Smart Inverter Grid Probing for Learning Loads: Part I – Identifiability Analysis

Siddharth Bhela, *Student Member, IEEE,* Vassilis Kekatos, *Senior Member, IEEE,* and Sriharsha Veeramachaneni

Abstract—Distribution grids currently lack comprehensive real-time metering. Nevertheless, grid operators require precise knowledge of loads and renewable generation to accomplish any feeder optimization task. At the same time, new grid technologies, such as solar photovoltaics and energy storage units are interfaced via inverters with advanced sensing and actuation capabilities. In this context, this two-part work puts forth the idea of engaging power electronics to probe an electric grid and record its voltage response at actuated and metered buses, to infer non-metered loads. Probing can be accomplished by commanding inverters to momentarily perturb their power injections. Multiple probing actions can be induced within a few tens of seconds. In Part I, load inference via grid probing is formulated as an implicit nonlinear system identification task, which is shown to be topologically observable under certain conditions. The conditions can be readily checked upon solving a max-flow problem on a bipartite graph derived from the feeder topology and the placement of probed and non-metered buses. The analysis holds for single- and multi-phase grids, radial or meshed, and applies to phasor or magnitude-only voltage data. Using probing to learn non-constant-power loads is also analyzed as a special case.

Index Terms—Smart inverters, topological observability, Jacobian matrix, generic rank, distribution grids, ZIP loads.

I. INTRODUCTION

Low-voltage distribution grids have been plagued with limited observability, due to limited instrumentation, low investment interest in the past, and their sheer extent [1]. Traditionally, utility operators monitor distribution grids by collecting measurements infrequently and only from a few critical buses. This mode of operation has been functional due to the under-utilization of distribution grids and the availability of historical data. Nevertheless, with the advent of distributed energy resources (DERs), electric vehicles, and demand-response programs, there is a critical need to reliably estimate the system state and learn non-metered loads to optimally dispatch the grid on a frequent basis (say 20 min). To this end, the communication capabilities of grid sensors together with the actuation and sensing features of power inverters found in solar panels, energy storage units, and electric vehicles could be utilized toward unveiling loads.

Although estimating loads or the grid state has heavily relied on pseudo-measurements, such measurements may not

be available or accurate under the current mode of operation [2], [3]. On the other hand, the widespread deployment of digital relays, phasor measurement units (PMUs), and inverter-interfaced DERs provide excellent opportunities for improving distribution grid observability [4], [5]. In addition, regular polling and on-demand reads of customer loads and voltages via smart meters have enhanced the accuracy of distribution system state estimation [6], [7]. Ignoring network information, a kernel-based scheme for learning loads is reported in [8]. Since the previous schemes collect data on a hourly basis, they are of limited use for real-time optimization.

Rather than passively collecting grid readings to infer non-metered loads, this works advocates engaging inverters to probe the grid and thus actively collect feeder data. We define *probing* as the technique of perturbing an electric grid for the purpose of finding unknown parameters. The idea of probing has been previously suggested towards estimating the electro-mechanical oscillation modes in power transmission systems [9], [10]. Perturbing the voltage and/or current of a single inverter has been adopted in the power electronics community to determine the grid-equivalent Thevenin impedance of inverters [11]. Moreover, modulating the primary droop control loop of inverters has been recently suggested for learning loads and topologies in direct-current grids [12]. Graph algorithms and identifiability conditions for recovering feeder topologies using inverter probing data have been devised in [13], [14].

Beyond their standard energy conversion functionality, smart inverters are being utilized for reactive power control and other feeder optimization tasks [15]. In fact, the grid voltage response to inverter injection changes has been used as a means to solve optimal power flow tasks in a decentralized and/or communication-free fashion; see for example [16], [17], [18], [19], [20]. Leveraging exactly this voltage response, grid probing attributes smart inverters a third functionality towards monitoring rather than grid control.

The contribution of Part I of this work is on three fronts. First, we formulate our *Probing-to-Learn* (P2L) technique in Section II. Exploiting the stationarity of non-metered loads during probing and assuming noiseless inverter readings in Part I, the P2L problem is posed as a coupled power flow task. Second, we provide intuitive and easily verifiable graph-theoretic conditions under which probing succeeds in finding non-metered loads under phasor (Section III) and non-phasor data (Section IV). Lastly, Section V extends probing to infer non-constant-power (ZIP) loads.

The results of Part I significantly extend our previous work of [21] in four directions: *i)* The analysis extends non-trivially to multiple rather than only two probing actions; *ii)* Probing setups with voltage magnitude and/or angle data are studied in

a unified fashion; *iii)* We use the feeder connectivity to upper bound the number of probing actions beyond which there is no identifiability benefit; and *iv)* A new proving technique generalizes the analysis from radial to meshed grids, thus covering the timely topic of loopy multiphase distribution grids and transmission systems.

Regarding *notation*, column vectors (matrices) are denoted by lower- (upper-) case boldface letters and sets by calligraphic symbols. The cardinality of set $\mathcal{X}$ is denoted by $|\mathcal{X}|$, and its complement by $\bar{\mathcal{X}}$. The operators $(\cdot)^\top$ and $(\cdot)^H$ stand for (complex) transposition; the floor and ceiling functions are denoted by $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$; $\mathrm{dg}(\mathbf{x})$ defines a diagonal matrix having $\mathbf{x}$ on its main diagonal; and $\mathbf{I}_N$ is the $N \times N$ identity matrix. The notation $\mathbf{x}_\mathcal{A}$ denotes the sub-vector of $\mathbf{x}$ indexed by $\mathcal{A}$; and $\mathbf{X}_{\mathcal{A},\mathcal{B}}$ is the matrix obtained by sampling the rows and columns of $\mathbf{X}$ indexed respectively by $\mathcal{A}$ and $\mathcal{B}$.

II. GRID PROBING

Albeit not every bus is metered in a distribution grid, some buses are equipped with sensors recording voltage magnitudes and/or angles, actual powers, and power factors. Moreover, the power injections in solar panels and energy storage devices can be instantly controlled using advanced power electronics. Building on the physical law that perturbing power injections at different buses is reflected on voltage changes across the grid, the key idea here is to engage power electronics to probe the grid with the purpose of learning non-metered loads.

To formally describe grid probing, let us briefly review a feeder model. Consider a feeder represented by a graph $\mathcal{G} = (\mathcal{N}^+, \mathcal{L})$ where the nodes in $\mathcal{N}^+ := \{0, \dots, N\}$ correspond to buses, and the edges in $\mathcal{L}$ to distribution lines. Let $\mathbf{Y} := \mathbf{G} + j\mathbf{B}$ be the grid bus admittance matrix and $\mathbf{G}$ (resp. $\mathbf{B}$) be the bus conductance (resp. susceptance) matrix. By definition, the entries B_{nm} and G_{nm} for $n \neq m$ are non-zero only if $(n, m) \in \mathcal{L}$. Let us express the voltage phasor at bus $n \in \mathcal{N}^+$ in Cartesian and polar coordinates as

$$v_n = v_{r,n} + j v_{i,n} = u_n e^{j\theta_n}.$$

The substation is indexed by $n = 0$, its voltage remains fixed at $1 + j0$, and the remaining buses comprise the set $\mathcal{N}$. If $\mathbf{v}_r := [v_{r,0} \; \cdots \; v_{r,N}]^\top$ and $\mathbf{v}_i := [v_{i,0} \; \cdots \; v_{i,N}]^\top$, define the system state as $\mathbf{v} := [\mathbf{v}_r^\top \; \mathbf{v}_i^\top]^\top$. Apparently, for each bus $n \in \mathcal{N}^+$, the squared voltage magnitude and the net power injections are quadratic functions of $\mathbf{v}$, whereas the voltage angle is a trigonometric function of $\mathbf{v}$ [22, Ch. 3]

$$u_n(\mathbf{v}) = u_n^2 = v_{r,n}^2 + v_{i,n}^2 \tag{1a}$$

$$p_n(\mathbf{v}) = v_{r,n} \sum_{m=0}^{N} (v_{r,m} G_{nm} - v_{i,m} B_{nm})$$
$$+ v_{i,n} \sum_{m=0}^{N} (v_{r,m} B_{nm} + v_{i,m} G_{nm}) \tag{1b}$$

$$q_n(\mathbf{v}) = v_{i,n} \sum_{m=0}^{N} (v_{r,m} G_{nm} - v_{i,m} B_{nm})$$
$$- v_{r,n} \sum_{m=0}^{N} (v_{r,m} B_{nm} + v_{i,m} G_{nm}) \tag{1c}$$

$$\theta_n(\mathbf{v}) = \arctan\left(\frac{v_{r,n}}{v_{i,n}}\right). \tag{1d}$$

With the proliferation of grid sensors and inverters, the distribution grid operator may have access to all four quantities $(u_n, \theta_n, p_n, q_n)$ on a subset of buses. Different from the conventional power flow (PF) setup with PQ and PV buses, we partition $\mathcal{N}^+$ into the subsets:

- The set $\mathcal{M}$ of metered buses for which $(u_n, \theta_n, p_n, q_n)$ are known and their power injections are possibly controllable. This set includes the substation and buses equipped with smart sensors and/or inverters. Its cardinality is denoted by $M := |\mathcal{M}|$.
- The set $\mathcal{O}$ of non-metered buses where no information is available. Its cardinality is denoted by $O := |\mathcal{O}|$, and apparently, $N + 1 = M + O$.

The inverters interfacing DERs are typically modeled as constant-power generators [15], [16], [18]: Internal control loops can reach setpoints for (re)-active power injections within microseconds. The setpoints should comply with solar irradiance and the rating of the inverter.

Remark 1. *We emphasize $p_n + jq_n$ is the net complex injection. If bus n hosts a smart inverter and a non-controllable load, it is henceforth assumed that the operator measures $p_n + jq_n$ and the voltage at the point of common coupling, and controls the complex injection from the inverter. This assumption is reasonable since smart inverters are usually equipped with sensors; e.g., the Pecan Street project measures both the net and inverter injections [23].*

Given the feeder topology captured in $\mathbf{Y}$ and the specifications $\{(u_n, \theta_n, p_n, q_n)\}_{n \in \mathcal{M}}$, our goal is to recover the power injections at non-metered buses $\{(p_n, q_n)\}_{n \in \mathcal{O}}$. Lacking a direct mapping from $\{(u_n, \theta_n, p_n, q_n)\}_{n \in \mathcal{M}}$ to $\{(p_n, q_n)\}_{n \in \mathcal{O}}$, the problem of finding the non-metered loads boils down to the task of recovering the underlying state $\mathbf{v}$ first. Collecting the grid data $\{(u_n^t, \theta_n^t, p_n^t, q_n^t)\}_{n \in \mathcal{M}}$ at time t and assuming for now these data are noiseless, we get the specifications

$$u_n(\mathbf{v}_t) = u_n^t, \qquad\qquad \forall n \in \mathcal{M} \tag{2a}$$
$$\theta_n(\mathbf{v}_t) = \theta_n^t, \qquad\qquad \forall n \in \mathcal{M} \tag{2b}$$
$$p_n(\mathbf{v}_t) = p_n^t, \; q_n(\mathbf{v}_t) = q_n^t \qquad \forall n \in \mathcal{M} \tag{2c}$$

which involve $4M$ equations over $2(N + 1)$ unknowns. A necessary condition for solving (2) is $4M \geq 2(N + 1)$. Since $N + 1 = M + O$, the condition simplifies to

$$M \geq O. \tag{3}$$

In other words, the metered buses must be at least as many as the non-metered ones.

To relax this condition on M, one may consider jointly processing the data $\{(u_n^t, \theta_n^t, p_n^t, q_n^t)\}_{n \in \mathcal{M}}$ collected across multiple times $t \in \mathcal{T}$ with $\mathcal{T} := \{1, \dots, T\}$. This approach does not improve the observability of the equations in (2), simply because the equations are independent over $\mathcal{T}$. Moreover, both the $4MT$ equations and the $2(N + 1)T$ state variables $\{\mathbf{v}_t\}_{t=1}^T$ scale with T. One way to relate power flow
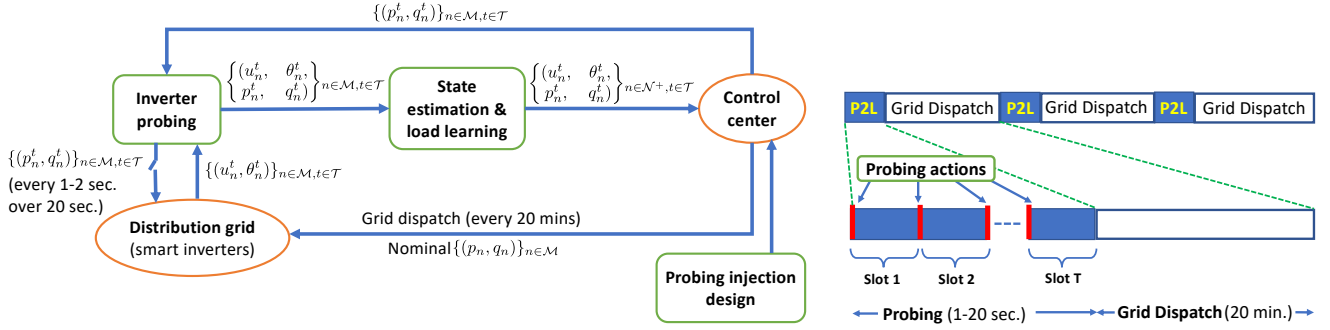
Fig. 1. Overview of the P2L framework: (a) block diagram depicting the P2L task with phasor data; (b) temporal organization of grid operation.

specifications across time is to assume that the non-metered loads remain invariant across $\mathcal{T}$, that is

$$p_n(\mathbf{v}_t) = p_n(\mathbf{v}_{t+1}), \qquad \forall n \in \mathcal{O}, t \in \mathcal{T}' \quad (4a)$$
$$q_n(\mathbf{v}_t) = q_n(\mathbf{v}_{t+1}), \qquad \forall n \in \mathcal{O}, t \in \mathcal{T}' \quad (4b)$$

where $\mathcal{T}' := \{1, \ldots, T-1\}$. In this way, we obtain the additional $2O(T-1)$ equations and couple the states $\{\mathbf{v}_t\}_{t=1}^{T}$.

Even though there may be an observability advantage in coupling specifications across time, the timespan of $\mathcal{T}$ is critical: For non-metered loads to remain unchanged, the timespan of $\mathcal{T}$ should be relatively short. But if the duration of $\mathcal{T}$ is too short, the metered injections in the buses of $\mathcal{M}$ may not change either. In this case, the grid state remains identical over $\mathcal{T}$, the scheme degenerates to the setup of (2) for $T = 1$, and there is no advantage by coupling specifications.

At this point, smart inverters come to our rescue: The timespan of $\mathcal{T}$ can be made sufficiently short so that the non-metered loads in the buses of $\mathcal{O}$ remain invariant over $\mathcal{T}$, whereas the power injections from smart inverters vary. The key point here is to couple power flow specifications through what we term *grid probing*. Probing can be accomplished by commanding inverters to change their power injections for one second. An inverter can curtail its solar output; (dis)-charge an energy storage unit; and/or change its power factor. Multiple probing actions can be instructed within tens of seconds. By intentionally perturbing inverter injections, the grid transitions across different states $\{\mathbf{v}_t\}_{t=1}^{T}$ depending on the probing injections and non-metered loads. Recording voltages $\{u_n^t, \theta_n^t\}_{t=1}^{T}$ over $n \in \mathcal{M}$ could unveil non-metered loads.

The metered buses in $\mathcal{M}$ can be classified into probing buses and metered but non-controllable buses. Although grid data $(u_n^t, \theta_n^t, p_n^t, q_n^t)$ are collected on both probed and metered buses, the operator can control only the probing buses. To simplify the presentation, we will henceforth assume that all metered buses are probing buses, although the analysis and algorithms apply to the more general setup.

Probing postulates two assumptions on non-metered loads: *a1)* They remain constant throughout $\mathcal{T}$; and *a2)* are modeled as of constant power. Assumption *a1)* may be reasonable over the short duration of probing. Regarding *a2)*, one could alternatively adopt a ZIP load model for bus $n \in \mathcal{O}$ [24]

$$-p_n^t(u_n^t) = \alpha_{p_n}(u_n^t)^2 + \beta_{p_n}u_n^t + \gamma_{p_n} \quad (5a)$$
$$-q_n^t(u_n^t) = \alpha_{q_n}(u_n^t)^2 + \beta_{q_n}u_n^t + \gamma_{q_n}. \quad (5b)$$

The parameters $(\alpha_{p_n}, \beta_{p_n}, \gamma_{p_n})$ correspond to the constant-impedance, constant-current, and constant-power components of active load; likewise $(\alpha_{q_n}, \beta_{q_n}, \gamma_{q_n})$ for reactive load. A non-metered ZIP load is then described by six rather than two parameters. Moreover, despite the model for load n does not change across $\mathcal{T}$, its power injection $p_n + jq_n$ does change for varying u_n. Then, the coupling equations are not valid in the form of (4) anymore. If the ZIP parameters are assumed invariant over $\mathcal{T}$, the power flow equations can still be coupled across $\mathcal{T}$, yet the identifiability analysis and the associated become perplex. To bypass this complexity, Section V copes with ZIP loads by resorting to single-slot probing.

Figure 1 depicts how probing can be incorporated into grid operation: Suppose a utility operates a demand-response program; manages energy storage; or controls smart inverters for reactive power control on a 20-min basis. To solve the optimal power flow problem, the operator needs to know the injections at non-metered buses. To do so, a probing interval lasting few tens of seconds precedes the feeder dispatch. This interval $\mathcal{T}$ is divided into T probing slots indexed by $t = 1, \ldots, T$. During each probing slot t, every inverter $n \in \mathcal{M}$ changes its injections to the setpoints (p_n^t, q_n^t) and reads voltage data (u_n^t) or (u_n^t, θ_n^t). At the end of interval $\mathcal{T}$, each inverter $n \in \mathcal{M}$ sends the collected data $\{u_n^t\}_{t\in\mathcal{T}}$ or $\{u_n^t, \theta_n^t\}_{t\in\mathcal{T}}$ back to the utility, and switches its setpoints back to their nominal values. The utility processes the collected data, infers the non-metered loads, and dispatches the grid for the next 20-min period. Some implementation details follow.

Remark 2. *The probing setpoints (p_n^t, q_n^t) for all t and $n \in \mathcal{M}$ are decided by the utility prior to $\mathcal{T}$ and communicated to all inverters via two-way communication links. This is to ensure that probing complies with voltage constraints and for improved load estimation accuracy; see Part II. The commanded setpoints are attained by simple PID controllers. Further, the inverters act synchronously along probing slots. Since potential delays may raise synchronization issues, developing protocols where inverters probe asynchronously is of interest.*

Remark 3. *The proposed probing scheme aims at recovering loads assuming the feeder topology is known. The topology includes bus connectivity and line impedances, phase assignments, and the statuses of capacitors and voltage regulators; see Remark 4. Although small errors in line impedances and*

regulator tap settings could be modeled as measurement noise, grid probing is as sensitive to topology errors as power flow equations are. However, probing can be also used for inferring grid topologies and line parameters without knowing non-metered loads [13], [14]. Moreover, phase assignments can be inferred from smart meter data; see e.g., [25]. Such techniques could precede P2L to find or calibrate feeder models.

Remark 4. *Feeders are equipped with voltage-control devices, such as regulators and capacitor banks, which respond to voltage excursions by changing their taps and switching on/off with time delays of around 30-90 seconds [24]. Since $\mathcal{T}$ lasts 20 sec or less, probing is not expected to trigger voltage control actions per se. Nonetheless, there are still chances for these actions to occur during $\mathcal{T}$ due to load fluctuations. If the utility does not monitor these devices in real-time or it cannot override their settings during probing, the topology learning techniques of Remark 3 could be possibly used. Voltage control actions and topology reconfigurations will be ignored in this work. Interestingly though, such actions could be used towards grid probing too.*

Grid probing can be now formally stated as follows.

Definition 1 (Probing-to-Learn task with phasor data)**.** *Given* **Y** *and probing data* $(u_n^t, \theta_n^t, p_n^t, q_n^t)$ *for all* $n \in \mathcal{M}$ *and* $t \in \mathcal{T}$, *the probing-to-learn (P2L) task entails solving the equations in* (2) *for* $t \in \mathcal{T}$ *jointly with the coupling equations in* (4).

The P2L task involves $4MT + 2O(T-1)$ equations in $2(N+1)T$ unknowns. A necessary condition for solving it is

$$M \geq \frac{O}{T} \tag{6}$$

which coincides with the condition in (3) for $T = 1$. For $T \geq 2$ however, it improves upon (3) if probing over multiple time instances is allowed. In [21], we have derived conditions under which the P2L task recovers non-metered loads for $T = 2$. The analysis there was further confined to non-phasor grid data $\{(u_n^t, p_n^t, q_n^t)\}_{n \in \mathcal{M}, t \in \mathcal{T}}$ and radial grids. The conference work of [26] extended the previous claims (without proofs) to meshed networks. Here, we broaden the scope to study the identifiability of the P2L task with phasor data $\{(u_n^t, \theta_n^t, p_n^t, q_n^t)\}_{n \in \mathcal{M}}$ over $\mathcal{T}$, and show that the analysis with non-phasor data can be seen as a special case of the former.

III. IDENTIFIABILITY OF P2L WITH PHASOR DATA

As customary in identifiability analysis, data will be assumed noiseless; noisy data are considered in Part II [27]. The relationship between the inputs $\{u_n^t, \theta_n^t, p_n^t, q_n^t\}_{n \in \mathcal{M}}$ and the outputs $\{p_n^t, q_n^t\}_{n \in \mathcal{O}}$ of the P2L task is implicit since the PF equations involve $\{\mathbf{v}_t\}_{t=1}^T$ as nuisance variables. Because of this, P2L is tackled in two steps. The first step of finding $\{\mathbf{v}_t\}_{t=1}^T$ is the challenging one. In the second step, one simply evaluates $(p_n(\mathbf{v}_t), q_n(\mathbf{v}_t))$ for all $n \in \mathcal{O}$ and $t = 1$. For numerical stability, one can recover the unknown injections by averaging as $\frac{1}{T}\sum_{t=1}^T p_n(\mathbf{v}_t)$ and $\frac{1}{T}\sum_{t=1}^T q_n(\mathbf{v}_t)$ for all $n \in \mathcal{O}$. Hence, if the system states $\{\mathbf{v}_t\}_{t=1}^T$ can be recovered by solving (2) and (4), the P2L task is deemed successful.
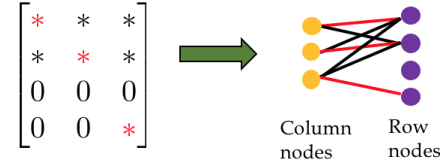

Fig. 2. The sparsity pattern of **E** and its bipartite graph $\mathcal{G}_E$: Column nodes are linked to row nodes depending on the entries of **E**. The perfect matching is marked in red. From Lemma 1, any matrix with this sparsity pattern is generically full rank. Had $E_{4,3} = 0$, no perfect matching would exist.

Granted the P2L equations are non-linear, identifiability can be ensured only within a neighborhood of the nominal $\{\mathbf{v}_t\}_{t=1}^T$. Upon invoking the inverse function theorem, a necessary and sufficient condition for locally solving P2L is that the Jacobian matrix $\mathbf{J}(\{\mathbf{v}_t\})$ related to the nonlinear equations of (2) and (4) is full rank. Because $\mathbf{J}(\{\mathbf{v}_t\})$ depends on $\{\mathbf{v}_t\}$, characterizing its column rank for any $\{\mathbf{v}_t\}$ is challenging.

To tackle this issue, we resort to the *generic rank* of a matrix defined as the maximum possible rank attained if the non-zero entries of the matrix are allowed to take arbitrary real values [28], [29]. If the generic rank of an $M \times N$ matrix **E** with $M \geq N$ equals N, matrix **E** is said to be of *full generic rank*. The generic rank of a matrix is related to a graph constructed by the sparsity pattern of the matrix, that is the locations of its (non)-zero entries. To explain this link, some graph-theoretic concepts are needed.

A graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ is bipartite if $\mathcal{N}$ can be partitioned into disjoint subsets $\mathcal{N}_1$ and $\mathcal{N}_2$, such that $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$, and every $\ell \in \mathcal{L}$ connects a node in $\mathcal{N}_1$ to a node in $\mathcal{N}_2$. A subset of edges $\mathcal{L}' \subseteq \mathcal{L}$ is termed a perfect matching of $\mathcal{N}_1$ to $\mathcal{N}_2$, if every vertex in $\mathcal{N}_1$ is incident to exactly one edge in $\mathcal{L}'$. The degree $\delta_n(\mathcal{G})$ of node n is defined as the number of edges incident to node n in $\mathcal{G}$. Given a matrix $\mathbf{E} \in \mathbb{R}^{M \times N}$, construct a bipartite graph $\mathcal{G}_E$ having $M + N$ nodes: Each column of **E** is mapped to a column node and each row of **E** to a row node. An edge runs from the n-th column node to the m-th row node only if $E_{mn} \neq 0$; see Fig. 2. Based on $\mathcal{G}_E$, we will use next claim.

Lemma 1 ([28], [29])**.** *An $M \times N$ matrix **E** has full generic rank if and only if the bipartite graph $\mathcal{G}_E$ features a perfect matching from the column nodes to its row nodes.*

According to Lemma 1 (proved in [29, Th. 12.10]), the generic identifiability of P2L relies on the sparsity pattern of $\mathbf{J}(\{\mathbf{v}_t\})$. The goal is to match every column node (state) of $\mathbf{J}(\{\mathbf{v}_t\})$ to a unique row node (equation). The non-zero entries of $\mathbf{J}(\{\mathbf{v}_t\})$ are the available links.

To characterize the sparsity pattern of $\mathbf{J}(\{\mathbf{v}_t\})$, consider the Jacobian matrices $\mathbf{J}^u(\mathbf{v})$, $\mathbf{J}^\theta(\mathbf{v})$, $\mathbf{J}^p(\mathbf{v})$, and $\mathbf{J}^q(\mathbf{v})$, associated accordingly with the squared voltage magnitudes and voltage angles, and the (re)active power injections over all buses. Matrix $\mathbf{J}(\{\mathbf{v}_t\})$ consists of stacked row-sampled submatrices of $\mathbf{J}^u(\mathbf{v}_t)$, $\mathbf{J}^\theta(\mathbf{v}_t)$, $\mathbf{J}^p(\mathbf{v}_t)$, and $\mathbf{J}^q(\mathbf{v}_t)$ corresponding to (2) and (4) for $t \in \mathcal{T}$. The matrices obtained by selecting the rows of $\mathbf{J}^u(\mathbf{v}_t)$ associated with buses in $\mathcal{M}$ and $\mathcal{O}$ are respectively denoted by $\mathbf{J}_{\mathcal{M}}^u(\mathbf{v}_t)$ and $\mathbf{J}_{\mathcal{O}}^u(\mathbf{v}_t)$. Similar notation is used for
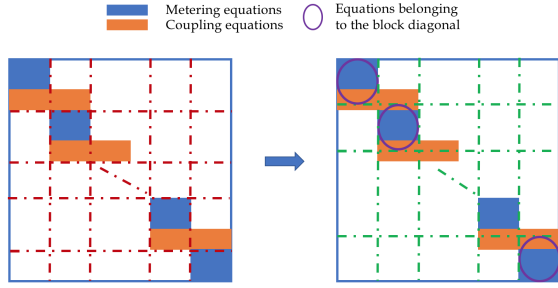
Fig. 3. *Left:* Sparsity pattern of $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$ [cf. (8)]. *Right:* block tridiagonal $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$ revealed after splitting each block row of coupling equations.

Fig. 4. Matchings on the IEEE 34-bus grid for $T = 4$ and $O = 6$ for the P2L task with phasor data.

$\mathbf{J}^\theta(\mathbf{v}_t)$, $\mathbf{J}^p(\mathbf{v}_t)$, and $\mathbf{J}^q(\mathbf{v}_t)$. Let us define

$$\mathbf{J}_{\mathcal{M}}(\mathbf{v}_t) := \begin{bmatrix} \mathbf{J}^u_{\mathcal{M}}(\mathbf{v}_t) \\ \mathbf{J}^\theta_{\mathcal{M}}(\mathbf{v}_t) \\ \mathbf{J}^p_{\mathcal{M}}(\mathbf{v}_t) \\ \mathbf{J}^q_{\mathcal{M}}(\mathbf{v}_t) \end{bmatrix} \quad \text{and} \quad \mathbf{J}_{\mathcal{O}}(\mathbf{v}_t) := \begin{bmatrix} \mathbf{J}^p_{\mathcal{O}}(\mathbf{v}_t) \\ \mathbf{J}^q_{\mathcal{O}}(\mathbf{v}_t) \end{bmatrix}.$$

Every $\mathbf{J}_{\mathcal{M}}(\mathbf{v}_t)$ corresponds to $4M$ metering equations, and every $\mathbf{J}_{\mathcal{O}}(\mathbf{v}_t)$ to $2O$ coupling equations. Having defined $\mathbf{J}_{\mathcal{M}}(\mathbf{v}_t)$ and $\mathbf{J}_{\mathcal{O}}(\mathbf{v}_t)$, the entire Jacobian matrix $\mathbf{J}(\{\mathbf{v}_t\})$ can be row-permuted as

$$\begin{bmatrix} \mathbf{J}_{\mathcal{M}}(\mathbf{v}_1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{J}_{\mathcal{O}}(\mathbf{v}_1) & -\mathbf{J}_{\mathcal{O}}(\mathbf{v}_2) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\mathcal{M}}(\mathbf{v}_2) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\mathcal{O}}(\mathbf{v}_2) & -\mathbf{J}_{\mathcal{O}}(\mathbf{v}_3) & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_{\mathcal{M}}(\mathbf{v}_3) & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{J}_{\mathcal{O}}(\mathbf{v}_T) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_{\mathcal{M}}(\mathbf{v}_T) \end{bmatrix}. \quad (8)$$

This row-permuted version of $\mathbf{J}(\{\mathbf{v}_t\})$ will be denoted by $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$, and has been obtained by interleaving block rows of metering and coupling equations.

Matrix $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$ features the sparsity pattern of a block tridiagonal matrix. To reveal this structure, split each block row of coupling equations into two block rows. The top block row will be grouped with the previous block row of metering equations. The bottom block row will be grouped with the next block row of metering equations as in Fig. 3.

Focus now on the blocks lying on the main diagonal of $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$. These blocks will be denoted by $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ for $t \in \mathcal{T}$. If for each $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$, its columns can be perfectly matched to its rows, then a perfect bipartite matching for the entire $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$ has been obtained. Then, Lemma 1 guarantees that $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$ and $\mathbf{J}(\{\mathbf{v}_t\})$ are generically full rank.

Our goal is to assign coupling equations to blocks so that every block $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ enjoys a perfect bipartite matching. There are $2O(T-1)$ coupling equations to be assigned to T blocks. A uniform allocation should assign $\frac{2O(T-1)}{T}$ coupling equations per block. With this allocation, block t will have $4M$ metering equations and $\frac{2O(T-1)}{T}$ coupling equations over its $2(N+1) = 2M+2O$ states in $\mathbf{v}_t$. For a perfect bipartite matching to exist, we need $4M + \frac{2O(T-1)}{T} \geq 2M + 2O$.

The last requirement coincides with the necessary condition of (6) for $T \geq 2$; but it is not enough: Every coupling equation
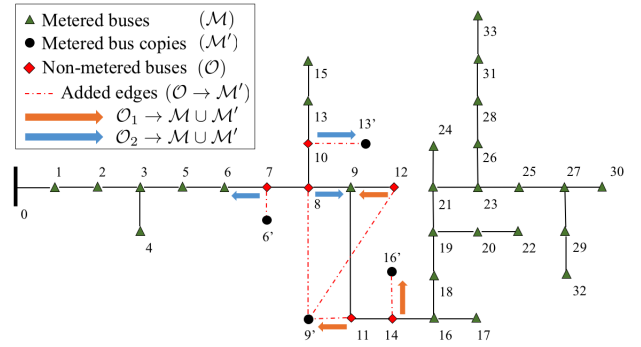
can be assigned to exactly one between two specific blocks; see Fig. 3. For example, a coupling equation in the block row involving $\mathbf{J}_{\mathcal{O}}(\mathbf{v}_2)$ and $-\mathbf{J}_{\mathcal{O}}(\mathbf{v}_3)$ can be grouped either with $\mathbf{J}_{\mathcal{M}}(\mathbf{v}_2)$ or $\mathbf{J}_{\mathcal{M}}(\mathbf{v}_3)$. Partitioning the coupling equations into groups of $\frac{2O(T-1)}{T}$ while adhering to the latter requirement is the crux of the identifiability analysis. To allocate coupling equations, let us first define the bipartite grid graph $\mathcal{G}_b$.

Definition 2 (Bipartite grid graph). *Consider the graph obtained from $\mathcal{G}$ upon maintaining only the edges between $\mathcal{M}$ and $\mathcal{O}$. Replicate the node set $\mathcal{M}$ to form $\mathcal{M}'$, and connect the nodes in $\mathcal{M}'$ to nodes in $\mathcal{O}$ by replicating the $\mathcal{M}$–$\mathcal{O}$ edges. The obtained bipartite graph will be denoted by $\mathcal{G}_b$.*

The identifiability of P2L relies on a matching in $\mathcal{G}_b$.

Theorem 1. *If $\mathcal{O}$ can be partitioned into $\{\bar{\mathcal{O}}_k\}_{k=1}^{\lceil T/2 \rceil}$ so that each one of them independently can be perfectly matched to $\mathcal{M} \cup \mathcal{M}'$ on $\mathcal{G}_b$, the Jacobian matrix $\mathbf{J}(\{\mathbf{v}_t\})$ related to the P2L task with phasor data is generically full rank.*

In essence, Theorem 1 provides sufficient conditions for successful probing. The proof of Theorem 1 relies on two lemmas shown in the appendix: Lemma 2 provides sufficient conditions for the coupling equations assigned to block t, so that $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ enjoys a bipartite matching. Lemma 3 explains when these conditions can be met simultaneously for all $t \in \mathcal{T}$. The analysis uses the concept of a *multi-set*. Different from a conventional set that contains unique elements, a multi-set is allowed to have multiple instances of elements. For example, we will override the definition of set union, so that $\{a, b\} \cup \{a, b\}$ does not yield $\{a, b\}$, but the multi-set $\{a, a, b, b\}$.

Lemma 2. *Partition $\mathcal{O}$ into $\mathcal{O}_t \cup \bar{\mathcal{O}}_t$ so that $|\mathcal{O} \cup \mathcal{O}_t| = 2O(T-1)/T$. Assume block $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ is assigned some coupling equations related to $\mathcal{O} \cup \mathcal{O}_t$. If the vertices in $\bar{\mathcal{O}}_t$ can be matched to the vertices in $\mathcal{M} \cup \mathcal{M}'$ on $\mathcal{G}_b$, the block $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ features a bipartite matching from its columns to its rows.*

Lemma 3. *Under the condition of Theorem 1, the coupling equations for two successive blocks $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ with $t = 2k - 1$ and $t = 2k$ share the same sparsity pattern of $\mathcal{O} \cup \mathcal{O}_k$ for $k = 1, \ldots, \lceil T/2 \rceil$.*

Theorem 1 follows as a direct consequence of Lemmas 2 and 3. To simplify the exposition, we will henceforth assume

Algorithm 1 Test for Successful Probing (phasor data)

1: Assign unit capacity to edges in $\mathcal{G}_b$ to define graph $\tilde{\mathcal{G}}_b$.
2: In $\tilde{\mathcal{G}}_b$, add source node n_s, and connect it to all nodes in $\mathcal{O}$. These edges are assigned unit capacity.
3: In $\tilde{\mathcal{G}}_b$, add destination node n_d, and connect it to all nodes in $\mathcal{M} \cup \mathcal{M}'$.
4: Initialize $T = 2$.
5: **while** $T \leq T_{\max}$ **do**
6:     The edges running between $\mathcal{M} \cup \mathcal{M}'$ and n_d are assigned capacities of $T/2$.
7:     Run a max-flow problem between n_s and n_d.
8:     **if** obtained n_s–n_d flow equals O **then**
9:         **return** *Probing setup is deemed successful for T.*
10:     **else**
11:         $T := T + 2$
12:     **end if**
13:     **return** *Probing setup is deemed unsuccessful.*
14: **end while**

even T. To appreciate the conditions of Theorem 1, examine the *probing setup*, that is the placement of non-metered and probed buses, of Figure 4. The black circles denote the copies $\mathcal{M}'$ of nodes in $\mathcal{M}$, and the dashed red lines show the added edges from $\mathcal{O}$ to $\mathcal{M}'$. The operator needs to infer the loads at the $O = 6$ non-metered buses marked by red diamonds. To study if probing this feeder over $T = 4$ slots is successful, the set $\mathcal{O}$ has to be partitioned into two subsets $\mathcal{O}_1$ and $\mathcal{O}_2$, so that the buses of each subset are matched to buses in $\mathcal{M} \cup \mathcal{M}'$ on $\mathcal{G}_b$. The orange and blue arrows show precisely these matchings. If the feeder were to be probed over $T = 2$ slots instead, probing would fail since buses $\{8, 11, 12\}$ cannot be uniquely matched to any buses in $\mathcal{M} \cup \mathcal{M}'$.

As illustrated through this example, to check the condition of Theorem 1 for a particular $(\mathcal{O}, \mathcal{M})$ probing setup, first one has to construct the bipartite graph $\mathcal{G}_b$ from $\mathcal{G}$. Then, given a number of probing actions T: *i)* the set $\mathcal{O}$ has to be partitioned into the subsets $\{\bar{\mathcal{O}}_k\}_{k=1}^{T/2}$; and *ii)* the nodes within each $\bar{\mathcal{O}}_k$ have to be mapped to the nodes in $\mathcal{M} \cup \mathcal{M}'$ on $\mathcal{G}_b$. Albeit these steps may seem computationally hard, they can be solved by a linear program as detailed in Algorithm 1.

Given a probing setup, Algorithm 1 finds the maximum flow between nodes n_s and n_d over graph $\tilde{\mathcal{G}}_b$ constructed from $\mathcal{G}_b$. The edges in $\tilde{\mathcal{G}}_b$ are organized in three layers: The edges of the first layer connect n_s to $\mathcal{O}$ and have unit capacities. The edges of the second layer connect $\mathcal{O}$ to $\mathcal{M} \cup \mathcal{M}'$ and have unit capacities as well. The edges of the third layer connect $\mathcal{M} \cup \mathcal{M}'$ to n_d and have capacities of $T/2$. This is to ensure that each node in $\mathcal{M} \cup \mathcal{M}'$ is mapped to at most $T/2$ nodes in $\mathcal{O}$ through the second layer. If the maximum n_s–n_d flow equals O, all first-layer edges have been used to their capacity to map every node in $\mathcal{O}$ to exactly one node in $\mathcal{M} \cup \mathcal{M}'$.

The max-flow problem can be solved using the Ford-Fulkerson algorithm, whose complexity scales linearly with the number of graph nodes and edges [30]. Moreover, if all edge capacities are integers, the algorithm finds an integral maximal flow. If the maximum n_s–n_d flow is smaller than O, there is no matching for the tested T. Then, the edge
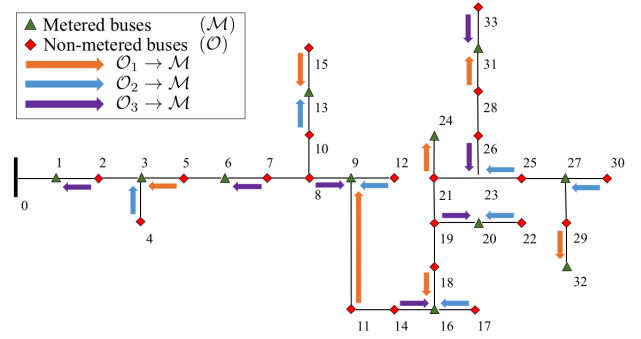

Fig. 5. Matchings on the IEEE 34-bus grid for $T = 6$ and $O = 21$ for the P2L task with non-phasor data.

capacities at the third layer can be increased and the process is repeated. Theorem 1 asserts that the chances of successful probing improve for larger T. This is because progressively smaller subsets of $\mathcal{O}$ need to be mapped to $\mathcal{M} \cup \mathcal{M}'$. Yet this gain in T is limited by the bus placement $(\mathcal{M}, \mathcal{O})$ as quantified next and shown in the appendix.

Lemma 4. *If $\delta_{\mathcal{M}}$ is the maximum node degree over $\mathcal{M}$ on $\tilde{\mathcal{G}}_b$, a probing setup with phasor data cannot turn into successful beyond $T_{\max} = \delta_{\mathcal{M}} - 1$.*

Lemma 4 implies that increasing T beyond $T_{\max}$ has no hope in making probing successful for a specific placement, and Algorithm 1 terminates with a negative answer. Once a $(\mathcal{M}, \mathcal{O})$ placement is deemed successful, there are two questions to be answered: *i)* how to select probing injections; and *ii)* how to recover the non-metered loads. Both questions along with numerical tests are deferred to Part II.

IV. IDENTIFIABILITY OF P2L WITH NON-PHASOR DATA

Since PMUs have limited penetration in distribution grids, requiring voltage phasor data at probing buses may be unrealistic. This section studies probing with non-phasor data.

Definition 3 (P2L task with non-phasor data). *Given $\mathbf{Y}$ and probing data (u_n^t, p_n^t, q_n^t) for $n \in \mathcal{M}$ and $t \in \mathcal{T}$, the P2L task entails solving the equations in (2a) and (2c) for $t \in \mathcal{T}$, jointly with the coupling equations in (4).*

A simple count of equations and unknowns dictates $M \geq \frac{2O}{T}$, which is clearly more restrictive than (6). We next provide a sufficient condition under which this task is solvable.

Theorem 2. *If $\mathcal{O}$ can be partitioned into $\{\bar{\mathcal{O}}_k\}_{k=1}^{\lceil T/2 \rceil}$ such that each one of them independently can be perfectly matched to $\mathcal{M}$ on $\mathcal{G}$, the Jacobian matrix $\mathbf{J}(\{\mathbf{v}_t\})$ related to the P2L task with non-phasor data is generically full rank.*

Dropping the voltage angle metering equations, matrix $\mathbf{J}_{\mathcal{M}}(\mathbf{v}_t)$ in (8) is replaced by

$$\mathbf{J}_{\mathcal{M}}(\mathbf{v}_t) := \begin{bmatrix} \mathbf{J}_{\mathcal{M}}^u(\mathbf{v}_t) \\ \mathbf{J}_{\mathcal{M}}^p(\mathbf{v}_t) \\ \mathbf{J}_{\mathcal{M}}^q(\mathbf{v}_t) \end{bmatrix}.$$

Similar to Theorem 1, it is not hard to see that the nodes in $\bar{\mathcal{O}}_t$ have to be matched to the nodes in $\mathcal{M}$, rather than $\mathcal{M} \cup \mathcal{M}'$.

Algorithm 2 Test for Successful Probing (non-phasor data)

1: Connect n_s to all nodes in $\mathcal{O}$ with unit-capacity edges.
2: Connect $\mathcal{O}$ to $\mathcal{M}$ based on $\mathcal{G}$ with unit-capacity edges.
3: Connect all nodes in $\mathcal{M}$ to n_d.
4: Initialize $T = 2$.
5: **while** $T \leq T_{\max}$ **do**
6: 　 Assign capacity $T/2$ to edges between $\mathcal{M}$ and n_d.
7: 　 Run a max-flow problem between n_s and n_d.
8: 　 **if** obtained n_s–n_d flow equals O **then**
9: 　　 **return** *Probing setup is deemed successful for T.*
10: 　 **else**
11: 　　 $T := T + 2$
12: 　 **end if**
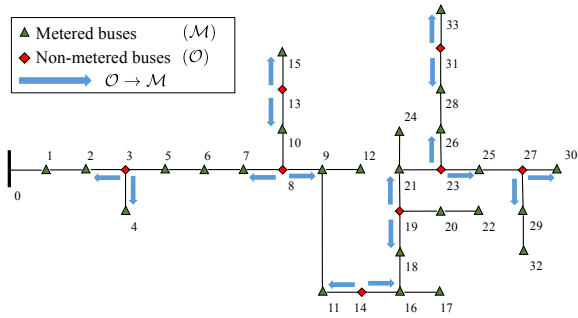13: 　 **return** *Probing setup is deemed unsuccessful.*
14: **end while**

Fig. 6. Matchings on the IEEE 34-bus grid for single-slot probing ($T = 1$) with non-phasor data and $O = 8$. Single-slot probing waives assumption *a2)* on exclusively constant-power loads.

Consider for example the probing setup of Figure 5. To infer the loads at $O = 21$ non-metered buses with $T = 6$ probing slots, the set $\mathcal{O}$ has to be partitioned into three subsets $\mathcal{O}_1$, $\mathcal{O}_2$, and $\mathcal{O}_3$, so that the buses of each subset are matched to $\mathcal{M}$. The orange, blue, and purple arrows in the figure show these matchings. Because non-metered buses are divided into three subsets, up to three non-metered buses can be matched to the same probed bus. For example, buses $\{14, 17, 18\}$ are all matched to the probed bus 16. Probing the same feeder over $T = 2$ or $T = 4$ rather than $T = 6$ slots would fail.

The condition of Theorem 2 can be easily tested by Algorithm 2 and up to the value of $T_{\max}$ provided next.

Corollary 1. *If $\delta_{\mathcal{M}}$ is the maximum degree of the nodes in $\mathcal{M}$ on the graph constructed by Alg. 2, a probing setup with non-phasor data cannot turn into successful beyond $T_{\max} = 2(\delta_{\mathcal{M}} - 1)$.*

Corollary 1 is proved as part of the proof of Lemma 4. Compared to Theorem 2, the condition of Theorem 1 provided more flexibility towards attaining a bipartite matching since probed buses can be used twice. If a probing setup is successful for non-phasor data, it is also successful for phasor data. Interestingly, the matchings in Theorems 1 and 2 depend solely on the sparsity pattern of $\mathbf{G}$ and the probing setup, so the claims here apply to even meshed (e.g., multiphase) grids.

V. SINGLE-SLOT PROBING

The analysis so far depends on assumption *a2)* of constant-power loads. Under the ZIP load model of (5), the coupling equations in (4) are no longer valid, and thus, the metering equations decouple across $\mathcal{T}$. Can the non-metered loads $p_n + jq_n$ for $n \in \mathcal{O}$ still be recovered upon collecting data on $\mathcal{M}$? This answer can be on the affirmative with *single-slot probing*, that is $T = 1$. Leveraging the tools of Sections III and IV, we next study the observability of single-slot probing. The ensuing two results (proven in the appendix) provide conditions for successful load recovery using (non)-phasor data.

Theorem 3. *If each bus in $\mathcal{O}$ can be matched to one unique bus in $\mathcal{M}$ on $\mathcal{G}$, the Jacobian matrix $\mathbf{J}(\mathbf{v}_1)$ related to single-slot probing ($T = 1$) with phasor data has full generic rank.*

Theorem 4. *If each bus in $\mathcal{O}$ can be matched to two unique buses in $\mathcal{M}$ on $\mathcal{G}$, the Jacobian $\mathbf{J}(\mathbf{v}_1)$ related to single-slot probing ($T = 1$) with non-phasor data has full generic rank.*

The conditions of Th. 3 and 4 can be tested by Algorithm 2 by fixing $T = 2$ and $T = 1$, respectively. Figure 6 shows a successful placement per Theorem 4.

If the conditions of Th. 3 and 4 are met, the non-metered loads $p_n + jq_n$ for $n \in \mathcal{O}$ can be recovered using single-slot probing, regardless if these loads are constant-power or not. However, the operator may also want to estimate their ZIP parameters in (5). Estimating these parameters directly with multi-slot probing becomes complicated. Instead, one could adopt multi-slot probing in a two-step process as follows: First, the feeder is probed over $\mathcal{T}$ with $|\mathcal{T}| = T > 3$. Under Th. 3 and 4, the operator obtains estimates $(\hat{u}_n^t, \hat{p}_n^t, \hat{q}_n^t)$ for all non-metered buses $n \in \mathcal{O}$ and $t \in \mathcal{T}$. Secondly, the ZIP parameters for active load n can be estimated through the least-squares (LS) fit

$$[\hat{\alpha}_{p_n}\ \hat{\beta}_{p_n}\ \hat{\gamma}_{p_n}]^{\top} := (\hat{\mathbf{U}}_n^{\top}\hat{\mathbf{U}}_n)^{-1}\hat{\mathbf{U}}_n^{\top}\hat{\mathbf{p}}_n \qquad (10)$$

where $\hat{\mathbf{p}}_n := [\hat{p}_n^1\ \ldots\ \hat{p}_n^T]^{\top}$ and the t-th row of matrix $\hat{\mathbf{U}}_n$ is $[(\hat{u}_n^t)^2\ \hat{u}_n^t\ 1]$ for $t = 1, \ldots, T$. Similar LS fits can be performed for the reactive ZIP load parameters. A major concern here is that all entries of the Vandermonde matrix $\hat{\mathbf{U}}_n$ are close to unity in compliance with voltage regulation. For $T = 3$, the determinant of $\hat{\mathbf{U}}_n$ is calculated as $(\hat{u}_n^1 - \hat{u}_n^2)(\hat{u}_n^1 - \hat{u}_n^3)(\hat{u}_n^2 - \hat{u}_n^3)$ [31], which yields $|\hat{\mathbf{U}}_n| = -2 \cdot 10^{-3}$ even for bus voltages as widely spread as $\hat{u}_n^1 = 0.9$, $\hat{u}_n^2 = 1.0$, and $\hat{u}_n^3 = 1.1$. This reveals that the task of estimating ZIP parameters from voltage/power data is *ill-posed*. This is germane to the task itself rather than the method (here probing) used to collect the data.

Finally, note that Th. 3 and 4 hold even when data are not collected via probing, e.g., smart meter data. Therefore, our observability analysis covers the general setup where voltage and active/reactive power data or specifications are given only for $\mathcal{M}$. Similar conditions were derived in [32], but were confined to radial grids.

VI. CONCLUSIONS

The novel technique of intentionally probing an electric grid using inverters to recover non-metered loads has been put forth

in the first part of this two-part work. The technique leverages the actuation capabilities of smart inverters, the data collected at probed buses, and the stationarity of non-metered loads, to formulate a power flow problem coupled over multiple times. Sufficient conditions that can be easily verified by solving a max-flow problem on a grid graph have been provided to test if a probing placement is successful. Beyond probing, the pertinent task of finding loads using data from a subset of buses has also been cast as a special case. Assuming a probing setup satisfies these conditions, Part II explains how inverter probing setpoints can be designed to improve load estimation accuracy, and provides numerical tests on the IEEE 34-bus feeder using a semidefinite program relaxation.

APPENDIX

Proof of Lemma 2: It can be easily verified that the sparsity patterns of $\mathbf{J}^u(\mathbf{v}_t)$ and $\mathbf{J}^\theta(\mathbf{v}_t)$ coincide with the sparsity pattern of $[\mathbf{I}_{N+1} \ \mathbf{I}_{N+1}]$. The sparsity patterns of $\mathbf{J}^p(\mathbf{v}_t)$ and $\mathbf{J}^q(\mathbf{v}_t)$ coincide with the sparsity pattern of $[\mathbf{G} \ \mathbf{G}]$ where $\mathbf{G}$ is the bus conductance matrix; see [22, Table 3.2]. From (8), the sparsity pattern of $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$ is

$$
\begin{bmatrix}
\mathbf{I}_{\mathcal{M},\mathcal{N}^+} & \mathbf{I}_{\mathcal{M},\mathcal{N}^+} \\
\mathbf{I}_{\mathcal{M},\mathcal{N}^+} & \mathbf{I}_{\mathcal{M},\mathcal{N}^+} \\
\mathbf{G}_{\mathcal{M},\mathcal{N}^+} & \mathbf{G}_{\mathcal{M},\mathcal{N}^+} \\
\mathbf{G}_{\mathcal{M},\mathcal{N}^+} & \mathbf{G}_{\mathcal{M},\mathcal{N}^+} \\
\mathbf{G}_{\mathcal{O},\mathcal{N}^+} & \mathbf{G}_{\mathcal{O},\mathcal{N}^+} \\
\mathbf{G}_{\mathcal{O}_t,\mathcal{N}^+} & \mathbf{G}_{\mathcal{O}_t,\mathcal{N}^+}
\end{bmatrix}
\tag{11}
$$

where the first block row relates to voltage magnitudes; the second to voltage angles; the third and fourth to probing injections; while the fifth and sixth to coupled injections.

To create a bipartite matching for block $\tilde{\mathbf{J}}_t(\mathbf{v}_t)$, unfold the sparsity pattern in (11) column-wise using $\mathcal{N}^+ = \mathcal{M} \cup \mathcal{O}$ as

$$
\begin{bmatrix}
\boxed{\mathbf{I}_{\mathcal{M},\mathcal{M}}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} & \mathbf{I}_{\mathcal{M},\mathcal{M}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} \\
\mathbf{I}_{\mathcal{M},\mathcal{M}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} & \boxed{\mathbf{I}_{\mathcal{M},\mathcal{M}}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} \\
\mathbf{G}_{\mathcal{M},\mathcal{M}} & \mathbf{G}_{\mathcal{M},\mathcal{O}} & \mathbf{G}_{\mathcal{M},\mathcal{M}} & \color{red}{\mathbf{G}_{\mathcal{M},\mathcal{O}}} \\
\mathbf{G}_{\mathcal{M},\mathcal{M}} & \mathbf{G}_{\mathcal{M},\mathcal{O}} & \mathbf{G}_{\mathcal{M},\mathcal{M}} & \color{red}{\mathbf{G}_{\mathcal{M},\mathcal{O}}} \\
\mathbf{G}_{\mathcal{O},\mathcal{M}} & \boxed{\mathbf{G}_{\mathcal{O},\mathcal{O}}} & \mathbf{G}_{\mathcal{O},\mathcal{M}} & \mathbf{G}_{\mathcal{O},\mathcal{O}} \\
\mathbf{G}_{\mathcal{O}_t,\mathcal{M}} & \mathbf{G}_{\mathcal{O}_t,\mathcal{O}} & \mathbf{G}_{\mathcal{O}_t,\mathcal{M}} & \color{red}{\mathbf{G}_{\mathcal{O}_t,\mathcal{O}}}
\end{bmatrix}.
\tag{12}
$$

The first block column in (12) relates to variables $\{v_{r,n}^t\}_{n\in\mathcal{M}}$, and can be matched to the first block row via $\mathbf{I}_{\mathcal{M},\mathcal{M}}$. Similarly, the third block column relates to variables $\{v_{i,n}^t\}_{n\in\mathcal{M}}$, and can be matched to the second block row. The second block column relates to variables $\{v_{r,n}^t\}_{n\in\mathcal{O}}$, and can be matched to the fifth block row via the main diagonal of $\mathbf{G}_{\mathcal{O},\mathcal{O}}$.

To achieve a bipartite matching, the fourth block column related to variables $\{v_{i,n}^t\}_{n\in\mathcal{O}}$ has to be matched to the union of the third, fourth, and sixth block rows. Lacking a simple diagonal matching now, we leverage the sparsity pattern of $\mathbf{G}$. It suffices to match the column nodes in $\mathcal{O}$ to the row nodes in $\mathcal{M} \cup \mathcal{M} \cup \mathcal{O}_t$. Because $\mathcal{O} = \mathcal{O}_t \cup \bar{\mathcal{O}}_t$, the column nodes $\mathcal{O}_t$ can be matched to the row nodes $\mathcal{O}_t$ via some diagonal entries of $\mathbf{G}_{\mathcal{O}_t,\mathcal{O}}$. Then, the column nodes $\bar{\mathcal{O}}_t$ have to be matched to the row nodes $\mathcal{M} \cup \mathcal{M}$. This can be accomplished based on the hypothesis of this Lemma, thus completing its proof. ∎

Proof of Lemma 3: The pair of blocks $\tilde{\mathbf{J}}_{2k-1}(\mathbf{v}_{2k-1})$ and $\tilde{\mathbf{J}}_{2k}(\mathbf{v}_{2k})$ will be jointly indexed by k. Define also

$$
\mathcal{R}_k := \bigcup_{\tau=1}^{k} \bar{\mathcal{O}}_\tau.
\tag{13}
$$

In addition to the claim of this lemma, we will also prove that when passing from pair $k-1$ to pair k, a set of coupling equations represented by $\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1}$ have not been assigned to block $2k-2$, and are free to be assigned to block $2k-1$.

Proving by induction, we start with the base case. The pair indexed by $k=1$ consists of $\tilde{\mathbf{J}}_1(\mathbf{v}_1)$ and $\tilde{\mathbf{J}}_2(\mathbf{v}_2)$. The active and reactive equations coupling these two blocks can be represented by $\mathcal{O} \cup \mathcal{O} = \mathcal{O}_1 \cup \mathcal{O}_1 \cup \bar{\mathcal{O}}_1 \cup \bar{\mathcal{O}}_1$. Let us assign $\mathcal{O}_1 \cup \mathcal{O}_1 \cup \bar{\mathcal{O}}_1$ to block 1. With this assignment, the coupling equations for $\tilde{\mathbf{J}}_1(\mathbf{v}_1)$ get the sparsity pattern of $\mathcal{O} \cup \mathcal{O}_1$. The remaining coupling equations in $\bar{\mathcal{O}}_1$ are assigned to block 2.

Block 2 shares with block 3 the coupling equations $\mathcal{O} \cup \mathcal{O}$, which are again expressed as $\mathcal{O}_1 \cup \mathcal{O}_1 \cup \bar{\mathcal{O}}_1 \cup \bar{\mathcal{O}}_1$. From this new set of coupling equations, assign $\mathcal{O}_1 \cup \mathcal{O}_1$ to block 2. Hence, the coupling equations for $\tilde{\mathbf{J}}_2(\mathbf{v}_2)$ have the sparsity pattern of $\mathcal{O}_1 \cup \mathcal{O}_1 \cup \bar{\mathcal{O}}_1 = \mathcal{O} \cup \mathcal{O}_1$. The unused coupling equations are represented by $\bar{\mathcal{O}}_1 \cup \bar{\mathcal{O}}_1 = \mathcal{R}_1 \cup \mathcal{R}_1$.

Suppose the claim holds for the block pair $k-1$. It is next shown that the claim holds for the block pair k consisting of blocks $2k-1$ and $2k$. Starting with the odd block $2k-1$, the unused equations $\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1}$ that couple blocks $2k-2$ and $2k-1$ are assigned to block $2k-1$. Block $2k-1$ is also coupled to block $2k$ via $\mathcal{O} \cup \mathcal{O}$ equations, which can be expressed as $\mathcal{O}_k \cup \mathcal{O}_k \cup \bar{\mathcal{O}}_k \cup \bar{\mathcal{O}}_k$. The key point here is that by the definition of $\mathcal{R}_{k-1}$ and because $\bar{\mathcal{O}}_k$'s are mutually exclusive by the hypothesis of this lemma, it holds that

$$
\mathcal{R}_{k-1} \cap \bar{\mathcal{O}}_k = \emptyset \text{ and } \mathcal{R}_{k-1} \subset \mathcal{O}, \text{ so that } \mathcal{R}_{k-1} \subseteq \mathcal{O}_k.
$$

Therefore, the set $\mathcal{O}_k$ can be partitioned into $\mathcal{R}_{k-1}$ and $\mathcal{O}_k \setminus \mathcal{R}_{k-1}$. From the equations coupling blocks $2k-1$ and $2k$, the equations $(\mathcal{O}_k \setminus \mathcal{R}_{k-1}) \cup (\mathcal{O}_k \setminus \mathcal{R}_{k-1}) \cup \bar{\mathcal{O}}_k$ are assigned to block $2k-1$. In this way, the coupling equations for block $2k-1$ have the sparsity pattern

$$
\underbrace{\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1}}_{\text{with block } 2k-2} \cup \underbrace{(\mathcal{O}_k \setminus \mathcal{R}_{k-1}) \cup (\mathcal{O}_k \setminus \mathcal{R}_{k-1}) \cup \bar{\mathcal{O}}_k}_{\text{with block } 2k} = \mathcal{O} \cup \mathcal{O}_k.
$$

The unused equations coupling blocks $2k-1$ and $2k$ are $\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1} \cup \bar{\mathcal{O}}_k$.

Moving to block $2k$ of pair k, the unused equations $\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1} \cup \bar{\mathcal{O}}_k$ coupling block $2k$ with block $2k-1$ are assigned to block $2k$. Block $2k$ is also coupled with block $2k+1$ through $\mathcal{O} \cup \mathcal{O} = \mathcal{O}_k \cup \mathcal{O}_k \cup \bar{\mathcal{O}}_k \cup \bar{\mathcal{O}}_k$. From this new set of coupling equations, assign equations $(\mathcal{O}_k \setminus \mathcal{R}_{k-1}) \cup (\mathcal{O}_k \setminus \mathcal{R}_{k-1})$ to block $2k$. Hence, the coupling equations assigned to block $2k-1$ have the sparsity pattern

$$
\underbrace{\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1} \cup \bar{\mathcal{O}}_k}_{\text{with block } 2k-1} \cup \underbrace{(\mathcal{O}_k \setminus \mathcal{R}_{k-1}) \cup (\mathcal{O}_k \setminus \mathcal{R}_{k-1})}_{\text{with block } 2k+1} = \mathcal{O} \cup \mathcal{O}_k.
$$

The unused equations coupling blocks $2k$ and $2k+1$ are

$$
\mathcal{R}_{k-1} \cup \mathcal{R}_{k-1} \cup \bar{\mathcal{O}}_k \cup \bar{\mathcal{O}}_k = \mathcal{R}_k \cup \mathcal{R}_k.
$$

For the last block pair, the coupling equations already assigned to block $T-1$ have the sparsity pattern $\mathcal{O} \cup \mathcal{O}_{T/2}$. The last block T differs from the previous blocks as it only gets the $\mathcal{R}_{T/2-1} \cup \mathcal{R}_{T/2-1} \cup \bar{\mathcal{O}}_{T/2}$ unused coupling equations between blocks $T-1$ and T. Under the condition of Th. 1, $\mathcal{R}_{T/2-1} \cup \bar{\mathcal{O}}_{T/2} = \mathcal{O}$ and because $\mathcal{O} = \mathcal{O}_{T/2} \cup \bar{\mathcal{O}}_{T/2}$, we also have $\mathcal{R}_{T/2-1} = \mathcal{O}_{T/2}$. Hence, the sparsity pattern of the last block is also given by $\mathcal{O} \cup \mathcal{O}_{T/2}$. Since every pair of successive blocks has the same structure, the diagonal blocks of $\tilde{\mathbf{J}}(\{\mathbf{v}_t\})$ will exhibit $\lceil T/2 \rceil$ distinct sparsity patterns. ∎

Proof of Lemma 4: Consider node $m \in \mathcal{M}$ in $\tilde{\mathcal{G}}_b$ with degree δ_m. This node is connected to node n_d via an edge having capacity $T/2$, and to $\delta_m - 1$ nodes in $\mathcal{O}$ via unit-capacity edges. The maximum flow that can pass through the second-layer edges to m is $\delta_m - 1$. This flow will be funneled through edge (m, n_d). Then, there is no advantage for this edge to have capacity larger than $\delta_m - 1$, so that $T/2 \leq \delta_m - 1$. Considering all $m \in \mathcal{M} \cup \mathcal{M}'$, there is no point in testing for values of T beyond $T \leq 2(\delta_{\mathcal{M}} - 1)$.

The bound can be improved, since the previous argument assumed that all $\delta_m - 1$ edges between $\mathcal{O}$ and $m \in \mathcal{M}$ have reached their capacity. That will not happen since the $\mathcal{O}$ nodes adjacent to m on the feeder, can be shared between m and its copy $m' \in \mathcal{M}'$ on $\tilde{\mathcal{G}}_b$. Hence, the flow passing jointly through m and m' cannot exceed $\delta_m - 1$. Then, the capacity of edge (m, n_d) plus the capacity of edge (m', n_d) can be safely limited to $\delta_m - 1$, implying $T \leq \delta_m - 1$. ∎

Proof of Theorem 3: The sparsity pattern of $\mathbf{J}(\mathbf{v}_t)$ can be derived from (11)–(12) by eliminating the blocks related to coupling equations

$$\begin{bmatrix} \boxed{\mathbf{I}_{\mathcal{M},\mathcal{M}}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} & \mathbf{I}_{\mathcal{M},\mathcal{M}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} \\ \mathbf{I}_{\mathcal{M},\mathcal{M}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} & \boxed{\mathbf{I}_{\mathcal{M},\mathcal{M}}} & \mathbf{I}_{\mathcal{M},\mathcal{O}} \\ \mathbf{G}_{\mathcal{M},\mathcal{M}} & \overset{\ulcorner}{\mathbf{G}_{\mathcal{M},\mathcal{O}}} & \mathbf{G}_{\mathcal{M},\mathcal{M}} & \mathbf{G}_{\mathcal{M},\mathcal{O}} \\ \mathbf{G}_{\mathcal{M},\mathcal{M}} & \mathbf{G}_{\mathcal{M},\mathcal{O}} & \mathbf{G}_{\mathcal{M},\mathcal{M}} & \overset{\ulcorner}{\mathbf{G}_{\mathcal{M},\mathcal{O}}} \end{bmatrix}. \quad (14)$$

Variables $\{v_{r,n}^t\}_{n \in \mathcal{M}}$ corresponding to the first block column and $\{v_{i,n}^t\}_{n \in \mathcal{M}}$ to the third block column can be matched respectively to the first and second block row via $\mathbf{I}_{\mathcal{M},\mathcal{M}}$. To complete the bipartite matching, the second and fourth block columns (variables $\{v_{r,n}^t\}_{n \in \mathcal{O}}$ and $\{v_{i,n}^t\}_{n \in \mathcal{O}}$) can be matched to the third and fourth block rows, accordingly. Hence, $\mathbf{J}(\mathbf{v}_t)$ is generically full rank if there exists a perfect matching in $\mathbf{G}_{\mathcal{M},\mathcal{O}}$, that is every non-metered node in $\mathcal{O}$ is mapped to a unique node in $\mathcal{M}$; see also Lemma 1. ∎

Proof of Theorem 4: Given non-phasor data, the second block row related to voltage angles in (14) is dropped. Following the arguments to the proof for Theorem 3, matrix $\mathbf{J}(\mathbf{v}_t)$ can be shown to be generically full rank if there exists a perfect matching in $\mathbf{G}_{\mathcal{M},(\mathcal{O} \cup \mathcal{O})}$, that is every node in $\mathcal{O}$ is mapped to two unique nodes in $\mathcal{M}$; see also Lemma 1. ∎

REFERENCES

[1] M. E. Baran, "Challenges in state estimation on distribution systems," in *Power Engineering Society Summer Meeting*, Vancouver, British Columbia, Jul. 2001.

[2] C. Klauber and H. Zhu, "Distribution system state estimation using semidefinite programming," in *Proc. North American Power Symposium*, Charlotte, NC, Oct. 2015.

[3] A. Gómez-Expósito, C. Gómez-Quiles, and I. Džafić, "State estimation in two time scales for smart distribution systems," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 421–430, Jan. 2015.

[4] A. Primadianto and C. N. Lu, "A review on distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3875–3883, Sep. 2017.

[5] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power systems synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1006–1013, May 2016.

[6] M. Baran and T. E. McDermott, "Distribution system state estimation using ami data," in *IEEE/PES Power Systems Conference and Exposition*, Seattle, WA, Mar. 2009.

[7] G. R. Gray, J. Simmins, G. Rajappan, G. Ravikumar, and S. A. Khaparde, "Making distribution automation work: Smart data is imperative for growth," *IEEE Power Energy Mag.*, vol. 14, no. 1, pp. 58–67, Jan. 2016.

[8] J. Yu, Y. Weng, and R. Rajagopal, "Mapping rule estimation for power flow analysis in distribution grids," 2017, (submitted). [Online]. Available: https://arxiv.org/abs/1702.07948

[9] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. Mittelstadt, "Electromechanical mode online estimation using regularized robust RLS methods," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1670–1680, Nov. 2008.

[10] D. Trudnowski and J. Pierre, "Signal processing methods for estimating small-signal dynamic properties from measured responses," in *Inter-area Oscillations in Power Systems*, ser. Power Electronics and Power Systems, A. R. Messina, Ed. Springer, 2009, pp. 1–36.

[11] M. Jaksic, Z. Shen, I. Cvetkovic, D. Boroyevich, R. Burgos, and P. Mattavelli, "Wide-bandwidth identification of small-signal dq impedances of ac power systems via single-phase series voltage injection," in *European Conf. on Power Electronics and App.*, Geneva, Switzerland, Sep. 2015.

[12] M. Angjelichinoski, A. Scaglione, P. Popovski, and . Stefanovi, "Decentralized DC microgrid monitoring and optimization via primary control perturbations," *IEEE Trans. Signal Processing*, vol. 66, no. 12, pp. 3280–3295, Jun. 2018.

[13] G. Cavraro and V. Kekatos, "Inverter probing for power distribution network topology processing," *IEEE Trans. Control of Network Systems*, vol. PP, no. 99, 2019.

[14] ——, "Graph algorithms for topology identification using power grid probing," *IEEE Control Systems Letters*, vol. 2, no. 4, pp. 689–694, Oct. 2018.

[15] K. Turitsyn, P. Sulc, S. Backhaus, and M. Chertkov, "Options for control of reactive power by distributed photovoltaic generators," *Proc. IEEE*, vol. 99, no. 6, pp. 1063–1073, Jun. 2011.

[16] M. Farivar, L. Chen, and S. Low, "Equilibrium and dynamics of local voltage control in distribution systems," in *Proc. IEEE Conf. on Decision and Control*, Florence, Italy, Dec. 2013, pp. 4329–4334.

[17] V. Kekatos, L. Zhang, G. B. Giannakis, and R. Baldick, "Voltage regulation algorithms for multiphase power distribution grids," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3913–3923, Sep. 2016.

[18] E. Dall'Anese and A. Simonetto, "Optimal power flow pursuit," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 942–952, Mar. 2018.

[19] Y. Tang, K. Dvijotham, and S. Low, "Real-time optimal power flow," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2963–2973, May 2017.

[20] D. B. Arnold, M. D. Sankur, M. Negrete-Pincetic, and D. S. Callaway, "Model-free optimal coordination of distributed energy resources for provisioning transmission-level services," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 817–828, Jan. 2018.

[21] 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.

[22] A. Gómez-Expósito, A. J. Conejo, and C. Canizares, Eds., *Electric Energy Systems, Analysis and Operation*. Boca Raton, FL: CRC Press, 2009.

[23] (2013) Pecan Street Inc. [Online]. Available: dataport.pecanstreet.org/

[24] W. H. Kersting, *Distribution System Modeling and Analysis*. New York, NY: CRC Press, 2001.

[25] Y. Liao, Y. Weng, G. Liu, Z. Zhao, C.-W. Tan, and R. Rajagopal, "Unbalanced three-phase distribution grid topology estimation and bus phase identification," 2018. [Online]. Available: https://arxiv.org/abs/1809.07192

[26] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Power grid probing for load learning: Identifiability over multiple time instances," in *Proc. IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc.*, Curaçao, Dutch Antilles, Dec. 2017.

[27] ——, "Smart inverter grid probing for learning loads: Part II – Probing injection design," *IEEE Trans. Power Syst.*, vol. PP, no. 99, 2019.

[28] W. T. Tutte, "The factorization of linear graphs," *Journal of the London Mathematical Society*, vol. 22, no. 2, pp. 107–111, 1947.

[29] A. B. Tucker, *Computer Science Handbook, Second Edition*. London, UK: Chapman & Hall/CRC, 2004.

[30] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[31] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 2013.

[32] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Power distribution system observability with smart meter data," in *Proc. IEEE Global Conf. on Signal and Information Process.*, Montreal, Canada, Nov. 2017.

Vassilis Kekatos (SM'16) is an Assistant Professor with the Bradley Dept. of ECE at Virginia Tech. He obtained his Diploma, M.Sc., and Ph.D. from the Univ. of Patras, Greece, in 2001, 2003, and 2007, respectively. He is a recipient of the NSF Career Award in 2018 and the Marie Curie Fellowship. He has been a research associate with the ECE Dept. at the Univ. of Minnesota, where he received the postdoctoral career development award (honorable mention). During 2014, he stayed with the Univ. of Texas at Austin and the Ohio State Univ. as a visiting researcher. His research focus is on optimization and learning for future energy systems. He is currently serving in the editorial board of the IEEE Trans. on Smart Grid.

Siddharth Bhela (S'07) received his B.S. degree in electrical engineering from Texas A&M University, College Station, TX, in 2007, where he was the recipient of a departmental scholarship for over 3 years. He received his M.Sc. degree in electrical engineering from Virginia Tech, Blacksburg, VA in 2015, and is now working towards his PhD. He has worked for General Electric Power Systems and CP&Y Inc. and is a registered professional engineer in the State of Texas. His research interests are focused on the application of optimization and graph-theoretic techniques, and development of algorithmic solutions for smart power systems.

Sriharsha Veeramachaneni is a researcher at NextEra Analytics Inc., a subsidiary of NextEra Energy Inc., specializing in statistics, machine learning and optimization. He holds a Ph.D. in computer engineering from the Rensselaer Polytechnic Institute.