

Enhance High Impedance Fault Detection and Location Accuracy via μ -PMUs

Qiushi Cui, *Member, IEEE*, and Yang Weng, *Member, IEEE*

Abstract—The high impedance fault (HIF) has random, irregular and unsymmetrical characteristics, making such a fault difficult to detect in distribution grids via conventional relay measurements with relatively low resolution and accuracy. This paper proposes a stochastic HIF monitoring and location scheme using high-resolution time-synchronized data in μ -PMUs for distribution network protection. Specifically, we systematically design a process based on feature selections, semi-supervised learning (SSL), and probabilistic learning for fault detection and location. For example, a wrapper method is proposed to leverage output data in feature selection to avoid overfitting and reduce communication demand. To utilize unlabeled data and quantify uncertainties, an SSL-based method is proposed using the Information Theory for fault detection. For location, a probabilistic analysis is proposed via moving window total least square based on the probability distribution of the fault impedance. For numerical validation, we set up an experiment platform based on the real-time simulator, so that the real-time property of μ -PMU can be examined. Such experiment shows enhanced HIF detection and location, when compared to the traditional methods.

I. INTRODUCTION

High impedance faults normally exist in distribution power systems with voltages ranging from 4 kV to 34.5 kV. Upon the occurrence of HIF, its immediate vicinity is imposed with potential danger, which is hazardous to public safety. For such reasons, researchers and engineers have been exploring novel ways to detect HIFs since the 1970s. Many algorithms, e.g., [1], aim at enhancing conventional relays at the early stage. They propose methods such as proportional relaying algorithm [2], impedance-based method [3] and PC-based fault locating and diagnosis algorithm [4], etc. However, two critical issues hinder HIF detection performance: measurement accuracy and information extraction capability. Due to these reasons, it was revealed in one case study that conventional protection cleared only 17.5% of staged HIFs [5]. Specifically, existing commercial microprocessor-based relays rely on threshold constructed by direct physical measurements, making detection logic exhibit unsatisfactory performances.

For measurement accuracy, μ -PMUs are becoming available in distribution grids, capable of providing HIF detection devices with high-precision and high-resolution measurements [6]. For example, multiple utility and campus locations have μ -PMUs installed in the distribution systems to enable further research [7]. With the data streams, phasor data concentrators (PDCs) can conduct data analytics algorithms for capturing distinct characteristics of HIFs. Therefore, this paper designs a framework for highly accurate data-driven HIF monitoring and location scheme based on μ -PMUs for distribution network protection. Specifically, we systematically design a process based on feature selections, SSL, and probabilistic learning for fault detection and location.

For feature selection, we aim at reducing feature numbers to avoid overfitting and reduce communication demand. As

μ -PMU data provides high-precision time-stamped measurements and synchronized data aggregation from multiple locations, we utilize the data to extract hidden information by considering three factors in the features: (1) how large the physical quantity becomes, (2) how fast the signal changes, and (3) how strong the harmonics are. By “how large”, we mean that the selected features need to include the signal magnitude from the discrete Fourier transform (DFT) over cycles. “How fast” considers the first-order derivative of times series data. As for “how strong each harmonic is”, it means the adopted feature needs to embed harmonic information derived from the PMU data. After information extraction, we rank features via a wrapper method instead of the filter method to leverage the output data in feature selection to keep a balance between information gain on selected features and the complexity for avoiding overfitting.

After feature selection, one can conduct supervised learning directly for obtaining the HIF results. For example, [8], [9] introduces expert systems based on artificial intelligence. [10] uses the wavelet transform to extract data and uses Bayes classifier to differentiate fault cases in the Bayesian framework. [11] uses the nearest neighborhood rule for classification simplicity. Decision tree learning method is proposed in [12] to provide a white box for interpretations. Neural network type of methods is proposed for highly nonlinear function approximation in the decision process [13], [14]. The fuzzy inference to detect HIFs can also be seen in the literature [15], [16]. However, these supervised learning methods have a strong requirement on the labeled dataset, which may not be widely available for HIF database. In addition, there are always fault or non-fault events that are never seen by the classifier. Finally, recording the events and correctly labeling them are a slow and mostly non-existent [5] process in utility companies. Therefore, we focus on utilize unlabeled data and quantify uncertainties for HIF in this paper.

For this purpose, an SSL-based method is proposed using the Information Theory for maximizing the accuracy on HIF fault detection. The work in [17] also utilizes SSL, but for the multi-class classification among four general types of events including HIFs. However, in this paper, we investigate the mathematical model of the HIF and then propose a binary classification method designated for traditional HIF detection with a fault current ranging from 10 to 50 amps [5]. Moreover, this paper adopts the wrapper approach to extract features from a systematically-designed pool with many practical and reliable features widely used in microprocessor-based relays.

For fault location, a variety of methods are proposed based on μ -PMUs data [18]–[22]. For example, [21] proposes to use compensation theorem in circuit theory to identify the source of different types of power events. [22] utilizes synchrophasor-based state estimation to locate the faults. In this paper, we

introduce an HIF location method according to the probability output of the SSL method and the HIF impedance probability distribution. For example, a probabilistic analysis is proposed via a moving window total least square based on the probability distribution of the fault impedance.

Finally, we use the real-time simulator for validation, which helps to verify the effectiveness of using real-time μ -PMU measurements. For example, we set up an experiment platform based on the real-time OPAL-RT simulator, so that the real-time property of μ -PMU can be examined. Such experiment shows enhanced HIF detection and location capability when compared to the traditional methods.

In summary, this paper makes three contributions. First, we build a feature pool and formulate a feature selection method to avoid overfitting and reduce communication bandwidth. Second, we propose an SSL method to increase data availability for HIF learning. In such a learning process, we utilize the mutual information metric to quantify the uncertainties from HIF fault and output detection confidence values for further HIF location. Third, within the probabilistic framework, we develop a HIF location method based on the anti-diode DC-source HIF model along with its fault impedance probability distribution.

This paper is organized as follows: Section II elucidates the proposed method including the feature selection method for HIF detection, the information-theoretic SSL method, and the probabilistic method for HIF location. The three methods are then combined and an implementation scheme is proposed in Section III. The experiment results are presented in Section IV, followed by the Discussions in Section V. Section VI explains the hardware experiment. Section VII draws the conclusions.

II. HIF MONITORING AND ALARM SCHEME

The proposed HIF monitoring and alarm scheme comprises three key factors in HIF detection and location: feature engineering, data classification, and fault location.

A. Feature Selection for HIF Detection

For regular fault detection in power systems, the voltage and current are ideal features since they capture most of the variance from the CT and PT measurements. Unlike regular faults, HIFs face challenges in identifying effective features due to the randomness of impedance. Therefore, past works focus on proposing different features from the time/frequency domain for indicating HIF [1], [23], [24]. However, these studies could not evaluate the effectiveness of their proposed features in contrast with other features.

To resolve this issue, we create a feature pool using various features from different proposals and form a large dataset containing HIF and non-HIF event data (the details of the dataset can be found in Section III-A). To determine a relevant subset of features, we conduct feature subset selection. The objective is to find an optimal subset of features that generates the highest possible classification accuracy. For example, we plot in Fig. 1 the percent variability explained by the top ten principal components, which provides only 52% of the data information. This means that many of the proposed

detection features are not useful when competing in the feature pool, calling an enhanced feature selection method for HIF detection.

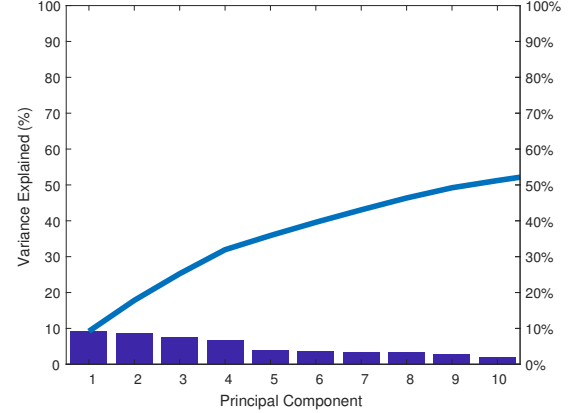


Fig. 1. First ten principal components.

Since HIFs are associated with different contact surfaces, moisture, temperatures, and locations, there is an extremely large number of parameter combinations when translated to the mathematical or simulation models of HIFs. For example, 200 features have 2^{200} possibilities for the feature set. Therefore, it is impossible to conduct simulations and experiments to cover all scenarios of HIFs and label them accordingly. This means that if we adopt the filter approach in feature selection, e.g., look into the input only, it is extremely expensive. Therefore, we adopt another key approach in feature selection, namely the wrapper approach [25]. Specifically, Fig. 2(a) shows that the filter approach attempts to assess the merits of features only from the data, ignoring the induction algorithm, which is used to induce a classifier that maps the space of feature values to the set of class values. For example, the classifier can be decision trees, support vector machines, etc. The resulting feature subset cannot guarantee good performance of the induction algorithm.

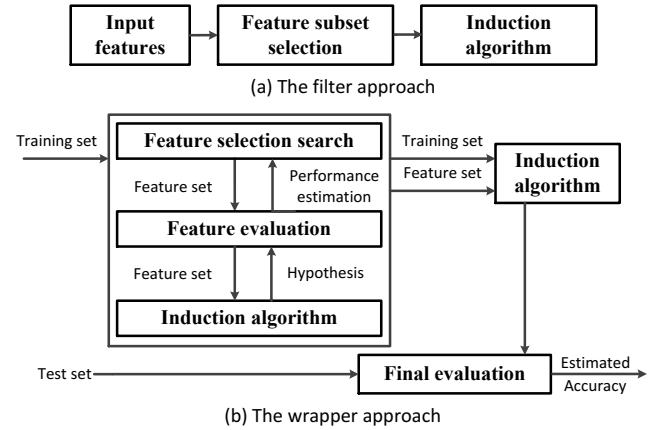


Fig. 2. Two approaches to feature subset selection [25].

Meanwhile, the wrapper approach in Fig. 2(b) evaluates the selection of feature subset using the induction algorithm itself as part of the function evaluating feature subsets. During data training, different feature sets are evaluated with the help of the feature selection search, feature evaluation, and

induction algorithm blocks. Feature evaluation estimates the performance is a certain feature set and updates the feature selection search algorithm with the current best feature set. It is noteworthy that the performance estimation refers to the accuracy estimation technique that estimates the accuracy of the induction algorithm. In this paper we adopt the cross-validation (CV) technique, the details of which are explained in Section III-A(2). Furthermore, the induction algorithm prepares the feature evaluation with the optimal parameters in a corresponding hypothesis, which is a function that can best classify the training data. Since the induction algorithm is wrapped into the selection procedure, this approach is more suitable to our setup.

B. Semi-supervised Learning for HIF Detection

After feature extraction, we can apply a supervised learning approach for HIF detection. However, the performance of supervised learning relies on the number of labeled HIF event in the past, which may not be sufficient. For this reason, we propose to employ semi-supervised to incorporate data from unseen events, so that only a small number of events needs to be labeled. Table I shows the comparison of required dataset among unsupervised learning, supervised learning, and SSL. As a highlight, SSL only requires a few labeled observations and can improve performance significantly by adding a large number of unlabeled observations, which is cheap to obtain.

TABLE I
THREE MACHINE LEARNING CATEGORIES FOR LABEL AVAILABILITY.

Category	Input dataset ^a	Labeling
Unsupervised Learning	$X = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$	$Y \in \emptyset$
Supervised Learning	$X = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$	$Y \in \mathbb{R}^{n \times 1}$
Semi-Supervised Learning	$X = [\mathbf{x}_1, \dots, \mathbf{x}_l, \dots, \mathbf{x}_{l+u}]^T$ ^b	$Y \in \mathbb{R}^{l \times 1}$

^a $\mathbf{x}_i = [x_1, x_2, \dots, x_d]$, $i \in N$, $X \in \mathbb{R}^{n \times d}$, where n denotes the number of observations and d denotes the number of features.

^b $l + u = n$, usually $l < u$, where l denotes the number of labeled observations and u denotes the number of unlabeled observations.

To utilize both labeled and unlabeled dataset, there are many SSL methods. Two popular methods are self-training and co-training approaches, where self-training makes the classifier use its own predictions to train itself, while co-training employs two classifiers to train each other with the most confident prediction labels they feel. In this paper, we choose self-training since HIF detection does not need two classifiers in co-training. For such training, we also need to choose among generative probabilistic models and discriminative models. For generative probabilistic, Gaussian mixture model is calculated between the observation and the label. The objective function attempts to learn the parametric vector θ through $\hat{\theta} = \arg \max_{\theta} (\log P(Y_L | X_L; \theta) + \alpha \log P(X_U; \theta))$, where $X_L = [\mathbf{x}_1, \dots, \mathbf{x}_l]^T$, $X_U = [\mathbf{x}_{l+1}, \dots, \mathbf{x}_{l+u}]^T$, $Y_L = [\mathbf{y}_1, \dots, \mathbf{y}_l]^T$. The first term is related to the posterior class probability of the labeled data. The second term introduces the log likelihood of the unlabeled data, with the weight α .

However, such an approach requires joint distribution, which is not available and hard to obtain due to the complicated relationship between input feature data and the output labels of HIF. This is because discriminative models focus on the conditional probability and attempt to “discriminate” the value of the label given the observation. For the discriminative model, we adopt the information-theoretic SSL method [26], because 1) it is built upon probabilistic description for the randomness of impedance in HIF, 2) it uses powerful metric such as mutual information based on previously selected features, and 3) it guarantees the globally optimal solution which means dependability in protection. Specifically, the deployed method relies on the square-loss mutual information regularization. Two useful objectives are offered in the method, namely the analytic expression of the globally optimal solution and the probabilistic values to determine HIFs. Consider a d -dimensional dataset $X \subseteq \mathbb{R}^d$ and $Y = \{1, \dots, c\}$ where c denotes the number of classes for HIF. Assume i.i.d. $\{(\mathbf{x}_i, y_i)_{i=1}^l\}$ and $\{(\mathbf{x}_i)_{i=l+1}^n\}$ with $n = l + u$ and $l < u$, where the former represents the labeled dataset, and the later represents the dataset without labels.

The goal of the proposed SSL method for HIF is to classify any $x \in X$ achieving $\hat{y} = \arg \max_{y \in Y} p(y | \mathbf{x})$. Initially, a uniform class-prior probability $p(y) = 1/c$ is assumed. Then, the class-posterior probability is approximated with $q(y | \mathbf{x}; \alpha) := \langle \mathbf{K}^{-1/2} \Phi_n(\mathbf{x}), \mathbf{D}^{-1/2} \alpha_y \rangle$, where kernel $k : X \times X \mapsto \mathbb{R}$, kernel matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$, the empirical kernel map $\Phi_n : X \mapsto \mathbb{R}^n$, $\mathbf{x} \mapsto (k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_n))^T$, the degree matrix $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ with $d_i = \sum_{j=1}^n k(\mathbf{x}_i, \mathbf{x}_j)$, the model parameter $\alpha = \{\alpha_1, \dots, \alpha_c\}$ and $\alpha_u = (\alpha_{y,1}, \dots, \alpha_{y,n})^T$. Consequently, the optimization problem in this particular SSL method is formulated as $\min_{\alpha_1, \dots, \alpha_c \in \mathbb{R}^n} \Delta(p, q) - \gamma \widehat{SMI} + \lambda \sum_{y \in Y} \frac{1}{2} \|\alpha_y\|_2^2$ (*), where the loss function is defined as the squared-difference of probabilities p and q : $\Delta(p, q) = \frac{1}{2} \int_X \sum_{y \in Y} (p(y | \mathbf{x}) - q(y | \mathbf{x}; \alpha))^2 p(\mathbf{x}) d\mathbf{x}$, the squared-loss mutual information (SMI) is given by $\widehat{SMI} = \frac{c}{2n} \text{tr}(\mathbf{A}^T \mathbf{D}^{-1/2} \mathbf{K} \mathbf{D}^{-1/2} \mathbf{A}) - \frac{1}{2}$, and $\text{tr}(\cdot)$ is the trace operator, $\mathbf{A} = (\alpha_1, \dots, \alpha_c) \in \mathbb{R}^{n \times c}$, $\gamma, \lambda > 0$ are the regularization parameters.

C. A Probabilistic Method for HIF Location

After SSL-based detection, we need to locate the fault. The location function (Fig. 3) utilizes the probability output of the SSL-based HIF detection algorithm, then ranks the prediction probability to narrow down the faulty zones. Through the probabilistic fault location estimation method which will be elucidated below, the estimated HIF location range is obtained using the μ -PMU data.

Specifically, we model HIF with the extensively used the anti-parallel dc-source model [24], as shown in Fig. 4b. Two variable resistors are both changing randomly, modeling the dynamic arcing resistance. Two sets of diodes and DC sources are connected in an anti-parallel configuration. The two DC sources are stationary but with different values, which model the asymmetric nature of HIF. The positive half cycle of HIF current is achieved when $V_{ph} > V_p$, while negative half cycle when $V_{ph} < V_n$. When $V_n < V_{ph} < V_p$, the current equals to zero, which represents the period of arc extinction. Test results

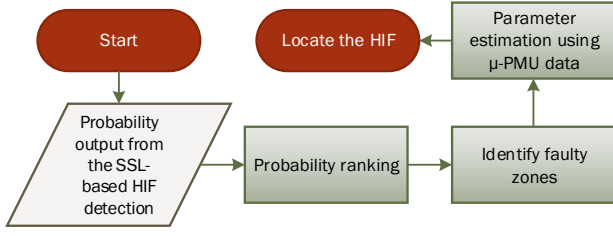


Fig. 3. The flowchart for the HIF location method.

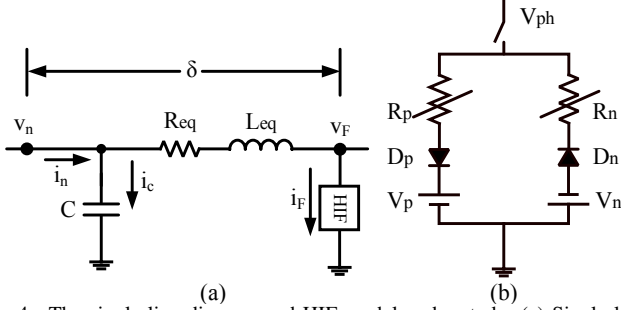


Fig. 4. The single line diagram and HIF model under study: (a) Single line model for the HIF location analysis [28]; (b) Two anti-diode HIF model [24].

of this HIF model reveal a good modeling performance and are validated in the simulation [24] and field test results [27].

Originally, we formulate the relationship between the one-terminal measurement and fault location using a constant-impedance constant-DC-source HIF model. Since constant impedance and DC source are assumed, the deployed HIF model in [28] cannot well represent the random phenomenon of fault impedance during arcing. Moreover, conventional solutions such as the one in [28] have limitations on measurement location, device, and accuracy [29], [30]. To solve these issues, we introduce randomness to the HIF model and sets up the HIF location system in the μ -PMU environment.

The single line diagram of the one terminal measurement system is demonstrated in Fig. 4a. The left-hand side terminal of Fig. 4a has the voltage values of v_n , denoting the n^{th} μ -PMU measurement. Depending on the location of the HIF fault, the equivalent line impedance $R_{eq} + j\omega L_{eq}$ is equal to the distance δ times the per-unit-length line resistance, inductance. To simplify the calculation, the shunt capacitance is lumped at the measurement terminal, consuming the current of i_C . v_F and i_F are the HIF voltage and current.

Assuming that R , L , and C represent the per-unit-length resistance, inductance, and capacitance, therefore, if we apply Kirchhoff's voltage law (KVL) in Fig. 4a circuit, we have $v_n = \delta R(i_n - i_C) + \delta L \frac{d(i_n - i_C)}{dt} + v_F$, where v_n and i_n are terminal voltage and current measurements, and v_F is the voltage at the fault point. According to the HIF model in Fig. 4b, we have the fault voltage in two cases:

$$v_F = \begin{cases} i_F R_p + V_p, & i_F \geq 0, \\ i_F R_n - V_n, & i_F < 0, \end{cases} \quad (1)$$

In underground cable networks, the line capacitance C is the major source of error in fault location [31]. However, in overhead line applications where HIF is associated with, the shunt capacitance can be neglected without generality in the distribution lines [32]. Therefore $i_C = \delta C \frac{dv_n}{dt} \approx 0$. In

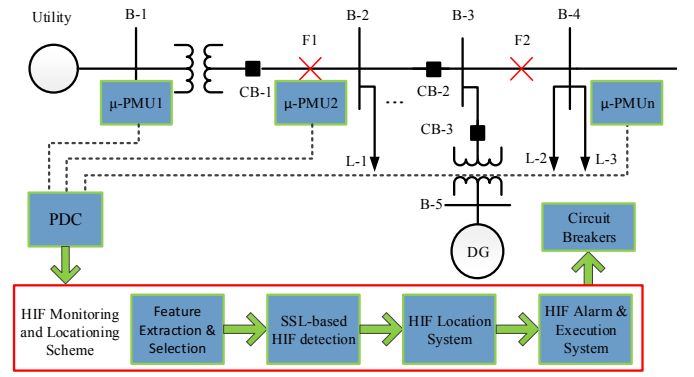


Fig. 5. The proposed HIF monitoring and alarm scheme with four function blocks: feature extraction and selection, SSL-based HIF detection, HIF location system, and HIF alarm & execution.

addition, if combining the two cases in one equation, and replacing (1) in the previously mentioned KVL equation, we have $v_n = \delta(Ri_n + L \frac{di_n}{dt}) + R_F i_F + V_{DC}$ (**), where R_F and V_{DC} have their positive and negative cycle values of R_p , R_n , and V_p , ($-V_n$) respectively.

We adopt the fault current (i_F) estimation method in [28] to estimate the distance value of the HIF, then utilize the moving window total least square method [28], [33] to estimate the variables of δ , R_F , and V_{DC} at the positive and negative cycles. Note that v_n and i_n are directly from μ -PMU data, and that $\frac{di_n}{dt}$ is derived from μ -PMU current data. From (**) we can get:

$$\begin{cases} R_p = c_{p1}\delta + c_{p0}, & i_F > 0, \\ R_n = c_{n1}\delta + c_{n0}, & i_F < 0, \end{cases} \quad (2)$$

where $c_{p1} = -\frac{1}{i_F}(Ri_n + L \frac{di_n}{dt})$, $c_{p0} = \frac{v_n - V_p}{i_F}$, $c_{n1} = \frac{1}{i_F}(Ri_n + L \frac{di_n}{dt})$, and $c_{n0} = -\frac{v_n + V_n}{i_F}$.

To better understand how the fault location is computed, we provide a simple example in Appendix B.

III. IMPLEMENTATION SCHEME

Fig. 5 shows the implementation scheme of the proposed tools. In such a scheme, μ -PMUs are installed along the feeder and phasor data concentrators (PDCs) connect the data for the red box. In the red box, an HIF scheme is shown with four function blocks: feature extraction and selection, SSL-based HIF detection, HIF location system, and HIF alarm & execution.

A. Feature Extraction and Selection

The collected data from PDC is normalized first and then goes through the algorithm of feature extraction and selection.

1) *Feature Extraction*: When the data resolution is low, although it is possible to correlate dependence between time slots, the dependence gets too weak due to the large time elapse in between. μ -PMUs provide higher sampling rate and measurement resolution comparing to traditional PMUs. Under this circumstance, some features such as harmonics are captured with high fidelity during the transient period upon a fault. This characteristic of μ -PMUs is important under the stringent detection requirements upon the HIF that has a very low fault current.

TABLE II
TYPICAL FEATURES IN THE FEATURE POOL.

Type	Example Feature
Time series feature	V_{abc}, I_{abc}
DFT-based original	$P, Q, I_{012}, V_{012}, \phi$
DFT-based 1 st order deri.	$df/dt, dV/dt, dP/dt, dH_{V1}/dt$
KF-based original	$KF-V-cos, KF-V-sin (H_1 \sim H_6)$
Other	$\theta_{V_2} - \theta_{V_0}, dV/dP, df/dP, KF-V-DC$

In the following, we show one example of how we extract features sequentially for a physical power system. We first apply DFT to estimate the voltage and current, and calculate multiple physical quantities to quantify how large their magnitudes are. Then, we take the derivative of some variables to quantify how fast the parameter changes. Such a process can capture small change which may not trigger traditional HIF detection methods. Thirdly, the harmonic magnitude coefficients (magnitudes) are estimated through KF by decomposing different harmonics and learn the transition matrix¹ (coefficient). Such features are presented as the first-order harmonics of in-phase and in-quadrature voltage components, i.e. $KF-V-cos-H_1$ and $KF-V-sin-H_1$. Finally, we convert power expert information to capture some unconventional phenomena, e.g., $\theta_{V_2} - \theta_{V_0}$, the angle difference between the negative and zero sequence voltage. Empirically, we find that such a feature is a good indicator for the unbalance level in distribution grids caused by HIF. It is noteworthy that, in the HIF monitoring and alarm scheme, the features are taken from several micro-PMUs in different locations to provide the synchronized spatial data. We adopt 200 ms² as the general window size of time-series measurement. The window size is adjustable according to the event under study. With such a systematic design, a feature pool is generated in Table II.

2) *Feature Selection*: This paper adopts the wrapper approach as its feature evaluator to solve the binary classification problem in HIF detection. Unlike the filter approach, the training set in the wrapper approach goes through three steps before it is sent to the ultimate induction algorithm: feature selection search, feature evaluation, and induction algorithm.

The best-first search engine is employed in the feature selection search, due to its robustness comparing with the hill-climbing search engine. Best-first search is a method that does not just terminate when the performance starts to drop but keeps a list of all attribute subsets evaluated so far, sorted in order of the performance measure so that it can revisit an earlier configuration instead [36]. It searches the space of attribute subsets by greedy hill-climbing augmented with a backtracking facility. Its algorithm is shown in Algorithm 1.

The evaluation function used here is five-fold CV. The assumption is that all folds are independent while training. The

¹The transition matrix here refers to the state transition matrix F . Assuming we have measurement $z(k)$ of the incoming signal $x(k)$. With the following Kalman filter model, the state variable $\mathcal{X}(k)$ can be recursively estimated: $\mathcal{X}(k+1) = F\mathcal{X}(k)$, $z(k) = H\mathcal{X}(k) + v(k)$, where F is the transition matrix and $v(k)$ is the measurement noise. Refer to [34] for the detailed implementation of the Kalman filter approach.

²In a μ -PMU with 120 frame/sec output streaming [35], its length is equivalent to 24 samples.

Algorithm 1 Best-first algorithm

- 1: Put the initial state on the *OPEN* list.
- 2: *CLOSED* list $\leftarrow \phi$, *BEST* $\leftarrow$ initial state.
- 3: Let $v = \arg \max_{w \in OPEN} f(w)$ (get the state from *OPEN* with maximal $f(w)$).
- 4: Remove v from *OPEN*, add v to *CLOSED*.
- 5: **if** $f(v) - \varepsilon > f(BEST)$ **then**
- 6: *BEST* $\leftarrow v$.
- 7: Expand v : apply all operators to v , giving v 's children.
- 8: For each child not in the *CLOSED* or *CLOSED* list, evaluate and add to the *CLOSED* list.
- 9: **if** *BEST* changed in the last k expansions **then**
- 10: goto 3.
- 11: **return** *BEST*.

number of repetitions is determined on the fly by evaluating the standard deviation of the accuracy estimate. If 1% of the standard deviation of the accuracy estimate is violated and five CVs have not been executed, an extra CV is initiated. This heuristic method works well from applicability to efficiency when dealing with large datasets. Exemplary CV applications using power system simulation data can be found in [17], [37], and [38]. Simply speaking, with the best-first search method, the exploration of feature space does not just terminate when the performance (estimated by CV) starts to drop but keeps a list of all feature subsets evaluated so far, sorted in order of the performance measure, so that it can revisit an earlier feature combination instead [39]. If we have not found an improved node in the last k expansions, we terminate the search. An improved node is defined as a node with an accuracy estimation at least ϵ higher than the best one found so far. In remaining of the paper, k is set to five and ϵ is 0.1%.

B. Classification Between HIF and non-HIF

By collecting various feature data from each μ -PMU, the HIF detection can be formulated as a large-scale binary classification problem, with the goal of accurately classifying each μ -PMU's measurement data as either 'HIF' or 'non-HIF'. This paper utilizes the SSL-based detection approach. The information-theoretic SSL method outputs the probability values for each observation based on local μ -PMU measurement. In other words, each μ -PMU is corresponding to one classification model based on its local measurement. Specifically, we get the model in (*) of Section II-B trained with the best parameters chosen by CV. This model minimizes the loss function regarding the posterior probability and maximizes the mutual information between the unseen data x_i and the label y . After this optimization process, this SSL method provides a probability matrix $Y_{prob}(i, j)$, which indicates the probability that data set x_i belongs to class j – either HIF or non-HIF.

The generalization error bounds of the proposed SSL-based detection problem can be established through its accessible analytical solution. The inequality is provided below since it guarantees a tight upper bound of the proposed protective scheme based on the inductive Rademacher complexity. The proof of which can be found in [26]. If the real class labels of the dataset X_U are available for classification (even though they are not used for training), for any $\eta > 0$ and

$0 < \epsilon < 1$, with the probability at least $1 - \epsilon$, the estimation expectation of the indicator loss function (defined as $\ell(z) = (1 - \text{sign}(z))/2$) with respect to the correct labeled data is bounded by $\mathbb{E}\ell(yf(x)) \leq \frac{1}{n} \sum_{i=1}^n \ell_\eta(y_i f(x_i)) + \frac{2B_k B_F}{\eta\sqrt{n}} + \min(3, 1 + \frac{4B_k^2 B_F'}{\eta}) \sqrt{\frac{\ln(2/\epsilon)}{2n}}$, where the output label y is now reformulated as ± 1 , multiplied by the decision function $f(x)$ that provides a closed form solution to label y as well using the optimal solution – the vector α^* ³ obtained from (*) in Section II-B, $\ell_\eta(z) = \min(1, \max(1 - z/\eta))$ is the surrogate loss function, B_k , B_F , and B_F' are the model parameters associated with α^* . The error bound equation shows how we benefit from unlabeled data by a lower empirical error when n increases. The necessity of introducing SSL in the HIF detection problem is therefore proved.

C. HIF Location Function

Once the HIF is detected, the HIF location information will be beneficial to the utility's dispatch center. The challenge of HIF location comes from the deviation of the fault impedance. Mathematically, the standard deviation of the fault impedance depends on the fault current range under study, along with the characteristics of the the ground (soil) or grounded object (concrete, tree, etc.). As mentioned in Section I, an HIF current usually ranging from 10 to 50 amps [5], which is the “blind zone” of the conventional relays. To have a larger detection range, we increase the upper bound from 50 amps to 100 amps. In order to generate a fault current between 10 and 100 A in a 25 kV benchmark system, we then calculate the fault range of 200 ~ 1,500 Ω . More details of the model settings are provided in Appendix A. Under the circumstances, we propose two types of probability distribution models to capture the deviation of the fault impedance: normal distribution and uniform distribution.

1) *Normal distribution of the HIF impedance*: In the first scenario, we have the HIF impedance R_F (including R_p and R_n) that follows $R_F \sim \mathcal{N}(\mu, \sigma^2)$. Thus, the fault location δ follows $\delta \sim \mathcal{N}(\frac{\mu - c_0}{c_1}, (\frac{\sigma}{c_1})^2)$, where both c_0 and c_1 contain their positive and negative coefficients. Furthermore, the confidence interval of the fault location estimation can be easily quantified.

2) *Uniform distribution of the HIF impedance*: In the second scenario, the HIF impedance is uniformly distributed within an impedance range. Assuming that R_F falls in the range of (R_{min}, R_{max}) , where $0 < R_{min} < R_{max}$. The range of the fault location estimation therefore belongs to $(\frac{R_{min} - c_0}{c_1}, \frac{R_{max} - c_0}{c_1})$.

D. High Impedance Fault Alarm and Execution System

Once the HIF location range is computed, the HIF Alarm and Execution System sends an assertion signal regarding the detection of an HIF to the control center of the system operator. The resulting action, depending on the local grid code, can be either an alarm, or the tripping signal sent to the execution system, which is part of the control system of the corresponding circuit breaker.

³For the reference of the audience, the analytical solution is $\alpha^* = \frac{n(nD^{-1/2}K^{1/2}BB^TK^{1/2}D^{-1/2} + \lambda nI_n - \gamma l c D^{-1/2}K D^{-1/2})^{-1} D^{-1/2} K^{1/2} B y$.

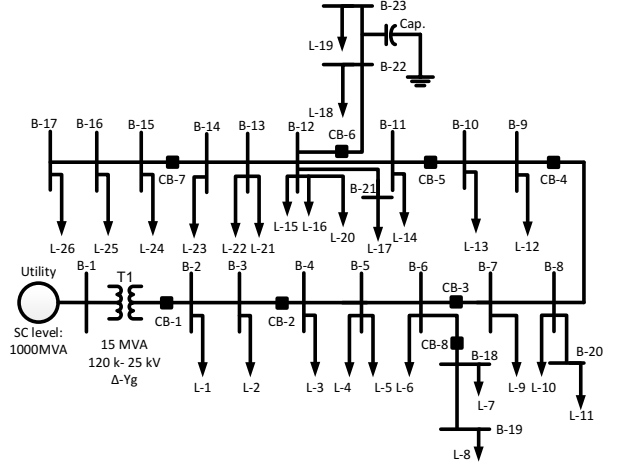


Fig. 6. Benchmark system [40].

IV. EXPERIMENT RESULTS

A. Benchmark System

The benchmark system (Fig. 6) under study is developed by McGill Electric Energy Systems Laboratory [40]. It models a 25 kV distribution feeder from a rural Canadian community. This feeder has a 120 kV-25 kV transformer connecting the utility and the feeder. The benchmark system runs in a real-time simulator which will be part of the hardware experiment platform that is explained in the next Section.

B. Feature Selection in HIF Detection

By far, we include 246 features in the feature pool. Many other features in the literature can be added to the feature pool in Table II. Therefore the feature pool is flexible and expandable to maximize the HIF detection accuracy. The best-first search algorithm is employed and the search is terminated if an improved node in the last k expansions is not found. An improved node is defined as a node with an accuracy estimation at least ϵ higher than the best one found so far. In the remainder of this paper, k is set to five and ϵ is 0.1%.

Since the wrapper approach considers how the algorithm and the training set interact, experiments have been done on real and artificial datasets to evaluate two different families of induction algorithms: decision trees (DT) and Naive-Bayes (NB). Literature indicates that both induction algorithms have achieved significant performance improvement on the same datasets [25]. Here, we provide a comparison between them in Table III. It can be seen that DT is outperforming NB in terms of the merit of test subset, interpretability and hardware applicability. This paper chooses DT as the induction algorithm due to its visualizability, interpretability, and applicability. The final selected feature group is exhibited in Table IV. All features in the second column belong to the selected feature group. 10 features are selected in this case. To clearly demonstrate the effectiveness of the selection, the feature group is categorized by the references.

The selected feature group includes sequence components of voltage and current (V_0 , I_2 , I_0), angle difference between sequence components ($\theta_{V_2} - \theta_{V_0}$, $\theta_{I_2} - \theta_{I_0}$), and the harmonic components derived from discrete Fourier transform and KF harmonic decomposition [44]. Especially for the harmonic

TABLE III
A COMPARISON BETWEEN DT AND NB INDUCTION ALGORITHM.

Algo.	# of features (original, final)	Merit of best subset found	Interpretability	Hardware applicability
DT	246, 10	1.000	High	[37]
NB	246, 14	0.999	Low	Not available

TABLE IV
THE SELECTED FEATURE GROUP RESULTING FROM WRAPPER METHOD
AND ITS REFERENCE (HIF MODEL: ANTI-PARALLEL DC-SOURCE).

Reference	Feature in the selected feature group
[1], [41]	V_0, I_2, I_0
[42], [1]	$\theta_{V_2} - \theta_{V_0}, \theta_{I_2} - \theta_{I_0}$ $V\text{-}THD\text{-}3rd,$
[1], [43], [44]	$V\text{-}THD\text{-}5th, KF\text{-}V\text{-}sin\text{-}H1,$ $KF\text{-}V\text{-}cos\text{-}H5, KF\text{-}V\text{-}cos\text{-}H1$

components, it contains the total harmonic distortion (THD) of the third and fifth order voltage signals, along with the in-phase (cos) and in-quadrant (sin) part of the KF estimated first ($H1$) and fifth ($H5$) order voltage signals. Since the PMUs cannot provide the harmonic components, voltage and current samples are collected by the PDC first, then the HIF monitoring and alarm scheme calculates the harmonics using the DFT- and KF-based signal processing techniques for multiple locations, as shown in Fig. 5. The details of both harmonic estimation methods can be found in [45] and [34].

Table V demonstrates the HIF detection performance comparison with other feature groups. Experiment data is obtained from 198 HIF events and 256 non-HIF events using MATLAB Simulink. By far, we conduct supervised learning to evaluate the binary classification performance. Meanwhile, we employ the following evaluation index:

- Precision = $\frac{TP}{TP+FP}$, where TP is the number of the true positive instances, and FP is the number of false positive instances, specifically, it is the number of incorrectly detected fault events but they are actually non-fault events. It indicates the percent of correctly predicted HIF instances over the total number of the predicted HIF instances.
- Recall = $\frac{TP}{TP+FN}$, where FN is the number of false negative instances, specifically, it is the number of incorrectly detected non-fault events but they are actually fault events. It measures the fraction between the correctly predicted HIF instances and the total number of HIF instances in the dataset.
- F1 score = $\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$. This index is the harmonic mean of the precision and recall, providing an overall evaluation of the classifier.

C. HIF Detection Performance

With the selected feature group, the proposed SSL algorithm is further tested in the benchmark systems. To test the performance of the proposed SSL algorithm, 14,580 HIF and non-HIF events are simulated and labeled. These events include a wide range of scenarios, such as HIFs with different

TABLE V
PERFORMANCE COMPARISON WITH OTHER PROPOSED FEATURES.

Method	# of Features	Precision	Recall	F1 score
[1]	2	0.519	0.203	0.292
Filter approach (top 10 features)	10	0.968	0.959	0.964
Selected feature group	10	0.979	1.000	0.989

Note: The top 10 features in filter method include: $\theta_{I_2} - \theta_{I_0}, I_2, I_0, V_2, V_0, KF\text{-}I\text{-}cos\text{-}H3, KF\text{-}V\text{-}sin\text{-}H3, KF\text{-}V\text{-}cos\text{-}H3, KF\text{-}V\text{-}cos\text{-}H5, KF\text{-}V\text{-}sin\text{-}H5$.

fault impedance, HIFs with different fault location, HIFs with different fault types, load switching, capacitor switching, load variation, etc. In order to demonstrate the concept of SSL, we pick up two typical features – df/dt and dV/dt – to visualize the complexity of the classification task in a two-dimension plot and to depict the SSL classification process of the unlabeled data. Here, the two selected features can be any other features, since the figure below focuses on the way how SSL works. Initially, 25% of the data is served as the training data, which is indicated in Fig. 7a. Concretely, these data are labeled with either circle, representing the non-fault events, or diamond, representing the fault events. At the meantime, the 75% unlabeled data is marked with “x”. It can be seen that overlapping between the non-fault and fault data is prevailing, which indicates the necessity of deploying a viable large-scale binary classification algorithm. Meanwhile, Fig. 7b shows the results of the SSL algorithm, two extra categories are added to highlight the predicted classes, marked with filled circles and diamonds. It is clear that the proposed method learns from the training data and conducts prediction on the unlabeled dataset. Moreover, although multi-dimension data is hard to visualize, Fig. 7b intuitively indicates, to some extent, that the predicted points are spatially consistent with the labeled data in Fig. 7a.

To further investigate the accuracy of the SSL-based method, we demonstrate the performance of this method in terms of regularization coefficients, precision, recall, and F1 score in Table VI. This table further compares the classification performance under two device precision: 2 digits representing the measurement from a commercial digital relay, and 4 digits representing a state-of-the-art μ -PMU’s voltage and current measurement. In this case study, the model parameters associated with α^* have the following values: $B_k = 10.0000$, $B_F = 11.0031$, and $B'_F = 147.7377$. On one hand, by comparing between 2 digits and 4 digits results, Table VI reveals and quantifies the improvement on fault detection. Take the metric of precision as an example, the enhancement on measurement precision has improved the precision by 2.82%, 0.74%, 9.69%, and 0.00% when there are 6.25%, 12.5%, 25%, and 50% of labeled data. On the other hand, statistically, the F1 score in average rises as the percentage of labeled data increases. However, to maintain a fair comparison, we randomly select the labeled data, which means that the quality of data during SSL varies case to case. However, despite of the performance oscillation due to the goodness of data, the precision, recall and F1 score have a trend to

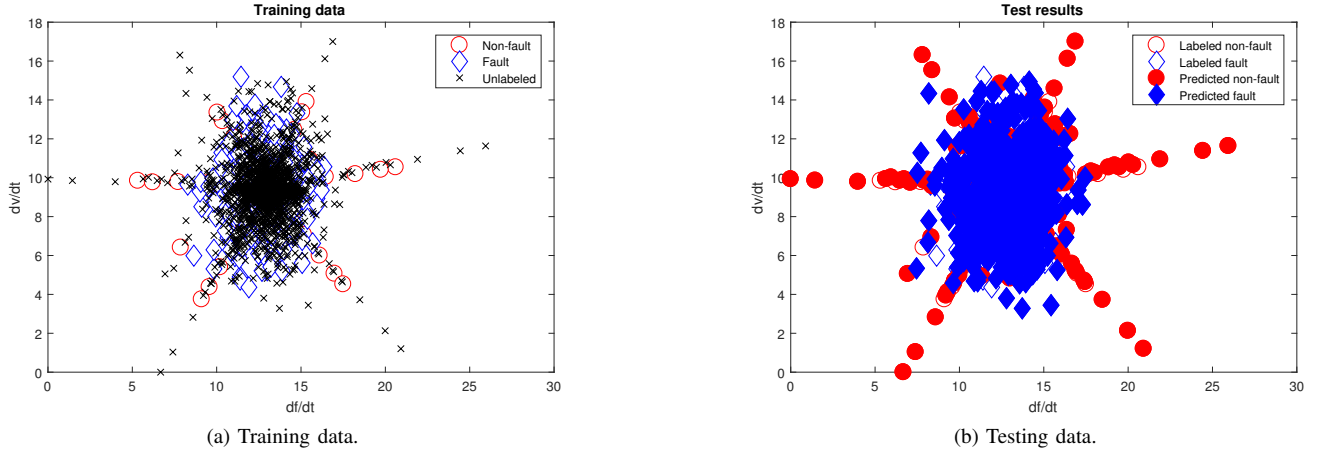


Fig. 7. Visualization of training and testing data in the SSL-based method. We utilize two typical features of df/dt and dV/dt to visualize the complexity of the classification task in a two-dimension plot and to depict the SSL classification process of the unlabeled data. The corresponding data is rescaled and standardized using the following equation: $Y = \frac{X - X_{min}}{\sigma}$, where X is the vector of the original data, Y is the rescaled and standardized vector, X_{min} is the minimum data value in the corresponding vector, and σ is the standard deviation of X . In this figure, 25% data are labeled.

TABLE VI
BINARY CLASSIFICATION RESULTS USING THE INFORMATION-THEORETIC SEMI-SUPERVISED LEARNING METHOD.

% of labeled data	Device precision (2 digits)					Device precision (4 digits)				
	γ	λ	precision	recall	F1 score	γ	λ	precision	recall	F1 score
6.25	0.01	1e-5	0.8396	0.9604	0.8959	0.01	1e-5	0.8678	0.9685	0.9154
12.5	0.1	1e-5	0.8479	0.9865	0.9120	0.01	1e-5	0.8553	0.9800	0.9134
25	0.01	1e-5	0.8031	0.9552	0.8726	0.01	1e-5	0.9000	0.9554	0.9269
50	0.1	1e-5	0.9981	0.9981	0.9981	0.1	1e-5	0.9981	0.9962	0.9972

increase as the percentage of labeled data grows as shown in Table VI. The decreasing of F1 score in the 2 digits results from 12.5% to 25% of labeled data indicates that it happens when more percentage of data is labeled but they can hardly provide more mutual information for the SSL. Nevertheless, the overall performance increases as there are more labeled data. Similarly, for the case where the F1 score in the 2 digits result is higher than that in the 4 digits results, since the percentage of labeled data is already high (50%), the device precision no longer plays a dominant role. Instead, the quality of data becomes more important, resulting in the F1 score difference of 0.0009. It comes to a conclusion that the performance of the proposed method maintains a larger than 85% precision rate when the percent of labeled data is no lower than 6.25% using the μ -PMUs, the recall rate of which is even higher. Besides, it highlights the advantage of the adoption of high precision μ -PMU over conventional relays in such protection task.

D. HIF Location

As discussed previously, one of the advantages of the information-theoretic SSL method is the probabilistic output of the data classification results [26]. This advantage enables us to narrow down the faulty zones by collecting the probabilistic output from each μ -PMU. We have simulated 18 events to test the probabilistic output of all μ -PMUs that are monitoring, from upstream, the HIF events for six critical lines: Line 2-3, 4-5, 7-8, 9-10, 11-12, 15-16. For each line, there are three fault locations tested. Clearly, under this assumption, the HIF

location is observable only at these six lines. In Table VII, we demonstrate one example on the case when the HIF is applied on Line 7-8 at 3.0 km downstream B-7 (total length of Line 7-8: 15 km). Then we rank the probability output of each μ -PMU using the averaged probability output values and determine the faulty line. As indicated by B-7 μ -PMU, the algorithm is 99.39% sure that the HIF is located at Line 7-8.

TABLE VII
PROBABILISTIC OUTPUT OF ALL μ -PMUs WHEN THE HIF IS APPLIED AT LINE 7-8 – 1.0 km DOWNSTREAM B-7.

Location of μ -PMU	B-2	B-4	B-7	B-9	B-11	B-15
Avg. prob. output over a 200 ms window size	0.2178	0.4041	0.9939	0.3808	0.1716	0.1775

Prior to presenting the location performance, the location error is defined as: $\text{Location error} = \frac{\text{estimated distance} - \text{actual distance}}{\text{line length}} \times 100\%$ [28]. The HIF location system is tested in multiple locations along the distribution feeder. Fig. 8 exhibits the testing results on the line between bus 2 and 3 (Line 2-3) of Fig. 6. Compared with the estimation error of the linear least square estimation (LSE) global behavior in [28]. The line length in [28] is 0.6 km, whereas the line tested in the benchmark system has 4.167 km. Therefore, there are more data points in the results of the proposed method. The errors in the proposed method tend to decrease when fault distance increases from 0 to 0.4 km. This is because the variation of the fault impedance plays a prominent role when the fault is

close to the measurement, and when the fault is further, the location estimation error is affected less by the fault impedance variation.

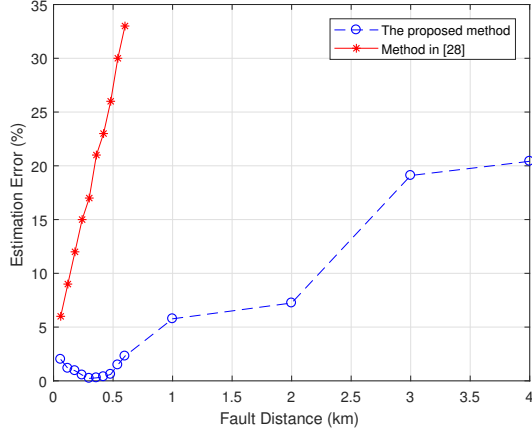


Fig. 8. Fault location error comparison. HIF impedance follows the uniform distribution in the proposed method. Line length in the proposed method: 4.167 km. Line length in [28]: 0.6 km.

Comparing with common low-impedance faults whose location can be quite accurate, the proposed fault location estimation method actually shows a significant improvement of performance. For example, for short lines less than 1 km, its location error is less than 6% comparing to [28]; while for lines up to 4 km, the method in [28] cannot location the HIF due to its limitations on measurement devices and the method. This paper aims to bridge the gap on the longer distribution line HIF location. Current practice is that once the HIF is detected, the utility begins a public address advising of the danger, and also dispatches service personnel to search for the downed conductor [5]. With the proposed method, 80% of line length does not require the searching of a line crew, and the removal of safety hazards is, therefore, more efficient.

Fig. 9 shows the fault location estimation when a fault that is 1.0 km downstream of bus 11 occurs. The distance estimation using moving window total least square method exhibits two behaviors under two types of the probability distribution of the fault impedance. When the blue band covers 85% of the estimation samples, Fig. 9a has a wider confidence interval band than Fig. 9b due to its large standard deviation of the fault impedance. In addition, in comparison with the work in [28], Table VIII is shown to highlight the advantage of the proposed HIF location method. As indicated in the table, the proposed solution is capable of handling the anti-parallel DC-source model with random fault impedance. Meanwhile, the utilization of μ -PMU devices enables higher measurement accuracy.

V. DISCUSSIONS

In this section, we first discuss the effect of μ -PMU placement, and then carry out more general evaluations considering fault scenarios in the context of measurement noise and DGs.

A. The Effect of μ -PMU Placement

It is reported that HIF instances are significantly less than those trigger the operation of the conventional protection [5].

TABLE VIII
COMPARISON WITH BENCHMARK HIF LOCATION METHOD.

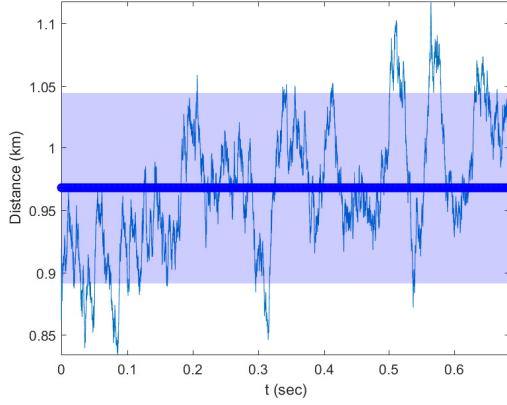
Method	HIF model	Randomness of fault impedance	Meas. device	Meas. accuracy
[28]	Anti-parallel anti-DC source	No	Disturbance monitoring equipment	0.01
[46]	Constant resistance	No	Power line communication	0.01
The proposed solution	Anti-parallel anti-DC source	Yes	μ -PMU	0.0001

Therefore, the location function of HIF is an add-on function to achieve. Similar to many other fault location methods that utilize PMUs or other devices [28], [22], the admittance matrix is the key to them. For HIF location in this paper, it is assumed that μ -PMUs are placed at each bus to get the full observability of the system. In reality, we suggest μ -PMUs are placed at the upstream of critical lines, then those lines in particular become observable for HIF location. If any load or distributed generations (DG) is at the upstream of HIFs, the proposed fault detection method can work but the location method cannot due to the unobservability issue. To determine where and how many μ -PMUs need to be placed, more work should be done on the techno-economic analysis.

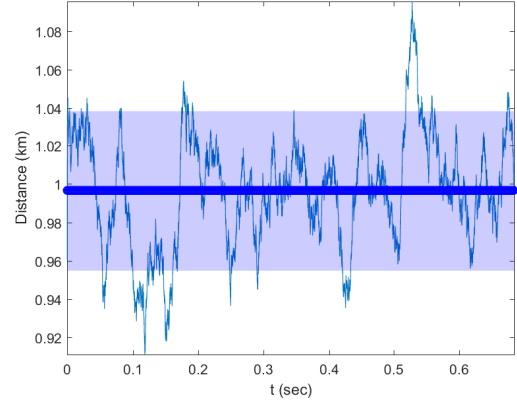
B. Performance Under Other Scenarios

1) *Considering DGs and More Fault scenarios:* To investigate the effects of environmental noises on the feature selection and the accuracy of the proposed method, the performance of the feature selection and fault location is tested under white Gaussian noise with mean 0 and variance of 10^{-6} and 10^{-4} [47]. We test the HIF detection performance on six lines (named after from bus - to bus) as can be found in Fig. 6. The HIF scenarios include 30%, 60%, and 90% of the length of each line listed. It is concluded that the selected features are not affected by environmental noise. Concretely, the superimposed noise does not change the results in Table IV. However, the accuracy of the fault location is affected by the noise. As shown in Table IX, the noise with a variance value of 10^{-4} deteriorates the fault location performance, and it gets worse when the HIF occurs at a further location. When the noise variance drops to 10^{-6} , no notable alteration in the location results is observed. Meanwhile, we integrate two DGs: one is a synchronous generator on B-8, and the other is an inverter-interfaced wind turbine on B-17. The results listed in Table IX contain the effect of DGs. For critical lines that are observable to the μ -PMUs, since the HIF cannot cause obvious voltage change, the fault location performance is hardly compromised before and after the DGs are integrated.

2) *Measurement Accuracy:* Commercial μ -PMUs have presented a satisfying accuracy in reality. It is reported that off-the-shelf μ -PMUs devices achieve an accuracy of 0.001° resolution for phasor angle, 0.0002% for phasor magnitude, and 0.01% for Total Vector of Error (TVE) [35], [48]. Since an HIF does not cause grid frequency deviation or even overcurrent



(a) HIF impedance probability distribution: normal. Average distance: 0.9679. Blue band covers 85% of the estimation samples.



(b) HIF impedance probability distribution: uniform. Average distance: 0.9967. Blue band covers 85% of the estimation samples.

Fig. 9. Estimated fault location when the empirical fault location is 1.0 km. Upstream measurement location: CB-5 (refer to Fig. 6). Line length: 1.59 km. Line 11-12 (from-bus to to-bus).

TABLE IX
FAULT LOCATION ERRORS IN DIFFERENT SCENARIOS WITH DGs
INTEGRATED IN THE SYSTEM.

Line	Noise var.	Errors (%) when HIF is located at		
		30%	60%	90%
2-3 (4.167 km)	0	7.2	15.8	19.9
	10^{-4}	8.5	16.8	20.6
4-5 (2.04 km)	0	6.1	12.4	15.6
	10^{-4}	9.8	17.7	21.7
7-8 (15.0 km)	0	5.2	10.8	17.7
	10^{-4}	1.8	6.9	25.8
9-10 (1.59 km)	0	4.1	10.6	12.2
	10^{-4}	9.0	12.0	18.4
11-12 (1.05 km)	0	2.5	6.2	7.3
	10^{-4}	11.8	16.0	19.0
15-16 (0.098 km)	0	3.9	5.9	8.4
	10^{-4}	11.2	14.0	16.6

[5], the measurement accuracy of μ -PMUs is not compromised in most of the time. However, it is noteworthy that decaying dc components may be present in fault currents, therefore, not each μ -PMU has high accuracy in such a condition. Other sources of inaccurate μ -PMUs data could be dropouts, packet loss, measurement bias, loss of GPS synchronization, etc., by which the robustness of the proposed method can be severely jeopardized. Regular testing and bad data detection mechanism are required to avoid the data quality issue.

VI. HARDWARE EXPERIMENT PLATFORM

The authors have set up a hardware experiment platform to validate the proposed HIF monitoring and alarm scheme. This platform is shown in Fig. 10. The real-time simulator emulates the 25 kV distribution feeder. Several μ -PMUs are connected to the analog output of the simulator through amplifiers. Each μ -PMU takes three-phase voltage and current signal of a certain bus and outputs the sample data to the PDC. The detection algorithm is embedded in the PC, which is programmed with all the functions of the proposed HIF monitoring and location

algorithm. The results on the detection time are shown in Table X. We test the HIF detection performance on six lines (named after from bus - to bus) in Fig. 6. The detection time is averaged over the measurement results by introducing HIFs at 30%, 60%, and 90% of the length of the line under study.

The experiment platform utilizes two μ -PMUs to collect data. As discussed in Section IV-D, current practice is to send dispatches service personnel to search for the downed conductor. Therefore, the response and processing time is not a concern for the HIF detection application. The proposed scheme actually works as backup protection to catch the HIFs that overcurrent relays cannot detect. The requirement on the response time is not strict. In fact, some applications [22], [49] have achieved the overall latency below 200 ms. It is noteworthy that the detection output of the proposed scheme is not sent back to the real-time simulator. Future work can be done on “closing the loop” by introducing the scheme output signal to the input of the simulator at the bus where circuit breaker tripping is anticipated.

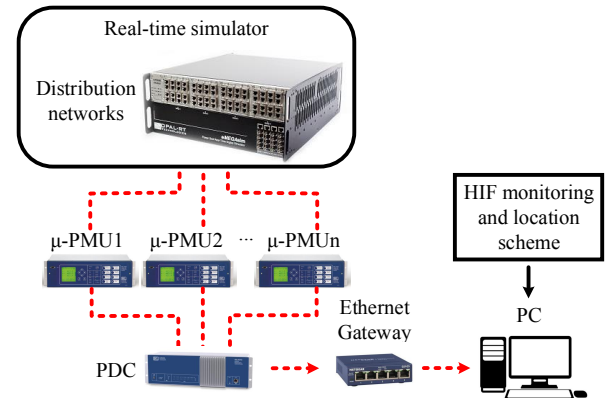


Fig. 10. The hardware experiment platform. The red dotted lines indicate physical connection among real-time simulator, μ -PMUs, PDC, Ethernet gateway and PC.

VII. CONCLUSIONS

This paper proposes a HIF monitoring and alarm scheme using μ -PMUs. Results illustrate that through the feature

TABLE X
HARDWARE EXPERIMENT RESULTS OF THE HIF DETECTION TIME.

Location of line	2-3	4-5	7-8	9-10	11-12	15-16
Avg. detection time (ms)	469	482	488	514	491	470

extraction and selection function we can have a flexible feature pool and generate a feature group that performs best under a certain induction algorithm after searching the entire feature space. The proposed SSL method not only addresses the issue of high labeling cost but also demonstrates the possibility of using the unlabeled data to help the overall classification with high accuracy. After narrowing down the faulty zones, we developed an HIF location method to locate the fault considering its probability distribution, the results of which indicate a small estimation error with the help of μ -PMUs, in comparison with previous work.

Similar to many other fault location methods that utilize PMUs or other devices, the admittance matrix is the key to them. For HIF location, it is suggested that μ -PMUs are placed at the upstream of critical lines. If any load or DG is at the upstream of HIFs, the proposed fault detection method can work but the location method cannot due to the unobservability issue. To determine where and how many μ -PMUs need to be placed, more work should be done on the techno-economic analysis. The HIF location error is mainly contributed by the variation of the fault impedance (the arcing) and the estimation of fault current. To further mitigate this error, the investigation needs to focus on a better modeling of the arcing and the fault current estimation. Future work on a model-less data-driven method to assist the HIF location can be an anticipated solution.

APPENDIX A PARAMETERS OF THE HIF MODEL

In order to generate a fault current between 10 and 100 A in a 25 kV benchmark system, the model settings are provided in Table XI.

TABLE XI
HIF MODEL SETTINGS.

Component	Value range	Values change every
V_p	5 ~ 6 kV	0.1 ms
V_n	7 ~ 8 kV	0.1 ms
R_p	200 ~ 1,500 Ω	0.1 ms
R_n	200 ~ 1,500 Ω	0.1 ms

APPENDIX B USING LSE FOR CALCULATING HIF LOCATION

We use one simplified example that views fault impedance as a constant value to show how fault location is computed. The discussion considering the probability distribution of the fault impedance can be found in Section III-C. According to [28], equation (**) in Section II-C can be reformulated into

$$v_n = \begin{bmatrix} Ri_n + L \frac{di_n}{dt} & i_F & sg_+(i_F) & sg_-(i_F) \end{bmatrix} \cdot \begin{bmatrix} \delta \\ R_F \\ V_p \\ V_n \end{bmatrix}, \quad (3)$$

where

$$sg_+(i_F) = \begin{cases} 1, & i_F > 0, \\ 0, & i_F \leq 0, \end{cases} \quad (4)$$

$$sg_-(i_F) = \begin{cases} 0, & i_F \geq 0, \\ -1, & i_F < 0. \end{cases} \quad (5)$$

In a typical LSE method, the solution to the parameter estimation of θ in $y = X\theta + \xi$ is given by

$$\hat{\theta} = (X^T X)^{-1} X^T y, \quad (6)$$

where ξ represents the environment noise. Similarly, the fault location can be also computed this way. Assume we utilize N samples of μ -PMU output streaming data (including the derived data $\frac{di_n}{dt}$, each element in equation (3) can be expanded into

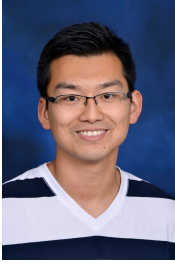
$$\begin{bmatrix} v_{n,1} \\ v_{n,2} \\ \vdots \\ v_{n,N} \end{bmatrix} = \begin{bmatrix} (Ri_n + L \frac{di_n}{dt})_1 & i_{F,1} & sg_+(i_{F,1}) & sg_-(i_{F,1}) \\ (Ri_n + L \frac{di_n}{dt})_2 & i_{F,2} & sg_+(i_{F,2}) & sg_-(i_{F,2}) \\ \vdots & \vdots & \vdots & \vdots \\ (Ri_n + L \frac{di_n}{dt})_N & i_{F,N} & sg_+(i_{F,N}) & sg_-(i_{F,N}) \end{bmatrix} \cdot \begin{bmatrix} \delta \\ R_F \\ V_p \\ V_n \end{bmatrix} \quad (7)$$

The transposed vector in equation (3), which contains the fault location variable δ , can, therefore, be computed using (6).

REFERENCES

- [1] D. C. Yu and S. H. Khan, "An adaptive high and low impedance fault detection method," *Power Delivery, IEEE Transactions on*, vol. 9, no. 4, pp. 1812–1821, Oct 1994.
- [2] J. Carr, "Detection of high impedance faults on multi-grounded primary distribution systems," *IEEE Transactions on Power Apparatus and Systems*, vol. 4, no. PAS-100, pp. 2008–2016, 1981.
- [3] M.-S. Choi, S.-J. Lee, D.-S. Lee, and B.-G. Jin, "A new fault location algorithm using direct circuit analysis for distribution systems," *Power Delivery, IEEE Transactions on*, vol. 19, no. 1, pp. 35–41, 2004.
- [4] J. Zhu, D. L. Lubkeman, and A. A. Girgis, "Automated fault location and diagnosis on electric power distribution feeders," *Power Delivery, IEEE Transactions on*, vol. 12, no. 2, pp. 801–809, 1997.
- [5] J. Tengdin, R. Westfall *et al.*, "High impedance fault detection technology report," *PSRC Working Group D15*, 1996.
- [6] A. von Meier, E. Stewart, A. McEachern, M. Andersen, and L. Mehrmanesh, "Precision micro-synchrophasors for distribution systems: A summary of applications," *Smart Grid, IEEE Transactions on*, vol. 8, no. 6, pp. 2926–2936, Nov 2017.
- [7] E. M. Stewart, A. Liao, and C. Roberts, "Open pmu: A real world reference distribution micro-phaser measurement unit data set for research and application development," *IEEE Power Engineering Letters*, Tech. Rep., Oct 2016.
- [8] C. L. Benner and B. D. Russell, "Practical high-impedance fault detection on distribution feeders," *Industry Applications, IEEE Transactions on*, vol. 33, no. 3, pp. 635–640, 1997.
- [9] C. Kim, B. D. Russell, and K. Watson, "A parameter-based process for selecting high impedance fault detection techniques using decision making under incomplete knowledge," *Power Delivery, IEEE Transactions on*, vol. 5, no. 3, pp. 1314–1320, 1990.
- [10] A.-R. Sedighi, M.-R. Haghifam, O. Malik, and M.-H. Ghassemin, "High impedance fault detection based on wavelet transform and statistical pattern recognition," *Power Delivery, IEEE Transactions on*, vol. 20, no. 4, pp. 2414–2421, 2005.

- [11] T. Lai, L. Snider, E. Lo, and D. Sutanto, "High-impedance fault detection using discrete wavelet transform and frequency range and rms conversion," *Power Delivery, IEEE Transactions on*, vol. 20, no. 1, pp. 397–407, 2005.
- [12] Y. Sheng and S. M. Rovnyak, "Decision tree-based methodology for high impedance fault detection," *Power Delivery, IEEE Transactions on*, vol. 19, no. 2, pp. 533–536, 2004.
- [13] A. F. Sultan, G. W. Swift, and D. J. Fedirchuk, "Detection of high impedance arcing faults using a multi-layer perceptron," *Power Delivery, IEEE Transactions on*, vol. 7, no. 4, pp. 1871–1877, 1992.
- [14] C. Kim and B. D. Russell, "A learning method for use in intelligent computer relays for high impedance faults," *Power Delivery, IEEE Transactions on*, vol. 6, no. 1, pp. 109–115, 1991.
- [15] M. Haghifam, A. Sedighi, and O. Malik, "Development of a fuzzy inference system based on genetic algorithm for high-impedance fault detection," *IEEE Proceedings Generation Transmission and Distribution*, vol. 153, no. 3, p. 359, 2006.
- [16] A. Etemadi and M. Sanaye-Pasand, "High-impedance fault detection using multi-resolution signal decomposition and adaptive neural fuzzy inference system," *Generation, Transmission & Distribution, IET*, vol. 2, no. 1, pp. 110–118, 2008.
- [17] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5152–5162, 2018.
- [18] S. Kantra, H. A. Abdelsalam, and E. B. Makram, "Application of pmu to detect high impedance fault using statistical analysis," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.
- [19] M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Location identification of high impedance faults using synchronized harmonic phasors," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2017, pp. 1–5.
- [20] B. C. de Oliveira, J. L. R. Pereira, I. D. de Melo, M. A. de Souza, and G. de Oliveira Alves, "A new methodology for high impedance fault detection, classification and location using pmus," in *2018 Simposio Brasileiro de Sistemas Eletricos (SBSE)*. IEEE, 2018, pp. 1–6.
- [21] M. Farajollahi, A. Shahsavari, E. M. Stewart, and H. Mohsenian-Rad, "Locating the source of events in power distribution systems using micro-pmu data," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6343–6354, 2018.
- [22] M. Pignati, L. Zanni, P. Romano, R. Cherkaoui, and M. Paolone, "Fault detection and faulted line identification in active distribution networks using synchrophasors-based real-time state estimation," *IEEE Transactions on Power Delivery*, vol. 32, no. EPFL-ARTICLE-218391, pp. 381–392, 2017.
- [23] K. Sagastabeitia, I. Zamora, A. Mazón, Z. Aginako, and G. Buigues, "Low-current fault detection in high impedance grounded distribution networks, using residual variations of asymmetries," *Generation, Transmission & Distribution, IET*, vol. 6, no. 12, pp. 1252–1261, 2012.
- [24] S. Gautam and S. M. Brahma, "Detection of high impedance fault in power distribution systems using mathematical morphology," *Power Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 1226–1234, 2013.
- [25] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1, pp. 273–324, 1997.
- [26] G. Niu, W. Jitkrittum, B. Dai, H. Hachiya, and M. Sugiyama, "Squared-loss mutual information regularization: A novel information-theoretic approach to semi-supervised learning," in *International Conference on Machine Learning*, 2013, pp. 10–18.
- [27] N. I. Elkalashy, M. Lehtonen, H. A. Darwish, M. A. Izzularab, and I. T. Abdel-maksoud, "Modeling and experimental verification of high impedance arcing fault in medium voltage networks," *Dielectrics and Electrical Insulation, IEEE Transactions on*, vol. 14, no. 2, pp. 375–383, 2007.
- [28] L. U. Iurinic, A. R. Herrera-Orozco, R. G. Ferraz, and A. S. Bretas, "Distribution systems high-impedance fault location: A parameter estimation approach," *Power Delivery, IEEE Transactions on*, vol. 31, no. 4, pp. 1806–1814, Aug 2016.
- [29] M. Farajollahi, M. Fotuhi-Firuzabad, and A. Safdarian, "Deployment of fault indicator in distribution networks: A mip-based approach," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2259–2267, 2018.
- [30] J. Nunes, A. Bretas, N. Bretas, A. Herrera-Orozco, and L. Iurinic, "Distribution systems high impedance fault location: A spectral domain model considering parametric error processing," *International Journal of Electrical Power & Energy Systems*, vol. 109, pp. 227–241, 2019.
- [31] S. Jamali and V. Talavat, "Fault location method for distribution networks using 37-buses distributed parameter line model," in *Eighth IEEE International Conference on Developments in Power System Protection*, vol. 1, Apr 2004, pp. 216–219 Vol.1.
- [32] W. H. Kersting and W. H. Phillips, "Distribution feeder line models," *Industry Applications, IEEE Transactions on*, vol. 31, no. 4, pp. 715–720, Jul 1995.
- [33] L. Ding, T. Bi, and D. Zhang, "Transmission line parameters identification based on moving-window tls and pmu data," in *2011 International Conference on Advanced Power System Automation and Protection*, vol. 3, Oct 2011, pp. 2187–2191.
- [34] I. Kamwa, S. Samantaray, and G. Joos, "Wide frequency range adaptive phasor and frequency pmu algorithms," *Smart Grid, IEEE Transactions on*, vol. 5, no. 2, pp. 569–579, 2014.
- [35] Power Standards Lab. (2014) Micro-phasor measurement unit brochure. [Online]. Available: <https://www.powerstandards.com/download-center/micropmu/>
- [36] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [37] Q. Cui, K. El-Arroudi, and G. Joos, "Islanding detection of hybrid distributed generation under reduced non-detection zone," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5027–5037, 2018.
- [38] U. B. Parikh and B. R. Bhalja, "Svr-based current zero estimation technique for controlled fault interruption in the series-compensated transmission line," *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1364–1372, 2013.
- [39] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [40] D. Zhuang, "Real time testing of intelligent relays for synchronous distributed generation islanding detection," Master's thesis, McGill University, Dec 2012.
- [41] H. Laaksonen and P. Hovila, "Method for high-impedance fault detection," *CIREN - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 1295–1299, 2017.
- [42] S. Huang, L. Luo, and K. Cao, "A novel method of ground fault phase selection in weak-infeed side," *Power Delivery, IEEE Transactions on*, vol. 29, no. 5, pp. 2215–2222, Oct 2014.
- [43] A. A. Girgis, W. Chang, and E. B. Makram, "Analysis of high-impedance fault generated signals using a kalman filtering approach," *Power Delivery, IEEE Transactions on*, vol. 5, no. 4, pp. 1714–1724, 1990.
- [44] I. Kamwa, S. R. Samantaray, and G. Joos, "Compliance analysis of pmu algorithms and devices for wide-area stabilizing control of large power systems," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1766–1778, May 2013.
- [45] T. Radil and P. M. Ramos, "Methods for estimation of voltage harmonic components," in *Power Quality*. IntechOpen, 2011.
- [46] A. N. Milioudis, G. T. Andreou, and D. P. Labridis, "Detection and location of high impedance faults in multiconductor overhead distribution lines using power line communication devices," *Smart Grid, IEEE Transactions on*, vol. 6, no. 2, pp. 894–902, 2015.
- [47] J. Zhao and L. Mili, "A framework for robust hybrid state estimation with unknown measurement noise statistics," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 1866–1875, 2018.
- [48] M. Jamei, A. Scaglione, C. Roberts, E. Stewart, S. Peisert, C. McParland, and A. McEachern, "Anomaly detection using optimally-placed μ pmu sensors in distribution grids," *IEEE Transactions on Power Systems*, 2017.
- [49] M. Li, A. Pal, A. G. Phadke, and J. S. Thorp, "Transient stability prediction based on apparent impedance trajectory recorded by pmus," *International Journal of Electrical Power & Energy Systems*, vol. 54, pp. 498–504, 2014.



Qiushi Cui (S'10-M'18) received the M.Sc. degree from Illinois Institute of Technology, and the Ph.D. degree from McGill University, both in Electric Engineering. Currently, he is a Postdoctoral Scholar in the Ira A. Fulton Schools of Engineering of Arizona State University (ASU). Prior to joining ASU, he was a Research Engineer at OPAL-RT Technologies Inc. from Dec. 2015 to Dec. 2017.

His research interests are in the areas of machine learning and big data applications in power systems, power system protection, smart cities, microgrid,

EV integration, renewable energies, and real-time simulation in power engineering. Dr. Cui won the Best Paper Award at the 13th IET International Conference in Developments in Power System Protection in Edinburgh, UK, in 2016. He was the winner of the Chunhui Cup Innovation and Entrepreneurship Competition for Overseas Chinese Scholars in Energy Sector in 2018.

Dr. Cui received the Postdoctoral Research Scholarship from Le Fonds de recherche du Québec – Nature et technologies, and held a Canada MITACS Accelerate Research Program Fellowship.



Yang Weng (M'14) received the B.E. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China; the M.Sc. degree in statistics from the University of Illinois at Chicago, Chicago, IL, USA; and the M.Sc. degree in machine learning of computer science and M.E. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University (CMU), Pittsburgh, PA, USA.

After finishing his Ph.D., he joined Stanford University, Stanford, CA, USA, as the TomKat Fellow for Sustainable Energy. He is currently an Assistant Professor of electrical, computer and energy engineering at Arizona State University (ASU), Tempe, AZ, USA. His research interest is in the interdisciplinary area of power systems, machine learning, and renewable integration.

Dr. Weng received the CMU Deans Graduate Fellowship in 2010, the Best Paper Award at the International Conference on Smart Grid Communication (SGC) in 2012, the first ranking paper of SGC in 2013, Best Papers at the Power and Energy Society General Meeting in 2014, ABB fellowship in 2014, and Golden Best Paper Award at the International Conference on Probabilistic Methods Applied to Power Systems in 2016.