Finding Faults in PV Systems: Supervised and Unsupervised Dictionary Learning Approaches with SSTDR

Ayobami S. Edun*, *Student Member, IEEE*, Cody LaFlamme*, *Student Member, IEEE*, Samuel Kingston†, *Student Member, IEEE*, Harsh Vardhan Tetali*, *Student Member, IEEE*, Evan Benoit†, *Student Member, IEEE*, Cynthia M. Furse †‡, *Fellow, IEEE*, Michael Scarpulla†, *Senior Member, IEEE*, and Joel B. Harley*, *Member, IEEE*

* Department of Electrical & Computer Engineering, University of Florida, Gainesville, USA.

† Department of Electrical & Computer Engineering, University of Utah, Salt Lake City, Utah, USA.
‡ Livewire Innovation

***Abstract*—This paper explains the use of supervised and unsupervised dictionary learning approaches on spread spectrum time domain (SSTDR) data to detect and locate disconnections in a PV array consisting of five panels. The aim is to decompose an SSTDR reflection signature into different components where each component has a physical interpretation, such as noise, environmental effects, and faults. In the unsupervised dictionary learning approach, the decomposed components are inspected to detect and localize faults. The maximum difference between actual and predicted location of the fault is 0.44 m on a system with five panels connected to an SSTDR box with a leader cable of 59.13 m and total length of 67.36 m including the effective length of the panels. In the supervised dictionary learning approach, the dictionary components are used to classify the SSTDR data to their respective fault types. Our results show a 97% accuracy using the supervised learning approach.**

***Index Terms*—Dictionary Learning, Sparse Coding, SSTDR, K-SVD, Solar Panels, Faults**

I. Introduction and Motivation

Solar arrays consist of solar panels connected in series and parallel configurations. The wires connecting the panels can be treated as transmission lines. In such a setup, the transmission line is susceptible to faults, such as line-to-line faults [1] and line to ground faults [2]. Faults in photovoltaics (PV) include ground faults [3], arc faults [4], and several other types of faults [5], [6]. PV faults often occur due to panel short circuits, open circuits, degradation, shading, and hotspot effects [7]. These faults create an impedance discontinuity and if not quickly corrected, they can impair PV reliability, durability, and subsequently lead to system breakdown [8]. To detect these faults, protection devices such as ground fault detection and interruption (GFDI) have been developed, but they cannot localize faults. Several other methods have been studied [9].

In this paper, we demonstrate the use of a dictionary learning algorithm to detect, classify, and locate disconnects in a solar array with spread spectrum time domain reflectometry (SSTDR) without no dependence on baseline measurements.

Prior work has shown that reflectometry can be used to accurately locate faults in solar panels and their connecting wires [6], measure capacitance of PV cells [10], characterize lumped elements in the middle of a transmission line [11], detect ground faults [12], detect arc faults [13], etc. In reflectometry schemes, an incident signal is sent through the electrical system. This signal reflects at points of impedance mismatch and returns to the test point. SSTDR uses a modulated pseudo-noise (PN) code as the incident signal, which is then correlated with the reflected PN code. From this correlated signature, the time delay between the incident and reflected responses gives the distance to the fault. The magnitude and phase of the reflection gives the characteristics and strength of the impedance mismatch [14], [15]. Impedance mismatch could result from cable connectors, soft faults (such as cable wear), disconnects, and other kinds of faults. The multiple reflections from all the possible impedance mismatches in transmission lines increase the complexity of these schemes and make the identification of faults difficult to visually distinguish in the signals.

To reduce the effect of these complexities in the reflected signals (reflection signature), baseline subtraction is often used in the fault detection scheme [16], [17]. The reflection signature of a healthy transmission line/PV system is stored to be used as a baseline. The reflection signature of a faulty setup is then subtracted from the baseline to remove the reflections from normal parts of the system (connectors, junctions, etc.) and make the reflections from the faults more clear. A major challenge with this baseline approach is that PV systems are subject to diverse environmental conditions, such as stress, humidity, corrosion, dust, etc. [18], [19], which can change the impedance and hence reflection response of the system [20]. This makes the baseline unstable, as the reflection signature of a healthy system will change over time. This can appear as noise in the reflection responses and exacerbates the already complex nature of the reflection signatures, necessitating the need for more advanced fault detection algorithms. In this paper, we demonstrate the use of dictionary learning, which has the following advantages. First, it is robust to noise and has been used for denoising [21], [22], [23]. Secondly, unlike other learning algorithms, dictionary learning decomposes a

signal into several components that represent different physical aspects of the system. Hence, dictionary learning can help separate the effects of the environment from that of a fault.

The goal of dictionary learning is to learn a low-dimensional representation of data. Signal reflections from photovoltaics lie in a high-dimensional space and possess high variability. Similar high-dimensional data are pervasive in fields such as machine learning, computer vision, image processing, and pattern recognition. Examples of such data include microarrays, which measure gene expression [24], images [25], [26], [27], and videos [28]. Inherent in many of these high-dimensional data is a low-dimensional representation, which contains a bases that can be used to represent the observed data. Reflection signatures can be represented as a linear or affine combination of the extracted basis signals. However, there often exist infinitely many combinations of ways in which such data can be represented as a combination of basis signals. Hence the representation is not unique.

To mitigate this challenge, a sparse representation of the data is enforced by dictionary learning. That is, the data is forced to be represented as a combination of only a few data points in the dictionary. Sparse representations have been used in image processing problems, such as image denoising [29] and image compression [30], but there exists little or no work that applies this theory to assessing faults in transmission lines and photovoltaics (PV) other than to predict short-term solar PV production [31].

In this work, we combine the ability of spread spectrum time domain reflectometry (SSTDR) to measure impedances on energized systems with the ability of dictionary learning and sparse representations, to identify and localize faults in a PV array. The goal is to learn basis signals that characterize faults, and environmental conditions while also using the learned dictionary for classifying new signals for the various kinds of faults. We consider only open circuits faults (disconnects). Our results show an overall accuracy of 97% for detecting and classifying these faults.

The rest of this article is organized as follows. In Section II, we give an overview on the theory of sparse coding, unsupervised and supervised dictionary learning using K-SVD and DK-SVD, respectively. We describe our experimental setup in Section III. In section IV, we discuss the results of applying these algorithms to detect and locate faults in the reflection signatures obtained from the experiments. Finally, in section V, we give our conclusions and future work.

II. Sparse Coding and Dictionary Learning

Sparse coding allows us to understand the contributions of different components in our signals, which may be the baseline or the result of environmental changes. The goal of sparse coding or a sparse representation is to express a given signal of dimension N as a linear combination of a small number of the basis signals stored in a dictionary. These basis signals, called atoms, are usually of unit norm and are stored column-wise in a dictionary matrix. We denote the dictionary as $\mathbf{D}$ and the atoms as $\boldsymbol{d}_k$ where $k = 1, ..., K$ and K represents the size of the dictionary. Hence $\mathbf{D} \in \mathbb{R}^{N \times K}$. The dictionary is said to be overcomplete if $K > N$ and undercomplete otherwise.

Mathematically, we denote the representation as:

$$\mathbf{Y} = \mathbf{D}\mathbf{X} \tag{1}$$

$$\mathbf{y}_i = \sum_{k=1}^{N} \mathbf{d}_k x_i^k \tag{2}$$

where $\boldsymbol{x}_i$, the i^{th} column of $\mathbf{X}$, of dimension K, denotes the sparse coefficients for the i^{th} data in $\mathbf{Y}$ and x_i^k is the k^{th} index of the vector $\mathbf{x}_i$. Hence $\mathbf{X} \in \mathbb{R}^{K \times M}$ and $\mathbf{Y} \in \mathbb{R}^{N \times M}$ where M is the number of input signals. Representing $\mathbf{Y}$ as a linear combination of atoms gives infinitely many combinations when the dictionary is undercomplete. Hence we enforce sparse constraints on $\mathbf{X}$ and relax the requirement of obtaining an exact representation. The goal is to represent an input signal $\boldsymbol{y}_i$, the i^{th} column of $\mathbf{Y}$, as a sparse linear combination of the dictionary atoms with a representation error less than some threshold ξ. In other words, we choose atoms (or basis) in the dictionary that best explain the reflection signature.

Mathematically, the problem formulation is:

$$\min_x \|\mathbf{x}_i\|_0 \quad \text{s.t. } \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_F^2 \leq \xi \tag{3}$$

where $\|\cdot\|_0$ denotes the L_0 pseudo-norm and $\|\cdot\|_F$ denotes the Frobenius norm. The solution to (3) is NP-hard (non-deterministic polynomial-time hard), because the L_0 norm is known to be non-convex. This necessitates the need for approximation algorithms. There exist several polynomial time approximation algorithms that iteratively solve for the sparse vector $\mathbf{X}$ in a greedy manner. Examples include matching pursuit (MP) [32] and orthogonal matching pursuit (OMP) [33].

There are several algorithms to learn a dictionary from a set of input signals. Examples include union of orthonormal basis [34], maximum a posteriori probability approaches [35], maximum likelihood methods [36], K-singular value decomposition (K-SVD) [37], and the method of optimal directions (MOD) [38]. Note that the dictionary can be learned in a supervised or unsupervised way. In this paper, we use discriminative K-singular value decomposition (DK-SVD) [39], a supervised learning approach for classification of faults and K-SVD as an unsupervised technique for explaining the physical interpretation of each dictionary atom.

A. K-SVD Algorithm

To learn the dictionary or low-dimensional signals that the reflection signatures consist of, we used the K-SVD algorithm.

The problem formulation is:

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \quad \text{subject to} \quad \|\mathbf{x}_i\|_0 \leq T, \forall i \tag{4}$$

The goal is to obtain a dictionary that minimizes the reconstruction error defined as $\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F$ where $\|\cdot\|_F$ denotes the Frobenius norm. In this algorithm, as shown in (4), we need to solve for $\mathbf{D}$, the dictionary and $\mathbf{X}$, the sparse coefficients. To achieve this, the dictionary update and sparse coding are done iteratively but treated as disjoint problems. As a result, any sparse coding algorithm can be used for obtaining the sparse coefficients, and that choice is independent of the dictionary learning algorithm. Orthogonal matching pursuit (OMP) was

chosen for our case. First we initialize $\mathbf{D}$ with some random complex matrix and then find $\mathbf{X}$ – the matrix of sparse coefficients. We then proceed to update the dictionary atoms using singular value decomposition (SVD).

Readers are referred to [37] for a more rigorous detail of the K-SVD algorithm. In this paper, we use K-SVD as our dictionary learning algorithm because of its high efficiency.

B. DK-SVD Algorithm

If we have labels for the reflection signatures representing each type of faults, then a supervised dictionary learning algorithm will be able to cluster our data as well as automatically identify faults. The K-SVD algorithm learns an adaptive dictionary, but it can not be used for supervised classification. Therefore, we use DK-SVD to learn a discriminative dictionary and classifier. The classifier is then used to classify a reflection signature as a reflection during the daylight, night, and with disconnections at different locations in the PV setup. This algorithm has a training and a testing stage.

1) Training Stage: The dictionary $\mathbf{D}$ and classifier $\mathbf{W}$ are learned at this stage. The problem formulation is:

$$\min_{\mathbf{D},\mathbf{X},\mathbf{W}} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 + \alpha\,\|\mathbf{H}-\mathbf{W}\mathbf{X}\|_F^2 \quad \text{subject to } \|\mathbf{x}_i\|_0 \le T, \forall i \tag{5}$$

where $\mathbf{Y}$ is a matrix of training data, $\mathbf{H}$ is a matrix of training data labels passed as an input to the algorithm, $\mathbf{W}$ is the classifier we aim to learn, and α is a regularization parameter. The variable α controls the trade-off between the reconstruction error ($\mathbf{Y}-\mathbf{D}\mathbf{X}$) and classification error ($\mathbf{H}-\mathbf{W}\mathbf{X}$). First, we create a matrix of labels $\mathbf{H} \in \mathbb{R}^{C\times M}$, where C is the number of classes and M is the number of signals. The columns of $\mathbf{H}$ will have a value of 1 in the corresponding index signifying the class the input (reflection signature) belongs to.

To solve this problem, we create new matrices $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{D}}$:

$$\widetilde{\mathbf{Y}} = \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\mathbf{H} \end{pmatrix} \tag{6}$$

$$\widetilde{\mathbf{D}} = \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{W} \end{pmatrix} \tag{7}$$

we then use K-SVD with $\widetilde{\mathbf{Y}}$ as input as shown:

$$\widetilde{\mathbf{D}}, \mathbf{X} = \arg\min_{\widetilde{\mathbf{D}},\mathbf{X}} \left\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{D}}\mathbf{X}\right\|_F^2 \quad \text{subject to } \|\mathbf{x}_i\|_0 \le T, \forall i. \tag{8}$$

the dictionary $\mathbf{D}$ and classifier $\mathbf{W}$ can then be extracted from $\widetilde{\mathbf{D}}$. Where $\widetilde{\mathbf{D}} \in \mathbb{R}^{(N+C)\times K}$, $\mathbf{D} \in \mathbb{R}^{N\times K}$ and $\mathbf{W} \in \mathbb{R}^{C\times K}$. After extracting $\mathbf{D}$, and $\mathbf{W}$, we need to normalize them. The dictionary and classifier are normalized column-wise as:

$$\mathbf{D} = \begin{bmatrix} \frac{\mathbf{d}'_1}{\|\mathbf{d}'_1\|_2} & \frac{\mathbf{d}'_2}{\|\mathbf{d}'_2\|_2} & \cdots & \frac{\mathbf{d}'_k}{\|\mathbf{d}'_k\|_2} \end{bmatrix} \tag{9}$$

$$\mathbf{W} = \begin{bmatrix} \frac{\mathbf{w}'_1}{\|\mathbf{w}'_1\|_2} & \frac{\mathbf{w}'_2}{\|\mathbf{w}'_2\|_2} & \cdots & \frac{\mathbf{w}'_k}{\|\mathbf{w}'_k\|_2} \end{bmatrix} \tag{10}$$

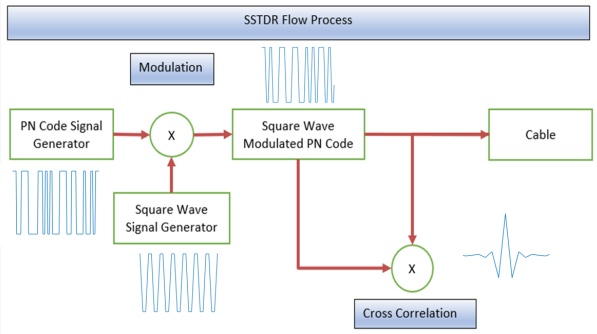


Fig. 1: *SSTDR block diagram.*

where $\mathbf{d}'_i$ and $\mathbf{w}'_i$ are the i^{th} column of $\widetilde{\mathbf{D}}$ and $\widetilde{\mathbf{W}}$ respectively. This completes the training stage.

2) Testing Stage: The classification of test signals is based on the dictionary $\mathbf{D}$ and classifier $\mathbf{W}$ that was learned in the training stage. To determine what class a test data (test reflection signature) belongs to, we first obtain its sparse coefficient $\mathbf{x}_i$ using OMP. The sparse coefficients are then multiplied by the classifier we learned in the training stage.

$$\mathbf{c}_i = \mathbf{W}\mathbf{x}_i \tag{11}$$

where $\mathbf{W} \in \mathbb{R}^{C\times K}$ and $\mathbf{x}_i \in \mathbb{R}^{K\times 1}$ hence $\boldsymbol{c}_i$ is a vector of dimension C corresponding to the number of classes. The index with the largest value is selected as the corresponding label of the test data.

III. Experimental Setup

SSTDR [40] is based on the reflection of an incident signal through a transmission line. When a signal is sent down a transmission line, reflections occur from every point of impedance mismatch [15]. SSTDR transmits a modulated pseudo-noise (PN) spread spectrum signal into a transmission line and then correlates the reflected and incident signals to create the reflection response shown in Fig. (1) [41]. The SSTDR signal has statistical characteristics that are similar to Gaussian noise, giving it the advantage that it can be used on live or energized systems without interference [15]. This fault detection technique does not depend on the panel current, voltage or maximum power point tracking (MPPT). The SSTDR hardware from Livewire innovation [42] comes with different modulation frequencies, which range from 187.5 kHz to 48.0 MHz. The incident wave is sent down the transmission line and is reflected at points of impedance mismatch. The

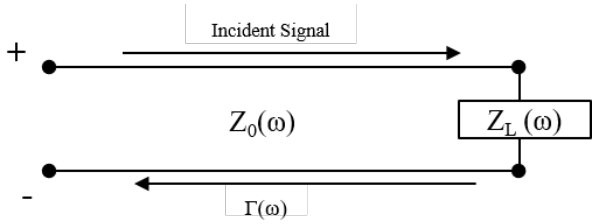


Fig. 2: *Transmission line at a load.*

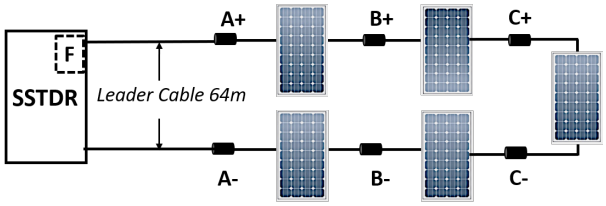


Fig. 3: *Experimental Setup.*

reflected signal is then correlated with the incident signal to produce the reflection signature. To calculate the distance to fault, the time domain reflection signature must be multiplied by the velocity of propagation to convert it to distance [41].

Figure (2) shows a transmission line with a characteristic impedance $Z_0(\omega)$ and a load $Z_L(\omega)$. The complex frequency reflection coefficient at the input is defined as

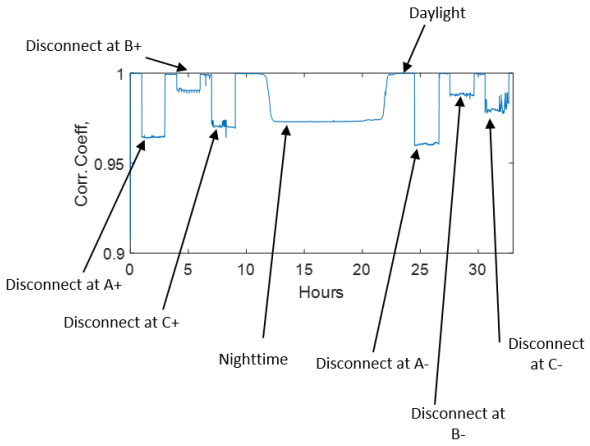
$$\Gamma(\omega) = \frac{Z_L(\omega) - Z_0(\omega)}{Z_L(\omega) + Z_0(\omega)} \tag{12}$$

When the end of the line is an open circuit, we expect a perfect reflection with a complex frequency-domain reflection coefficient of 1. Similarly, a short circuit gives a complex frequency-domain reflection coefficient of -1. More generally, negative reflection coefficients represent loads with an impedance less than the transmission line characteristic impedance while positive reflection coefficients signify a load with higher impedance than the transmission line characteristic impedance.

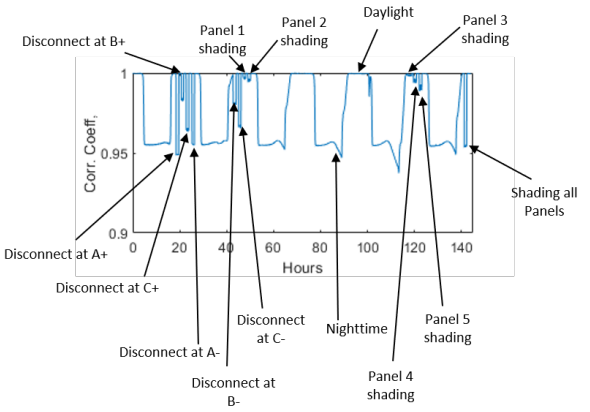
To validate the use of dictionary learning, we conducted two experiments with the exact same setup. The first experiment was conducted on the 12th of July 2019, while the second experiment was conducted on the 1st of October 2019. In each experiment, the SSTDR box was connected to five Renogy PV panels in series via a 59.13 m AWG 10 PV cable. The characteristics of the panels are shown in Table I. Two 29.56 m cables were connected using an MC4 connector for the 59.13 m leader cable. Similarly, each panel is connected using MC4 connectors. The length of the cable connecting one panel to the other is 0.91 m. These connections are illustrated in Fig. 3. We have labelled each panel connection for a clear description and analysis of the experiment. Note that SSTDR has two test modes – static and intermittent. A static test is used to test the PV connection at an instant of time while intermittent tests can be used to monitor the whole system. In an intermittent test, SSTDR signal is sent at a fixed interval of time (usually 256 signals a second), and the resulting wire

TABLE I: *Characteristics of the panel*

Solar Panel Characteristics	
Module Type	RNG-100D
Maximum Power (Pmax) at STC	100 W
Open-circuit Voltage (Voc)	22.5 V
Maximum System Voltage (Vmp)	18.9 V
Optimum Operating Current (Imp)	5.29 A
Optimum Operating Voltage (Vmp)	18.9 V
Max System Voltage	600 V DC (UL)
Short Circuit Current (Isc)	5.75 A
Max Series Fuse Rating	15 A



(a)



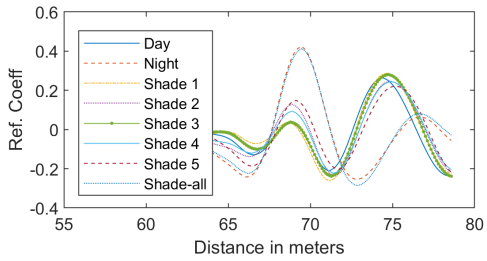
(b)

Fig. 4: *(a) Correlation coefficient of all measurements from the experiment on 12th July 2019 with the first measurement. (b) Correlation coefficient of all measurements from the experiment on 1st October 2019 with the first measurement.*

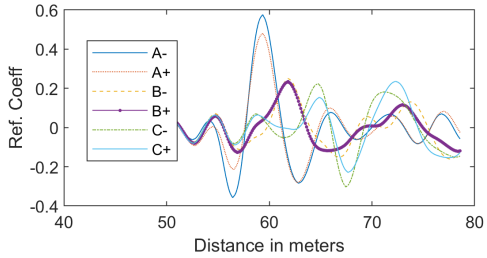
scan data is saved for future use. This test mode was used throughout our experiments.

In the first experiment, the intermittent test was carried out for two (2) days to observe the changes in the SSTDR signal during daylight in contrast to the night time. While the intermittent test was running, we induced open circuit faults at the locations A+, A-, B+, B-, C+, and C- on different days during daylight. The respective MC4 connectors were disconnected for about 1 hour in each case. A data set of 1,732,000 reflection responses was generated for the two days of the experiment. By taking every 1000th sample, we extract training data set of 1732 measurements. This data set was used as the training data set for the DK-SVD algorithm. Each measurement was stored column-wise to get matrix $\mathbf{Y}$ and labelled accordingly to get matrix $\mathbf{H}$.

Similarly, in the second experiment, the intermittent test was carried out for six (6) days. While the intermittent test was running, we induced open circuit faults at the locations A+, A-, B+, B-, C+ and C- on different days during daylight. The respective MC4 connectors were disconnected for about 1 hour in each case to give a long run of measurements such



(a)



(b)

Fig. 5: *(a) Time domain reflection signature of daylight, nighttime, shading of each and shading of all panels. (b) Time domain reflection signature of disconnects at locations A, B and C.*

that it will be visible in the plot. We induced full shading on all panels together and on each of the panels individually at different times in an attempt to detect a shading condition and what panel is shaded. Using K-SVD, we could detect that a panel is shaded but further analysis is needed to locate the exact panel that is shaded. For this reason, we do not show the result of shading. For our classification task, we have a total of 8 classes: daylight, night, disconnect at A+, disconnect at A-, disconnect at B+, disconnect at B-, disconnect at C+, disconnect at C-. A data set of 7,642,308 reflection signatures was generated for the six days of the experiment's run. Each reflection signature is stored column-wise in a matrix. Since the data set is large, all reflection signatures cannot be plotted in a single plot. We correlated every measurement with the first SSTDR measurement received. This data set was used as the test set for both K-SVD and DK-SVD.

Figure 4 shows a minimum correlation of 0.94 between all measurements and the first measurement. In the correlation plot in Fig. 4b, we observe a daily periodic pattern in the plot. Our investigations revealed the lower valleys represent the data at night while the regions with a correlation of approximately one represent the data during the day. The measurements corresponding to the disconnects and shading have been marked on the correlation plot.

A. Time-Domain Reflection Signatures

Figure 5 shows the measured time-domain reflection signatures for all classes. Reflection signatures from the day, night, shading of the panels are shown in Fig. 5a. The reflection signature from the night is clearly distinguishable and similar to shading of all panels. We see that the reflection signatures from shading of single panels overlap and are harder to distinguish. Similarly, Fig. 5b shows reflection signatures form the six possible disconnections. Notice that the reflection from A+ and A- could be expected to be the same, because at the same distance on the transmission line, but they are actually quite different. This is because SSTDR has a filter on its positive terminal, which reduces the amplitude of reflections. This also applies to the case of [B+, B-] and [C+, C-] [43].

IV. Results

A. K-SVD

The goal of this paper is to use our dictionary learning for fault detection, classification and localization. We learned a dictionary of 30 atoms with a sparsity of 4. Each atom represents a physical aspect of the system, such as the location of strong reflection, as will be described below. A sparsity of 4 means the reflection signature is a linear combination of 4 of the possible 30 dictionary atoms. We have chosen a small dictionary size for ease of analysis. In building the dictionary, we chose 200 as our number of iterations. At the end of the iterations, the algorithm outputs a dictionary matrix and corresponding sparse coefficients for the measurements. Once the dictionary is learned, it is fixed and can be used for detecting and locating faults.

Figure 6 shows the coefficient map of how the dictionary atoms are combined across the measurements. The x-axis shows the 7643 measurements we passed as input to the K-SVD algorithm, while the y-axis shows the 30 dictionary atoms numbered from 0 to 29. The coefficient map shows the magnitude of the coefficients for each measurement. Each column of the coefficient map shows the contribution of each atom in representing a measurement. Comparing Fig. 4b and Fig. 6, we observe a periodic pattern in the coefficient map. We observe that the 7^{th} atom corresponds to the dips we see in Fig. 4b. These dips represent the data at night time. Similarly, we observe that the 4^{th} atom conveys information about the daylight. All measurements are expressed as a combination of daylight or night time and three other dictionary atoms.

To detect a fault, we inspect the sparse coefficients corresponding to each measurement. We refer to the chosen

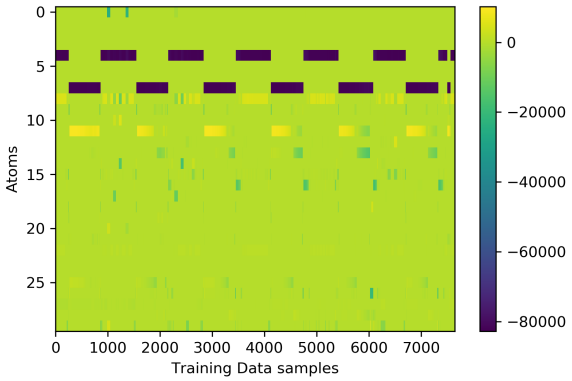


Fig. 6: *K-SVD dictionary coefficient map showing the major components of each reflection signature from the experiment conducted on the 1st of October 2019*

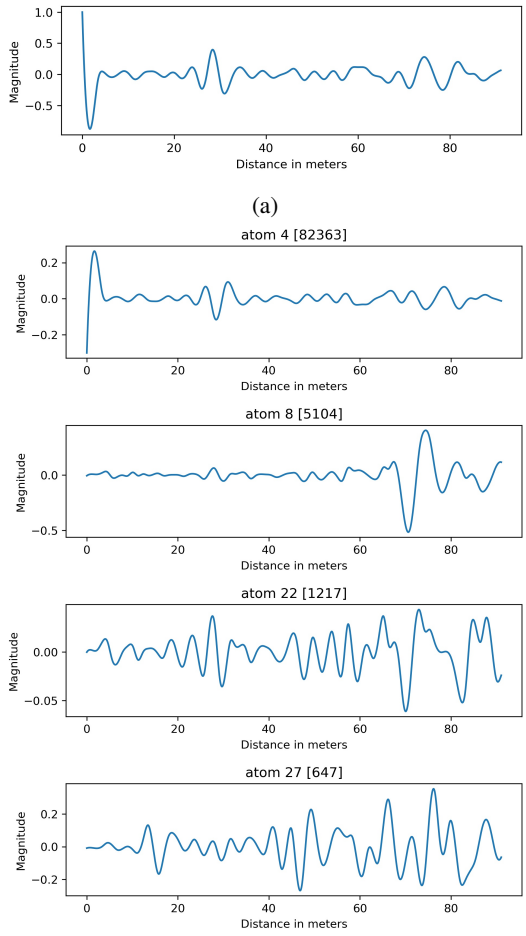


Fig. 7: *(a) Reflection signature during daylight. (b) The decomposition of the daylight reflection signature into several components*

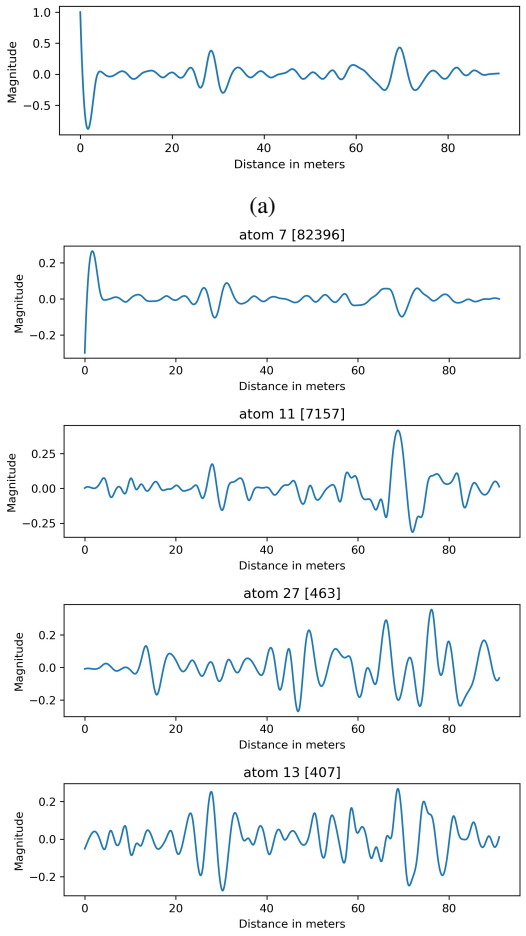


Fig. 8: *(a) Reflection signature during night time. (b) The decomposition of night time reflection signature into several components*

dictionary atoms as the prominent components of a reflection signature (with 4 major components in this case, since we chose a sparsity of 4). Each atom is then ranked based on the weight of the coefficients from highest to lowest. The atom with the highest coefficient is called the first prominent component and so on. We observed that the first prominent component (the atom with the highest magnitude) captures the baseline of the experimental system setup, while the second prominent component (the atom with the second highest magnitude) captured the fault in the system, and the third and fourth prominent component convey some other information.

We compare the decomposition of data from daylight and night time. Fig. 7 shows the decomposition of a daylight data. Observe that atom 4 has the highest coefficient as shown in the figure. Hence this atom captures the baseline reflection signature during the day. Atom 8 has the second highest coefficient and this atom captures information about the reflection of the SSTDR signal from all the panels in the setup. Atoms 22 and 27 have very small coefficients, and they capture noise and environmental conditions.

Similarly, Fig. 8 shows the decomposition of a night time data, and atom 7 has the highest coefficient as shown in the figure. This atom captures the baseline reflection signature during the night time. Atom 11 again captures information about the reflection of the SSTDR signal from all the panels in the setup. We observe there are more variations in the magnitude of the signal from 0m until the end of the signal relative to atom 8. This shows that the reflection from the panels and cables changes with the time of day. This is probably because the panels have cooled down, and their effective impedance is different than in the day time. Atoms 13 and 27 capture noise and possibly some other environmental information. K-SVD was able to learn a separate and representative baseline for a reflection signature during daylight and at night time.

Similar analysis was done for reflection signatures when there is a fault. We examine a reflection signature that corresponds to a disconnection at A-. Due to space constraints, we will show results only for the negative side of the experimental setup. In Fig. 9, we show the decomposition of the reflection signature of a disconnect at A-. The strongest component is atom 4 which corresponds to the baseline signal during the day. Atom 0, the second prominent component, gives information about the fault in the system. To locate the fault, we inspect the location of the peak of this signal – 59.13

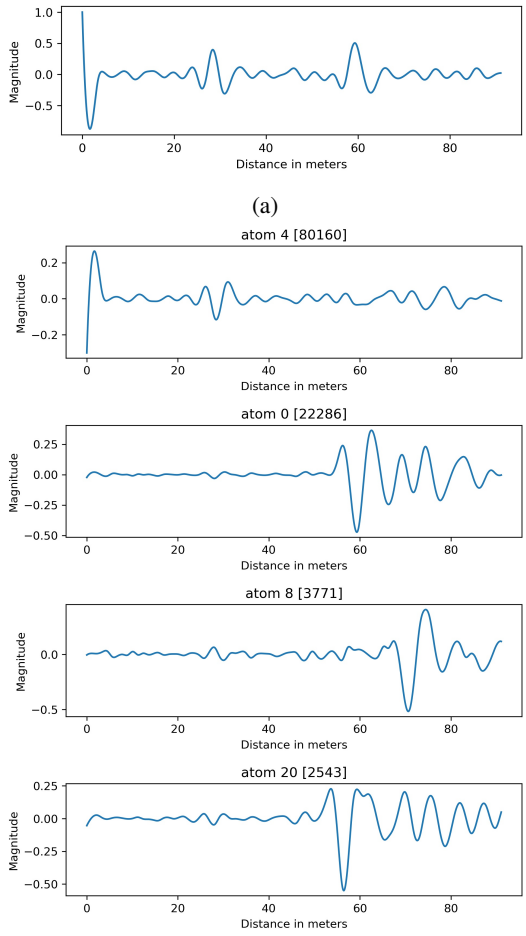


Fig. 9: *(a) Reflection signature with disconnect at A-. (b) The decomposition of the reflection signature into several components*

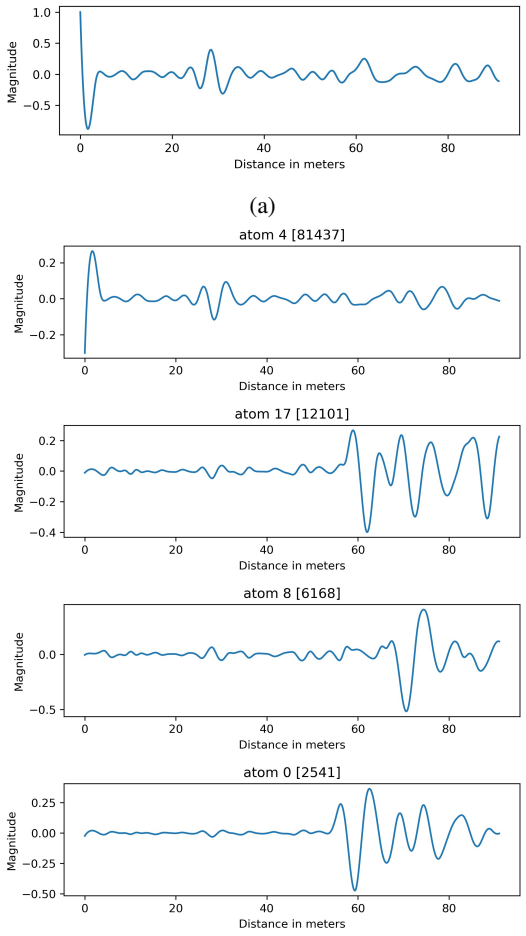


Fig. 10: *(a) Reflection signature with disconnection at B-. (b) The decomposition of the reflection signature into several components*

m (194 ft) which is the location of the disconnect as shown in Fig. 3. This approach was used to inspect measurements (reflection signatures) that correspond to a disconnection at B- and C-, and the results are shown in Table II. The maximum and average difference between the disconnect location and the estimated location are 3.22 m and 1.8 m respectively. Note that the difference in localization between disconnects location and estimated location by the algorithm is due to the effective length contributed by the panel. We observed that each panel contributes a length of approximately 1.83 m (6 ft), which means the initial location of disconnects should be corrected by adding $1.83p$ to the initial location where p here is the number of panels before that location. After correcting the location of the disconnects, the maximum and average difference between the corrected location of disconnects and the estimated location drops to 0.44 m and 0.26 m respectively. By inspecting the peak of the second prominent component, we can detect and locate disconnections in a PV system.

TABLE II: *Results of the localization of disconnects*

Disconnection Localization			
Label	Disconnect location (m) (cables only)	Corrected disconnect location (m)	Estimated location (m)
A-	59.13	59.13	59.33
A+	59.13	59.13	59.33
B-	60.05	61.87	62.02
B+	60.05	61.87	62.02
C-	60.96	64.62	64.18
C+	60.96	64.62	64.18

B. DK-SVD

If we have training data for each fault, then we can utilize DK-SVD. In this method, as explained in section IVB, we learn both a dictionary $\mathbf{D}$ and a classifier $\mathbf{W}$ using the training measurements from the first experiment and the class labels. Recall that there are 8 labels in our training data. These are Day, Night, disconnect at A+, disconnect at A-, disconnect at B+, disconnect at B-, disconnect at C+, disconnect at C-

1) Training stage: Each measurement from the training data was normalized by the maximum value of the measurement and stored column-wise to get matrix $\mathbf{Y}$, and the matrix $\mathbf{H}$ was obtained from the corresponding labels. Using matrix $\mathbf{Y}$ and the matrix $\mathbf{H}$ as input, a dictionary of 110 atoms with a

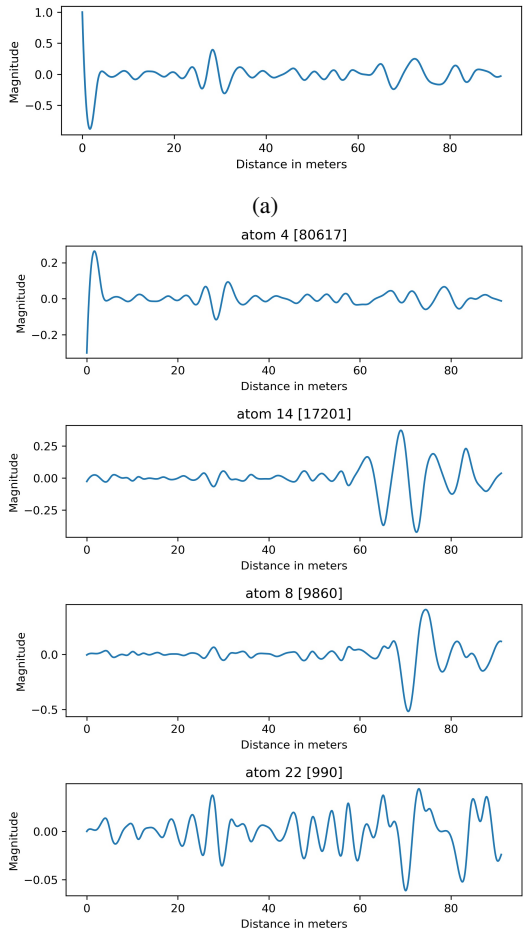


Fig. 11: *(a) Reflection signature with disconnection at C-. (b) The decomposition of the reflection signature into several components*

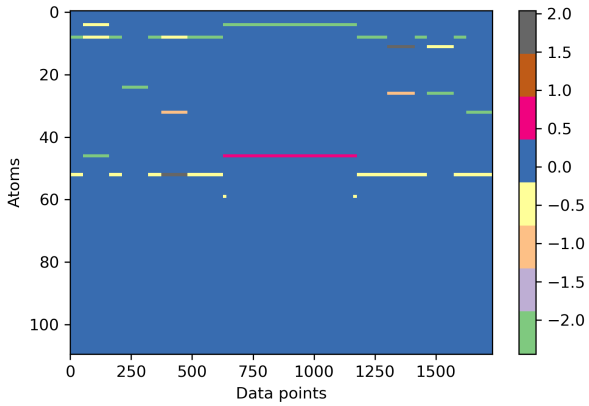


Fig. 12: *DK-SVD dictionary coefficient map showing the major components of each reflection signature of the experiments conducted 1st July,2019*

sparsity of 4 was trained. This means a reflection signature is decomposed as a linear combination of 4 atoms within the dictionary. For regularization, an $\alpha = 4$ was used during training. In our training, we chose 100 as our number of iterations.

Figure 12 shows the coefficient map of the training reflection signatures with respect to how the dictionary atoms were used. The x-axis shows the 1732 training measurements which were passed as input to the DK-SVD algorithm, while the y-axis shows the dictionary atoms numbered from 0 to 109. The coefficient map shows the magnitude of the coefficients for each measurement. Each column of the coefficient map shows the contribution of each atom in representing a measurement. Comparing Fig. 4a and Fig. 12, we observe that the 4^{th} and 46^{th} atom corresponds to night time in Figure. 4b. Also, the 8^{th} and 52^{nd} atom are the major components of the daylight measurements. Similarly, the 11^{th}, 24^{th} and 32^{nd} atoms are the major components of the disconnects at A-, B- and C-, respectively. All measurements are expressed as a combination of four dictionary atoms.

C. Testing stage:

The data generated in the second experimental setup was used as the testing data set. Note that the experiment was conducted 3 months after the first experiment for the training data. We use the dictionary **D** and classifier **W** learned during the training stage to classify each test data. First, we used OMP to determine the sparse coefficients of each measurement with respect to the dictionary atoms we learned during the training phase. OMP produces a vector of length 110 (size of the dictionary) with 4 non-zero elements. Each test data was then classified as explained in section IIB.

Figure 13 is a confusion matrix that shows the accuracy of classification into the different kinds of disconnects. The accuracy of classification of the disconnects at daylight, A+, B+,B- and C- is 100%. However we see some confusions between classification of the daylight and nighttime. This is because there is no proper definition of when the time daylight begins or when it ends. Also observe that disconnects at C+ and A- were classified as C- and A+ respectively. This is quite acceptable because as shown in [11], the reflection at either side of the wire should be the same. Overall, the DK-SVD algorithm achieved a 97% classification accuracy.

We also tested the DK-SVD algorithm with test data from shading each panel (a single panel at a time), and all panels. Note that shading data was not included in the training data. As a result, the algorithm classified the shading of each panel as a data from daylight. This is expected since the plots in Fig. 5a reveal how similar data from shading is to data from daylight. Also, the data from shading all panels was classified as nightime which again is expected as revealed in Fig. 5a.

V. Conclusion and Future Work

Spread spectrum time domain reflectometry is a viable means of locating faults on energized transmission lines. This work has introduced a supervised and unsupervised learning algorithm to analyze SSTDR data from a PV array.

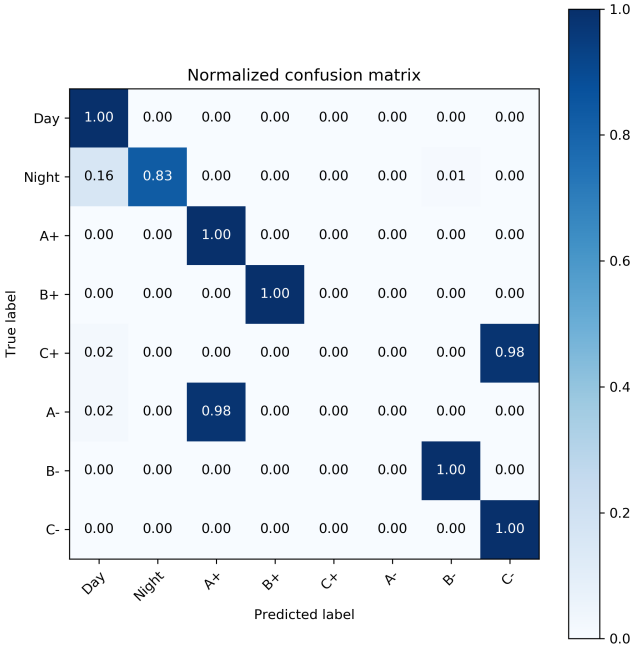


Fig. 13: *Classification of test measurements into different disconnects based on location with DK-SVD*

Using K-SVD, we show that we can decompose an SSTDR reflection signature into several components where each component conveys different information. The component with the highest coefficient represents the baseline, while the component with the highest coefficient conveys whether there is a fault or not. This component was also used to locate the distance to fault. The predicted location of the fault using K-SVD closely matches the actual location of the faults.The maximum and average difference between the corrected location of disconnects and the estimated location drops to 0.44 m and 0.26 m respectively. For the case where labels of the data are available and used as an input to the algorithm, we were able to directly classify a reflection signature into the different classes considered during the training stage. Results show a 97% accuracy for classifying faults to the corresponding faults types based on the location of disconnect. As a future work, other faults such as ground and arc faults will be considered.

VI. Disclosure

Dr. C.M. Furse is a co-founder of Livewire Innovation, Inc. which is commercializing SSTDR technology, and therefore has a financial conflict of interest with this company.

References

[1] A. Eskandari, J. Milimonfared, M. Aghaei, A. K. Vidal de Oliveira, and R. Rüther, "Line-to-line faults detection for photovoltaic arrays based on I-V curve using pattern recognition," in *Proc. of the IEEE Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 0503–0507.

[2] D. S. Pillai and N. Rajasekar, "An MPPT-based sensorless line–line and line–ground fault detection technique for PV systems," *IEEE Transactions on Power Electronics*, vol. 34, no. 9, pp. 8646–8659, 2019.

[3] J. Flicker and J. Johnson, "Photovoltaic ground fault detection recommendations for array safety and operation," *Solar Energy*, vol. 140, pp. 34 – 50, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X16304819

[4] Q. Xiong, S. Ji, X. Liu, X. Feng, F. Zhang, L. Zhu, A. L. Gattozzi, and R. E. Hebner, "Detecting and localizing series arc fault in photovoltaic systems based on time and frequency characteristics of capacitor current," *Solar Energy*, vol. 170, pp. 788 – 799, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X18305450

[5] S. R. Madeti and S. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," *Solar Energy*, vol. 158, pp. 161 – 185, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X17307508

[6] M. U. Saleh, C. Deline, E. Benoit, S. Kingston, A. S. Edun, N. K. T. Jayakumar, J. B. Harley, C. Furse, and M. Scarpulla, "An overview of spread spectrum time domain reflectometry responses to photovoltaic faults," *IEEE Journal of Photovoltaics*, vol. 10, no. 3, pp. 844–851, 2020.

[7] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A comparative evaluation of advanced fault detection approaches for PV systems," *IEEE Journal of Photovoltaics*, vol. 9, no. 2, pp. 513–527, 2019.

[8] K. Blunt and R. Gold, "PG&E knew for years its lines could spark wildfires, and didn't fix them," July 2019. [Online]. Available: https://www.wsj.com/articles/pg-e-knew-for-years-its-lines-could-spark-wildfires-and-didnt-fix-them-11562768885

[9] C. M. Furse, M. Kafal, R. Razzaghi, and Y. Shin, "Fault diagnosis for electrical systems and power networks: A review," *IEEE Sensors Journal*, pp. 1–1, 2020.

[10] S. Kingston, N. K. T. Jayakumar, M. U. Saleh, E. Benoit, A. S. Edun, R. Sun, C. Furse, M. Scarpulla, and J. B. Harley, "Measurement of capacitance using spread spectrum time domain reflectometry (SSTDR) and dictionary matching," *IEEE Sensors Journal*, pp. 1–1, 2020.

[11] A. S. Edun, N. K. T. Jayakumar, S. Kingston, C. Furse, M. Scarpulla, and J. B. Harley, "Spread spectrum time domain reflectometry with lumped elements on asymmetric transmission lines," *IEEE Sensors Journal*, pp. 1–1, 2020.

[12] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "PV ground-fault detection using spread spectrum time domain reflectometry (SSTDR)," in *Proc. of the IEEE Energy Conversion Congress and Exposition*, 2013, pp. 1015–102.

[13] M. K. Alam, F. H. Khan, J. Johnson, and J. Flicker, "PV arc-fault detection using spread spectrum time domain reflectometry (SSTDR)," in *Proc. of the IEEE Energy Conversion Congress and Exposition (ECCE)*, 2014, pp. 3294–3300.

[14] C. Furse, Y. C. Chung, C. Lo, and P. Pendayala, "A critical comparison of reflectometry methods for location of wiring faults," *Smart Structures and Systems*, vol. 2, no. 1, pp. 25–46, 01 2006.

[15] C. Furse, P. Smith, C. Lo, Y. C. Chung, P. Pendayala, and K. Nagoti, "Spread spectrum sensors for critical fault location on live wire networks," *Structural Control and Health Monitoring*, vol. 12, no. 3-4, pp. 257–267, 2005.

[16] M. U. Saleh, C. Deline, S. Kingston, N. K. T. Jayakumar, E. Benoit, J. B. Harley, C. Furse, and M. Scarpulla, "Detection and localization of disconnections in PV strings using spread-spectrum time-domain reflectometry," *IEEE Journal of Photovoltaics*, vol. 10, no. 1, pp. 236–242, 2020.

[17] M. Kafal, F. Mustapha, W. Ben Hassen, and J. Benoit, "A non destructive reflectometry based method for the location and characterization of incipient faults in complex unknown wire networks," in *Proc. of the IEEE AUTOTESTCON*, 2018, pp. 1–8.

[18] Y. Lyu, A. Fairbrother, M. Gong, J. H. Kim, X. Gu, M. Kempe, S. Julien, K.-T. Wan, S. Napoli, A. Hauser, G. O'Brien, Y. Wang, R. French, L. Bruckman, L. Ji, and K. Boyce, "Impact of environmental variables on the degradation of photovoltaic components and perspectives for the reliability assessment methodology," *Solar Energy*, vol. 199, pp. 425 – 436, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X20301237

[19] M. Mussard and M. Amara, "Performance of solar photovoltaic modules under arid climatic conditions: A review," *Solar Energy*, vol. 174, pp. 409 – 421, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X18308417

[20] A. Bala Subramaniyan, R. Pan, J. Kuitche, and G. TamizhMani, "Quantification of environmental effects on PV module degradation: A physics-based data-driven modeling method," *IEEE Journal of Photovoltaics*, vol. 8, no. 5, pp. 1289–1296, 2018.

[21] Chaozhu Zhang and Liang Zhao, "Image de-noising based on learned dictionary," in *Proc. of the International Conference on Multimedia Technology*, July 2011, pp. 3101–3104.

[22] C. Shen and M. Zhang, "A image denoising algorithm based on sparse dictionary," in *Proc. of the IEEE International Conference on Electronics*

Information and Emergency Communication (ICEIEC), 2017, pp. 124–127.

[23] K. Su, H. Fu, B. Du, H. Cheng, H. Wang, and D. Zhang, "Image denoising based on learning over-complete dictionary," in *Proc. of the International Conference on Fuzzy Systems and Knowledge Discovery*, 2012, pp. 395–398.

[24] R. Boscolo, J. Liao, and V. Roychowdhury, "An information theoretic exploratory method for learning patterns of conditional gene coexpression from microarray data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 1, pp. 15–24, Jan 2008.

[25] Kuang-Chih Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, May 2005.

[26] A. M. Karimi, J. S. Fada, N. A. Parrilla, B. G. Pierce, M. Koyutürk, R. H. French, and J. L. Braid, "Generalized and mechanistic PV module performance prediction from computer vision and machine learning on electroluminescence images," *IEEE Journal of Photovoltaics*, vol. 10, no. 3, pp. 878–887, 2020.

[27] A. M. Karimi, J. S. Fada, M. A. Hossain, S. Yang, T. J. Peshek, J. L. Braid, and R. H. French, "Automated pipeline for photovoltaic module electroluminescence image processing and degradation feature classification," *IEEE Journal of Photovoltaics*, vol. 9, no. 5, pp. 1324–1335, 2019.

[28] Z. Huang, R. Wang, S. Shan, and X. Chen, "Projection metric learning on grassmann manifold with application to video based face recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 140–149.

[29] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.

[30] O. Bryt and M. Elad, "Improving the K-SVD facial image compression using a linear deblocking method," in *Proc. of the IEEE Convention of Electrical and Electronics Engineers in Israel*, Dec 2008, pp. 533–537.

[31] P. Shamsi, M. Marsousi, H. Xie, W. Fries, and C. Shaffer, "Dictionary learning for short-term prediction of solar PV production," in *Proc. of the IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5.

[32] S. G. Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[33] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1.

[34] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, "Learning unions of orthonormal bases with thresholded singular value decomposition," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, March 2005, pp. v/293–v/296 Vol. 5.

[35] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, Feb 2003.

[36] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, no. 2, pp. 337–365, Feb 2000.

[37] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[38] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, vol. 5, March 1999, pp. 2443–2446 vol.5.

[39] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2691–2698.

[40] P. Smith, C. Furse, and J. Gunther, "Analysis of spread spectrum time domain reflectometry for wire fault location," *IEEE Sensors Journal*, vol. 5, no. 6, pp. 1469–1478, 2005.

[41] C. Furse, N. K. T. Jayakumar, E. Benoit, M. U. Saleh, J. LaCombe, M. Scarpulla, J. Harley, S. Kingston, B. Waddoups, and C. Deline, "Spread spectrum time domain reflectometry for complex impedances: Application to PV arrays," in *Proc. of the IEEE AUTOTESTCON*, Sep. 2018, pp. 1–4.

[42] L. I. Inc., "Live cable fault detection by livewire innovation," http://www.livewireinnovation.com.

[43] M. U. Saleh, J. B. Harley, N. K. Tumkur Jayakumar, S. Kingston, E. Benoit, M. A. Scarpulla, and C. Furse, "Reflectometry on asymmetric transmission line systems," *Progress In Electromagnetics Research M*, vol. 89, pp. 121–130, 2020.