbol{h^l}(j)|\}$
3. update the support:
   $\boldsymbol{SUP^{l+1}} = \boldsymbol{SUP^l} \cup \boldsymbol{sup^l}$
4. update signal estimate:
   $\boldsymbol{s^{l+1}} = \boldsymbol{argmin}_{z:SUP(z) \subseteq SUP^{l+1}} \|\boldsymbol{d} - \boldsymbol{Dz}\|_2$
   $\boldsymbol{r^{l+1}} = \boldsymbol{d} - \boldsymbol{As^{l+1}}$
   $\boldsymbol{l} = \mathrm{l} + \mathbf{1}$

Until stopping criterion met

Output: $\widehat{\boldsymbol{s}} = \boldsymbol{s^l}$

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 [38], [39], and [44] (please refer to these references for full details and publicly available models and data set). The sampling frequency was set as 9.6 kHz to follow the practical instructions in developing PMU-embed 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 to set the signal to noise ratio at 10–100 dB randomly, based on the current PMU standards.

[3]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 CPEs AND THEIR SPECIFICATIONS

Event Name	Specifications
Capacitor Switching (CS)	Refer to ref. [40]
Arc Furnace (AF)	Refer to ref. [40]
Induction Motor Start-Up (IM)	Refer to ref. [40]
Lightening (LT)	Refer to ref. [40]
Transformer Energizing (TE)	Refer to ref. [40]
Three Phase Nonlinear Load (3L)	Refer to ref. [40]
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 $\boldsymbol{Y} \in \mathbb{R}^{M\times N}$, test sample $\boldsymbol{y}^{\boldsymbol{test}} \in \mathbb{R}^{M}$

1. Extract the feature matrix from the training data, using a random transformation matrix $\boldsymbol{F} \in \mathbb{R}^{L\times M} : \boldsymbol{D} = \boldsymbol{FY} \in \mathbb{R}^{L\times N}$.
2. Extract features of the testing data from the test sample, using the matrix $\boldsymbol{F}$ used in step 1: $\boldsymbol{d}^{\boldsymbol{test}} = \boldsymbol{F}\boldsymbol{y}^{\boldsymbol{test}}$.
3. Calculate the approximated training dictionary $\hat{D}$ using Algorithm 3.
4. Solve P_1 or NP_1 for sparse vector $\widehat{s_1}$ using OMP (Alg. 1)
5. Compute J purified vectors $\widehat{s_1}^j \, for j = 1, \ldots, J$, using indicator function $g(\widehat{s_1}) : \in \mathbb{R}^L \rightarrow \mathbb{R}^L$, such that $\widehat{s_1}^j = g(\widehat{s_1})$, is a new vector whose only nonzero entries are the entries in $\widehat{s_1}$ that are associated with class c_j.
6. Compute residual $r_j = \|\boldsymbol{d}^{\boldsymbol{test}} - D^j \boldsymbol{s}_1^j\|_2 \; for \; j = 1, \ldots, J$.
7. $j^* = \underset{j=1,\ldots,j}{argmin} \, r_j$.

output: Classify $(\boldsymbol{y}^{\boldsymbol{test}}) \triangleq c_{j^*}$

B. Featurewise Comparison

In the first case study, a set of 500 CPE from each of the 16 CPE classes in addition to a normal three-phase sinusoidal signal (total 17 classes) have been selected and the corresponding feature dictionary $D \in \mathbb{C}^{(M\times T)\times N}$ has been formed by the concatenation of the associated time–frequency feature images resulted from the most popular features in the literature [22]–[32]. The five feature tensors, spectrogram of STFT, scalogram of Morlet wavelet, ST, S-method, and HSS-transform, 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 Informative Data Samples Selection

input: Dimensionality optimized training dictionary $D \in \mathbb{R}^{L\times N}$

1- Initiate $\hat{D}$ with any arbitrary extreme point of D.
2- Find the best element that minimizes the Hausdorff distance.
$$a_j^* = \underset{a_j \in D\setminus\hat{D}}{\text{argmin}} \, d_H(\hat{D} \cup a_j, D).$$
3- $\hat{D} \leftarrow \hat{D} \cup a_j^*$
4- Return to step if the desired $\hat{N}$ or ϵ is not achieved.

output: Approximated training dictionary $\hat{D} \in \mathbb{R}^{L\times \hat{N}}$

TABLE II
FEATUREWISE COMPARISON OF CLASSIFICATION ACCURACY OF TSIC FOR 18 CLASSES OF CP EVENTS: ALL TRAINING SAMPLES (VERSUS *Informative Samples*)

CP event (all training samples)	TISC Classification Rate (%)				
	SPEC	SCA	ST	SM	HSST
CS	90(85)	90(83)	**94(90)**	93(84)	**94(90)**
AF	84(80)	84(81)	90(87)	90(87)	**90(88)**
IM	90(86)	89(85)	95(90)	95(78)	**98(97)**
LT	75(72)	73(72)	78(75)	81(79)	**83(81)**
TE	88(84)	88(82)	93(92)	93(91)	**95(94)**
3L	92(83)	92(85)	95(88)	94(75)	**96(95)**
MM	80(79)	82(80)	88(87)	87(82)	**91(90)**
AM	73(71)	70(66)	79(76)	80(81)	**82(80)**
HD	93(85)	92(88)	**97(93)**	96(93)	96(93)
LLG	92(90)	92(89)	**98(92)**	97(90)	**98(94)**
LL	91(90)	89(89)	**97(94)**	96(89)	**97(93)**
MJ	89(83)	87(84)	**94(91)**	92(82)	93(90)
Normal	89(88)	92(90)	95(90)	94(83)	**97(94)**
OB	82(79)	85(80)	**93(90)**	91(85)	**93(91)**
SLG	88(84)	90(83)	94(92)	93(92)	**97(94)**
FR	86(81)	81(80)	91(89)	90(87)	**93(90)**
FJ	65(61)	63(60)	78(77)	79(75)	**82(79)**
PJ	82(75)	79(74)	93(89)	92(87)	**96(92)**
Average	84.94 (80.77)	84.33 (80.6)	91.2 (87.88)	90.52 (84.44)	**93.16 (90.72)**

Algorithm 3 as informative training TF feature images (which varies from 10% for SLG faults to 80% for arc furnace) and revaluated 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 a limited set of informative samples. One may conclude that the HSST images are wiser choices and resulted in better classification accuracy compared with the state-of-the-art TF features that have been used in the literature. It is worth noting that these TFR have been mostly used to extract features from single-phase events in previous works.

C. Comparison Versus State-of-the-Art Techniques

ANNs and SVM are the most popular methods that have been vastly utilized in a variety of pattern recognition problems. ANNs, SVMs, and their derivations 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

TABLE III
CP EVENT ALL TRAINING SAMPLES (VERSUS INFORMATIVE SAMPLES)

CP event (all training samples)	Classification Rate (%) using HSST		
	RBF-ANN	RBF-SVM	TISC
CS	92.8(91.6)	**94.4(90.8)**	94.0(90.4)
AF	86.6(85.4)	88.2(88)	**90.2(88.8)**
IM	94.4(90.6)	98.2(92.6)	**98.6(97.8)**
LT	72.4(68.0)	75.8(70.6)	**83.2(81.6)**
TE	94.0(91.5)	**95.2(88.8)**	**95.2(94.8)**
3L	97.4(95.2)	**97.6(91.4)**	96.8(95.4)
MM	86.4(84.6)	90.0(88.2)	**91.6(90.2)**
AM	79.8(75.8)	**82.2(80.6)**	82.0(80.2)
HD	**96.2**(85.4)	94.4(88.8)	95.6(**93.4**)
LLG	97.4(88.8)	98.2(92.4)	**98.6(94.6)**
LL	96.8(87.6)	96.8(94.0)	**97.4(94.2)**
MJ	90.6(83.6)	91.6(88.6)	**93.0(90.4)**
Normal	98.2(95.2)	**97.8(94.8)**	97.2(94)
OB	91.4(88.4)	**93.8**(90.1)	93.6(**91.2**)
SLG	94.2(84.6)	95.6(91.4)	**96.0(92)**
FR	91.6(86.0)	**94.8**(90.6)	93.8(**91.8**)
FJ	74.8(70.8)	78.6(72.8)	**82.4(79.8)**
PJ	95.8(91.4)	97.4(90.4)	**97.8(92.4)**
Average	90.54 (85.8)	92.25 (86.49)	**93.16 (90.72)**

that in high-dimensional feature spaces, the usage of radial basis functions (RBFs) projects the initial feature space into an infinite-dimensional Hilbert space, which can solve the nonlinearity problem with a high accuracy [40]. As such, we have designed and evaluated an RBF 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 [40] (for details, please refer to [41]). Moreover, we implemented an RBF 1v1 SVM (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 [42]. We determined the penalty factor in addition to the adjustable parameter of the RBF using an improved ant colony optimization algorithm introduced in [42]. Table III shows the identification accuracy rate for TISC-CPE versus RBF-NN [43] and RBF-SVM [42] 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.

The average performance of all approaches is slightly over 90% with TISC faintly led by 1% margin. This is a notable achievement by 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 that 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 this article 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 that contains simultaneous three-phase patterns at once (2). Single-phase methods should be implemented separately on individual phases, which results in loosing possible relative coupling between-phase information such as unbalance.
2) According to the sparse recovery literature [36], 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 [37].
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 the TISC-CPE formulation, but it also improves the overall recovery performance by enhancing the underdeterminedness of the feature tensor by adding more columns.
5) Table IV [25] compares the time complexity of a two-layer ordinary ANN versus 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 signal s (Section IV-A).

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 [45] revealed that in three-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 monitoring central unit (MCU) with 150 MIPS, FPU, 512-kB Flash, EMIF, 12-b 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

TABLE IV
COMPUTATIONAL COMPLEXITY COMPARISON

Algorithm	ANN	Sparse
Training	$I \times \mathcal{O}(M \times J \times N)$	NA
Test	$\mathcal{O}\left(p\left(J+\left(\frac{M}{2}\right)^2\right)\right)$,	$\mathcal{O}\left(k \times (M+N)\right)$

of machine learning and AI, a deep neural network should outperform any known classification technique.

VI. CONCLUSION

In this article, we explored the problem of CPEs classification for islanding detection in power inverters using a new modified three-phase TFR called in addition to the sparse representation-based classifier. TFRs are the most popular techniques in studying nonstationary 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 CPEs 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 versus state-of-the-art algorithms.

REFERENCES

[1] S. M. Muyeen and S. Rahman, *Communication, Control and Security Challenges for the Smart Grid*. Edison, NJ, USA: IET, 2017.

[2] J. H. R. Enslin and P. J. M. Heskes, "Harmonic interaction between a large number of distributed power inverters and the distribution network," *IEEE Trans. Power Electron.*, vol. 19, no. 6, pp. 1586–1593, Nov. 2004.

[3] J. C. Balda, A. Mantooth, R. Blum, and P. Tenti, "Cybersecurity and power electronics: Addressing the security vulnerabilities of the Internet of Things," *IEEE Power Electron. Mag.*, vol. 4, no. 4, pp. 37–43, Dec. 2017.

[4] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proc. IEEE*, vol. 100, no. 1, pp. 210–224, Jan. 2012.

[5] H. H. Zeineldin, E. F. El-Saadany, and M. M. A. Salama, "Impact of DG interface control on islanding detection and nondetection zones," *IEEE Trans. Power Del.*, vol. 21, no. 3, pp. 1515–1523, Jul. 2006.

[6] F. Harirchi, M. G. Simoes, M. Babakmehr, A. AlDurra, S. M. Muyeen, and A. Bubshait, "Multi-functional double mode inverter for power quality enhancement in smart-grid applications," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, Oct. 2016, pp. 1–8.

[7] W. Hu and Y.-L. Sun, "A compound scheme of islanding detection according to inverter," in *Proc. Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Mar. 2009, pp. 1–4.

[8] K.-K.-R. Choo, M. M. Kermani, R. Azarderakhsh, and M. Govindarasu, "Emerging embedded and cyber physical system security challenges and innovations," *IEEE Trans. Dependable Secure Comput.*, vol. 14, no. 3, pp. 235–236, May 2017.

[9] R. Arghandeh and Y. Zhou, Eds., *Big Data Application in Power Systems*. Amsterdam, The Netherlands: Elsevier, 2017.

[10] R. Nicolosi, L. Piegari, and A. Benigni, "A smart PV inverter controller with PMU capability," in *Proc. Clemson Univ. Power Syst. Conf. (PSC)*, Mar. 2016, pp. 1–7.

[11] A. Khamis, H. Shareef, E. Bizkevelci, and T. Khatib, "A review of islanding detection techniques for renewable distributed generation systems," *Renew. Sustain. Energy Rev.*, vol. 28, pp. 483–493, Dec. 2013.

[12] C.-I. Chen and Y.-C. Chen, "Intelligent identification of voltage variation events based on IEEE Std 1159-2009 for SCADA of distributed energy system," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2604–2611, Apr. 2015.

[13] R. S. Kunte and W. Gao, "Comparison and review of islanding detection techniques for distributed energy resources," in *Proc. 40th North Amer. Power Symp.*, Sep. 2008, pp. 1–8.

[14] K. El-Arroudi, G. Joos, I. Kamwa, and D. T. McGillis, "Intelligent-based approach to islanding detection in distributed generation," *IEEE Trans. Power Del.*, vol. 22, no. 2, pp. 828–835, Apr. 2007.

[15] K. N. E. Ku Ahmad, J. Selvaraj, and N. A. Rahim, "A review of the islanding detection methods in grid-connected PV inverters," *Renew. Sustain. Energy Rev.*, vol. 21, pp. 756–766, May 2013.

[16] Y. Wang, J. Ravishankar, and T. Phung, "Wavelet transform-based feature extraction for detection and classification of disturbances in an islanded micro-grid," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 11, pp. 2077–2087, Jun. 2019.

[17] E. Shahryari, M. Nooshyar, and B. Sobhani, "Combination of neural network and wavelet transform for islanding detection of distributed generation in a small-scale network," *Int. J. Ambient Energy*, vol. 40, no. 3, pp. 263–273, Apr. 2019.

[18] P. K. Ray, N. Kishor, and S. R. Mohanty, "Islanding and power quality disturbance detection in grid-connected hybrid power system using wavelet and *S*-transform," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1082–1094, Sep. 2012.

[19] S. Suja and J. Jerome, "Pattern recognition of power signal disturbances using S transform and TT transform," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 1, pp. 37–53, Jan. 2010.

[20] C. Xue, L. Hui-Jin, Z. Quan-Ming, Y. Li-Ming, and L. Qing-Fen, "Power quality disturbances detection and location using mathematical morphology and complex wavelet transformation," in *Proc. 3rd IEEE Conf. Ind. Electron. Appl.*, Jun. 2008, pp. 2263–2268.

[21] S. Raza, H. Mokhlis, H. Arof, J. A. Laghari, and L. Wang, "Application of signal processing techniques for islanding detection of distributed generation in distribution network: A review," *Energy Convers. Manage.*, vol. 96, pp. 613–624, May 2015.

[22] R. R. Ch and K. H. Reddy, "Islanding detection techniques for grid integrated DG—A review," *Int. J. Renew. Energy Res.*, vol. 9, no. 2, pp. 960–977, 2019.

[23] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[24] M. Majidi, M. S. Fadali, M. Etezadi-Amoli, and M. Oskuoee, "Partial discharge pattern recognition via sparse representation and ANN," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 2, pp. 1061–1070, Apr. 2015.

[25] M. Babakmehr, H. Sartipizadeh, and M. G. Simoes, "Compressive informative sparse representation-based power quality events classification," *IEEE Trans Ind. Informat.*, vol. 16, no. 2, pp. 909–921, Feb. 2020.

[26] F. Harirchi and M. G. Simoes, "Enhanced instantaneous power theory decomposition for power quality smart converter applications," *IEEE Trans. Power Electron.*, vol. 33, no. 11, pp. 9344–9359, Nov. 2018.

[27] V. L. Merlin, R. C. Santos, A. P. Grilo, J. C. M. Vieira, D. V. Coury, and M. Oleskovicz, "A new artificial neural network based method for islanding detection of distributed generators," *Int. J. Electr. Power Energy Syst.*, vol. 75, pp. 139–151, Feb. 2016.

[28] H. R. Baghaee, D. Mlakic, S. Nikolovski, and T. D. Dragicevic, "Support vector machine-based islanding and grid fault detection in active distribution networks," *IEEE J. Emerg. Sel. Topics Power Electron.*, to be published.

[29] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, Apr. 1996.

[30] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67, Apr. 1992.

[31] R. A. Brown, M. L. Lauzon, and R. Frayne, "A general description of linear time-frequency transforms and formulation of a fast, invertible transform that samples the continuous S-transform spectrum nonredundantly," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 281–290, Jan. 2010.

[32] L. Stankovic, T. Thayaparan, and M. Dakovic, "Signal decomposition by using the S-method with application to the analysis of HF radar signals in sea-clutter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4332–4342, Nov. 2006.

[33] S. R. Mohanty, N. Kishor, P. K. Ray, and J. P. S. Catalao, "Comparative study of advanced signal processing techniques for islanding detection in a hybrid distributed generation system," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 122–131, Jan. 2015.

[34] B. Biswal, P. K. Dash, and B. K. Panigrahi, "Non-stationary power signal processing for pattern recognition using HS-transform," *Appl. Soft Comput.*, vol. 9, no. 1, pp. 107–117, Jan. 2009.

[35] A. Ashrafian, M. Rostami, and G. B. Gharehpetian, "Hyperbolic S-transform-based method for classification of external faults, incipient faults, inrush currents and internal faults in power transformers," *IET Gener., Transmiss. Distrib.*, vol. 6, no. 10, pp. 940–950, 2012.

[36] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[37] H. Sartipizadeh and T. L. Vincent, "Computing the approximate convex hull in high dimensions," 2016, *arXiv:1603.04422*. [Online]. Available: http://arxiv.org/abs/1603.04422

[38] T. Becejac, P. Dehghanian, and M. Kezunovic, "Analysis of PMU algorithm errors during fault transients and out-of-step disturbances," in *Proc. IEEE PES Transmiss. Distrib. Conf. Exposit.-Latin Amer. (PES T D-LA)*, Sep. 2016, pp. 1–6.

[39] J. Luszcz, Ed., *Power Quality Issues in Distributed Generation*. Rijeka, Croatia: InTech, 2015. [Online]. Available: https://www.intechopen.com/books/power-quality-issues-in-distributed-generation/a-comprehensive-modeling-and-simulation-of-power-quality-disturbances-using-matlab-simulink

[40] M. Valtierra-Rodriguez, R. de J. Romero-Troncoso, R. A. Osornio-Rios, and A. Garcia-Perez, "Detection and classification of single and combined power quality disturbances using neural networks," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2473–2482, May 2014.

[41] W.-M. Hung and W.-C. Hong, "Application of SVR with improved ant colony optimization algorithms in exchange rate forecasting," *Control Cybern.*, vol. 38, no. 3, pp. 863–891, 2009.

[42] C.-C. Liao, "Enhanced RBF network for recognizing noise-riding power quality events," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 6, pp. 1550–1561, Jun. 2010.

[43] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[44] S. Wang, P. Dehghanian, and L. Li, "Power grid online surveillance through PMU-embedded convolutional neural networks," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, Sep./Oct. 2019, pp. 1–8.

[45] H. Guzman, M. Bermúdez, C. Martín, F. Barrero, and M. Durán, "Application of DSP in power conversion systems—A practical approach for multiphase drives," in *Applications of Digital Signal Processing through Practical Approach*, vol. 28. London, U.K.: IntechOpen, Oct. 2015, p. 161. [Online]. Available: https://idus.us.es/bitstream/handle/11441/36155/Capitulo_OA_Barrero_2015_application.pdf?sequence=1&isAllowed=y

[46] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

Mohammad Babakmehr (Member, IEEE) received the B.S. degree in electrical engineering from Central Tehran University, Tehran, Iran, in 2008, the M.Sc. degree in biomedical-bioelectric engineering from the Amirkabir University of Technology, Tehran, in 2011, and the M.Sc. and Ph.D. degrees in electrical engineering from the Colorado School of Mines, Golden, CO, USA, in 2016 and 2017, respectively.

He was with Clemson University, Charleston, SC, USA, as a Research Associate. He is currently a Data Scientist at Ford Motor Company, Dearborn, MI, USA. His research interests include artificial intelligence, machine learning, smart grid technologies, signal processing, and big data analytics.

Farnaz Harirchi (Member, IEEE) received the B.S. degree in electrical engineering from Central Tehran University, Tehran, Iran, in 2008, the M.Sc. degree in electrical and electronic engineering from the Iran University of Science and Technology, Tehran, in 2011, and the Ph.D. degree in electrical engineering with specialties in power electronics from the Department of Electrical Engineering and Computer Science, Colorado School of Mines (CSM), Golden, CO, USA, in December 2017.

She has been with the Restoration Center, Clemson University, Charleston, SC, USA, since 2018. Her research interests include renewable energy aggregation, smart microgrids, power electronics, and intelligent control for high-power electronics applications.

Payman Dehghanian (Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2009, the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2011, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, Texas, USA, in 2017.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, George Washington University, Washington, DC, USA. His research interests include power system protection and control, power system reliability and resiliency, asset management, and smart electricity grid applications.

Dr. Dehghanian was a recipient of the 2013 IEEE Iran Section Best M.Sc. Thesis Award in Electrical Engineering, the 2014 and 2015 IEEE Region 5 Outstanding Professional Achievement Awards, and the 2015 IEEE-HKN Outstanding Young Professional Award.

Johan H. Enslin (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Rand Afrikaans University (RAU), Johannesburg, South Africa, in 1981, 1983, and 1988 respectively.

He has combined a 39-year career with leadership in industry and academia, in USA, Europe, and South Africa. He served as a senior executive for private business operations and a professor in electrical engineering. He is currently the Duke Energy Endowed Chaired Professor in Smart Grid at Clemson University, Charleston, SC, USA, and the Executive Director for the Energy Systems Program at the Zucker Family Graduate Education Center, North Charleston, SC, USA.

Dr. Enslin is a fellow of South African Institute of Electrical Engineers (SAIEE) and a registered Professional Engineer in South Africa.