Transforming Conventional Waveform Measurements into Synchro-waveforms: A Data-Driven Method for Event Signature Alignment and Synchronization Operator Estimation

Zong-Jhen Ye, *Student Member, IEEE*, and Hamed Mohsenian-Rad, *Fellow, IEEE*

***Abstract*—A series of methodologies are proposed to transform conventional waveform measurements from legacy power quality meters into synchro-waveforms. This study is motivated by the presence of thousands of legacy power quality meters in operation worldwide that provide event-triggered waveform measurements but lack time-synchronization among their data. Consequently, the waveform measurements from these legacy meters cannot be directly used as synchro-waveforms, limiting their applicability in the promising synchro-waveform applications that have been introduced in the literature in recent years. We address this issue without requiring legacy power quality meters to be equipped with GPS receivers or other time-synchronization hardware. Our data-driven methods operate in two steps: first, they perform optimization-based event signature alignment, and then they use the results to estimate a synchronization operator between any two legacy meters. The proposed methods are accurate, robust, and computationally efficient. All case studies presented in this paper are based on real-world waveform measurements.**

***Keywords*: Synchro-waveforms, synchronization operator, conventional waveform sensors, data-driven method, real-world measurements, event signature alignment, optimization.**

I. Introduction

Synchro-waveforms, i.e., time-synchronized waveform measurements, have emerged recently as a new measurement technology with diverse applications in power system monitoring and situational awareness; e.g., see the overview in [1]. The device to measure synchro-waveforms is often referred to as a Waveform Measurement Unit (WMU). The common approach to achieve time-synchronization among WMUs is to use the Global Positioning System (GPS); see [2], [3], [4, Section 4.6].

A. Motivation: A Practical Challenge

WMU installations are still very rare in practice. In fact, although there is already a wide range of sensor technology deployments that *do* provide voltage and current waveform measurements, such measurements are *not* time synchronized.

One of such conventional technologies that is currently widely used in practice is the power quality sensor. The majority of the existing power quality sensors that have been deployed by many utilities over the past two decades are not equipped with time-synchronization capabilities, such as GPS; due to the high cost. In fact, it was observed in a recent study in [5] that the time stamps from conventional power quality sensors can be off by several milliseconds to several seconds.

As a result, while many utilities do have various conventional power quality sensors to measure voltage and current waveforms, they are unable to use such existing measurements as synchro-waveforms due to lack of proper time-synchronization. Thus, most utilities are still currently unable to support the emerging applications of synchro-waveforms.

The authors are with the University of California, Riverside, CA, USA. The corresponding author is H. Mohsenian-Rad. E-mail: hamed@ece.ucr.edu.

B. Summary of Contributions

In this paper, we seek to address the above open problem. The contributions in this paper can be summarized as follows: (a) Through an automated process, we take waveform measurements from existing conventional power quality meters that are not time-synchronized and convert them into synchro-waveforms. Our proposed algorithm is data-driven, working solely by analyzing the event signatures in event-triggered waveform measurements. (b) The first step in the proposed algorithm includes novel methodologies to align event signatures among unsynchronized sensors. Different optimization-based approaches are developed to achieve computational efficiency and robustness. (c) The second step in the proposed algorithm estimates the unknown synchronization operators among legacy power quality meters. Here, synchronization operators are the offsets in time that are applied to the local clocks at the ordinary power quality meters to time-synchronize their waveform measurements. (d) Two approaches are developed to extend the methodology to be applicable to any number of sensors, either with minimum computational burden or with the option to utilize redundancy to enhance performance.

C. Related Literature

Three recent overview papers about the emerging field of synchronized waveform measurements are published in [1]–[3]. The link to a related IEEE Power and Energy Society Task Force is available in [6]. Some of the applications of synchro-waveforms have been recently discussed in the literature, such as for event and fault location identification [7]–[9], wildfire monitoring [5], power system oscillation monitoring [10], [11], power system protection [12], [13], and dynamic modeling of the sub-cycle behavior of inverter-based resources [14], [15].

The use of GPS to time-synchronize waveform measurements is discussed in [3]. The use of communications protocols for time-synchronization is discussed in [16]. The

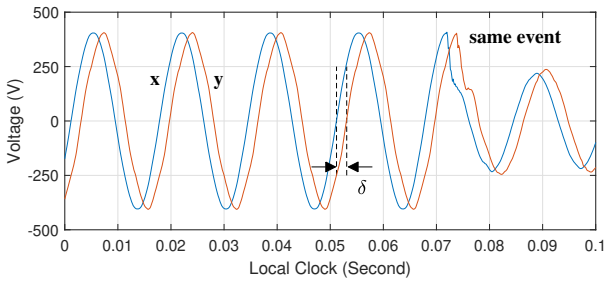

Fig. 1. Two time-series of voltage waveform measurements of the same phase of the power system but from two different power quality sensors at two different locations. The two sensors are *not* time synchronized.

difficulties in achieving time alignment among unsynchronized measurements is discussed in [17], with focus on phasor measurements (not waveform measurements). In [18], it is explained that, in practice, it is rare to have access to time-synchronized waveform measurements due to the high cost of time-synchronization equipment. Some papers focus on the analysis of waveform data at only one sensor, such as in [19].

In another line of research, the focus has been on addressing temporary loss of time synchronization, such as when the GPS signal is momentarily lost. This issue has been discussed in the literature for Phasor Measurement Units (PMUs), such as for fault location identification based on model-based methods [20], or for anomaly sequence detection and damping control based on data-driven methods [21], [22]. Other causes of latency or time offset among PMUs is discussed in [23]. None of the studies in [20]–[22] is related to waveform measurements; as they focus on phasor measurements. In [24], the problem of addressing momentary loss of time synchronization is discussed for both PMUs and WMUs, specifically for the problem of fault location identification. The proposed method is model-based and it requires access to the circuit model.

Compared to our conference paper in [25], the methods in this journal paper are completely redesigned to significantly improve performance, as we will see through performance comparisons. The new methods are also more robust. Furthermore, they can be applied to an arbitrary number of conventional power quality sensors, while the analysis in [25] is limited to only two sensors. Hence, the approach in this journal paper are overall more effective and more practical.

II. Problem Statement

Consider two ordinary (unsynchronized) power quality sensors on a power network. They record voltage and current waveforms during power quality events. The waveform measurements that are captured by these two sensors are event-triggered [4, Section 4.2]. In practice, a wide range of events can trigger waveform-capture by power quality sensors, such as voltage sags and swells, faults, and changes in harmonic distortions. Although each sensor has a local clock to time-stamp its own waveform measurements, the two local clocks are not synchronized with each other. Therefore, the waveform measurements from the two sensors are not time-synchronized.

An example is shown in Fig. 1. Only one phase of the waveform measurements are shown in this figure. Notations $\mathbf{x}$ and $\mathbf{y}$ denote the two time-series of the voltage waveform measurements that are recorded by two ordinary power equality sensors during the *same* physical event in the power system:

$$\mathbf{x} = x[1], \ldots, x[n] \quad \text{and} \quad \mathbf{y} = y[1], \ldots, y[n], \tag{1}$$

where n is the number of measurement samples. For the measurements in Fig. 1, we have $n = 6 \times 128 = 768$ samples; because 6 cycles of waveforms are captured at 128 samples per cycle. Due to lack of time synchronization between the two power quality sensors that provide the measurements, there is $\delta = 2$ milliseconds offset between the voltage waveform measurements from the two sensors. We refer to δ as the *synchronization operator* between the two measurements.

If δ is known, then we can shift one of the two time-series $\mathbf{x}$ and $\mathbf{y}$ by δ and turn the two time-series into synchro-waveforms. However, the challenge is that δ is *not* known.

The objective in this paper is to examine the waveform measurements that are recorded by unsynchronized power quality sensors to estimate the offsets among their local clocks, and hence to time-synchronize their waveform measurements.

Furthermore, we seek to expand our analysis to also time-synchronize waveform measurements from an arbitrary number of unsynchronized ordinary power quality meters.

Throughout this paper, we focus on the analysis of voltage waveform measurements, rather than the analysis of current waveform measurements. Event signatures in current waveform measurements often vary significantly across the power system, depending on the location of each sensor and the nearby circuit elements. Therefore, current waveform measurements are typically not suitable for the analysis that is conducted in this paper. However, once data-driven time synchronization is achieved among a group of sensors by examining their voltage waveform measurements, subsequent applications of time synchronization may involve both voltage and current waveform measurements from those sensors.

The fact that we specifically focus on the analysis of "events" in this paper directly stems from the fact that thousands of legacy power quality meters are currently in operation around the world that provide *event-triggered* waveform captures but without time synchronization. If no power quality event occurs in the power system, then there would be no event-triggered waveform captured by the legacy power quality meters. Accordingly, there would be no need for time synchronization, as there would be no waveform measurements to time-synchronize. However, in reality, power quality events occur regularly, often a few times each day. These events happen for a multitude of reasons, including utility equipment switching, major load switching, operation or misoperation of distributed energy resources, faults, and self-clearing incipient failures. Therefore, in reality, each legacy power quality meter regularly generates event-triggered waveform measurements that can be analyzed using the methodologies in this paper to achieve data-driven time synchronization among the waveform measurements from multiple legacy power quality meters.

Next, we will break down the above problem into two specific sub-problems. We will develop the methods to solve these two sub-problems in Sections III and IV. We will present the extension to the case with several sensors in Section V.

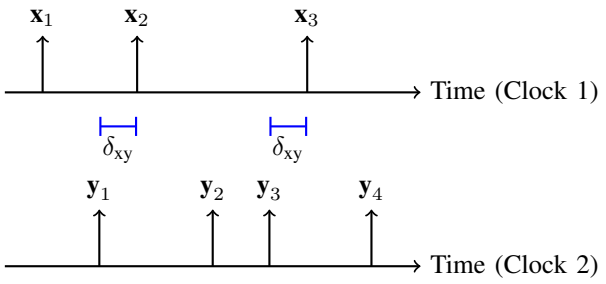

Fig. 2. Top arrows indicate the timing of the events and the corresponding time-series of the waveform captures at the first power quality sensor. Bottom arrows indicate the timing of the events and the corresponding time-series of the waveform captures at the second power quality sensor. The amount of the time-synchronization operator between the two sensors is shown by δ_{xy}.

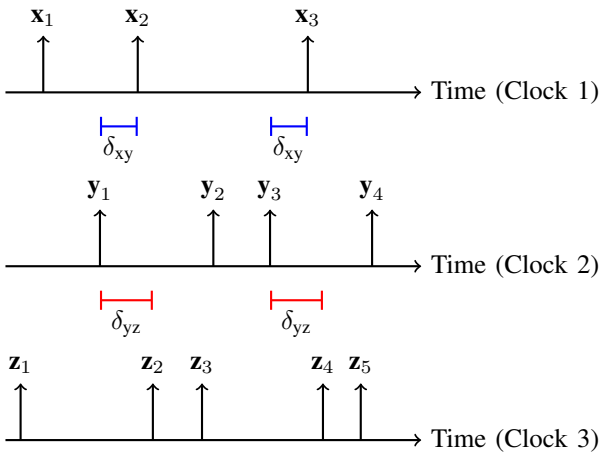

Fig. 3. Extending the example in Fig. 2 to the case with three power quality sensors. The unknown synchronization operator between the first and the second sensors is denoted by δ_{xy}. The unknown synchronization operator between the second and the third sensors is denoted by δ_{yz}.

A. Problem 1: Alignment of Unsynchronized Events

A power quality sensor records (captures) the waveforms only if it detects an event. It is possible that the power quality sensor at one location does capture an event while the power quality sensor at another location does not capture the same event. This is because the same physical event can affect the voltage waveforms at each location of the power system differently from the other locations on the same network. The strategy for event capture may also differ among different sensors. As a result, the set of events that are captured by one power quality sensor can be different from the set of the events that are captured by another power quality sensor.

Let $\mathcal{X}$ denote the set of time-series of the events that are captured by one power quality sensor. Let $\mathcal{Y}$ denote the set of time-series of the events that are captured by another power quality sensor on the same network during the same period of time. Take the case in Fig. 2 as an example. Three events are captured by the first power quality sensor, and four events are captured by the second power quality sensor. We have:

$$\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \quad \text{and} \quad \mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4\}. \tag{2}$$

The second event at the first sensor, i.e., $\mathbf{x}_2$, and the first event at the second sensor, i.e., $\mathbf{y}_1$, correspond to the same physical cause. This event is major enough to affect the waveforms at the locations of both sensors. The third event at the first sensor, i.e., $\mathbf{x}_3$, and the third event at the second sensor, i.e. $\mathbf{y}_3$, also correspond to the same physical cause. The rest of the events, i.e., $\mathbf{x}_1$, $\mathbf{y}_2$, and $\mathbf{y}_4$, are captured only by one of the two sensors. Accordingly, we need to align event $\mathbf{x}_2$ with event $\mathbf{y}_1$ and also align event $\mathbf{x}_3$ with event $\mathbf{y}_3$.

The above alignments are *not* known in advance in practice. Accordingly, Problem 1 is concerned with identifying which pairs of the waveform time-series from the two unsynchronized sensors should be aligned to be used for the purpose of estimating the synchronization operator. For the example in Fig. 2, solving Problem 1 would result in identifying the following pairs for event alignment between the two sensors:

$$\mathcal{X} \cap \mathcal{Y} = \{\mathbf{x}_2, \mathbf{y}_1\} \text{ and } \{\mathbf{x}_3, \mathbf{y}_3\}. \tag{3}$$

B. Problem 2: Estimation of the Synchronization Operator

Given the waveform measurements of the *pairs* of the events in set $\mathcal{X} \cap \mathcal{Y}$, the next step is to conduct a data-driven analysis to estimate the unknown synchronization operator δ. This requires estimating the time difference between the two time-series of the waveform measurements for each pair $\{\mathbf{x}_i, \mathbf{y}_j\} \in \mathcal{X} \cap \mathcal{Y}$. In principle, this can be done by matching the patterns between the waveform signatures in $\mathbf{x}_i$ and $\mathbf{y}_j$.

However, this is a challenging task. The same physical event can affect the voltage waveforms at each location differently from other locations on the same power network. Therefore, even if both sensors do capture the same event, the waveform signature of the event at one sensor may *not* be exactly the same as the waveform signature of the same event at another sensor. As an example, in Fig. 1, the two event signatures are caused by the same event. However, there are considerable differences between these two signatures. These differences can cause difficulties in matching the two signatures.

Another challenge is with regards to accurately estimating the start of the event in the waveform measurements at each power quality sensor based on the event signature that is captured by that sensor. Some events, such as major faults, can create severe event signatures for which the start of the event can be identified somewhat easily. Some other events, such as switching actions, may *not* create major signature. In such cases, identifying the start of the event can be difficult.

Furthermore, since an event can occur at any location in the power system, the electrical distance between the location of the event and the location of the sensor is not the same among the sensors. Therefore, due to the differences in waveform propagation delay, there can be some inherent (albeit not major) drift in estimating the synchronization operator. This needs to be addressed by developing a statistical method that can estimate the synchronization operator based on the waveform measurements from several pairs of events.

C. Extension to Several Sensors

If there are more than two sensors, then the above problems can be more complicated. An example is shown in Fig. 3. Here, the waveform measurements are captured by three

sensors that are not time synchronized. As we previously defined in (2), $\mathcal{X}$ is the set of the events that are captured by the first sensor, and $\mathcal{Y}$ is the set of the events that are captured by the second sensor. Similarly, we define $\mathcal{Z}$ as the set of the events that are captured by the third sensor. Five events are captured by the third power quality sensor. We have:

$$\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5\}. \tag{4}$$

Similar to the event alignment in (3), we can obtain:

$$\begin{aligned} \mathcal{X} \cap \mathcal{Z} &= \{\mathbf{x}_2, \mathbf{z}_2\} \text{ and } \{\mathbf{x}_3, \mathbf{z}_4\}, \\ \mathcal{Y} \cap \mathcal{Z} &= \{\mathbf{y}_1, \mathbf{z}_2\} \text{ and } \{\mathbf{y}_3, \mathbf{z}_4\}. \end{aligned} \tag{5}$$

From (3) and (5), we can also obtain:

$$\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z} = \{\mathbf{x}_2, \mathbf{y}_1, \mathbf{z}_2\} \text{ and } \{\mathbf{x}_3, \mathbf{y}_3, \mathbf{z}_4\}, \tag{6}$$

which is the set of the events that are captured by all the three power quality sensors. Of course, it is also possible that some of the events are captured only by a *subset* of the power quality sensors. In that case, there can be fewer events in set $\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z}$ than in sets $\mathcal{X} \cap \mathcal{Y}$, $\mathcal{X} \cap \mathcal{Z}$, and $\mathcal{Y} \cap \mathcal{Z}$.

The above alignments are not known in advance. Thus, Problem 1 is concerned with identifying which waveform time-series from the three (or many more) sensors should be aligned to use their alignment for the purpose of estimating the synchronization operators among the unsynchronized sensors.

With three unsynchronized power quality sensors, we can define the following *synchronization operators* among them:

$$\delta_{\text{xy}} = -\delta_{\text{yx}}, \quad \delta_{\text{xz}} = -\delta_{\text{zx}}, \quad \delta_{\text{yz}} = -\delta_{\text{zy}}. \tag{7}$$

Problem 2 is concerned with estimating the above unknown synchronization operators. The number of unknown synchronization operators increases as we increase the number of sensors. For example, if there are four sensors, then the number of synchronization operators would be six, i.e., the combinations of two out of four. Interestingly, the unknown synchronization operators are not independent. This can help us in solving Problem 2, as we will discuss in Section V.

III. Methodology: Step 1 - Data-Driven Optimization-Based Event Signature Alignment

In this section, we develop new methods to solve Problem 1 that we introduced in Section II-A. Throughout this section, we assume that the waveform measurements come from two ordinary power quality sensors that are not time-synchronized. Our focus is on identifying a subset of the event signatures from the two power quality sensors to pair and align them to help us estimate the unknown syncrhonization operator.

We propose three different approaches in this section. They are summarized in Table I. In each approach, we first assume that we know parameter $M = |\mathcal{X} \cap \mathcal{Y}|$, i.e., the total number of the pairs of the event signatures that need to be aligned. We then relax this auxiliary assumption to also estimate parameter M itself. Our goal is to ultimately achieve a methodology that is both computationally efficient and robust (in Approach 3).

TABLE I
Summary of Approaches for Event Signature Alignment

Approach	Optimization with Fixed M	Finding M	Computational Efficiency	Robustness
1	Eq. (19)	Eq. (20)	High	Low
2	Eq. (22)	Eq. (23)	Low	High
3	Eq. (19)	Eq. (23)	High	High

A. Metric to Compare Event Signatures

Consider two time-series of the waveform measurements that are captured (through event-triggered waveform recording mechanisms) by two unsynchronized power quality sensors. As we defined in (1), the time-series of the waveform measurements from the first power quality sensor is denoted by vector $\mathbf{x}$ and the time-series of the waveform measurements from the second power quality sensor is denoted by vector $\mathbf{y}$.

Let us derive the *differential waveforms* corresponding to time-series $\mathbf{x}$ and time-series $\mathbf{y}$ as follows [4, Section 4.2.5]:

$$\begin{aligned} \Delta\mathbf{x} &= \mathbf{x}[CT+1:n] - \mathbf{x}[1:n-CT], \\ \Delta\mathbf{y} &= \mathbf{y}[CT+1:n] - \mathbf{y}[1:n-CT], \end{aligned} \tag{8}$$

where T is the sampling rate of the power quality sensors, such as $T = 128$ samples per cycle for all the real-world case studies that we use in this paper. The differential waveforms in (8) are the difference between the original waveform measurements and the delayed versions of the same waveform measurements, where the delay is C cycles. In this paper, we set $C = 3$ cycles to assure separating the duplicate event signatures in differential waveforms; see [4, p. 168]. Comparing event signatures based on differential waveforms (instead of raw waveforms) has many advantages, such as removing the impact of the background load to focus only on the waveform signature that is caused by the event. Whenever needed, we can use the per-unit representation of the waveforms before extracting $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$ to have comparable event signatures.

Events in a correct event pair $\{\mathbf{x}_i, \mathbf{y}_j\}$ should show similar patterns between $\Delta\mathbf{x}_i$ and $\Delta\mathbf{y}_j$; because they are supposed to be captured during the same physical event in the system.

To quantify the similarity between the event signatures in a pair of differential waveforms in $\Delta\mathbf{x}_i$ and $\Delta\mathbf{y}_j$, we define:

$$\sigma_{ij}^{\text{RL}} = \max_{k=1:\lfloor \eta/2 \rfloor - 1} \text{Corr}\big(\Delta\mathbf{x}_i(k:\eta), \Delta\mathbf{y}_j(1:\eta-k+1)\big), \tag{9}$$

and

$$\sigma_{ij}^{\text{LR}} = \max_{k=1:\lfloor \eta/2 \rfloor - 1} \text{Corr}\big(\Delta\mathbf{x}_i(1:\eta-k+1), \Delta\mathbf{y}_j(k:\eta)\big), \tag{10}$$

where $\eta = n - CT$ denotes the length of the time-series of the differential waveforms, and $\text{Corr}(\cdot,\cdot)$ denotes the correlation coefficient between the two time-series. In (9), we compare a *right-moving* window in $\Delta\mathbf{x}_i$ with a *left-moving* window in $\Delta\mathbf{y}_j$, as shown in Fig. 4(a). The correlation coefficient is calculated at each step in such comparison; creating the correlation coefficient time-series in Fig. 4(c). The maximum in such correlation coefficient time-series is denoted by σ_{ij}^{RL}, at the point that is marked with the black dot on this figure.

Similarly, in (10), we compare a left-moving window in $\Delta\mathbf{x}_i$ with a right-moving window in $\Delta\mathbf{y}_j$, as shown in Fig.

4(b). The correlation coefficient is calculated at each step; creating the correlation coefficient time-series in Fig. 4(d). The maximum in such correlation coefficient time-series is denoted by σ_{ij}^{LR}, at the point that is marked on this figure.

In (9), we move $\Delta\mathbf{x}_i$ along $\Delta\mathbf{y}_j$. In (10), we move $\Delta\mathbf{y}_j$ along $\Delta\mathbf{x}_i$. Together, these two comparisons fully examine the similarities between $\Delta\mathbf{x}_i$ and $\Delta\mathbf{y}_j$. In practice, a typical power quality meter records the waveform measurements from $n/2$ samples before the event-trigger point to $n/2$ samples after the event-trigger point [4, p. 147][26, p. 154]. As a result, the recorded event signature always starts at the midpoint of the time-series of the captured waveform in (1). Therefore, we set the range of k in (9) and (10) based on $\eta/2$ instead of η. This reduces computation time and also ensures that correlation coefficient calculation always includes at least $\eta/2$ samples.

We then use the following metric to assess the similarity between the event signature time-series in $\mathbf{x}_i$ and in $\mathbf{y}_j$:

$$\sigma_{ij} = \max\{\sigma_{ij}^{\text{RL}}, \sigma_{ij}^{\text{LR}}\}. \tag{11}$$

This metric indicates the maximum correlation that can be achieved between $\mathbf{x}_i$ and $\mathbf{y}_j$ among all possible alignments between a window with a length of at least $\eta/2$ samples.

Importantly, the maximization in (9) and the maximization in (10) can also provide us with the positions (sample numbers) of the moving windows that result in achieving the highest correlation coefficients. In this regard, we can obtain:

$$k_{ij}^{\text{RL}} = \underset{k=1:\lfloor\eta/2\rfloor-1}{\arg\max}\ \text{Corr}\big(\Delta\mathbf{x}_i(k:\eta), \Delta\mathbf{y}_j(1:\eta-k+1)\big), \tag{12}$$

and

$$k_{ij}^{\text{LR}} = \underset{k=1:\lfloor\eta/2\rfloor-1}{\arg\max}\ \text{Corr}\big(\Delta\mathbf{x}_i(1:\eta-k+1), \Delta\mathbf{y}_j(k:\eta)\big). \tag{13}$$

The values of σ_{ij}^{RL} and k_{ij}^{RL} are marked respectively with the horizontal and the vertical dashed lines in Fig. 4(c). Similarly, the values of σ_{ij}^{LR} and k_{ij}^{LR} are marked respectively with the horizontal and the vertical dashed lines in Fig. 4(d). Similar to the definition of σ_{ij} in (11), we can define k_{ij} as follows:

$$\text{if } \sigma_{ij}^{\text{RL}} \geq \sigma_{ij}^{\text{LR}} \text{ then } k_{ij} = k_{ij}^{\text{RL}} \text{ else } k_{ij} = k_{ij}^{\text{LR}}. \tag{14}$$

Here, k_{ij} is the position of the moving windows that result in σ_{ij}. Next, we propose three approaches to utilize σ_{ij} and k_{ij} to conduct event signature alignment, i.e., to solve Problem 1.

B. Event Signature Alignment - Approach 1

Let b_{ij} denote a *binary* decision variable to indicate whether waveform $\mathbf{x}_i$ and waveform $\mathbf{y}_j$ correspond to the same physical event. This variable can be defined as follows:

$$b_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \text{ and } \mathbf{y}_j \text{ are the same event,} \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Since the pairing of the event signatures is inherently one-on-one (i.e. an event $\mathbf{x}_i$ can only be paired with one event $\mathbf{y}_j$), the binary variables need to comply with the constraints:

$$\sum_j b_{ij} \leq 1, \ \ \forall\, \mathbf{x}_i \in \mathcal{X}, \qquad \sum_i b_{ij} \leq 1, \ \ \forall\, \mathbf{y}_j \in \mathcal{Y}. \tag{16}$$

Next, we note that, although the two power quality sensors are *not* time-synchronized, the local clock at each sensor can help us determine the sequence of the waveform captures at each power quality sensor. For example, from (2), we know that $\mathbf{x}_1$ was captured before $\mathbf{x}_2$, and $\mathbf{x}_2$ was captured before $\mathbf{x}_3$. We must maintain these sequences when we identify which waveform from set $\mathcal{X}$ corresponds to which waveform (if any) from set $\mathcal{Y}$. For instance, it does not make sense to pair $\mathbf{x}_1$ and $\mathbf{y}_4$ while we also pair $\mathbf{x}_3$ and $\mathbf{y}_2$; because this would violate the sequence of the waveforms in sets $\mathcal{X}$ and $\mathcal{Y}$. To address this issue, for each i and each j, we must have:

$$b_{rs} \leq 1 - b_{ij}, \qquad \forall r, s : r < i, \ \ s > j. \tag{17}$$

If $b_{ij} = 1$, i.e., $\mathbf{x}_i$ and $\mathbf{y}_j$ are deemed to correspond to the same event, then (17) becomes $b_{rs} \leq 0$; which means $b_{rs} = 0$ for any pair of $\mathbf{x}_r$ and $\mathbf{y}_s$ for which $r < i$ and $s > j$. If $b_{ij} = 0$, i.e., $\mathbf{x}_i$ and $\mathbf{y}_j$ are *not* deemed to correspond to the same event, then (17) becomes $b_{rs} \leq 1$; which would *not* impose any constraint on binary variable b_{rs} for any r or s.

If we seek to pair exactly M waveform captures across the two unsynchronized power quality sensors, then we must have:

$$\sum_i \sum_j b_{ij} = M, \tag{18}$$

where M is an integer number that is upper bounded by the number of members in set $\mathcal{X}$ and the number of members in set $\mathcal{Y}$. That is, M is upper bounded by $\min\{|\mathcal{X}|, |\mathcal{Y}|\}$.

1) Optimization with Fixed M: First, suppose M is given. Accordingly, we need to pair exactly M waveform captures across the two power quality sensors. For notional simplicity,

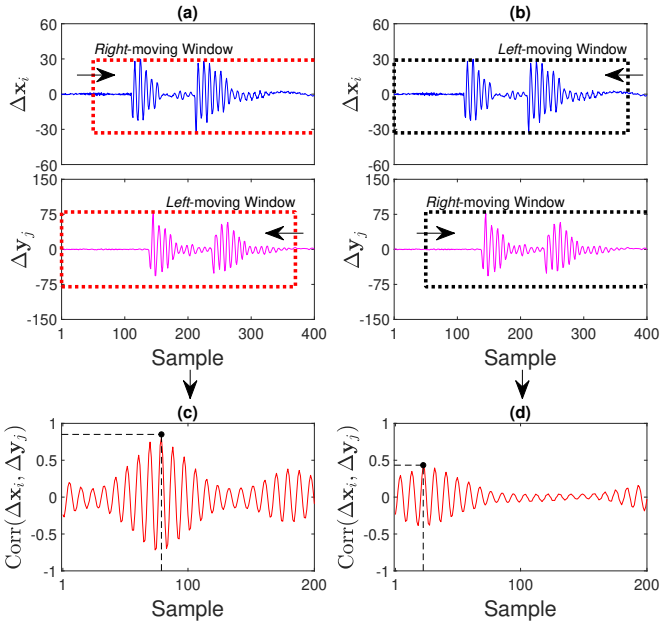

Fig. 4. The moving windows in (a) result in obtaining the time-series of the correlation coefficients in (c), based on the formulations in (9) and (12). The black dot and its dashed lines indicate the values of σ_{ij}^{RL} and k_{ij}^{RL}. Similarly, the moving windows in (b) result in obtaining the time-series of the correlation coefficients in (d), based on the formulations in (10) and (13). The black dot and its dashed lines indicate the values of σ_{ij}^{LR} and k_{ij}^{LR}. Accordingly, we can obtain σ_{ij} and k_{ij} based on (11) and (14).

we stack up b_{ij} for all $\mathbf{x}_i \in \mathcal{X}$ and all $\mathbf{y}_j \in \mathcal{Y}$ into vector $\mathbf{b}$. In this regard, we propose the following optimization problem:

$$\begin{aligned} \underset{\mathbf{b}}{\text{maximize}} \quad & \sum_i \sum_j b_{ij}\sigma_{ij} \\ \text{subject to} \quad & \text{Eqs. (15) and (16) and (17) and (18).} \end{aligned} \tag{19}$$

The above optimization problem is an integer linear program (ILP). It can be solved using solvers such as CPLEX [27]. The goal of the above optimization is to identify M pairs of event signatures across the two power quality sensors to maximize the total correlation among the paired waveform signatures.

It would be insightful to discuss a special case where two unrelated disturbances occur almost at the same time at two different locations. If the two events are sufficiently distanced in time (such as by at least 30 cycles for the SEL power quality meters that we have used in this study), then the two sensors would capture them as two separate events. As a result, each event would be examined separately; just like any other event. However, if the two events are very close to each other in time, then they would fall within a single event-triggered waveform capture. In that case, each sensor would record the two events as a single event, with an inherently *combined* event signature. The shape of the combined event signature at each sensor would depend on the timing of the two events, as well as the electrical distance between the location of each event and the location of each sensor. If the two combined event signatures match each other, then the two combined events would be paired as part of the solution in (19); just like any other pair of events. If the two combined event signatures match each other (such as when one sensor is much closer to the location of one event and the other sensor is much closer to the location of the other event), then the two combined event signatures would not be paired in (19) due to low similarity. In that case, the pair of the combined events from this special case would not be used for the purpose of estimating the synchronization operator. The latter outcome would be acceptable; because we can still use several other pairs of events to achieve time synchronization.

2) Finding M: For a given M, let us denote the optimal objective value of the optimization problem in (19) by $u(M)$. Increasing M results in adding more non-negative terms to the optimal objective value in (19). However, the amount of increase in $u(M)$ diminishes as we increase M; due to the decreasing similarities between the event signatures of the additional pairs of events. To find the right M, we can start from $M = 2$ and we repeatedly solve (19) by incrementing M until we either reach M's upper bound at $\min\{|\mathcal{X}|, |\mathcal{Y}|\}$ or we drop below a threshold on the amount of increase in the optimal objective value of the optimization problem in (19), i.e., until the following stopping criteria is reached:

$$\Phi(M) \triangleq |u(M) - u(M-1)| \leq \epsilon. \tag{20}$$

The value of threshold ϵ can be set based on the minimum required correlation between two event signatures for them to be deemed to correspond to the same physical event.

In practice, there could be disturbances for which the event signatures from different sensors do not align well, such as when two unrelated disturbances happen simultaneously at two different locations. Such events are likely not to be paired when solving the optimization in (19). However, this would be acceptable; because it is permissible if some event signatures are not paired for the ultimate purpose of time synchronization. What is most important in Step 1 is to have a stopping criterion that yields as many correctly aligned pairs of events as possible from the two sensors. This has a direct implication for the process of finding M: it is *not* necessary to always determine the exact value of M in Step 1. While the estimation of M should not exceed its true value, it is acceptable to choose an M that is slightly less than the true value. This is because if the estimation of M exceeds its true value, it will result in the inclusion of incorrect event pairs, thereby compromising the accuracy of the synchronization operator estimation in Step 2.

C. Event Signature Alignment - Approach 2

The method in Approach 1 in Section III-B only utilizes the values of σ_{ij}, which are based on (9)-(11). It does not utilize the values of k_{ij}, which are based on (12)-(14). In this section, we propose Approach 2 to utilize k_{ij} instead of σ_{ij}.

Let κ_i denote the sample number in time-series $\Delta\mathbf{x}_i$ that corresponds to the starting sample of the moving window that resulted in obtaining k_{ij}. Similarly, let κ_j denote the sample number in time-series $\Delta\mathbf{y}_j$ that corresponds to the starting sample of the moving window that resulted in obtaining k_{ij}.

Next, suppose t_i denote the time-stamp of sample number κ_i based on the local clock at the first power quality sensor. Similarly, suppose t_j denote the time-stamp of sample number κ_j based on the local clock at the second power quality sensor. Accordingly, for each pair $\{\mathbf{x}_i, \mathbf{y}_j\}$, we can define:

$$\lambda_{ij} = t_i - t_j. \tag{21}$$

In essence, λ_{ij} is an estimation of δ_{xy} that is obtained solely based on pairing event-triggered time-series $\mathbf{x}_i$ from the first sensor with event-triggered time-series $\mathbf{y}_j$ from the second sensor. Whether or not λ_{ij} is a good estimation of δ_{xy} depends on whether or not $\mathbf{x}_i$ and $\mathbf{y}_j$ correspond to the same event.

1) Optimization with Fixed M: For a given M, we propose to replace (19) with the following optimization problem:

$$\begin{aligned} \underset{\mathbf{b}}{\text{minimize}} \quad & \sum_{i,r} \sum_{j,s} b_{ij}b_{rs} \left|\lambda_{ij} - \lambda_{rs}\right| \\ \text{subject to} \quad & \text{Eqs. (15) and (16) and (17) and (18),} \end{aligned} \tag{22}$$

where the vector of binary optimization variables $\mathbf{b}$ is identical to the one in (19). The constraints in (22) are also identical to the constraints in (19). However, the objective function in (22) is different from the objective function in (19). Notice that while we maximize the objective function in (19), we minimize the objective function in (22). To understand this new formulation, we note that if $\mathbf{x}_i$ and $\mathbf{y}_j$ are paired together and $\mathbf{x}_r$ and $\mathbf{y}_s$ are also paired together, then $b_{ij}b_{rs} = 1$. In that case, the objective function would include $|\lambda_{ij} - \lambda_{rs}|$, which is the difference between the estimation of synchronization operator δ based on pairing $\mathbf{x}_i$ and $\mathbf{y}_j$ and the estimation of synchronization operator δ based on pairing $\mathbf{x}_r$ and $\mathbf{y}_s$.

Problem (22) is an integer nonlinear program (INLP). The non-linearity is due to the multiplication of variable b_{ij} with variable b_{rs}. An INLP can be solved using GUROBI [28].

2) Finding M: For a given M, let us denote the optimal objective value of the optimization problem in (22) by $v(M)$. Increasing M results in adding more non-negative terms to the optimal objective value in (22). The amount of increase in $v(M)$ depends on the amount of discrepancies among the estimations of the synchronization operator that are obtained from different pairs of event signatures by solving the optimization problem in (22). If we exceed the correct number of event signatures that need to be paired, i.e., if we exceed the correct choice of M, then there will be a sudden and major jump in the optimal objective value in (22). To find the right M, we can start from $M = 2$ and repeatedly solve (22) by incrementing M until we either reach M's upper bound at $\min\{|\mathcal{X}|, |\mathcal{Y}|\}$ or we exceed a threshold on the amount of increase in the optimal objective value of the optimization problem in (22), i.e., until the following stopping criteria is reached:

$$\Psi(M) \triangleq |v(M) - v(M-1)| \geq \varepsilon. \tag{23}$$

As we will see in the case studies, the above stopping criteria is robust, i.e., it is *not* sensitive to the choice of ε. The different direction of the inequalities in (23) and (20) is due to the differences between the objective functions in the minimization in (22) versus the maximization in (19).

D. Event Signature Alignment - Approach 3

Approach 1 *is* computationally efficient; because the optimization in (19) is linear. However, Approach 1 is not robust; because the stopping criteria in (20) is sensitive to the choice of its threshold. On the contrary, Approach 2 is not computationally efficient; because the optimization in (22) is nonlinear. However, Approach 2 *is* robust; because the stopping criteria in (23) is not sensitive to the choice of its threshold.

Approach 1 only utilizes the values of σ_{ij} from (9)-(11). Approach 2 only utilizes the values of k_{ij} from (12)-(14).

In this section, we propose a new approach that is both computationally efficient and robust. It utilizes both the values of σ_{ij} from (9)-(11) and the values of k_{ij} from (12)-(14).

1) Optimization with Fixed M: This aspect in Approach 3 is similar to Approach 1. To obtain vector $\mathbf{b}$, we solve the optimization problem in (19). Recall that this optimization problem is linear, and it utilizes σ_{ij} from (9)-(11).

2) Finding M: Once we obtained $\mathbf{b}$ by solving the optimization problem in (19), we use the resulting $\mathbf{b}$ to calculate the objective value based on the optimization problem in (22). To see this, let us denote the solution from (19) by $\mathbf{b}^{(19)}$. Let us also denote the solution from (22) by $\mathbf{b}^{(22)}$. The below list shows the differences and similarities between the stopping criteria in Approach 3 versus in Approach 1 and Approach 2.

- Approach 1: Use $\mathbf{b}^{(19)}$ to obtain $u(M)$ to check Eq. (20).
- Approach 2: Use $\mathbf{b}^{(22)}$ to obtain $v(M)$ to check Eq. (23).
- Approach 3: Use $\mathbf{b}^{(19)}$ to obtain $v(M)$ to check Eq. (23).

Here, we ensure computational efficiency; because we solve the ILP in (19) instead of the INLP in (22). Nevertheless, we still use the formulation of the objective function in (22) to find M in a robust fashion. As a result, Approach 3 inherits the key advantages from both Approach 1 and Approach 2.

To better clarify the differences among the above different approaches, let us compare the solutions from Problem (19) with the solutions from Problem (22). To do so, we need to consider the choice of M. For a given M that is *less* than the true value of M, the solutions from (19) may or may not match the solutions from (22). In principle, it is possible that both solutions provide correctly aligned pairs for the given choice of M; yet, they provide different subsets of the correctly aligned pairs. If the given M is *equal* to the true value of M, then a mismatch between the solution from (19) and the solution from (22) would indicate the presence of at least one incorrect alignment. However, we did not encounter any such case in our various case studies. If the given M is *greater* than the true value of M, then the solutions from (19) and (22) would both include incorrect alignments. In this last scenario, the key question is whether the algorithm would stop due to choice of the stopping criteria in each approach. The stopping criteria in Approaches 2 and 3 based on $v(M)$ are more robust than the stopping criterion in Approach 1 in based on $u(M)$. We will see one such example in a case study in Section VI-B.

IV. Methodology: Step 2 - Statistical Estimation of the Synchronization Operator

Recall from Section III-C that λ_{ij} in (21) provides an estimation of the synchronization operator δ_{xy} based on pairing the event-triggered time-series $\mathbf{x}_i$ from the first power quality sensor with the event-triggered time-series $\mathbf{y}_j$ from the second power quality sensor. However, as we will discuss in this section, there are multiple (often stochastic) factors that can undermine the accuracy of estimating δ_{xy} by using only one pair of time-series $\mathbf{x}_i$ and $\mathbf{y}_j$. Therefore, we need to conduct a statistical analysis based on several pairs of such time series.

A. Factors Affecting the Estimation of Parameter δ

Even if $\mathbf{x}_i$ and $\mathbf{y}_j$ indeed correspond to the same event, we may still experience some discrepancies between the estimated value λ_{ij} and the true value δ_{xy}. Broadly speaking, we have:

$$\begin{aligned}\lambda_{ij} = \delta_{\text{xy}} &+ \text{Difference in Propagation Delay} \\ &+ \text{Error in Data-Driven Event Alignment.}\end{aligned} \tag{24}$$

The difference in propagation delay often exists between any two sensor locations, relative to the location of the event. When an event, such as a fault, occurs, its impact may not be seen at the same time if the electrical distance is not the same between the location of the event and the locations of the two sensors. Importantly, the difference in propagation delay is not fixed. It varies depending on the location of each event.

As for the error in data-driven event alignment, it comes from various sources. Recall from Section III-A that we use correlation to align the event signatures from different sensors. In practice, it is unlikely that we achieve precise alignment, given the data-driven nature of the analysis. For example, although we extract the differential waveforms to minimize the impact of the background load and other unrelated factors, we cannot fully eliminate such impact. Thus, the event signatures are still distorted. Furthermore, the same physical

event can affect the voltage waveforms at each location of the system differently. Hence, the event signatures from different power quality sensors can have major differences; making it practically impossible to exactly align those signatures.

B. Statistical Estimation of Parameter δ

In this section, we propose a statistical approach to estimate δ_{xy}, while considering the challenges that we discussed in Section IV-A. Our goal is to extract δ_{xy} from the following results which come from the methods in Section III:

$$\lambda_{ij}, \quad \forall\{\mathbf{x}_i, \mathbf{y}_j\} \in \mathcal{X} \cap \mathcal{Y}. \tag{25}$$

Suppose we stack up all the above values in one vector, denoted by $\boldsymbol{\lambda}$. The number of rows in $\boldsymbol{\lambda}$ is equal to the cardinality of set $\mathcal{X} \cap \mathcal{Y}$. For example, for the scenario in Fig. 2, and is based on set $\mathcal{X} \cap \mathcal{Y}$ in (3), we have:

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_{2,1} & \lambda_{3,3} \end{bmatrix}^T. \tag{26}$$

In practice, vector $\boldsymbol{\lambda}$ can be a long vector; because set $\mathcal{X} \cap \mathcal{Y}$ can have several members due to pairing several event signatures over a period of several hours or few days.

One option to estimate δ_{xy} based on the expected value:

$$\hat{\delta}_{\text{xy}} = \mathbf{1}^T \boldsymbol{\lambda} \,/\, |\boldsymbol{\lambda}|, \tag{27}$$

where $|\boldsymbol{\lambda}|$ denotes the length of vector $\boldsymbol{\lambda}$ and $\mathbf{1}$ is a vector with the same length as $\boldsymbol{\lambda}$ with all entries equal to one.

To eliminate outliers, we can use the Median Absolute Deviation (MAD) [29], which is calculated as follows:

$$\text{MAD} = \text{median}\left(|\boldsymbol{\lambda} - \text{median}\left(\boldsymbol{\lambda}\right)\mathbf{1}|\right). \tag{28}$$

We can then exclude all entries in $\boldsymbol{\lambda}$ that are more than γMAD away from median($\boldsymbol{\lambda}$), where γ is often set to 1.4826 [29].

C. The Overall Algorithm

The combined process across Step 1 and Step 2 is summarized in Algorithm 1. The input to Algorithm 1 is the time-series of all the event-triggered waveform captures at each of the two unsynchronized sensors. The output from Algorithm 1 is the result of estimating the corresponding synchronization operator. Step 1 is implemented from Line 3 to Line 25. Step 2 is implemented from Line 26 to Line 30. A "for loop" is used from Line 7 to Line 25 to determine M. Inside the for loop, we have implemented Approach 1, Approach 2, and Approach 3, which are separated using "if then" structures. Stopping criteria are implemented in Lines 16 and 22, depending on the approach. The outcome of Step 1 is $\mathbf{b}^\star$ and $M^\star$, which are updated in Line 24 inside the "for loop" every time that M does not trigger the stopping criteria. The final values of $\mathbf{b}^\star$ and $M^\star$ are then used in Step 2 to obtain and refine $\boldsymbol{\lambda}$ in Lines 27 to 30. The algorithm ends by estimating δ_{xy} from $\boldsymbol{\lambda}$.

Running Algorithm 1 inherently requires access to sets $\mathcal{X}$ and $\mathcal{Y}$; which are the inputs to the algorithm in Line 1. Therefore, we need to first accumulate at least a handful of event-triggered waveform measurements before we can run Algorithm 1. In practice, accumulating one or two days of event-triggered waveform measurements is usually sufficient to construct $\mathcal{X}$ and $\mathcal{Y}$, enabling the execution of Algorithm 1.

Algorithm 1 Combining Step 1 and Step 2

1: **Input** : Event sets $\mathcal{X} = \{\mathbf{x}_i\}$ and $\mathcal{Y} = \{\mathbf{y}_j\}$.
2: **Output** : Synchronization operator δ_{xy}.
3: \\ **Step 1**
4: Obtain $\Delta\mathbf{x}_i$ and $\Delta\mathbf{y}_i$ by using (8).
5: Obtain σ_{ij} by using (9)-(11).
6: Obtain k_{ij} by using (12)-(14).
7: **For** $M = 1, \ldots, \min\{|\mathcal{X}|, |\mathcal{Y}|\}$ **Do**
8: **If** Approach 1 or Approach 3 **Then**
9: Obtain $\mathbf{b}$ by solving the maximization in Problem (19).
10: **Else** \\ Approach 2
11: Obtain $\mathbf{b}$ by solving the minimization in Problem (22).
12: **End If**
13: Set $u(M) = \sum_i \sum_j b_{ij}\sigma_{ij}$.
14: Set $v(M) = \sum_{i,r} \sum_{j,s} b_{ij}b_{rs} \left|\lambda_{ij} - \lambda_{rs}\right|$.
15: **If** $M \geq 2$ **Then**
16: **If** Approach 1 **Then**
17: Obtain $\Phi(M)$ by using $u(M)$ and $u(M-1)$ in (20).
18: **If** $\Phi(M) \leq \epsilon$ **Then** Go To Line 25 **End If**
19: **Else** \\ Approach 2 or Approach 3
20: Obtain $\Psi(M)$ by using $v(M)$ and $v(M-1)$ in (23).
21: **If** $\Psi(M) \geq \varepsilon$ **Then** Go To Line 25 **End If**
22: **End If**
23: **End If**
24: Set $\mathbf{b}^\star = \mathbf{b}$ and $M^\star = M$.
25: **End For**
26: \\ **Step 2**
27: Use $\mathbf{b}^\star$ and $M^\star$ to obtain $\mathcal{X} \cap \mathcal{Y}$.
28: Use $\mathcal{X} \cap \mathcal{Y}$ to obtain $\boldsymbol{\lambda}$.
29: Exclude outliers in $\boldsymbol{\lambda}$ by using (28).
30: Obtain and return δ_{xy} by using (27).

As needed, Algorithm 1 can be used on a regular basis to re-calibrate time-synchronization. This can be necessary especially if the local clock at some sensors in unreliable and likely to drift, such as due aging, losing power for several hours, or major change in operational temperature [30]. The interval for recalibration can be set by the operator, such as to be daily, weekly, or monthly, depending on the type and age of the sensors and the operational needs. Recalibration can be set up in different ways. One option is to reset $\mathcal{X}$ and $\mathcal{Y}$ in Line 1 to include only the events that have occurred since the previous recalibration. Another option is to give higher weights to the more recent entries in $\boldsymbol{\lambda}$ when we estimate δ_{xy} in Line 30.

V. Extension to Multiple Sensors

So far, we have assumed that the event-triggered waveform measurements come from only two unsynchronized power quality sensors. In this section, we extend the analysis to an arbitrary number of unsynchronized power quality sensors.

Let $\mathbb{S}$ denote a set that includes all the sets of the time-series of the event-triggered waveform measurements that are captured by all the unsynchronized power quality sensors. For example, for the scenario in Fig. 3, we have:

$$\mathbb{S} = \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}, \tag{29}$$

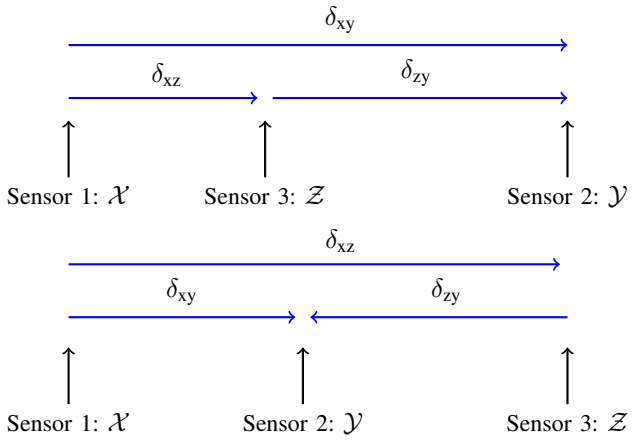

Fig. 5. The relationships among synchronization operators. Two scenarios for the sequence of the waveform measurements from three unsynchronized sensors are shown. In the top figure, the third sensor has a clock that is behind the clock for the first sensor and ahead of the clock for the second sensor. In the bottom figure, the third sensor has a clock that is behind the clocks for the first sensor and the second sensor. In both scenarios, as well as in every other possible sequence among these three sensors, we have: $\delta_{xy} = \delta_{xz} + \delta_{zy}$.

where $\mathcal{X}$ and $\mathcal{Y}$ are defined in (2) and $\mathcal{Z}$ is defined in (4).

We can apply any of the methods in Sections III and IV to the waveform signatures in any two members in set $\mathbb{S}$. Such pairwise analysis will lead to the following results:

$$\hat{\delta}_{xy}, \quad \forall \mathcal{X}, \mathcal{Y} \in \mathbb{S}. \tag{30}$$

For example, for the scenario in Fig. 3, we can obtain:

$$\hat{\delta}_{xy},\ \hat{\delta}_{xz},\ \hat{\delta}_{yz}. \tag{31}$$

However, the above estimations are *not* independent. To see this, first, we note that from (7), we have:

$$\hat{\delta}_{xy} = -\hat{\delta}_{yx}, \quad \hat{\delta}_{xz} = -\hat{\delta}_{zx}, \quad \hat{\delta}_{yz} = -\hat{\delta}_{zy}. \tag{32}$$

Next, we note that

$$\hat{\delta}_{xy} = \hat{\delta}_{xz} + \hat{\delta}_{zy} = \hat{\delta}_{xz} - \hat{\delta}_{yz}, \tag{33}$$

where the second equality is due to (32). The above relationship is illustrated in Fig. 5. The arrangements in this figure are based on the true synchronization operators δ_{xy}, δ_{xz}, and δ_{zy}, where $\delta_{xy} = \delta_{xz} + \delta_{zy}$. A similar arrangement must hold also among the estimated synchronization operators, as in (33).

The relationship in (33) can be generalized as follows:

$$\hat{\delta}_{xy} = \hat{\delta}_{xz} + \hat{\delta}_{zy}, \quad \forall \mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathbb{S}. \tag{34}$$

Next, we propose two methods that are both built upon (34). Both methods can estimate δ between any two unsynchronized power quality sensors. However, the first method seeks to achieve this goal with minimal computation time. Specifically, the first method runs Algorithm 1 for $|\mathbb{S}|-1$ times. Conversely, the second method seeks to take full advantage of the available redundancy to make the results more robust. The second method runs Algorithm 1 for $|\mathbb{S}| \times (|\mathbb{S}| - 1)/2$ times. If the number of sensors is relatively small, then it is recommended to use the second method. If the number of sensors is relatively large, then it is recommended to use the first method.

A. Method without Utilizing Redundancy

Suppose we are working with a large number of unsynchronized sensors, all of which provide event-triggered waveform:

$$\mathbb{S} = \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \ldots, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}. \tag{35}$$

In the first method, we propose the following:

- Run Algorithm 1 between $\mathcal{A}$ and $\mathcal{B}$ to estimate δ_{ab}.
- Run Algorithm 1 between $\mathcal{B}$ and $\mathcal{C}$ to estimate δ_{bc}.

  $\vdots$
- Run Algorithm 1 between $\mathcal{Y}$ and $\mathcal{Z}$ to estimate δ_{yz}.

Thus, we run the algorithm $|\mathbb{S}| - 1$ times to obtain:

$$\hat{\delta}_{ab}, \hat{\delta}_{bc}, \ldots, \hat{\delta}_{yz}. \tag{36}$$

Note that Algorithm 1 can be run in parallel for the above separate estimations. From (36) and (34), we can estimate δ between any two sensors. For example, for the first sensor, we estimated δ_{ab} in (36). Using (34), we can also estimate:

$$\begin{aligned} \hat{\delta}_{ac} &= \hat{\delta}_{ab} + \hat{\delta}_{bc} \\ &\vdots \\ \hat{\delta}_{az} &= \hat{\delta}_{ab} + \hat{\delta}_{bc} + \ldots + \hat{\delta}_{yz}. \end{aligned} \tag{37}$$

B. Method with Utilizing Redundancy

The method in Section V-A has low computational complexity to be applicable to a large number of sensors. Next, we propose another method that leverages redundancy for increased robustness, but with higher computational complexity.

From (34), we have two ways to estimate δ_{xy}:

- Directly from $\mathcal{X} \cap \mathcal{Y}$, and
- Indirectly from $\mathcal{X} \cap \mathcal{Z}$ and $\mathcal{Z} \cap \mathcal{Y}$, for *any* $\mathcal{Z} \in \mathbb{S} \backslash \{\mathcal{X}, \mathcal{Y}\}$.

Together, the above two options can provide us with a total of

$$1 + (|\mathbb{S}| - 2) = |\mathbb{S}| - 1 \tag{38}$$

different estimations for δ_{xy}. This provides us with significant redundancy in our estimation process as we increase the number of unsynchronized power quality sensors. For instance, we can take the following average as the ultimate estimation:

$$\frac{1}{|\mathbb{S}| - 1} \left(\hat{\delta}_{xy} + \sum\nolimits_{\mathcal{Z} \in \mathbb{S} \backslash \{\mathcal{X}, \mathcal{Y}\}} \hat{\delta}_{xz} + \hat{\delta}_{zy} \right). \tag{39}$$

As a special case, we can limit the estimations in (39) to only those events that have been captured by all the available unsynchronized power quality sensors. In that case, our focus would be only on the events that belong to the intersection of all the sets in set $\mathbb{S}$, such as $\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z}$ for the case in (35).

VI. Experiments

In this section, we use real-world waveform measurements to assess the performance of the proposed methods. Three Schweitzer Engineering Laboratories (SEL) 735 Power Quality Meters are used to conduct our experiments. All three sensors are equipped with GPS signal receivers, utilizing the SEL 2401 Satellite-Synchronized Clock to ensure precise time synchronization among the measurements. Throughout the case studies, we are consistently aware of the ground truth

TABLE II
COMPARISON BETWEEN THE PROPOSED METHOD (APPROACH 3) AND THE METHOD IN [25]

M	Reference [25]		Proposed Method	
	Event Pairs	Stop Criteria	Event Pairs	Stop Criteria
1	$\{\mathbf{x}_1, \mathbf{y}_1\}$	0.180	$\{\mathbf{x}_1, \mathbf{y}_1\}$	-
2	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$	0.385	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$	0
3	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$	0.610	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$	0.466
4	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_{21}\}$	0.868	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_{21}\}$	0.483
5	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_{21}\}$ $\{\mathbf{x}_{14}, \mathbf{y}_{22}\}$	1.205	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_{21}\}$ $\{\mathbf{x}_5, \mathbf{y}_{23}\}$	0.490
6	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_{14}\}$ $\{\mathbf{x}_5, \mathbf{y}_{21}\}$ $\{\mathbf{x}_{14}, \mathbf{y}_{22}\}$	1.583 **Stop:** $M = 5$	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_{14}, \mathbf{y}_{14}\}$ $\{\mathbf{x}_{15}, \mathbf{y}_{16}\}$ $\{\mathbf{x}_{18}, \mathbf{y}_{22}\}$	146422 **Stop:** $M = 5$

regarding the timing of all waveform measurement samples and the values of the synchronization operators. Two sensors are installed at 480 V (three-phase), and one sensor is installed at 12.47 kV (three-phase). The sampling rate for all waveform measurements is 128 samples per cycle. Each event capture contains 62 cycles, i.e., 7,936 measurement samples, based on a common setting for the SEL 735. The duration of the events varies, such as from only a quarter of a cycle to several cycles.

A. Comparison with the Method in [25]

In this section, we compare the performance of the proposed method with the state-of-the-art method in [25]. Recall from Section I-C that the method in [25] seeks to solve the same problem as in this paper, albeit it requires the number of unsynchronized power quality sensors to be limited to two. Accordingly, we consider a scenario with two unsynchronized sensors. The first sensor captures 28 events. The second sensor captures 53 events during the same period. We have:

$$|\mathcal{X}| = 28 \quad \text{and} \quad |\mathcal{Y}| = 53. \tag{40}$$

Among the two sensors, there are $|\mathcal{X} \cap \mathcal{Y}| = 5$ events that are paired, i.e., they were captured by both sensors. Set $\mathcal{X} \cap \mathcal{Y}$ comprises the following pairs of event-triggered time-series:

$$\{\mathbf{x}_1, \mathbf{y}_1\}, \{\mathbf{x}_2, \mathbf{y}_2\}, \{\mathbf{x}_3, \mathbf{y}_3\}, \{\mathbf{x}_4, \mathbf{y}_{21}\}, \{\mathbf{x}_5, \mathbf{y}_{23}\}. \tag{41}$$

Table II provides a detailed comparison between the proposed method (Approach 3) and the method in [25]. For both methods, we change M from 1 to 6, and provide the list of the events that are paired and the stopping criteria accordingly.

First, notice that the stopping criteria in both methods correctly identify the number of pairs to be $M = 5 = |\mathcal{X} \cap \mathcal{Y}|$. However, the process to find M is much more robust in the proposed method, where the stop index goes from 0.490 to 146422, versus in the method in [25], where the stop index goes from 1.205 to 1.583. The latter is only a 30% increase while the former is a 300,000 fold increase. Therefore, the proposed method is much more robust in its stopping criteria.

Furthermore, at $M = 5$, the outcome of the event alignment from the method in [25] is *incorrect*, while the outcome of the event alignment from the method in this paper is *correct*. In fact, the method in [25] identified $\{\mathbf{x}_{14}, \mathbf{x}_{22}\}$ as a pair, which is incorrect. This is marked with color red in Table II. The correct pair is rather $\{\mathbf{x}_5, \mathbf{x}_{23}\}$, as we previously saw in (41).

Regarding the last row in Table II, at $M = 6$, the event pairs are marked in gray. This is because the analysis in this row is only an auxiliary analysis that triggers the stopping criteria to identify $M = 5$. Accordingly, the event pairs in this row are not useful. Only the stopping criteria is useful in this row.

It is insightful to compare the runtime between our method (Approach 3) and the method in [25]. The runtime for our method (Approach 3) was 43 minutes. The runtime for the method in [25] was almost 60 days, i.e., almost 2000 times longer. In fact, it was impossible to use a single PC to run the method in [25]. We had to use multiple PCs as well as cloud resources, all in parallel, with a total runtime adding up to 60 days. The much better runtime for our method was due to using the new metric in Section III-A to compare the event signatures, instead of using the metric in Section IV-A in [25].

B. Comparison among Approaches 1, 2, and 3

Next, we compare the three methods that we proposed in Section III. We consider two unsynchronized sensors. The first sensor captures 10 events. The second sensor too captures 10 events during the same period. We have: $|\mathcal{X}| = |\mathcal{Y}| = 10$. Among the two sensors, there are $|\mathcal{X} \cap \mathcal{Y}| = 5$ events that are paired, i.e., they were captured by both sensors. Set $\mathcal{X} \cap \mathcal{Y}$ comprises the following pairs of event-triggered time-series:

$$\{\mathbf{x}_1, \mathbf{y}_1\}, \{\mathbf{x}_2, \mathbf{y}_2\}, \{\mathbf{x}_3, \mathbf{y}_3\}, \{\mathbf{x}_4, \mathbf{y}_8\}, \{\mathbf{x}_5, \mathbf{y}_{10}\}. \tag{42}$$

The above scenario is smaller than the scenario in Section VI-A. This is because Approach 2 is computationally complex, much more so than Approaches 1 and 3. Hence, in order to compare all three methods, we need to consider a smaller case.

Importantly, Approaches 1, 2, and 3 can all provide correct results for event alignment, i.e., they all achieve (42). However, these three methods are different in terms of their runtime and robustness. This is shown in Table III. First, we note that the runtime for Approaches 1 and 3 is very low, while the runtime for Approach 2 is high. For instance, at $M = 5$, the runtime for Approaches 1 and 3 is a small fraction of a second, while the runtime for Approach 2 is 3347 seconds ($\sim$56 minutes).

Next, let us consider the robustness in the stopping criteria. Approach 1 uses the criteria in (20), which requires $\Phi(M)$ to drop below a threshold. To assess robustness, we notice that at $M = 6$, we have: $\Phi(M-1)/\Phi(M) = 0.78/0.04 \approx 20$. Approaches 2 and 3 use the criteria in (23), which requires $\Psi(M)$ to exceed above a threshold. To assess robustness, we notice that at $M = 6$, we have: $\Psi(M)/\Psi(M-1) = 21114/1.23 \approx 17166$ for Approach 2, and $\Psi(M)/\Psi(M-1) = 146422/1.23 \approx 119042$ for Approach 3. Therefore, Approach 2 and Approach 3 are more robust; because they demonstrate a

TABLE III
COMPARING RUNTIME AND ROBUSTNESS ACROSS APPROACHES 1, 2, AND 3.

M	Approach 1			Approach 2			Approach 3		
	Event Pairs	Stop Criteria $\Phi(M)$	Runtime	Event Pairs	Stop Criteria $\Psi(M)$	Runtime	Event Pairs	Stop Criteria $\Psi(M)$	Runtime
1	$\{\mathbf{x}_1, \mathbf{y}_1\}$	-	-	$\{\mathbf{x}_1, \mathbf{y}_1\}$	-	-	$\{\mathbf{x}_1, \mathbf{y}_1\}$	-	-
2	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$	0.93	0.04	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$	0.01	650	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$	0.01	0.05
3	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$	0.91	0.04	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$	0.32	1537	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$	0.32	0.05
4	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$	0.80	0.04	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$	0.77	2442	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$	0.77	0.04
5	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$ $\{\mathbf{x}_5, \mathbf{y}_{10}\}$	0.78	0.05	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$ $\{\mathbf{x}_5, \mathbf{y}_{10}\}$	1.23	3347	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$ $\{\mathbf{x}_5, \mathbf{y}_{10}\}$	1.23	0.05
6	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_7, \mathbf{y}_5\}$ $\{\mathbf{x}_9, \mathbf{y}_6\}$ $\{\mathbf{x}_{10}, \mathbf{y}_7\}$	0.04 Stop:M=5	0.15	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_4, \mathbf{y}_8\}$ $\{\mathbf{x}_5, \mathbf{y}_9\}$ $\{\mathbf{x}_6, \mathbf{y}_{10}\}$	21114 Stop:M=5	4089	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_3, \mathbf{y}_3\}$ $\{\mathbf{x}_7, \mathbf{y}_5\}$ $\{\mathbf{x}_9, \mathbf{y}_6\}$ $\{\mathbf{x}_{10}, \mathbf{y}_7\}$	146422 Stop:M=5	0.33

The unit for runtime is second.

drastically higher change in their metric for stopping criteria; which makes it much easier to find the correct value of M.

Table III also shows the solutions from each method for pairing the events. Recall that Approaches 1 and 3 solve Problem (19) while Approach 2 solves Problem (22). When $M \leq 5$, the event pairing solutions are the same in each row. However, when $M = 6$, which is when the choice of M exceeds the true value of M, the event pairing solutions become different. In fact, all three approaches result in some incorrect event pairs in this last row; however, those incorrect results are not the same. To see this, notice that the last three (gray) pairs at $M = 6$ under Approaches 1 and 3 are different from the last three (gray) pairs at $M = 6$ under Approach 3. As a result, the value of $\Psi(M = 6)$ is different in Approach 2 from the the value of $\Psi(M = 6)$ in Approach 3.

C. Accuracy in Estimating δ

Once the event signatures are aligned, the next step is to estimate the synchronization operator. Consider the scenario in Section VI-B, where $|\mathcal{X}| = |\mathcal{Y}| = 10$ and $|\mathcal{X} \cap \mathcal{Y}| = 5$. The true value for the synchronization operator is set to $\delta_{\text{xy}} = 100$ milliseconds. Using $M = 5$ pairs of aligned event signatures, we obtain the following values for λ_{ij}, all in milliseconds:

$$106.34,\ 105.43,\ 138.23,\ 103.26,\ 120.11. \tag{43}$$

From (27), we can obtain $\hat{\delta}_{\text{xy}} = 114.7$ milliseconds, which has 14.7% error. From (28), and by setting $\gamma = 1.4826$, we can use MAD to exclude two outliers at 138.23 and 120.11 milliseconds. Accordingly, we obtain $\hat{\delta}_{\text{xy}} = 105.0$ milliseconds, which has 5% error. Thus, with as few as only five aligned event signatures, we can achieve reasonable accuracy in estimating δ. In practice, one can use event-triggered waveform measurements over multiple days; thus increasing the number of pairs of aligned waveform signatures. This can further improve accuracy due to increased redundancy in data.

The percentage of error in estimating δ can significantly reduce (improve) if the true value of δ increases. For example, if δ is 1 second, then the percentage error can be much less than 1%. This is because the amount of error in estimating δ is almost fixed, regardless of the value of δ. Such error depends on two types of factors, as we listed in (24) in Section IV-A. Both of those factors are within microseconds to few milliseconds. Therefore, for a higher δ, the value of $|\delta - \hat{\delta}|$ reduces in percentage; resulting in a smaller percentage error.

D. Importance of Sequence Constraints

The optimization-based methods in Section III use various constraints. The constraint in (17) is distinctly subtle. It holds the sequence of the events at each unsynchronized sensor.

In this section, we investigate the importance of including the constraint in (17) as part of the optimization-based method in Approach 3. The results are shown in Table IV. Here, we repeat the case study in Section VI-A for Approach 3, but this time we remove constraint (17) from the optimization problem formulation in (19). Interestingly, removing constraint (17) can significantly reduce the runtime in Approach 3. Recall from Section VI-A that Approach 3 *with* constraint (17) takes several minutes to run. However, the runtime for Approach 3 *without* constraint (17) is only a fraction of a second.

Nevertheless, removing constraint (17) from the optimization problem in (19) can lead to incorrect results. This is evident from the column on event pairs and the column on stop criteria in Table IV. Not only some of the paired events are incorrect, but also the choice of parameter M is incorrect. The incorrect results are marked in red in Table IV. Therefore, it is necessary to include the sequence constraints in (17) in the proposed optimization-based methods in Section III.

TABLE IV
APPROACH 3 WITHOUT THE SEQUENCE CONSTRAINTS IN (17)

M	Event Pairs	Stop Criteria	Runtime
1	-	-	-
2	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$	0.00	0.25
3	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_4, \mathbf{y}_{21}\}$	0.47	0.10
4	$\{\mathbf{x}_1, \mathbf{y}_1\}$ $\{\mathbf{x}_2, \mathbf{y}_2\}$ $\{\mathbf{x}_{14}, \mathbf{y}_{22}\}$ $\{\mathbf{x}_4, \mathbf{y}_{21}\}$	48792 **Stop:** $M = 3$	0.10

TABLE V
ALIGNED EVENTS AND ESTIMATION RESULTS FROM THREE SENSORS AS IN (45)

Aligned Event Signatures	λ_{ij} in milliseconds		
	$\{\mathbf{x}, \mathbf{y}\}$	$\{\mathbf{x}, \mathbf{z}\}$	$\{\mathbf{y}, \mathbf{z}\}$
$\{\mathbf{x}_3, \mathbf{y}_4, \mathbf{z}_1\}$	60.22	101.74	41.52
$\{\mathbf{x}_5, \mathbf{y}_6, \mathbf{z}_2\}$	60.13	117.72	57.46
$\{\mathbf{x}_7, \mathbf{y}_9, \mathbf{z}_4\}$	60.22	98.44	38.17
$\{\mathbf{x}_{10}, \mathbf{y}_{12}, \mathbf{z}_5\}$	60.22	131.76	71.63
$\{\mathbf{x}_{11}, \mathbf{y}_{13}, \mathbf{z}_6\}$	60.17	131.80	71.63
$\{\mathbf{x}_{12}, \mathbf{y}_{14}, \mathbf{z}_7\}$	60.17	114.47	54.29
$\{\mathbf{x}_{14}, \mathbf{y}_{16}, \mathbf{z}_9\}$	59.35	95.48	36.05
$\{\mathbf{x}_{15}, \mathbf{y}_{17}, \mathbf{z}_{10}\}$	74.60	112.51	37.91
$\{\mathbf{x}_{16}, \mathbf{y}_{18}, \mathbf{z}_{11}\}$	58.35	95.92	37.52
$\{\mathbf{x}_{17}, \mathbf{y}_{19}, \mathbf{z}_{12}\}$	74.68	110.21	35.40

E. Impact of Data-Driven Factors

As we discussed in Section IV-A, several factors can affect the data-driven estimation of the synchronization operator. For a given pair of correctly aligned event signatures $\{\mathbf{x}_i, \mathbf{y}_j\}$, we almost always have some mismatch between λ_{ij} in (21) and the true value of δ_{xy}. This was previously discussed in (24).

To see this, again consider the five pairs of event-triggered time-series in (42) that we previously discussed in Section VI-B. The corresponding values of λ_{ij} are obtained as:

$$\begin{aligned}
&\lambda_{1,1} \text{ based on } \{\mathbf{x}_1, \mathbf{y}_1\} \text{ is } 6.34 \text{ milliseconds},\\
&\lambda_{2,2} \text{ based on } \{\mathbf{x}_2, \mathbf{y}_2\} \text{ is } 5.43 \text{ milliseconds},\\
&\lambda_{3,3} \text{ based on } \{\mathbf{x}_3, \mathbf{y}_3\} \text{ is } 32.23 \text{ milliseconds},\\
&\lambda_{4,8} \text{ based on } \{\mathbf{x}_4, \mathbf{y}_8\} \text{ is } 3.26 \text{ milliseconds},\\
&\lambda_{5,10} \text{ based on } \{\mathbf{x}_5, \mathbf{y}_{10}\} \text{ is } 20.11 \text{ milliseconds}.
\end{aligned} \tag{44}$$

In theory, the above values must be equal; since they all estimate the synchronization operator between the same two sensors. However, in practice, these values are different. The differences are due to the various data-driven factors that we explained in Section IV-A. Using (27), we can combine all the results in (44) to obtain $\hat{\delta}_{\text{xy}} = 14.67$ milliseconds.

F. A Case with More Than Two Sensors

In this section, we use real-world waveform measurements from three sensors. The first sensor captures 17 events. The second sensor captures 19 events. The third sensor captures 12 events. Accordingly, we can construct set $\mathbb{S}$ as in (35), where:

$$|\mathcal{X}| = 17,\ |\mathcal{Y}| = 19,\ |\mathcal{X}| = 12,\ |\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z}| = 10. \tag{45}$$

The true values for the synchronization operator are as follows, all in milliseconds: $\delta_{\text{xy}} = 60$, $\delta_{\text{xz}} = 100$, and $\delta_{\text{yz}} = 40$.

Table V shows the results. The first column shows all the event signatures from the three sensors that are aligned with each other by Approach 3. The second column shows the data-driven estimation of λ_{ij} for the aligned events between the first and the second sensors. The third column shows the data-driven estimation of λ_{ij} for the aligned events between the first and the third sensors. The last column shows the data-driven estimation of λ_{ij} for the aligned events between the second and the third sensors. The definition of λ_{ij} is given (21).

By using the analysis in (28), we remove several outliers for each pair of sensors, including four entries of λ_{ij} for $\{\mathbf{x}, \mathbf{y}\}$, two entries of λ_{ij} for $\{\mathbf{x}, \mathbf{z}\}$, and four entries of λ_{ij} for $\{\mathbf{y}, \mathbf{z}\}$.

First, suppose we use the method without redundancy, as in Section V-A. This method would require running Algorithm 1 by $|\mathbb{S}| - 1 = 2$ times. From (27), we can use the results in Table V to estimate δ for $|\mathbb{S}| - 1 = 2$ pairs of sensors:

$$\hat{\delta}_{\text{xy}} = 60.19, \quad \hat{\delta}_{\text{yz}} = 37.76. \tag{46}$$

From (46), together with (37), we can also obtain

$$\hat{\delta}_{\text{xz}} = \hat{\delta}_{\text{xy}} + \hat{\delta}_{\text{yz}} = 97.95. \tag{47}$$

Next, suppose we use the method with redundancy, as in Section V-B. This method would require running Algorithm 1 by $|\mathbb{S}| \times (|\mathbb{S}| - 1)/2 = 3$ times. Two runs of Algorithm 1 would provide the same estimations as in (46). The third run, i.e., the additional run, will use Table V to further provide:

$$\hat{\delta}_{\text{xz}} = 105.81. \tag{48}$$

From (39), we combine the results in (46) and (48) to obtain:

$$(105.81 + 97.95)/2 = 101.88. \tag{49}$$

By comparing (47) and (49), We can see that both methods from Section V provide acceptable results for $\hat{\delta}_{\text{xz}}$; however, the results from the second method in (49) is more accurate than the results from the first method in (47) in this case.

G. Example Application: Monitoring DER Dynamics

Time-synchronizing waveform measurements from ordinary power quality sensors can be used in a wide range of applications in power systems, especially in power distribution systems where access to GPS-synchronized waveform measurements is limited. One such application is in studying the dynamic behavior of inverter-based distributed energy resources (DERs) in response to various disturbances in power systems. Such analysis can shed light on how different events may potentially disrupt the operation of DERs. This will allow distribution system operators to plan accordingly.

Consider the examples in Fig. 6. Here we examine time-synchronized waveforms from two power quality meters at two PV units during four different disturbances. Both PV units are DERs. They are interconnected at 480 V at two different locations. The inverters of the two PV units are from the same manufacturer; however, the two PV units have different sizes, with up to 100 kW and 200 kW rated generation capacity, respectively. The disturbances are as follows:

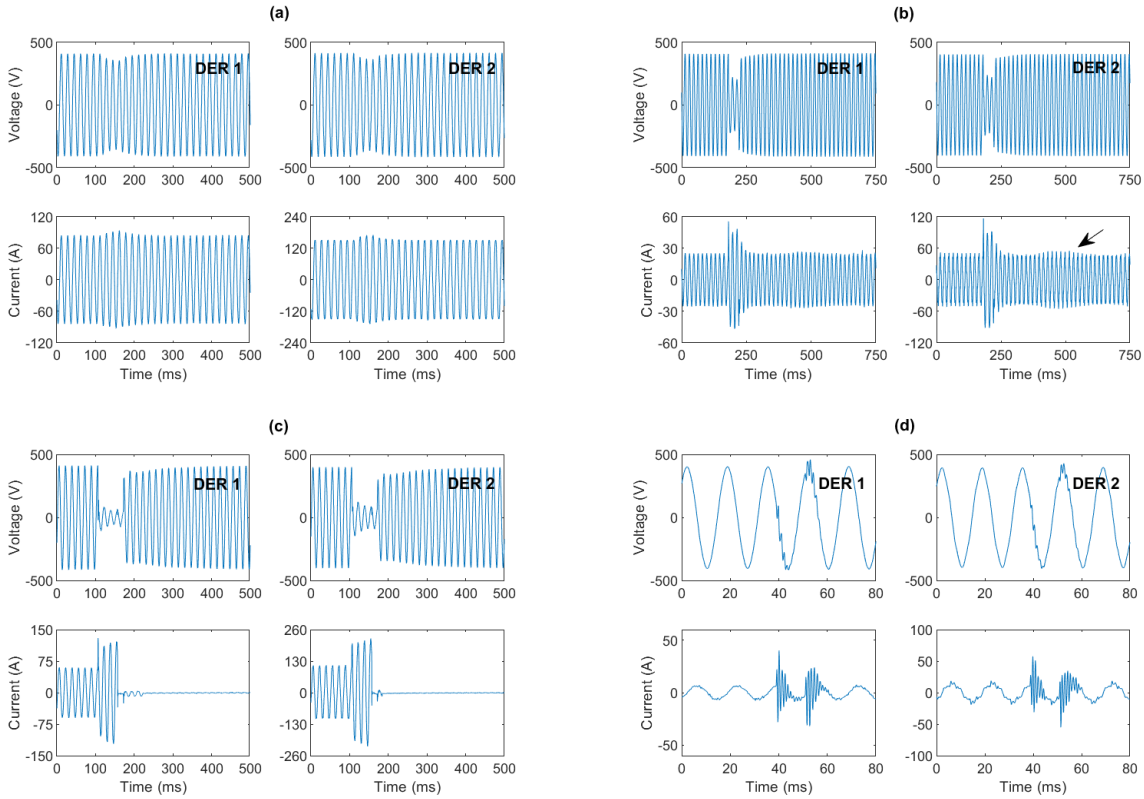

Fig. 6. Application of time-synchronized waveform measurements in benchmarking the behavior of DERs on power distribution systems during: (a) a voltage sag; (b) a fault where the DERs managed to ride through the fault; (c) a fault where the DERs could not ride through the fault; (d) sub-cycle oscillations.

- Voltage Sag: Fig. 6(a) shows the simultaneous responses of the two PV units to a voltage sag. Both PV units momentarily experience slight increases in their injected current; with no major impact on their normal operation.
- Fault *with* Ride-through: Fig. 6(b) shows the simultaneous responses of the two PV units to a fault in the power system, where both PV units experience significant agitation. Nevertheless they both manage to *ride through* the fault. Notice the "bump" in the dynamic response of DER 2 that is marked with an arrow. This shows the visible presence of low-frequency modulated oscillations at about 5 Hz that are likely developed in the inverter's phase locked loop control system due to the disturbance that the fault caused in the inverter's terminal voltages; e.g., see [31]. A similar phenomena also appears in the dynamic response of DER 1 to the same fault; but it is less significant (and harder to see) in DER 1.
- Fault *without* Ride-through: Fig. 6(c) shows the simultaneous responses of the two PV units to another fault. Both PV units *trip* during the fault. However, they do *not* trip exactly at the same time. They also demonstrate slightly different dynamic behavior during their tripping.
- Oscillations: Fig. 6(d) shows the simultaneous damping oscillations that occurred at the two PV units in response to another fault in the power system. The frequency of the oscillations at both PV units is around 1 kHz.

These and other similar analysis can be used not only to model the dynamic behavior of DERs, but also to conduct benchmarking and health monitoring of DERs. For instance, one can compare how different DERs (from the same vendor or from different vendors) within a certain distribution or sub-transmission system respond to the same disturbances. Such comparison can lead to early detection of DER malfunctions. It can also be used to fine-tune the control parameters of the DERs by using the desirable responses of other similar DERs as *reference* points. The results will also directly help network operators to better predict and prepare for the dynamic responses of DERs to various disturbances as the penetration of DERs continues to grow in their network.

Time-synchronizing waveform measurements can also support other applications in distribution systems, such as event and fault location identification on distribution feeders, wildfire monitoring in distribution networks, and power distribution system protection; e.g., see the studies in [5], [8], [12].

VII. Conclusions and Future Work

Novel data-driven optimization-based methods were proposed to transform the output of conventional unsynchronized event-triggered waveform sensors into synchro-waveforms. By developing a series of techniques, we achieved methods that are accurate, robust, and computationally efficient.

The proposed methods were tested by real-world waveform measurements from field sensors in a test site in California.

The methods in this paper can provide utilities with the practical means to benefit from the increasing applications of synchro-waveforms without the need to upgrade their

existing conventional waveform sensors. Since there are tens of thousands of unsynchronized waveform sensors in operation, such as power quality meters and fault recorders, the practical impact of the proposed methods can potentially be significant.

Since, by construction, the measurements from each event-triggered waveform capture contain an event, one should focus on analyzing the event signatures for the purpose of data-driven time synchronization. However, if, in the future, the legacy power quality meters are upgraded to capture waveforms more frequently, even in the absence of an event, one could also consider taking advantage of those additional "ambient" measurements to potentially enhance the performance in data-driven time synchronization.

REFERENCES

[1] H. Mohsenian-Rad and W. Xu, "Synchro-waveforms: A window to the future of power systems data analytics," *IEEE Power and Energy Magazine*, vol. 21, no. 5, pp. 68–77, Sep. 2023.

[2] W. Xu, Z. Huang, X. Xie, and C. Li, "Synchronized waveforms – a frontier of data-based power system and apparatus monitoring, protection, and control," *IEEE Trans. on Power Delivery*, vol. 37, no. 1, pp. 3–17, 2022.

[3] Alvaro Furlani Bastos, S. Santoso, W. Freitas, and W. Xu, "Synchrowaveform measurement units and applications," in *Proc. of the IEEE PES General Meeting*, Atlanta, GA, Jul. 2019.

[4] H. Mohsenian-Rad, *Smart Grid Sensors: Principles and Applications*. Cambridge University Press, UK, Apr. 2022.

[5] H. Mohsenian-Rad, A. Shahsavari, and M. Majidi, "Analysis of power quality events for wildfire monitoring: Lessons learned from a California wildfire," in *Proc. of the IEEE PES ISGT*, San Juan, Puerto Rico, 2023.

[6] https://ieee-synchrowaveform.engr.ucr.edu/.

[7] M. Mansour Lakouraj, H. Hosseinpour, H. Livani, and M. Benidris, "Waveform measurement unit-based fault location in distribution feeders via short-time matrix pencil method and graph neural network," *IEEE Trans. on Industry Applications*, vol. 59, no. 2, pp. 2661–2670, 2023.

[8] M. Izadi and H. Mohsenian-Rad, "Synchronous waveform measurements to locate transient events and incipient faults in power distribution networks," *IEEE Trans. on Smart Grid*, vol. 12, pp. 4295–4307, 2021.

[9] I. Khan, H. Sun, K. Kim, J. Guo, and D. Nikovski, "Combined detection and localization model for high impedance fault under noisy condition," in *Proc. of the IEEE PES General Meeting*, Orlando, FL, Jul. 2023.

[10] T. Zaman *et al.*, "Multimode synchronous resonance detection in converters dominated power system using synchro-waveforms," in *Proc. of the Int. Conference on Electricity Distribution*, Rome, Italy, Jun 2023.

[11] N. Ehsani, F. Ahmadi-Gorjayi, Z.-J. Ye, A. McEachern, and H. Mohsenian-Rad, "Sub-cycle event detection and characterization in continuous streaming of synchro-waveforms: An experiment based on gridsweep measurements," in *Proc. of the IEEE North American Power Symposium*, Asheville, NC, Oct 2023.

[12] B. Sharma, M. K. Jena, and S. Dash, "Assessment of signal similarity index based line protection scheme for systems with inverter based resources," *IEEE Systems Journal*, vol. 17, no. 3, 2023.

[13] H. Mohsenian-Rad, M. Kezunovic, and F. Rahmatian, "Synchro-waveforms in wide-area monitoring, control, and protection: Real-world examples and future opportunities," *IEEE Power and Energy Magazine*, vol. 22, no. 6, Nov. 2024.

[14] F. Ahmadi-Gorjayi and H. Mohsenian-Rad, "Data-driven models for sub-cycle dynamic response of inverter-based resources using WMU measurements," *IEEE Trans. on Smart Grid*, pp. 4125–4128, Sep 2023.

[15] H. Mohsenzadeh-Yazdi, F. Ahmadi, and H. Mohsenian-Rad, "Sub-cycle dynamics modeling of ibrs using lstm methods and synchro-waveform measurements," in *Proc. of the IEEE PES General Meeting*, Jul. 2024.

[16] M. Alberto *et al.*, "Newly implemented real-time PQ monitoring for transmission 4.0 substations," *Electric Power Systems Research*, vol. 204, p. 107709, Mar 2022.

[17] Z. Dai and J. E. Tate, "Emulating synchrophasor frequency measurements with transient stability simulation," *IEEE Trans. on Power Systems*, vol. 36, no. 5, pp. 4066–4074, 2021.

[18] M. MansourLakouraj, H. Hosseinpour, H. Livani, and M. Benidris, "Waveform measurement unit-based fault location in distribution feeders via short-time matrix pencil method and graph neural network," *IEEE Trans. on Industry Applications*, vol. 59, no. 2, pp. 2661–2670, 2023.

[19] Z. Long, F. Lin, F. Zhou, W. Li, Y. Liu, J. Fan, K. Hu, F. Li, and S. Xie, "Development and traceability method of a 100 ka reference impulse current measuring system," *IEEE Trans. on Industry Applications*, vol. 59, no. 2, pp. 2246–2253, 2023.

[20] R. F. Buzo, H. M. Barradas, and F. B. Leão, "A new method for fault location in distribution networks based on voltage sag measurements," *IEEE Trans. on Power Delivery*, vol. 36, no. 2, pp. 651–662, 2021.

[21] K. R. Mestav, X. Wang, and L. Tong, "A deep learning approach to anomaly sequence detection for high-resolution monitoring of power systems," *IEEE Trans. on Power Systems*, vol. 38, no. 1, pp. 4–13, 2023.

[22] Q. Li, S. Liu, and H. Chaoui, "A reinforcement learning based model-free wide-area damping control under random pmu time delays," in *Proc of the IEEE International Symposium on Industrial Electronics*, 2021.

[23] P. Castello, G. Gallus, P. Pegoraro, and S. Sulis, "Measurement platform for latency characterization of wide area monitoring, protection and control systems," *IEEE Trans. on Inst. and Meas.*, vol. 73, 2024.

[24] Z.-J. Ye, M. Izadi, M. Farajollahi, and H. Mohsenian-Rad, "A remedy to losing time synchronization at D-PMUs, H-PMUs, and WMUs in event location identification in power distribution systems," *IEEE Trans. on Smart Grid*, vol. 15, no. 1, pp. 651–654, 2024.

[25] Z.-J. Ye and H. Mohsenian-Rad, "A data-driven time-synchronization method to convert conventional power quality waveform measurements into synchro-waveforms," in *Proc. of the IEEE PES SGSMA*, May 2024.

[26] Schweitzer Engineering Laboratories, "Instruction Manual: SEL-735 Power Quality and Revenue Meter," 2024.

[27] https://www.ibm.com/products/ilog-cplex-optimization-studio.

[28] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2023. [Online]. Available: https://www.gurobi.com

[29] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiley, 2009.

[30] H. Zhou, "Adaptive high-accuracy timing module: Algorithms and performance bounds - MS Thesis," Ottawa, Canada, Nov. 2009.

[31] Y. Cheng *et al.*, "Real-world subsynchronous oscillation events in power grids with high penetrations of inverter-based resources," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 316–330, 2023.

Zong-Jhen Ye (S'20) received M.S. degree in environmental engineering from National Cheng Kung University, Tainan, Taiwan, and the second M.S. degree in electrical engineering from National Taipei University of Technology, Taipei, Taiwan. He is currently working toward the Ph.D. degree in electrical engineering at University of California, Riverside, CA, USA. His research interests are the applications of optimization techniques on transmission lines and distribution systems.

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 - 2022. He was the recipient of the National Science Foundation (NSF) CAREER Award, the Best Paper Award from the IEEE Power and Energy Society General Meeting, the Best Paper Award from the IEEE Conference on Smart Grid Communications, and a Technical Achievement Award from the IEEE Communications Society. He has been the PI or co-PI on sixteen million dollars research grants in the area of smart grid. He has served as Editor for the IEEE TRANSACTIONS ON POWER SYSTEMS, IEEE TRANSACTIONS ON SMART GRID and the IEEE POWER ENGINEERING LETTERS.