

# UC Riverside

## BCOE Research

### Title

Sparse Tracking State Estimation for Low-Observable Power Distribution Systems Using D-PMUs

### Permalink

<https://escholarship.org/uc/item/9rb1w00d>

### Authors

Akrami, Alireza  
Asif, Salman  
Mohsenian-Rad, Hamed

### Publication Date

2022

Peer reviewed

# Sparse Tracking State Estimation for Low-Observable Power Distribution Systems Using D-PMUs

Alireza Akrami, *Student Member, IEEE*, M. Salman Asif, and Hamed Mohsenian-Rad, *Fellow, IEEE*

**Abstract**—A new state estimation method is proposed for power distribution networks that suffer from low-observability. The proposed distribution system state estimation (DSSE) method leverages the high reporting rate of only a small number of distribution-level phasor measurement units (D-PMUs), a.k.a., micro-PMUs, to unmask and characterize sparsity among the state variables. The DSSE problem is formulated over differential synchrophasors as an adaptive group sparse recovery problem to track the changes that are made in the states of the system due to the events that are captured in D-PMU measurements. The formulated DSSE is further augmented to use adequate side information on the support of the vector of unknowns that is obtained from the outcome of an event-zone identification analysis prior to solving the DSSE problem. The sufficient conditions for the uniqueness of the obtained sparse recovery solution are derived with respect to the available side information. Moreover, a calibration mechanism is developed to address drifting in the tracking state estimation to enhance robustness.

**Index Terms**—Distribution system state estimation, low-observability, sparsity, adaptive group sparse signal recovery, differential synchrophasors, distribution synchrophasors, D-PMU.

## I. INTRODUCTION

### A. Background and Motivation

Distribution system state estimation (DSSE) is an important monitoring tool in power distribution system operation [1]. DSSE methods may use different types of available measurements in order to recover the state variables of the system, i.e., the nodal voltage phasors and branch current phasors [2].

The performance of DSSE is directly affected by the extent of observability in the power distribution system, which depends on the type, number, and location of sensors [3]. In practice, power distribution feeders often suffer from *low-observability*; because the number of sensors is much less than the number of state variables. The installation of smart meters and advanced metering infrastructure (AMI) have improved observability in power distribution networks. However, smart meters report measurements only once every 15 minutes to 1 hour [4]. These low reporting rates are *not* sufficient to capture the high dynamic of power distribution systems, which are caused by the growing penetration of distributed energy resources (DERs), the emergence of new types of loads, and the development of demand response programs [5]–[7].

The above issue can be resolved with the advent of the state-of-the-art distribution-level phasor measurement units (D-PMUs), a.k.a., micro-PMUs<sup>1</sup>, which report phasor measurements once every 8 to 100 milliseconds [9], [10]. D-PMUs

have been used in recent years at power distribution feeders for various tasks, such as to achieve situational awareness [9] and to support optimal grid reconfiguration [11].

D-PMUs can also be used in DSSE algorithms [10]. However, to resolve the low-observability issue to run a conventional DSSE, we need to install hundreds of D-PMUs on each feeder. This is cost prohibitive due to, not only the high cost of sensors, but also the cost of the communications infrastructure.

Therefore, finding a suitable solution to mitigate the low-observability in DSSE is an important practical challenge.

### B. Related Works

The common approach to compensate for low-observability in DSSE is to use pseudo-measurements. Pseudo-measurements are often constructed by using historical load data or real-time AMI data [12]. However, due to the uncertainty and the variability in distribution systems, pseudo-measurements are typically not accurate [13]. Lack of time synchronization is another factor that can negatively impact the accuracy of pseudo-measurements [14]. Inaccurate pseudo-measurements can create ill-conditioned mathematical optimization in the DSSE problem formulation; which may prevent it from converging to a reliable solution [13].

There have been efforts to make pseudo-measurements more robust against system uncertainties, through either statistical approaches or machine learning methods. In [15], a Gaussian mixture model is used for the load probability density function for inclusion in a conventional weighted least square (WLS)-DSSE. In [16], a two-stage data clustering method is used to construct the pseudo-measurements. In [17], artificial neural networks are trained to generate pseudo-measurements from limited measurements. However, statistical methods still require reliable statistical models to be accurate. As for the machine learning methods, they require large sets of reliable training data to generate accurate pseudo-measurements. The authors in [18] proposed a new data-driven DSSE method based on training a deep neural network model to solve the DSSE problem without adding pseudo-measurements; instead they added physical information of the underlying power distribution feeder, such as the parameters of the distribution lines to further increase the accuracy. While this method is very promising, as a data-driven method it naturally requires access a considerably large data set in order to train the machine learning model. Furthermore, as the size of network grows, it may become difficult to cover all the possible scenarios in the underlying physical system in the training process.

A different approach that has emerged recently is to address low-observability through *compressed sensing*. In this approach, instead of trying to make the low-observable DSSE

<sup>1</sup>The authors are with the Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA. This work is supported in part by UCOP grant LFR-18-548175. The corresponding author is H. Mohsenian-Rad, e-mail: hamed@ece.ucr.edu.

<sup>1</sup>Micro-PMU is a trademark of PSL [8]. Therefore, we use the term D-PMUs; since our study is *not* specific to any particular commercial technology.

TABLE I  
COMPARISON AMONG DSSE METHODS WHICH USE COMPRESSED SENSING

Reference	Grounds for Sparsity	State Variables	Solution Approach	Tracking	Analysis on Optimality
[19]	Correlated structure of nodal voltages	Voltage	$\ell_1$ norm minimization solved by Newton-Raphson	No	Yes
[20]	Neglecting load currents vs. the substation current	Current	Lasso minimization	No	No
[21]	Missing measurements	Voltage	Rank minimization	No	No
[22]	Temporal correlation of state variables	Voltage	Trace norm minimization	No	No
[23]	Correlated structure of nodal voltage	Voltages	Survey which includes several methods	No	No
[25]	Differential synchrophasors	Voltage + Current	Lasso minimization solved by ADMM	No	No
This paper	Grouped differential synchrophasors	Voltage + Current	Group Lasso minimization augmented by partial side information solved by ADMM	Yes	Yes

problem fully-observable through the above aforementioned methods, the goal is to rather extract and leverage the potential sparsity features in the DSSE problem so as to solve the original DSSE problem while it remains low-observable. In this regard, the authors in [19] used the correlation among nodal voltages to compress the measurements; and then they used techniques from compressed sensing to solve the DSSE problem. In [20], the authors developed a current-based sparse DSSE method that is built on the assumption that the load currents are negligible in comparison with the injected current from the substation. This consideration results in making the problem sparse. In [21], a method based on matrix completion is developed to estimate the missing values in the DSSE problem, i.e., to estimate those state variables that do not have direct measurements. In [22], a block tensor completion is proposed to estimate the nodal voltages. In this method, the correlations among the state variables are used in a tensor norm minimization problem; which is a generalization of the matrix completion method. Finally, in [23], the authors used the correlation among the nodal voltages, and they accordingly presented a comparison among the performance of four different compressed sensing techniques, where the robustness of sparse recovery against bad data is also discussed.

### C. Summary of Technical Contributions

In this paper, we propose a novel sparse tracking linear state estimation method for three-phase power distribution feeders with low-observability, where we take advantage of the high reporting rates of a small number of D-PMUs to turn the DSSE problem into a *group sparse signal recovery* problem. Our approach is fundamentally different from the previous studies which used compressed sensing; because we address low-observability in the context of *differential synchrophasors*, c.f. [24]. Unlike the studies that use the correlation among nodal voltages to compress the measurements and extract the sparsity features for a single time slot, in this paper, we propose a novel *tracking* DSSE method over sequential time slots to capture and track the changes that are made in the system due to various events. To the best of our knowledge, this paper is the first study to use differential phasors in a DSSE problem. Moreover, the way that we define sparsity in our model and the way that we formulate the problem are unique and novel.

Compared to the preliminary conference version of this work in [25], the current journal version has several new and important contributions; including new ideas and new results. First, the method in this paper is significantly more rigorous to address the grounds for sparsity. In particular, it systematically defines four distinct groups of state variables to characterize sparsity under different operating conditions. On the contrary, the model in [25] was a preliminary model; and it was meant to only introduce the basic idea in sparsity. Second, in this paper, we extract and utilize *side information* prior to solving the DSSE problem in order to augment the sparse signal recovery method. This augmentation is a key reason for the significant improvement in the performance in this paper. Third, the preliminary study in [25] does not address the conditions for optimality and uniqueness, drift identification, and calibration. As we will see in our case studies, the proposed method in this paper significantly outperforms the preliminary method in [25].

A brief comparison between the new method that is proposed in this paper versus the comparable methods that we discussed above is provided in Table I.

The main contributions of this paper are as follows:

- 1) Developing a novel sparse tracking DSSE method to leverage the measurements from only a small number of D-PMUs in a low-observable three-phase distribution feeder,
- 2) Establishing the grounds for sparsity in the DSSE problem in *differential mode* based on the engineering characteristics of power distribution systems as well as analyzing the events that occur on the distribution feeders and are captured by D-PMUs. We introduce four types of state variables in the context of *group sparsity*, thus solving the DSSE problem as an *adaptive group sparse signal recovery* problem,
- 3) Developing an event zone identification method that is used before running the sparse tracking DSSE. It provides the DSSE problem formulation with *side information* on the support of the sparse vector of the four types of state variables,
- 4) Deriving sufficient conditions for the *uniqueness* of the DSSE solution that is obtained from the sparse recovery problem with respect to the constructed side information,
- 5) Developing a mechanism to identify potential drifting, caused by accumulative estimation error of differential syn-

chrophasors; thus to calibrate the estimated voltage phasors.

## II. SPARSE LINEAR DSSE PROBLEM FORMULATION IN DIFFERENTIAL MODE

Consider a multi-phase power distribution network that is represented by a graph  $\mathcal{G} := (\mathcal{N}, \mathcal{L})$ , where  $\mathcal{N}$  denotes the set of nodes and  $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$  denotes the set of distribution lines. Let  $\varphi = \{A, B, C\}$  denote the set of phases. Suppose  $\mathbf{V}^t$  denotes the vector of all voltage phasors at time slot  $t$ , assuming that they are represented in the *rectangular* form:

$$\mathbf{V}^t := [ \Re(v_{n,\phi}^t) \Im(v_{n,\phi}^t) ]^\top, \quad \forall n \in \mathcal{N}, \quad \forall \phi \in \varphi, \quad (1)$$

where  $\Re(\cdot)$  denotes the real part of the phasor,  $\Im(\cdot)$  denotes the imaginary part of the phasor, and  $v_{n,\phi}^t$  denotes the voltage phasor at bus  $n$  on phase  $\phi$  at time slot  $t$ . The length of  $\mathbf{V}^t$  is  $|\mathcal{N}|$ . In a similar setting,  $\mathbf{I}^t$  denotes the vector of all the current phasors at time slot  $t$ :

$$\mathbf{I}^t := [ \Re(i_{kl,\phi}^t) \Im(i_{kl,\phi}^t) ]^\top, \quad \forall kl \in \mathcal{L}, \quad \forall \phi \in \varphi, \quad (2)$$

where  $i_{kl,\phi}^t$  denotes the current phasor at line  $kl$  on phase  $\phi$  at time slot  $t$ . The length of  $\mathbf{I}^t$  is  $|\mathcal{L}|$ . Since we study radial distribution feeders, we have  $|\mathcal{L}| = |\mathcal{N}| - 1$ .

### A. DSSE Problem Under Low-Observability Conditions

Let  $\mathbf{x}^t$  be a  $N \times 1$  vector which contains all state variables of the power distribution system at time interval  $t$ . Throughout this paper, we assume that the state variables include the vector of voltage phasors for *all* nodes, i.e.,  $\mathbf{V}^t$ , and the vector of current phasors for *all* lines, i.e.,  $\mathbf{I}^t$ . That is,

$$\mathbf{x}^t := [ (\mathbf{V}^t)^\top (\mathbf{I}^t)^\top ]^\top, \quad (3)$$

where we have  $N = |\mathcal{N}| + |\mathcal{L}|$ .

Also, let  $\mathbf{y}^t$  denote the vector of all available measurements. At each time slot  $t$ , we assume that the measurements come from only D-PMUs. Since our focus is on a challenging scenario where the distribution network is *not* fully-observable, we assume that D-PMUs are installed at only a few nodes. For example, one D-PMU can be installed at the substation; and one or at most two D-PMUs can be installed on each lateral, as we will see in the case studies in Section IV. D-PMUs measure voltage phasors on three-phases at nodes  $\mathcal{N}_m \subseteq \mathcal{N}$  and current phasors on three-phases at lines  $\mathcal{L}_m \subseteq \mathcal{L}$ .

D-PMUs provide two types of measurements, voltage phasors and current phasors. Accordingly, there are two types of equations that we need to include in the state estimation problem. First, there are equations that map the state variables to the voltage phasor measurements through identity mapping for each node  $n \in \mathcal{N}_m$ :

$$\Re(v_{n,\phi}^t) = \frac{1}{2} e_{n,\phi}^\top (v_n^t + \overline{v_n^t}), \quad (4)$$

$$\Im(v_{n,\phi}^t) = \frac{1}{2j} e_{n,\phi}^\top (v_n^t - \overline{v_n^t}), \quad (5)$$

where  $e_{n,\phi}$  is the  $\phi$ -th canonical basis vector and  $v_n^t$  is the vector of voltage phasors at all phases at bus  $n$  at time slot  $t$  in complex form. Second, there are equations which map

the state variables to the current measurements for each line segment  $kl \in \mathcal{L}_m$  through the KCL equations:

$$\Re(i_{kl,\phi}^t) = \frac{1}{2} e_{kl,\phi}^\top (Y_{kl}(v_k^t - v_l^t) + \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})), \quad (6)$$

$$\Im(i_{kl,\phi}^t) = \frac{1}{2j} e_{kl,\phi}^\top (Y_{kl}(v_k^t - v_l^t) - \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})), \quad (7)$$

where  $Y_{kl}$  is the admittance of line segment  $kl$ .

For the rest of this paper, we represent the equations in (4)-(7) in a compact form through measurement matrix  $\Psi$ . Thus, we have:

$$\mathbf{y}^t = \Psi \mathbf{x}^t. \quad (8)$$

Matrix  $\Psi$  represents  $|\mathcal{N}_m| + |\mathcal{L}_m|$  equations. Therefore, its size is  $(|\mathcal{N}_m| + |\mathcal{L}_m|) \times N$ .

Under the low-observability circumstances, we do *not* have enough measurements to solve the system of linear equations in (8). Thus, one can use the circuit laws to introduce additional equations, based on the so-called *virtual measurements*, to capture the relationships among the state variables at the line segments that are not equipped with sensors, i.e.,  $\forall kl \notin \mathcal{L}_m$ :

$$0 = \frac{1}{2} e_{kl,\phi}^\top (Y_{kl}(v_k^t - v_l^t) + \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})) - \Re(i_{kl,\phi}^t), \quad (9)$$

$$0 = \frac{1}{2j} e_{kl,\phi}^\top (Y_{kl}(v_k^t - v_l^t) - \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})) - \Im(i_{kl,\phi}^t). \quad (10)$$

Again, let us represent the equations in (9) and (10) in a compact form as:

$$\mathbf{0} = \Phi \mathbf{x}^t. \quad (11)$$

Matrix  $\Phi$  represents  $|\mathcal{L}| - |\mathcal{L}_m|$  equations. Therefore, its size is  $(|\mathcal{L}| - |\mathcal{L}_m|) \times N$ .

While the system of equations in (8) maps the *unknown phasors* to the *known phasors*, the system of equations in (11) relates the *unknown phasors* to other *unknown phasors*.

**Remark 1:** In a standard DSSE problem, if a bus or a line segment is already equipped with a sensor, then we do *not* need to consider the corresponding virtual measurement anymore. Because it does not provide any new independent equation. However, in a low-observable network, virtual measurements *can* sometimes provide new independent equations.

Now, let us concatenate the system of equations in (8) and (11) to obtain:

$$\mathbf{z}^t = \mathbf{H} \mathbf{x}^t \quad (12)$$

where

$$\mathbf{H} := [ \Psi \Phi ]^\top, \quad \mathbf{z}^t := [ \mathbf{y}^t \mathbf{0} ]^\top. \quad (13)$$

Matrix  $\mathbf{H}$  is  $M \times N$ , where  $M$  denote the rank of matrix  $\mathbf{H}$ , i.e., the number of independent rows in stacking of matrices  $\Psi$  and  $\Phi$ . For the rest of this paper, we assume that matrix  $\mathbf{H}$  is full-row-rank, i.e., it includes only the independent rows in  $\Psi$  and  $\Phi$ . The DSSE problem under the *low-observability* condition is the problem of solving (12) when

$$M < N. \quad (14)$$

In that case, the system of equations in (12) is *undetermined* and its solution is *unspecified*. The set of solutions are

$$\mathbf{H}^\top (\mathbf{H} \mathbf{H}^\top)^{-1} \mathbf{z}^t + \text{null}(\mathbf{H}), \quad (15)$$

where  $\text{null}(\cdot)$  denotes the null space of matrix  $\mathbf{H}$ . Therefore, the DSSE problem *cannot* be solved in its standard form.

## B. DSSE Problem in Differential Mode

Let us consider the same overall setting as in Section II-A, but let us represent all the D-PMU measurements and all the state variables in a *differential mode* as

$$\Delta \mathbf{z}^t := \mathbf{z}^t - \mathbf{z}^{t-1}, \quad (16)$$

$$\Delta \mathbf{x}^t := \mathbf{x}^t - \mathbf{x}^{t-1}. \quad (17)$$

We can write the relationship in (12) in differential mode as

$$\Delta \mathbf{z}^t = \mathbf{H} \Delta \mathbf{x}^t. \quad (18)$$

If (14) holds, then the system is still *not* fully-observable even in differential mode. Thus, in principle, we still cannot find a unique solution for the equations in (18). However, unlike in (12), we *can* solve the differential form of the DSSE problem in (18) by posing it as a *sparse signal recovery* problem [26]. The key is to show that the vector of state variables in the DSSE problem in differential mode, i.e.,  $\Delta \mathbf{x}^t$ , is a *sparse* vector. Next, we explain the grounds for sparsity.

## C. Grounds for Sparsity

In this section, we examine the engineering characteristics in power distribution systems and make the case that the DSSE problem in *differential mode* can be formulated as a sparse signal recovery problem. The fundamental observation throughout this section is as follows; given the *short time interval* in D-PMU measurements, the changes in the states of the system from time  $t - 1$  to time  $t$  are caused by only one major change in the components of the power distribution system. We refer to these major changes as *events*, such as load switching, a sudden change in a DER output, etc. [24].

When an event takes place on a distribution feeder, it may impact a *subset* (or all) of nodal voltages and branch currents. The impact of an event on the network can be studied by considering *compensation theorem* from circuit theory [27]; based on which we can replace the circuit element that is the source cause of an event with a *current source* that injects the same level of current at the location of the event. The advantage of using compensation theorem is that it provides us with a direct relationship and also an equivalent circuit model to study how an event may change the state variables.

1) *Partitioned Network Representation at an Event*: Consider the power distribution network in Fig. 1(a). Suppose a major load switching event occurs at bus 29 during the short time-period between time  $t - 1$  and time  $t$ . Consider the *path* between the substation and the event bus; as marked in red. Let us represent this path with  $\mathcal{T} := (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the set of buses and  $\mathcal{E}$  is the set of line segments on this path.

Next, let us *partition* the *rest* of the distribution system into four subgraphs, denoted by  $(\mathcal{V}_i, \mathcal{E}_i)$ ,  $\forall i \in \{1, \dots, 4\}$ . These partitions are marked in Fig. 1(a). Note that, together with path  $(\mathcal{V}, \mathcal{E})$ , these partitions cover the entire distribution system:

$$\bigcup_{i=1}^4 \mathcal{V}_i = \mathcal{N} \setminus \mathcal{V}, \quad \bigcup_{i=1}^4 \mathcal{E}_i = \mathcal{L} \setminus \mathcal{E}. \quad (19)$$

In our analysis, it is critical to figure out *under what circumstances* the changes in voltage or current are non-zero or

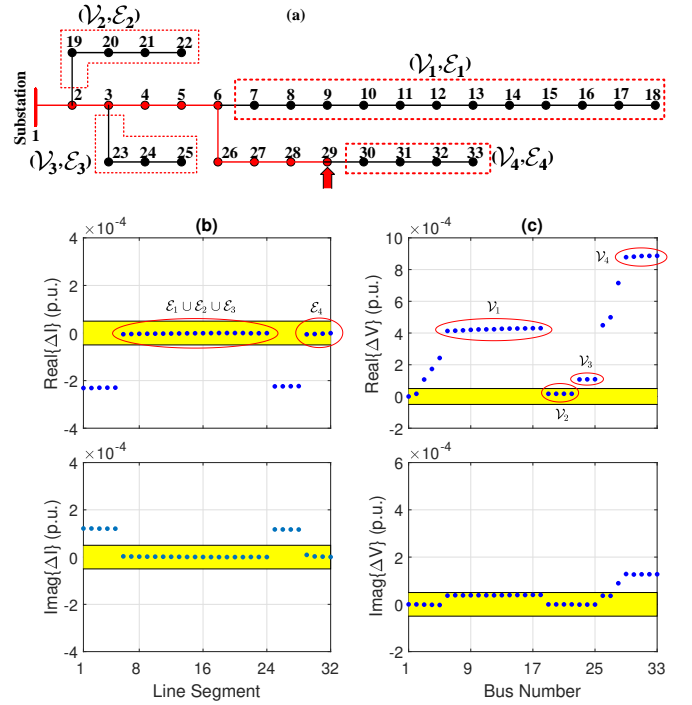


Fig. 1. An example to illustrate sparsity: (a) An instance of the IEEE 33-node test system right after a major load change at node 29; (b) The changes in line currents, i.e.,  $\Delta I$ ; (c) The changes in bus voltages, i.e.,  $\Delta V$ .

(approximately) zero, i.e., they lay outside or inside the *zero approximation region*; as marked with the yellow ribbon in Figs. 1(b) and (c). Note that, a key feature of sparse signal recovery is its capability to manage the width of the zero approximation region as a trade-off against estimation error.

2) *Impact of an Event on Line Current*: Following the compensation theorem [27], suppose we replace the event on node 29 in Fig. 1(a) by an equivalent current source. If we write the *Kirchhoff's current law* at all nodes in  $\mathcal{V}$ , we can show that the change in the current for all the line segments in set  $\mathcal{E}$  is equal to the amount of current that is injected by a current source, in the equivalent circuit of the power distribution system that is constructed by applying the compensation theorem; because the change in the injected current to all the nodes is zero; except for the node where the event occurs. Moreover, if we write the same system of equations for all the nodes in  $\mathcal{N} \setminus \mathcal{V}$ , we can show that the amount of change in the current for all the line segments in set  $\mathcal{L} \setminus \mathcal{E}$  is zero; otherwise it means that there is a non-zero change in the injected current to one of the nodes outside the red path, i.e., there is another event somewhere else on the distribution feeder which is in contradiction to our initial assumption that there is only one event at the current time.

In summary, once an event occurs somewhere on a power distribution feeder, the line segments on the red path may experience non-zero changes; while the line segments outside the red path do not experience any change in their current.

3) *Impact of an Event on Nodal Voltage*: When an event takes place, the nodal voltages on the nodes along the red path are *directly* affected. This is because of the changes in the current of the line segments that are along the red path. However, the voltages on the nodes outside the red path may or

may not change drastically depending on several factors, such as line parameters, network topology, and most importantly, how the event changes voltage across the nodes that are located on the boundaries of the red path and other partitions. To elaborate, suppose  $s$  is a node on the red path that has a neighboring node  $r$  outside the red path. For example  $s$  could be bus 2, 3, 6, or 29 in Fig. 1 (a). By writing down the Eq. (6) for line segment  $sr$  that lays outside the red path, one can show that the following equality holds:

$$\frac{R_{sr}}{X_{sr}} = \frac{\Re(\Delta v_s^t) - \Re(\Delta v_r^t)}{\Im(\Delta v_s^t) - \Im(\Delta v_r^t)}, \quad (20)$$

where  $R_{sr}$  and  $X_{sr}$  are the resistance and reactance of line  $sr$ , respectively. From (20), whether or not the differential voltage phasors for the nodes outside the red path, i.e.,  $\Delta v_r^t$ , is small enough to be replaced by zero; depends on the value of other parameters in (20). However, in either case, the same zero/non-zero condition would hold also for any other node in the *same* partition. As a result, there exists a *group sparsity* among the nodal voltages in differential mode for the buses that are in the *same* partition  $\mathcal{V}_i$ ,  $\forall i \in \{1, \dots, 4\}$  outside the red path. For example, in Fig. 1(c), the parameters in (20) are such that the differential voltage phasors for the nodes in partition  $\mathcal{V}_1$  are all non-zero and the differential voltage phasors for the nodes in partition  $\mathcal{V}_2$  are all approximately zero.

4) *Fundamental Conclusions:* We can combine the analysis in Sections II.C.1, II.C.2, and II.C.3 to conclude the following corollary; which builds a foundation for our analysis:

**Corollary 1:** At each time slot  $t$ , we can divide the state variables in differential mode into *four* types:

- 1) **Differential current phasors known to be zero:** For all the line segments in set  $\mathcal{L} \setminus \mathcal{E}$ ,  $\Delta I$  would be zero.
- 2) **Differential current phasors known to be non-zero:** For all the line segments in set  $\mathcal{E}$ ,  $\Delta I$  would be non-zero.
- 3) **Grouped differential voltage phasors:** For all buses in each of the sets  $\mathcal{V}_i$ ,  $\forall i \in \{1, \dots, 4\}$ , either  $\Delta V$  would be (approximately) zero or non-zero for all such buses.
- 4) **Differential voltage phasors known to be non-zero:** For all buses in set  $\mathcal{V}$ ,  $\Delta V$  would be non-zero. ■

The proof of Corollary 1 is provided in Appendix A.

**Remark 2:** The discussion on the grounds for sparsity in this paper are made based on the assumption that only one major event happens across the distribution feeder during each time slot of the proposed DSSE process. This is a reasonable assumption in practice; as it is observed in the previous studies that have looked at real-world D-PMU data, such as in [?], [24], [28]–[30]. However, for the occasional scenarios, where two or more events occur simultaneously during the same 100 msec time slot, one may need to modify the grounds for sparsity that we discussed in Section II-C. Addressing such cases could be the subject of an extension of this work in the future.

**Remark 3:** The grounds for sparsity that we discussed above are developed based on the assumption that the distribution feeder has a radial topology; which is very common in practice. The above discussions are not applicable to transmission systems, where the network has a meshed topology. Therefore,

the analysis in this paper is specific to the DSSE; because it takes advantage of the features of power distribution systems.

Note that, in principle, we do *not* know the actual type for any of the above state variables in advance; unless the location (i.e., the bus number) of the event is known to us.

#### D. Event Zone Identification

There are recent methods that use D-PMU measurements to identify the location of events on power distribution systems, e.g., see [24]. If such methods are *truly accurate*, then we can use them to identify exactly which state variables belong to each of the four groups that we introduced in Section II-C.

However, in practice, event location identification is *not* exact. This is particularly the case in *low-observable* networks, where the number of D-PMUs may not be enough to identify the exact location of all events. Interestingly, our approach here does *not* require knowing the exact location of the event.

Suppose we do *not* know the exact location of the event at bus 29 in Fig. 1(a), but we do know that the event occurs *somewhere* on the lateral that includes bus 26 to bus 33. In other words, suppose we only know the *zone* of the event. In that case, all the discussions in Section II-C on the features of voltage and current differential phasors for all nodes and line segments in  $(\mathcal{V}_1, \mathcal{E}_1)$  and  $(\mathcal{V}_2, \mathcal{E}_2)$  and  $(\mathcal{V}_3, \mathcal{E}_3)$  would still be valid. Moreover, all the previous analysis for the nodes and line segments on the path between the substation and the event zone up to node 29 would also still be valid.

The difference between knowing the exact event bus versus only knowing the zone of the event would affect our analysis only with respect to the differential voltage and current phasors for the nodes and line segments in  $(\mathcal{V}_4, \mathcal{E}_4)$ . Given that we do not know the exact location of the event at bus 29, we can simply merge set  $(\mathcal{V}_4, \mathcal{E}_4)$  to set  $(\mathcal{V}, \mathcal{E})$ . That is, we simply do not include set  $(\mathcal{V}_4, \mathcal{E}_4)$  in our sparsity assumptions; everything else will remain the same as in Section II-C.

As for how to obtain the event *zone*, we propose an extension of the event location identification method in [24]. First, we split the distribution feeder into several mutually exclusive zones and represent each zone by an index  $c$ . One simple way to form the zones is to consider all the nodes on the same lateral as one zone, provided that we have installed at least one D-PMU on each lateral. Then, we can simply replace Eq. (12) in [24] with the following:

$$c^* = \arg \min_c \frac{1}{n_c} \|\Delta \mathbf{V}^{c,f} - \Delta \mathbf{V}^{c,b}\|_2, \quad (21)$$

where  $n_c$  denotes the number of buses in zone  $c$ ;  $\Delta \mathbf{V}^{c,f}$  is the vector of estimated differential voltage phasors from the *forward sweep* in zone  $c$ ; and  $\Delta \mathbf{V}^{c,b}$  is the same vector from the *backward sweep*. For more details please refer to [24].

#### E. DSSE as a Sparse Signal Recovery Problem

We now go back to the DSSE problem in differential mode based on the system of linear equations in (18). Based on grounds for sparsity that we established in Section II-C, we can now formulate this problem in a form that can be solved using techniques in *sparse signal recovery*.

To recover a sparse solution of the system of equations in (18), we can solve the following *basis pursuit* problem which is a well-known convex relaxation of the original intractable  $\ell_0$ -norm minimization problem in sparse recovery [26]:

$$\min_{\Delta \mathbf{x}^t} \|\Delta \mathbf{x}^t\|_1 \quad \text{s.t.} \quad \Delta \mathbf{z}^t = \mathbf{H} \Delta \mathbf{x}^t, \quad (22)$$

where  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm. Next, we will modify (22) to incorporate the following ‘‘side information’’ so that we can solve the above optimization problem: 1) the four groups of state variables that we introduced in Section II-C; and 2) the zone of the event that we identified in Section II-D.

### III. SOLUTION METHOD, UNIQUENESS, AND CALIBRATION

#### A. Incorporating Side Information

To incorporate the side information that we listed in Section II-D, we modify (22) into an *adaptive group-sparse recovery* problem [31]. Recall from Section II-C that we divided all the state variables in differential mode into four types. Let us refer to them as Type 1, Type 2, Type 3, and Type 4. Accordingly, suppose vector  $\Delta \mathbf{x}^t$  is partitioned into  $P$  groups. Three partitions are formed based on the state variables of Type 1, Type 2, and Type 4, respectively. For the state variables of Type 3, the number of partitions is equal to the number of sets  $\mathcal{V}_i$ , as we defined in Section II-C. As we discussed in Section II-C, the state variables in each partition have similar tendency towards being (approximately) zero or non-zero.

Let us represent the state variables in each partition  $p$  by  $\Delta \mathbf{x}_p^t$ , where  $p = 1, \dots, P$ . Let  $w_p$  denote the *adaptive weight* for partition  $p$ . Set  $w_p = 1$  for Type 1 state variables because they can be replaced by zero. Set  $w_p = 0$  for Type 2 and Type 4 state variables; since they are expected to be non-zero. Set  $w_p$  for Type 3 state variables to be a number between 0 and 1 based on their distribution of historical zero/non-zero values.

To incorporate the weights, we reformulate the basis pursuit problem (22) as the following weighted  $\ell_1$ -norm minimization:

$$\min_{\Delta \mathbf{x}^t} \sum_{p=1}^P w_p \|\Delta \mathbf{x}_p^t\|_1 \quad \text{s.t.} \quad \Delta \mathbf{z}^t = \sum_{p=1}^P \mathbf{H}_p \Delta \mathbf{x}_p^t, \quad (23)$$

where  $\mathbf{H}_p$  denotes a submatrix of  $\mathbf{H}$  with only the columns associated with the state variables in partition  $p$ .

It is more common to solve the related unconstrained relaxation of (23), a.k.a., *weighted Lasso* problem [32] as

$$\min_{\Delta \mathbf{x}^t} \frac{1}{2} \|\Delta \mathbf{z}^t - \sum_{p=1}^P \mathbf{H}_p \Delta \mathbf{x}_p^t\|_2^2 + \lambda \sum_{p=1}^P w_p \|\Delta \mathbf{x}_p^t\|_1, \quad (24)$$

where  $\lambda > 0$  is a regularization parameter which controls the trade-off between the estimation error and the sparsity level of the unknown vector  $\Delta \mathbf{x}^t$ . By changing  $\lambda$  we change the width of the yellow ribbon in Figs. 1(b) and (c). The first term in (24) is the squared error loss and the second term is the weighted  $\ell_1$ -norm penalty. If  $\lambda$  is such that  $\lambda w_p > \|\mathbf{H}_p^T \Delta \mathbf{z}^t\|_\infty$ , then we force the corresponding estimated  $\Delta \mathbf{x}_p^t$  to be zero.

#### B. ADMM Solution

We use the *alternating direction method of multipliers* (ADMM) [33] to solve problem (24). We obtain the *augmented Lagrangian* for (24), under constraints  $\Delta \mathbf{x}_p^t - \boldsymbol{\mu}_p = 0$ , as

$$\begin{aligned} \mathcal{L}_\rho(\Delta \mathbf{x}^t, \boldsymbol{\mu}, \mathbf{u}) = & \frac{1}{2} \|\Delta \mathbf{z}^t - \sum_{p=1}^P \mathbf{H}_p \Delta \mathbf{x}_p^t\|_2^2 + \lambda \sum_{p=1}^P w_p \|\boldsymbol{\mu}_p\|_1 \\ & + \rho \langle \mathbf{u}, \Delta \mathbf{x}^t - \boldsymbol{\mu} \rangle + \frac{\rho}{2} \left\| \sum_{p=1}^P (\Delta \mathbf{x}_p^t - \boldsymbol{\mu}_p) \right\|_2^2, \end{aligned} \quad (25)$$

where  $\Delta \mathbf{x}^t$  and  $\boldsymbol{\mu}$  are the primal variables,  $\mathbf{u}$  is the dual variable, and  $\rho > 0$  is a Lagrangian parameter.

The ADMM method alternately updates all the primal and dual variables at every iteration as follows

$$\Delta \mathbf{x}^t = \arg \min_{\Delta \mathbf{x}^t} \mathcal{L}_\rho(\Delta \mathbf{x}^t, \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{u}}) \quad (26)$$

$$\boldsymbol{\mu} = \arg \min_{\boldsymbol{\mu}} \mathcal{L}_\rho(\Delta \mathbf{x}^t, \boldsymbol{\mu}, \tilde{\mathbf{u}}) \quad (27)$$

$$\mathbf{u} = \tilde{\mathbf{u}} + (\Delta \mathbf{x}^t - \boldsymbol{\mu}), \quad (28)$$

where  $(\Delta \tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{u}})$  and  $(\Delta \mathbf{x}, \boldsymbol{\mu}, \mathbf{u})$  denote the estimates from the previous and the current iteration, respectively. The solutions to (26) and (27) can be obtained by setting the derivatives of (25) w.r.t.  $\Delta \mathbf{x}^t$  and  $\boldsymbol{\mu}^t$ , respectively. The resulting closed form solutions for the updates can be described as

$$\Delta \mathbf{x}^t = (\mathbf{H}^T \mathbf{H} + \rho \mathbf{I})^{-1} (\mathbf{H}^T \Delta \mathbf{z}^t + \rho (\tilde{\boldsymbol{\mu}} - \tilde{\mathbf{u}})) \quad (29)$$

$$\boldsymbol{\mu}_p = \mathcal{S}_{w_p \lambda / \rho}(\Delta \mathbf{x}_p^t + \tilde{\boldsymbol{\mu}}_p) \quad \forall p \in P \quad (30)$$

$$\mathbf{u} = \tilde{\mathbf{u}} + (\Delta \mathbf{x}^t - \boldsymbol{\mu}) \quad (31)$$

where  $\mathcal{S}_{w_p \lambda / \rho}(\cdot)$  is a *soft-thresholding* operator [34]; and  $\boldsymbol{\mu}_p$ ,  $\Delta \mathbf{x}_p^t$ , and  $\tilde{\boldsymbol{\mu}}_p$  forms the  $p$ -th partition of the respective vectors.

Importantly, the convergence of the ADMM is guaranteed for the above sparse DSSE problem; because the objective function in the proposed problem formulation is the sum of several convex functions with linear constraints, c.f [33].

#### C. Uniqueness of the Sparse DSSE Solution

The conditions for the uniqueness of the solution in a sparse signal recovery problem typically depend on the structure of the measurement matrix  $\mathbf{H}$ , the number of non-zero entries, and the support and the sign sequence of the non-zero entries. Thus, to check the uniqueness of the sparse DSSE solution, we extend the dual certificate-based method in [35], where we incorporate the side information by assigning different values to  $w_p$  for different partitions  $p = 1, \dots, P$ .

Suppose the true sparse signal  $\Delta \mathbf{x}$  in (18) has non-zero entries on an index set  $\Gamma$ . Let us denote  $\Delta \mathbf{x}_\Gamma$  as the non-zero entries,  $\mathbf{s}_\Gamma = \text{sign}(\Delta \mathbf{x}_\Gamma)$ , and  $\mathbf{H}_\Gamma$  as a matrix that consists of the columns in  $\mathbf{H}$  that correspond to the indices in  $\Gamma$ . We can show that  $\Delta \mathbf{x}$  is also the *unique* solution of the  $\ell_1$ -norm minimization problem in (24) using the following theorem.

**Theorem 1:** A vector  $\Delta \mathbf{x}$  that satisfies the equations in (18) with support  $\Gamma$  and sign sequence  $\mathbf{s}_\Gamma$  can be recovered from the unique optimal point of the optimization problem in (24) if the following conditions are satisfied:

1. Submatrix  $\mathbf{H}_\Gamma$  is full-rank.
2. For all  $\gamma \notin \Gamma$ , the following inequality holds:

$$|\mathbf{H}_\gamma^\top \mathbf{H}_\Gamma (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma| < w_\gamma. \quad (32)$$

3.  $\lambda$  and  $\mathbf{W}$  are selected so that

$$\text{sign}(\Delta \mathbf{x}_\Gamma - \lambda (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma) = \mathbf{s}_\Gamma. \quad (33)$$

$\mathbf{H}_\gamma$  denotes the column of  $\mathbf{H}$  with index  $\gamma$ ,  $\mathbf{W}_\Gamma$  is a diagonal matrix whose diagonal entries are  $w_p$  for the partitions associated with the indices in  $\Gamma$ , and  $w_\gamma$  is the corresponding weight of the partition that  $\gamma$  belongs to. ■

The proof of Theorem 1 is provided in Appendix B. The conditions in (32) relax the tightness of similar conditions discussed in [35] for standard Lasso problem in which all weights  $w_\gamma$  are equal. From (32), if a  $w_\gamma$  is very large, then the corresponding condition becomes irrelevant and does not affect the uniqueness of the solution. This condition supports our claims that incorporating the side information can help us recover the correct solution with provable guarantees.

#### D. Computational Complexity

The main computational complexity for solving the Lasso problem in (24) comes from the complexity of solving the least-squares optimization problem in (26). This optimization problem is solved once at every iteration. The solution for this problem is provided in (29). We can pre-calculate and store the inverse matrix  $(\mathbf{H}^\top \mathbf{H} + \rho \mathbf{I})^{-1}$  and use it in every iteration, for as long as the topological configuration of our network remains unchanged. The typical computational cost for solving such least-square optimization problem for a dense matrix  $\mathbf{H}$  with size  $M \times N$  is  $\mathcal{O}(MN + N^2)$ . However, since  $\mathbf{H}$  in our problem is sparse, the cost of computing is reduced to  $\mathcal{O}(kM + N^2)$ , where  $k$  is the number of nonzero entries in each row of matrix  $\mathbf{H}$ . For the base case scenario in our case studies where we have only six D-PMUs in the system, we have:  $k \leq 8$ . As for the number of iterations in our case studies, in the worst case scenario, we approximately use a maximum of 1000 iterations in the process of solving the Lasso problem.

As for the WLS-DSSE method with linear equations, the cost of computation per iteration is  $\mathcal{O}(MN^2 + N^3)$ . Therefore, our proposed method is computationally less expensive than the conventional WLS-DSSE method. We will further discuss the computation time in the case studies in Section IV-A.

#### E. Sparse Tracking Linear DSSE

The solution for  $\Delta \mathbf{x}^t$  tells us *whether* and *how* the states of the power distribution system *change* during each measurement interval. This by itself is an important result. However, to make the DSSE results more useful to the utility operator, we need to convert the results from the differential mode back to the standard mode to obtain the standard state variables. This can be done in the context of *tracking state estimation*.

Similar to the real-world power distribution networks, we assume that there are other legacy meters available in addition to the D-PMUs, such as through a SCADA system and/or an AMI. Regardless of how the legacy meters are installed at the

primary side and/or the secondary side of the network, they are assumed to be sufficient to achieve full-observability at the moment when such legacy measurements become available. However, importantly, as we discussed in Section I-A, the low reporting rate of the legacy meters requires the power system operator to operate the power distribution system *without* updating the state variables for 30 to 60 minutes, i.e., until the next readings of the legacy meters become available [7], [18]. Accordingly, our goal in this paper is to take advantage of the high reporting rate of the D-PMUs to continue updating the state variables; yet addressing the *low-observability* in D-PMU measurements, i.e., the fact that in practice only a few D-PMUs can be available on each power distribution feeder.

In this setting, the measurements from legacy meters are assumed to be used in order to update the initial values for the tracking state estimation problem at time  $t = 0$ . The details on how such initial values are obtained by using the measurements from the legacy meters are *not* of concern in this study; because they are beyond the scope of this paper. For example, one can use any of the existing DSSE methods in the literature, such as WLS-DSSE [36], or use power flow analysis [3], or any other method to obtain the initial values for the state variables from the low-reporting-rate legacy meters.

Once the initial values are provided by the measurements from the legacy meters, we start the use of the proposed DSSE method to continue updating the state variables *without* access to any new measurement from the legacy meters. In this regard, we keep solving the *low-observable* DSSE problem in (18) once every 100 milliseconds *solely* based on the measurements from only a few D-PMUs for the next 30 to 60 minutes. In this 30 to 60 minutes period, we keep adding the estimated values that are obtained in differential mode to the initial values in order to obtain the nodal voltage phasors for the whole network at each iteration:

$$\mathbf{V}^t = \mathbf{V}^0 + \sum_{\tau=1}^{t-1} \Delta \mathbf{V}^\tau \quad (34)$$

where  $\mathbf{V}^0$  denotes the vector of the initial voltage phasors that is obtained at time  $t = 0$  by using the measurements from the legacy low-reporting-rate meters in the system. Note that, if needed, the time-step for the proposed tracking DSSE can be adjusted with respect to the time frame in which the dynamics in the power distribution system may change due to the events.

#### F. Drift Identification and Calibration

1) *Drift Identification*: By solving the adaptive group Lasso in (24), we minimize the error in estimating the differential voltage phasors at each time slot. However, as time goes by, we inevitably see *accumulative error* in estimating the voltage phasors in (34). Such accumulative error can ultimately result in a considerable drift in the DSSE results from the true values.

In order to identify a possible drifting in the state estimation results, we check the following condition:

$$\|\mathbf{z}^t - \mathbf{H}\mathbf{x}^t\|_2 \geq \beta, \quad (35)$$

where  $\beta$  is a threshold parameter. If condition (35) holds at any time  $t$ , then we trigger a *calibration* mechanism.



TABLE II  
COMPARING DIFFERENT DSSE METHODS IN ESTIMATING VOLTAGE PHASORS AND ESTIMATING DIFFERENTIAL VOLTAGE PHASORS

Method	Average MAPE	STD of MAPE	Average MAE $\times 10^{-3}$		STD of MAE $\times 10^{-3}$	
			No Calibration	Calibration	No Calibration	Calibration
The Proposed Sparse Tracking DSSE	4.51%	4.66	8.2	6.1	5.2	4.7
The DSSE Method in [25]	12.04%	12.74	9.1	-	4.6	-
The Conventional WLS-DSSE as in [36]	19.57%	9.65	28.2	-	17.4	-

2) *Calibration*: At each time interval  $t$ , D-PMUs can serve as *reference* to obtain voltage at the few specific locations where D-PMUs are installed. Accordingly, whenever calibration is needed, we use the available D-PMU measurements in each partition  $p$  as the reference point to calibrate the voltage estimation for all the nodes within the same partition through:

$$\mathbf{V}_{p,cal}^t := \mathbf{V}_p^t \times \zeta_p, \quad (36)$$

where  $\zeta_p$  is the ratio of the voltage phasor measured by the D-PMU to the voltage phasor estimated at the same bus in partition  $p$ . The ratio is calculated separately for the real part and the imaginary part of the voltage phasor.

#### IV. CASE STUDIES

The case studies are done by simulating the IEEE 33-bus distribution test network [37]. Tracking DSSE is performed upon receiving the D-PMU measurements at 100 milliseconds intervals during the one hour period between two consecutive AMI readings. The legacy meters' data provide the *initial states* for the proposed tracking DSSE. At most, one event is assumed to occur within each time slot between the two readings of D-PMUs. Each event can be a change in active/reactive power load, distributed generation, or capacitor bank switching; and it is assumed to be within  $\pm 50\%$  of its value at the previous time slot. The loads are assumed to be constant-power loads.

Unless we state otherwise, we consider five zones in the network. The longer lateral is divided into two zones. Each shorter lateral is one zone. One D-PMU is placed at the substation; and five D-PMUs are placed at the end of each zone, i.e. at buses 9, 18, 22, 25, and 33. Thus, the ratio of the available measurements to the unknowns is  $12/65 = 0.185$ . Error in D-PMU measurements have zero mean and standard deviation of  $\sigma_V = 0.1\%$  for voltage measurements and  $\sigma_I = 1\%$  for current measurements. A total of 120,000 random scenarios are generated in MATPOWER in MATLAB R2018b.

##### A. Performance Comparison

Performance comparison is done with two other methods. The first one is the conventional WLS-DSSE method [36]. For the WLS-DSSE method, we inevitably need to add pseudo-measurements to make the network fully-observable; otherwise the WLS-DSSE method cannot solve the undetermined system of equations for the DSSE problem. Thus, to have a consistent and fair comparison with our proposed method, we used the initial measurements from the legacy meters at time

$t = 0$  as pseudo-measurement for the rest of time slots. Of course, the pseudo-measurements that come from the legacy meters will be updated at the next interval when new measurements become available, i.e., after 30 to 60 minutes. The second one is our previous method in [25]. Since differential synchrophasors are small, we use mean absolute percentage error (MAPE) to assess the estimation of differential phasors:

$$\text{MAPE} = \frac{1}{2N} \sum_{i=1}^N \left( \left| \frac{\Re(\Delta \hat{x}_i - \Delta x_i)}{\Re(\Delta x_i)} \right| + \left| \frac{\Im(\Delta \hat{x}_i - \Delta x_i)}{\Im(\Delta x_i)} \right| \right) \times 100\%, \quad (37)$$

where  $\Delta \hat{x}_i$  denotes the estimated value and  $\Delta x_i$  denotes the true value of the  $i$ -th differential state variable, respectively. Also, we use the mean absolute error (MAE) to assess the estimation of the synchrophasors:

$$\text{MAE} = \frac{1}{2N} \sum_{i=1}^N \left( |\Re(\hat{x}_i - x_i)| + |\Im(\hat{x}_i - x_i)| \right), \quad (38)$$

where  $\hat{x}_i$  denotes the estimated value and  $x_i$  denotes the true value of the  $i$ -th state variable, respectively.

The statistical results for performance comparison are shown in Table II. We can see that the proposed method has a much lower MAPE and MAE; both in terms of its mean and its standard deviation. In 97.9% of all the scenarios, the MAPE is less for the proposed method than WLS-DSSE.

**Remark 4:** Given the sparsity in the vector of state variables, the true value for some of state variables is zero; this causes the denominator in (37) to be zero. As a result, if a state variable with zero value is estimated incorrectly with a non-zero value, then it would cause the MAPE to become infinity; which would make the MAPE useless. Therefore, for those rare scenarios, we consider the percentage error of the associated state variables to be 100% (as the maximum error level). Importantly, in our proposed DSSE method, the weighted Lasso optimization in (24) has an  $\ell_1$ -norm penalty term whose associated weights are assigned such that the above misestimation is avoided. On the contrary, the conventional WLS-DSSE cannot prevent such misestimation; because it is not designed to recover a sparse vector. This is one of the reasons that MAPE for the WLS-DSSE is quite larger than that of other two methods which use sparse signal recovery.

The average computation time, as well as the total number of divergent cases, are shown for each method in Table III. We can see that the computation time of the proposed sparse DSSE

TABLE III  
COMPARISON OF COMPUTATIONAL TIME AND  
CONVERGENCE RATIO OF DSSE METHODS

Method	Average Computational Time (ms)	Number of Divergence Cases
The Proposed Sparse Tracking DSSE	764	0
The Conventional WLS-DSSE	2234	3748

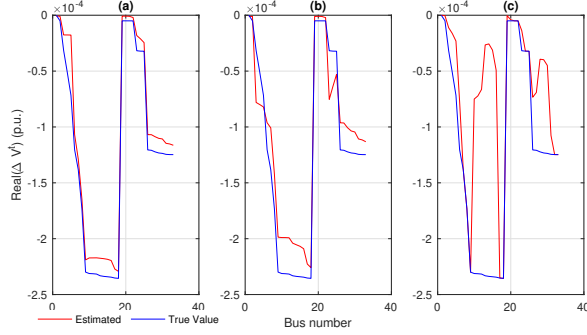


Fig. 2. The estimation of the real part of *differential* voltage phasors at all buses in a typical scenario: a) the proposed sparse tracking DSSE; b) the DSSE method in [25]; and c) The conventional WLS-DSSE method in [36].

method is much lower than that of the conventional WLS-DSSE. Also, unlike the WLS-DSSE method, the proposed method does not experience any case with divergence.

### B. State Estimation Profiles and Impact of Calibration

The state estimation results for an example scenario for differential voltage phasors and ordinary voltage phasors are shown in Figs. 2 and 3. Real parts of the estimated states are shown here. The results for the imaginary parts are similar and omitted due to space limitation. We can see that the proposed DSSE method significantly outperforms the other two methods. Moreover, the proposed calibration method has further enhanced the performance of voltage estimation in comparison to the case without calibration. Importantly, the calibration method is not applicable to the WLS-DSSE. That is why the curve for WLS-DSSE does not include the calibration.

Furthermore, the changes over time in the gross error in state estimation of the voltage phasors is shown in Fig. 4. As we can see, when the state estimation error hits the drift identification threshold  $\beta$  as in Eq. (35) at  $t = t_1$ , the proposed calibration mechanism is triggered and it helps the tracking DSSE method to reduce the gross error. No need to say that this calibration can improve the performance only to a certain limit and the gross error inevitably continues to rise; until the next cycle when the measurements from the legacy meters become available at  $t = t_2$ . At that moment, the system becomes *momentarily fully-observable*; thus, the state estimation error is reset to zero. It is worth clarifying that, based on the setup in our case studies, the difference between using and not using the proposed calibration method is visible in Fig. 4 only during the period between  $t = t_1$  and  $t = t_2$ . That is why the blue curve is visible only during this period.

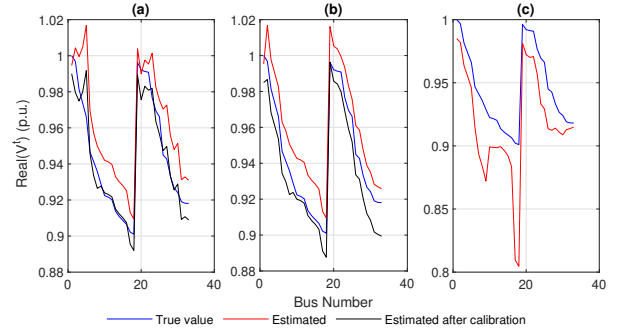


Fig. 3. The estimation of the real part of the *ordinary* voltage phasors at all buses: a) The proposed sparse tracking DSSE; b) The DSSE method in [25]; and c) The conventional WLS-DSSE method in [36].

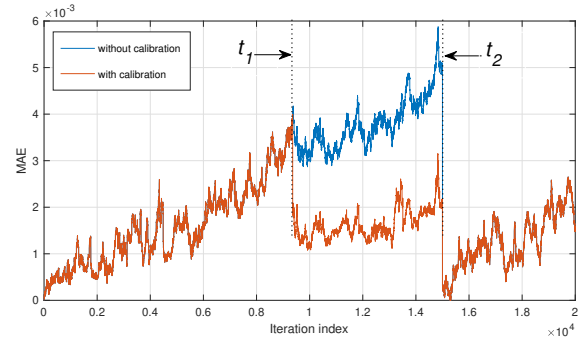


Fig. 4. The gross estimation error of voltage phasors over time for the proposed sparse tracking DSSE method.

### C. Importance of Side Information

Next, we repeat the scenarios in Section IV-A but this time we use *non-adaptive* sparse recovery, i.e., we do *not* use the side information. This results in using the same weights for all the unknowns in the problem formulation. The average MAPE increases from 4.51% to 15.58%; and the average MAE increases from  $8.2 \times 10^{-3}$  to  $12.4 \times 10^{-3}$ . The new error levels are still less than that of the conventional WLS-DSSE. These results show the importance of using the side information; which is one of the key contributions in this paper.

### D. Impact of Renewable Energy Resources

Without loss of generality, suppose the distribution network has four renewable generators at nodes 14, 20, 23, and 30. Fig. 5 shows the DSSE results. Increasing the penetration rate of the renewable energy resources results in decreasing the accuracy of the DSSE method. The degraded accuracy in the DSSE algorithm is due to the degraded accuracy in the event zone identification. This makes the classification of state variables more challenging; although the basic sparsity features remain the same. In other words, the unknown vector is still a sparse vector even under high penetration of renewable energy resources; however, obtaining the side information about the location of the zero/non-zero entries is more difficult. This is not a surprising result; because the increasing penetration of renewable energy resources is a major challenge in practically every DSSE method; e.g., see a similar concern in [38].

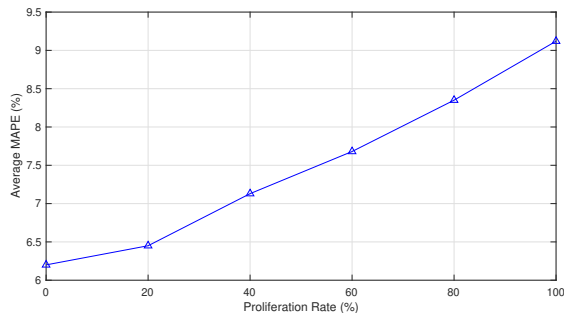


Fig. 5. Average MAPE versus the penetration rate of renewable generation.

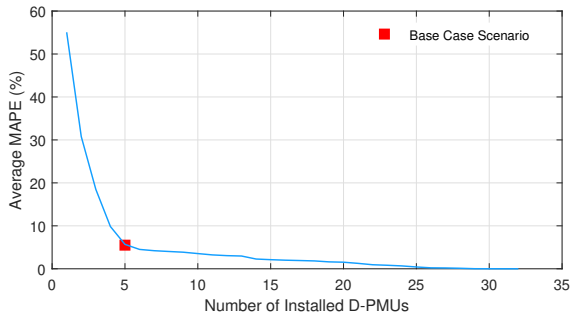


Fig. 6. Average MAPE versus the number of newly installed micro-PMUs.

### E. Importance of the Number of D-PMUs

So far, we have assumed using five D-PMUs. However, increasing the number of D-PMUs can improve the performance of the proposed DSSE method; even though the network may still remain low-observable. The average MAPE with respect to the number of installed D-PMUs is shown in Fig. 6. Note that, the number of D-PMUs also affects the number of zones and the way that we extract the side information. However, to have a consistent comparison, we do *not* change the way that the zones are defined and identified. Based on the results in Fig. 6, adding more sensors slightly lowers the average estimation error; however, we believe that for the defined sparse DSSE in a low-observable network, some points near the knee point on the curve can be assumed as the optimal number of sensors. One practical way to add new sensors is to simply equip buses which have the highest estimation error.

### F. Importance of the Location of Event

To figure out the effect of the location of event on the performance of the proposed method, the distribution of average MAPE versus the node of the event is plotted in Fig. 7, where the red bar shows the median and the blue box shows the distribution of the first to the third quartile of MAPE. As we can see, the higher MAPE belongs to the scenarios wherein the event occurs at locations which have greater distance from the substation. This lowers the sparsity level of the unknown vector. Thus, the capability of the proposed sparse DSSE method to recover the solution is compromised in such scenarios. Of course, the magnitude of the event also has impact on the MAPE. Events with lower magnitude cause such small changes in the operating points of the system that

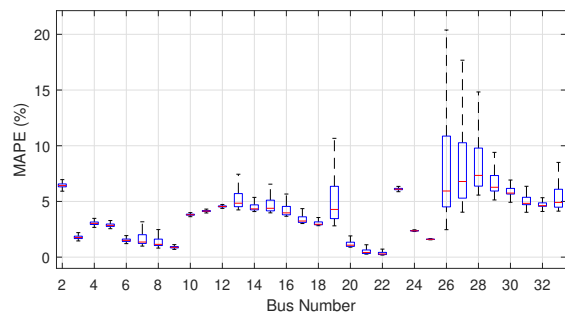


Fig. 7. Distribution of MAPE with respect to the node of event.

sparse DSSE method cannot track them. Once the magnitude is larger, the performance of proposed method is enhanced.

### G. Importance of Measuring Phase Angle

A unique feature of D-PMUs is their ability to measure *phase angle* [9]. In this section, we seek to answer the following question: is it critical is to use phase angle measurements? To answer this question, we repeat all the scenarios, but we assume that phase angel at all nodes is zero. This will cause the average MAPE to increase from 4.51% to 11.76% and the average MAE to increase from  $8.2 \times 10^{-3}$  to  $11.39 \times 10^{-3}$ . Furthermore, removing phase angle causes a negative effect on the calibration mechanism to the extend that it increases the average MAE from  $6.1 \times 10^{-3}$  to  $18.67 \times 10^{-3}$ . Therefore, the use of D-PMUs is indeed necessary in this method.

### H. Impact of Unbalanced Three-Phase Operation

In this section, we examine the IEEE 123-bus test system [39]; which is a highly *unbalanced* distribution network with Wye load connections. We use the percentage unbalance (PU) in current at the substation as the metric for unbalance in the network. We assume that only 19 D-PMUs are available; and they are installed at nodes 6, 11, 16, 18, 32, 39, 47, 56, 59, 66, 71, 75, 85, 86, 96, 101, 114, 250, and 450. This means that the ratio of the available measurements to the unknowns is  $38/245 = 0.156$ . All phasor measurements are three-phase.

We set the initial PU to 8%. Next, we modify the loads across the network to increase the PU level. As shown in Fig. 8, under various PU levels, the average MAPE is much lower for the proposed DSSE than for the conventional WLS-DSSE. Note that, since the DSSE formulation based on synchrophasor measurements is inherently linear; the unbalanced operation of the distribution system does *not* affect the core characteristics of the proposed sparse DSSE. The small increment in the MAPE is due to the increase in the neutral current as the unbalance level increases. Moreover, the estimation error in this case study is higher than that of the previous case studies based on the IEEE 33-bus test system. This is due to the higher number of small laterals for the IEEE 123-bus test system. It is not due to the unbalanced operation. The performance can be improved if we add more D-PMUs to the laterals; even while the system still remains low-observable.

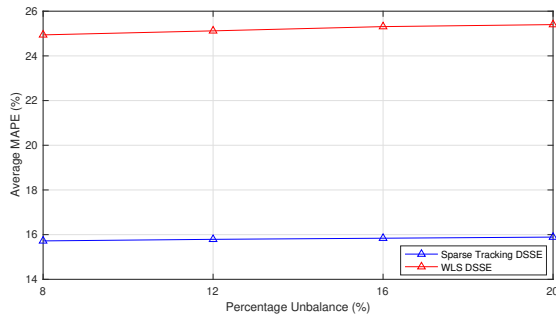


Fig. 8. Average MAPE versus the percentage phase unbalance.

### I. Examining the Uniqueness of Solution

As we discussed in Section III-C, the exactness of the solution in a sparse recovery problem directly depends on its uniqueness. Of interest in this context is the tightness of the theoretical sufficient conditions to assure that the obtained DSSE solution is unique and thus exact. Importantly, the sufficient conditions in [35] are satisfied only in 0.8% of the scenarios. However, the sufficient conditions that we developed in Theorem 1 are satisfied in 11.3% of the scenarios. Thus, the theoretical optimality conditions have significantly improved in our analysis; compared to the conventional sufficiency conditions for a generic sparse recovery problem without side information. This is an important theoretical achievement. Of course, since these theoretical results are based on sufficient conditions, in theory, the rest of the scenarios may or may not achieve the unique solution. Nevertheless, our numerical results confirm that the mean square error in state estimation is less than  $10^{-8}$  in 96.4% of the scenarios.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, a novel sparse tracking DSSE method is proposed to address the *low-observability* in power distribution networks, where the measurements come from only a small number of D-PMUs. The analysis is done using *differential synchrophasors* and by dividing the state variables into *four types* based on their *group sparsity* properties. Prior to solving the formulated DSSE problem, an *event zone* identification analysis is applied to augment the proposed sparse DSSE with the *side information* on the support of the unknown vector. The DSSE is modeled as an adaptive group sparse recovery problem to estimate the differential synchrophasors, which are added to the initial values to estimate the voltage phasors in standard mode at each time slot. Since the proposed method has a *tracking scheme*, a drift identification and *calibration* method is developed to enhance robustness. The proposed method is tested on the IEEE 33-bus and IEEE-123 bus power distribution feeder and the simulation results show the effectiveness of the proposed method over the conventional WLS DSSE method which is aided by the pseudo-measurements.

As a possible extension of this work in the future, one can further study the problem when multiple major events happen on the distribution feeder at the same time. Also, one can investigate how adding other types of measurements, such as load measurements, or power measurements that are obtained

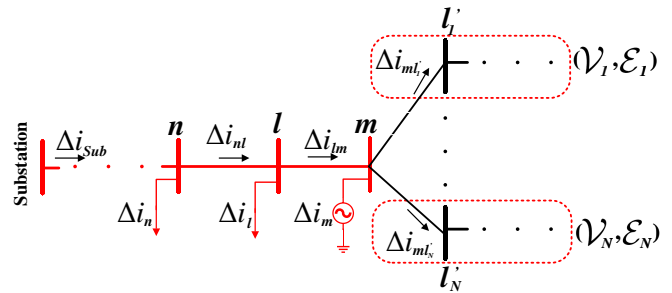


Fig. 9. Equivalent circuit of distribution feeder when an event happens at node  $m$ . This equivalent circuit is used in the proof in Appendix A.

from the voltage and current phasors at the D-PMUs, to the existing phasor measurements in our proposed tracking DSSE method may affect the problem formulation and the results.

## APPENDIX A PROOF OF COROLLARY 1

Consider the network in Fig. 9, which is the equivalent circuit of a power distribution feeder at the time of event that is obtained by applying the Compensation Theorem [27]. The event at bus  $m$  is replaced by a *current source* which injects the same level of current due to the event to event bus. By applying the Kirchhoff's Current Law (KCL) at the event bus, i.e., the last bus on the red path, we have:

$$\Delta i_m - \Delta i_{lm} + \sum_{j=1}^N \Delta i_{ml'_j} = 0. \quad (39)$$

We can write a similar equation at node  $l$  as

$$\Delta i_l + \Delta i_{nl} - \Delta i_{lm} = 0. \quad (40)$$

Based on the Compensation Theorem, since no event occurs at bus  $l$ , there is no change in the injected current to this bus, i.e., we have:  $\Delta i_l = 0$ . This is also true for all other buses in set  $\mathcal{V}$ . Thus,  $\Delta i_{nl} = \Delta i_{lm}$ . If we continue writing the same type of equations for the buses on the red path until we reach the substation, we can conclude that the change in the line current for all the line segments in set  $\mathcal{E}$  is equal to the change in the injected current that is drawn from the substation:

$$\Delta i_{sub} = \Delta i_{lm} \quad \forall lm \in \mathcal{E}. \quad (41)$$

Next, let us write the KCL at all nodes in set  $\mathcal{V}_1$ . For the same reason, the change in current for all the line segments in set  $\mathcal{E}_1$  is equal to  $\Delta i_{ml'_1}$ . Consider the whole set of  $(\mathcal{V}_1, \mathcal{E}_1)$  as a super node. If the value of injected current to this super node, i.e.,  $\Delta i_{ml'_1}$  is non-zero, then it means that this current is being injected to one of the inner nodes of the super node, which means that there is an event somewhere inside the super node that would be against our initial assumption. Therefore,  $\Delta i_{ml'_1} = 0$ . Accordingly, we have:

$$\sum_{j=1}^N \Delta i_{ml'_j} = 0. \quad (42)$$

By replacing (42) in (39), we obtain:

$$\Delta i_m = \Delta i_{sub}. \quad (43)$$

Therefore, when an even happens, the change in current for all the line segments in set  $\mathcal{E}$  is non-zero, and equal to the value of injected current from the substation node, or the value of the equivalent current source. Also, the change in current for all the line segments in set  $\mathcal{L} \setminus \mathcal{E}$  is zero.

Next, let us write the Ohm's law for line segment  $ml'_1$  as in (6). Since  $\Delta i_{ml'_1} = 0$ , we can obtain

$$\frac{R_{ml'_1}}{X_{ml'_1}} = \frac{\Re(\Delta v_m) - \Re(\Delta v_{l'_1})}{\Im(\Delta v_m) - \Im(\Delta v_{l'_1})}. \quad (44)$$

Thus, based on the values of the parameters, there are two scenarios to consider. First, if the impedance of line  $ml'_1$  and the differential voltage at node  $m$  are such that

$$\frac{R_{ml'_1}}{X_{ml'_1}} \approx \frac{\Re(\Delta v_m)}{\Im(\Delta v_m)}; \quad (45)$$

then, we can (approximately) replace  $\Delta v_{l'_1}$  by zero. This is the case that holds for the nodes in set  $\mathcal{V}_2$  in the illustrative example shown in Fig. 1(c). Second, if the corresponding parameters are not such that (45) holds, then  $\Delta v_{l'_1} \neq 0$ . ■

#### APPENDIX B PROOF OF THEOREM 1

The proof of Theorem 1 largely follows the procedure outlined in [40]. We will first discuss the necessary and sufficient conditions that any optimal point of the convex optimization problem in (24) must satisfy. Let us rewrite the problem in (24) in a compact form as

$$\min_{\Delta \mathbf{x}} \frac{1}{2} \|\Delta \mathbf{z} - \mathbf{H}\Delta \mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\Delta \mathbf{x}\|_1, \quad (46)$$

where  $\mathbf{W}$  is a diagonal matrix with weights  $w_\gamma$  for  $\gamma \in \{1, \dots, N\}$ . The optimality conditions can be obtained by using the first-order necessary KKT conditions that are also sufficient; the optimization problem in (24) is convex. A necessary condition for vector  $\Delta \mathbf{x}^*$  to be an optimal solution of (46) is that zero vector belongs to the subgradient of its objective. We can write this condition as

$$\mathbf{H}^\top (\mathbf{H}\Delta \mathbf{x}^* - \Delta \mathbf{z}) + \lambda \mathbf{W} \partial \|\Delta \mathbf{x}^*\|_1 = 0. \quad (47)$$

$\partial \|\Delta \mathbf{x}^*\|_1$  denotes the subgradient which can be defined as

$$\partial \|\Delta \mathbf{x}^*\|_1 \in \begin{cases} \text{sign}(\Delta \mathbf{x}^*) & \Delta \mathbf{x}^* \neq 0 \\ [-1, 1] & \Delta \mathbf{x}^* = 0. \end{cases}$$

We can write (47) in a compact form as

$$|\mathbf{H}_\gamma^\top (\mathbf{H}\Delta \mathbf{x}^* - \Delta \mathbf{z})| \leq \lambda w_\gamma.$$

One can show that the solution is a unique minimum if the following sufficient conditions are satisfied [33]:

$$\mathbf{H}_\gamma^\top (\mathbf{H}\Delta \mathbf{x}^* - \Delta \mathbf{z}) = -\lambda w_\gamma \text{sign}(\Delta \mathbf{x}_\gamma^*), \text{ if } \Delta \mathbf{x}_\gamma^* \neq 0 \quad (48)$$

$$|\mathbf{H}_\gamma^\top (\mathbf{H}\Delta \mathbf{x}^* - \Delta \mathbf{z})| < \lambda w_\gamma, \text{ if } \Delta \mathbf{x}_\gamma^* = 0. \quad (49)$$

We can now prove that if a vector  $\Delta \mathbf{x}$  with support  $\Gamma$  and sign sequence  $\mathbf{s}_\Gamma$  satisfies the set of equations  $\mathbf{H}\Delta \mathbf{x} = \Delta \mathbf{z}$  and the

conditions in (32) and (33), then we can explicitly write the unique solution of (46) supported on  $\Gamma$  as

$$\Delta \mathbf{x}_\Gamma^* = \Delta \mathbf{x}_\Gamma - \lambda (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma. \quad (50)$$

The sign condition in (33) implies that  $\text{sign}(\Delta \mathbf{x}_\Gamma^*) = \mathbf{s}_\Gamma$ .

To prove the optimality of  $\Delta \mathbf{x}^*$  in (50), we need to show that  $\Delta \mathbf{x}^*$  satisfies the necessary and sufficient conditions in (48) and (49). Let us first look at the condition in (48) as

$$\begin{aligned} & \mathbf{H}_\Gamma^\top (\mathbf{H}_\Gamma \Delta \mathbf{x}_\Gamma^* - \Delta \mathbf{z}) \\ &= \mathbf{H}_\Gamma^\top (\mathbf{H}_\Gamma [\Delta \mathbf{x} - \lambda (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma] - \Delta \mathbf{z}) \\ &= \mathbf{H}_\Gamma^\top (\underbrace{\mathbf{H}_\Gamma \Delta \mathbf{x} - \Delta \mathbf{z}}_{=0}) - \lambda \mathbf{W}_\Gamma \mathbf{s}_\Gamma \\ &= -\lambda \mathbf{W}_\Gamma \mathbf{s}_\Gamma = -\lambda \mathbf{W}_\Gamma \text{sign}(\mathbf{x}_\Gamma^*). \end{aligned} \quad (51)$$

For the condition in (49), we can show that for any  $\gamma \notin \Gamma$

$$\begin{aligned} & |\mathbf{H}_\gamma^\top (\mathbf{H}_\Gamma \Delta \mathbf{x}_\Gamma^* - \Delta \mathbf{z})| \\ &= |\mathbf{H}_\gamma^\top (\mathbf{H}_\Gamma [\Delta \mathbf{x} - \lambda (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma] - \Delta \mathbf{z})| \\ &= |\lambda \mathbf{H}_\gamma^\top \mathbf{H}_\Gamma (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma| \\ &< \lambda w_\gamma. \end{aligned} \quad (52)$$

The last inequality is due to the main condition in (32). The above inequality completes the proof; since  $\Delta \mathbf{x}_\Gamma^*$ , as defined in (50), satisfies the necessary and sufficient conditions to be the strict optimal solution of the problem in (46). ■

#### REFERENCES

- [1] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312–2322, Mar. 2019.
- [2] I. Dafi, R. A. Jabr, and T. Hrnji, "Hybrid state estimation in complex variables," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5288–5296, Sep. 2018.
- [3] A. Abur and A. Gomez-Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.
- [4] V. Madani *et al.*, "Distribution automation strategies, challenges and opportunities in a changing landscape," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2157–2165, Jul. 2015.
- [5] A. Akrami, M. Doostizadeh, and F. Aminifar, "Power system flexibility: an overview of emergence to evolution," *J. Modern Power Syst. Clean Energy*, vol. 7, no. 5, pp. 987–1007, Sep. 2019.
- [6] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4796–4805, Nov. 2019.
- [7] J. Song, E. Dall'Anese, A. Simonetto, and H. Zhu, "Dynamic distribution state estimation using synchrophasor data," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 821–831, Jan. 2020.
- [8] <https://www.powerstandards.com/download/micropmu-data-sheet/>.
- [9] H. Mohsenian-Rad, E. Stewart, and E. Cortez, "Distribution synchrophasors: Pairing big data with analytics to create actionable information," *IEEE Power Energy Mag.*, vol. 16, no. 3, pp. 26–34, May 2018.
- [10] P. Deoliveira-Dejesus, N. Rodriguez, D. Celeita, and G. Ramos, "PMU-based system state estimation for multigrounded distribution systems," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1071–1081, Mar. 2021.
- [11] A. Akrami, M. Doostizadeh, and F. Aminifar, "Optimal reconfiguration of distribution network using  $\mu$ PMU measurements: A data-driven stochastic robust optimization," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 420–428, Jan. 2020.
- [12] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Enhancing observability in distribution grids using smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5953–5961, Nov. 2018.
- [13] K. A. Clements, "The impact of pseudo-measurements on state estimator accuracy," in *Proc. IEEE PES General Gen. Meeting*, Detroit, MI, 2011.
- [14] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov. 2015.

- [15] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement," *IET Gen., Transm. Distrib.*, vol. 4, no. 1, pp. 50–59, Jan. 2009.
- [16] Y. R. Gahrooei, A. Khodabakhshian, and R. Hooshmand, "A new pseudo load profile determination approach in low voltage distribution networks," *IEEE Trans. Power Syst.*, vol. 33, pp. 463–472, Jan. 2018.
- [17] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1888–1896, Nov. 2012.
- [18] J. Ostrometzky, K. Berestizshevsky, A. Bernstein, and G. Zussman, "Physics-informed deep neural network method for limited observability state estimation," *IEEE Trans. Smart Grid*, Available: arXiv:1910.06401, 2019.
- [19] S. M. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1631–1642, Jul. 2014.
- [20] M. Majidi, M. Etezadi-Amoli, and H. Livani, "Distribution system state estimation using compressive sensing," *In. J. Elec. Power Energy Syst.*, vol. 88, pp. 175–186, Jun. 2017.
- [21] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, "Matrix completion for low-observability voltage estimation," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2520–2530, May 2020.
- [22] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Enhanced tensor completion based approaches for state estimation in distribution systems," *IEEE Trans. Ind. Inform.*, vol. 17, no. 9, pp. 5938–5947, Sep. 2021.
- [23] S. Dahale, H. S. Karimi, K. Lai, and B. Natarajan, "Sparsity based approaches for distribution grid state estimation - a comparative study," *IEEE Access*, vol. 8, pp. 198 317–198 327, 2020.
- [24] 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 Trans. Power Syst.*, vol. 33, no. 6, pp. 6343–6354, Nov. 2018.
- [25] A. Akrami, M. S. Asif, and H. Mohsenian-Rad, "Sparse distribution system state estimation: An approximate solution against low observability," in *Proc. IEEE PES ISGT Conf.*, Washington, DC, Feb. 2020.
- [26] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [27] K. S. Kumar, *Electric circuits and networks*. Pearson, India, 2009.
- [28] A. Shahsavari, M. Farajollahi, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, "Situational awareness in distribution grid using micro-PMU data: A machine learning approach," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6167–6177, Nov. 2019.
- [29] A. Aligholian, A. Shahsavari, E. Stewart, E. Cortez, and H. Mohsenian-Rad, "Unsupervised event detection, clustering, and use case exposition in micro-pmu measurements," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3624–3636, Jul. 2021.
- [30] A. Shahsavari, A. Sadeghi-Mobarakeh, E. M. Stewart, E. Cortez, L. Alvarez, F. Megala, and H. Mohsenian-Rad, "Distribution grid reliability versus regulation market efficiency: An analysis based on micro-pmu data," *IEEE Trans. on Smart Grid*, vol. 8, no. 6, pp. 2916–2925, Nov. 2017.
- [31] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," *J. Fourier Anal. Applic.*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [32] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [33] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–121, Jul. 2011.
- [34] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the Lasso and generalizations*. CRC press, 2015.
- [35] H. Zhang, W. Yin, and L. Cheng, "Necessary and sufficient conditions of solution uniqueness in 1-norm minimization," *J. Optim. Theory Appl.*, no. 164, pp. 109–122, 2015.
- [36] D. A. Haughton and G. T. Heydt, "A linear state estimation formulation for smart distribution systems," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1187–1195, May 2013.
- [37] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1994.
- [38] Y. Weng, R. Negi, and M. D. Ilic, "Probabilistic joint state estimation for operational planning," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 601–612, Jan. 2019.
- [39] <https://site.ieee.org/pes-testfeeders/resources>.
- [40] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.



**Alireza Akrami** (S'18) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2014, and the M.Sc. degree in electrical engineering from University of Tehran, Tehran, Iran, in 2018. He is currently a Ph.D. candidate in electrical engineering at the University of California, Riverside, CA, USA. He is specifically researching on distribution system monitoring. His research interests include smart grid initiatives, power system optimization, and applications of machine learning in power systems. He was a recipient of MPCE journal best paper award in the year 2019 and was selected among the 2019 Best Reviewers of the IEEE TRANSACTIONS ON SMART GRID.



**M. Salman Asif** is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of California, Riverside. Dr. Asif received his B.Sc. degree in 2004 from the University of Engineering and Technology, Lahore, Pakistan, and an M.S.E.E degree in 2008 and a Ph.D. degree in 2013 from the Georgia Institute of Technology, Atlanta, Georgia. He worked as a Senior Research Engineer at Samsung Research America, Dallas from August 2012 to January 2014 and as a Postdoctoral Researcher at Rice University from February 2014 to June 2016. He received NSF CAREER Award in 2021, Google Faculty Award in 2019, and UC Regents Faculty Fellowship Award in 2017. His research interests include compressive sensing, computational and medical imaging, and machine learning.



**Hamed Mohsenian-Rad** (M'09-SM'14-F'20) received the Ph.D. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada, in 2008. He is currently a Professor of electrical engineering and a Bourns Family Faculty Fellow at the University of California, Riverside, CA, USA. His research is on monitoring, data analysis, and optimization of power systems and smart grids. He is the author of the textbook *Smart Grid Sensors: Principles and Applications* by Cambridge University Press - 2021.

He was the recipient of the National Science Foundation (NSF) CAREER Award, the Best Paper Award from the IEEE Power & Energy Society General Meeting, and the Best Paper Award from the IEEE Conference on Smart Grid Communications. He has been the PI on ten million dollars research grants in the field of smart grid. He has served as Editor for the IEEE TRANSACTIONS ON SMART GRID and the IEEE POWER ENGINEERING LETTERS.