Synchrophasor Recovery and Prediction: A Graph-based Deep Learning Approach

James J.Q. Yu, *Member, IEEE*, David J. Hill, *Life Fellow, IEEE*, Victor O.K. Li, *Fellow, IEEE*, and Yunhe Hou, *Senior Member, IEEE*

***Abstract*—Data integrity of power system states is critical to modern power grid operation and control. Due to communication latency, state measurements are not immediately available at the control center, rendering slow responses of time-sensitive applications. In this paper, a new graph-based deep learning approach is proposed to recover and predict the states ahead of time utilizing the power network topology and existing measurements. A graph-convolutional recurrent adversarial network is devised to process available information and extract graphical and temporal data correlations. This approach overcomes drawbacks of the existing synchrophasor recovery and prediction implementation to improve the overall system performance. Additionally, the approach offers an adaptive data processing method to handle power grids of various sizes. Case studies demonstrate the outstanding recovery and prediction accuracy of the proposed approach, and investigations are conducted to illustrate its robustness against bad communication conditions, measurement noise, and system topology changes.**

***Index Terms*—Wide-area measurement system, communication latency, prediction system, deep learning, state estimation, internet of things.**

I. Introduction

WITH the gradual adoption of wide-area monitoring system (WAMS), power grids and energy systems are embracing the Internet of Things (IoT) technologies [1], [2]. Thanks to the distributed sensors across the grids, system operating states can be measured in a synchronous and frequent manner [3]. These synchrophasors enable advanced power system operations and control, including but not limited to the massive adoption of renewable generations and demand-side control, which were conventionally not possible with the traditional supervisory control and data acquisition system [4]. Much research has been performed on utilizing the information in a wide range of power system services, see [4]–[7] for some examples.

For applications using high-frequency power system state data, the quality-of-service of data transmissions is among the critical factors in system operation and control [8]. According to analyses by the North American Synchrophasor Initiative (NASPI), data quality issues in power applications can be characterized by data accuracy, availablity, and timeliness [9], [10]. While existing research relies heavily on the sampling devices in WAMS to guarantee data accuracy, data availability and timeliness concerns are generally neglected. It is commonly assumed that the communication infrastructure is error-free and enjoys zero-latency when addressing power system issues [11]. However, as analyzed in [8], [11], [12], stochastic packet drops and data transmission latency can significantly influence data integrity in power systems. The problem is becoming increasingly serious due to the expansion of power grids and introduction of new power electronics in the past few years [13], [14]. In many communication network implementations, data loss issues can be addressed by packet re-transmission. However, this can further increase the data transmission latency, rendering a slower system response [12], [15]. There exists previous work investigating solutions to handle missing PMU data, see [13], [16] for examples. The proposed approaches mainly utilize the low-rank property of synchrophasors or their Hankel matrix for missing data recovery. However, such low-rank-based methods are not guaranteed to handle scenarios with a large missing data volume or incoherence as available data is not guaranteed to satisfy the minimal obtained measurement requirement [13]. Additionally, scaling errors can be hardly detected since the system model information is not included in the solutions [16].

Recently, a synchrophasor recovery and prediction framework (SRPF) was proposed to address the data integrity problem led by communication latency in WAMS [8]. The framework employs recent deep neural network techniques to recover missing system measurements in the past and predict future ones in real-time. SRPF comprises two modular sub-systems, each of which can accommodate various implementations to meet different requirements of power applications, provided the input and output of the sub-systems are respected. Preliminary case studies demonstrate satisfactory system performance on a small-scale power system [8], [17].

However, there exists a research gap in addressing the data integrity problem in power systems. The original design of SRPF in [8] does not fully utilize the available information in the system dynamics, especially the power network topology. Similar to other learning-based research, e.g., [18], [19], system states are considered as Euclidean data vectors during processing without topological information. Nonetheless, this design requires the learning system to implicitly extract the hidden topology information from the input data, instead of explicitly utilizing it as another input. There are several shortcomings of this design. On the one hand, the recovery accuracy is potentially undermined since excessive computa-

This work was supported by the Theme-based Research Scheme of the Research Grants Council of Hong Kong, under Grant No. T23-701/14-N.

James J.Q. Yu is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China (email: yujq3@sustc.edu.cn). A part of this work was done at Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong.

David J. Hill, Victor O.K. Li, and Yunhe Hou are with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong.

tional effort is required to learn the topology information. On the other hand, additional neural network layers are necessary to identify system dynamics characteristics for large systems, rendering the framework less scalable. These drawbacks will be demonstrated by case studies in Section IV.

To bridge the research gap, in this work we propose a new graph-convolutional recurrent adversarial network (GRAN), which is employed to predict power system measurements. Combining the design principles of existing neural network structures in the literature, a new graph-convolutional recurrent layer is first constructed to process measurements as graphs, and data features can be simultaneously extracted both graphically and temporally. Furthermore, based on this neural network layer, we propose GRAN, which incorporates the recent generative adversarial model in its architecture. The proposed network takes power grid topology information and available measurements as inputs to predict future system states. This is among the pioneering work of using graph-based deep learning techniques in addressing power system issues.

The main contributions of this work are listed as follows:

- We propose a new neural network structure called graph-convolutional recurrent layer to simultaneously extract graphical and temporal characteristics of data. It can be adopted in addressing a wide range of problems with such data structures.
- We formulate GRAN to greatly improve the predictor performance in recovering and predicting synchrophasors. It overcomes drawbacks of the existing implementation.
- We carry out a series of comprehensive case studies to investigate the performance of the proposed system. The results demonstrate its efficacy and robustness to latency, data noise, and system changes.

The rest of this paper is organized as follows. We briefly introduce the communication latency problem in WAMS and SRPF in Section II. Section III elaborates the detailed formulation and design of the proposed graph-based deep learning approach, and discusses the training and prediction methodologies. In Section IV, case studies are recorded with analyses. Finally, this paper is concluded in Section V.

II. Synchrophasor Recovery and Prediction

As analyzed in the previous section and the literature, e.g., [8], communication latency is not negligible in WAMS. With the gradual deployment of phasor measurement units (PMUs) that develops millisecond-level system state samples, tens or hundreds of milliseconds latency significantly influences the response speed of time-sensitive power system applications [11]. Suppose PMUs $p \in \mathcal{P}$ in WAMS sample the state at a same time. These measurements need to be transmitted to a control center, resulting in stochastic latency d_p. While previous work assumes $d_p \equiv 0$, the latency must be respected in real-world implementation. As a result, at least $\max_{p\in\mathcal{P}} d_p$ waiting time is imposed on any subsequent control and protection applications. This problem becomes critical when the communication infrastructure experiences random latency spikes since the system needs to wait for the last measurement [8].

To resolve communication latency issues in WAMS, the authors proposed a synchrophasor recovery and prediction problem in [8], which can be summarized as follows. We first model the power network by an undirected graph $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the sets of buses and branches, respectively. We use $A = \{a_{ij}\} \in \mathbb{B}^{B\times B}$ to represent the adjacency matrix of the graph with diagonal entries set to ones. Matrix A represents the key topology information of the power network. If bus $i, j \in \mathcal{V}$ are connected by transmission lines or power electronics, the corresponding $a_{ij} = a_{ji} = 1$. For bus $i \in \mathcal{V}$, we denote the set of its connecting branches by $\mathcal{E}_i \subset \mathcal{E}$. In addition, neighboring buses of bus $i \in \mathcal{V}$ are defined by $\mathcal{N}_i = \{k \in \mathcal{V} | (i, k) \in \mathcal{E}_i\}$. We use symbol τ to denote investigated time instances in the time horizon $\mathcal{T}$, which represents the discrete time instances when PMUs sample system variables.

In the power system $G(\mathcal{V}, \mathcal{E})$, PMUs are installed on buses in the set $\mathcal{V}^{\mathrm{M}} \subseteq \mathcal{V}$. We use $\mathcal{V}^{\mathrm{N}} = \{k \in \mathcal{N}_i | i \in \mathcal{V}^{\mathrm{M}}\}$ to denote the neighboring buses connected to $\mathcal{V}^{\mathrm{M}}$. Similar to [8], we employ the widely adopted PMU model [20], [21] which measures the complex voltage phasor $V_{i,\tau}$ of its installation bus i at τ, and all complex current phasors $I_{ik,\tau}$ of branch $(i, k) \in \mathcal{E}_i$. Considering the lumped-circuit model for transmission lines, voltage phasors of buses $k \in \mathcal{N}_i$ can be calculated by employing transmission line parameters. Therefore, voltage phasors of $\mathcal{V}^{\mathrm{M}}$ can be sampled directly, while those of $\mathcal{V}^{\mathrm{N}}$ can be subsequently calculated.

Let $M_\tau = \{V_{i,\tau}, I_{ik,\tau} | i \in \mathcal{V}^{\mathrm{M}}, (i, k) \in \mathcal{E}_i\}$ be the measurements generated by PMUs on $\mathcal{V}^{\mathrm{M}}$ for a past time instance τ. Due to communication latency, only partial measurements, denoted by $M^-_\tau \subseteq M_\tau, \forall\tau \leq 0$, is received by the control center at the current time. Subsequently, a partial system state $S^-_\tau = h(M^-_\tau) = \{V_{i,\tau} | i \in \mathcal{V}^{\mathrm{MP}}_{\tau,t} \cup \mathcal{V}^{\mathrm{NP}}_{\tau,t}\}$ of the investigated time instance can be developed, where $h(\cdot)$ reflects the power grid topology and network parameter information, $\mathcal{V}^{\mathrm{MP}}_{\tau,t} \subseteq \mathcal{V}^{\mathrm{M}}$ and $\mathcal{V}^{\mathrm{NP}}_{\tau,t} = \{k \in \mathcal{N}_i | i \in \mathcal{V}^{\mathrm{MP}}_{\tau,t}\}$ are the buses whose measurements are received, and the neighbors of $\mathcal{V}^{\mathrm{MP}}_{\tau,t} \subseteq \mathcal{V}^{\mathrm{M}}$, respectively. The problem can be interpreted as recovering and predicting complete system states $S_\tau = \{V_{i,\tau} | i \in \mathcal{V}\}, \tau \in \mathbb{Z}$ using partially available measurements $\{M^-_\tau\}_{t\in\mathcal{T}}$.

Aiming at solving this problem, the authors proposed a delay-aware synchrophasor recovery and prediction framework in [8]. SRPF considers time-series system states as a data matrix, in which the entries correspond to measurements not yet received by the control center are missing. In the framework, two-subsystems, named *predictor* and *estimator*, cooperate to recover these missing entries. Additionally, the predictor can be further employed to make predictions on the system states in the near future, which represent a possible trend of power system dynamics. To achieve the objective, the predictor first finds the most recent time instance τ^* in the past such that all states before it are complete. These states are then used to make a prediction of the immediately next one:

$$S^+_{\tau^*+1} = \mathrm{Predictor}(S^-_{\tau^*}, S^-_{\tau^*-1}, \cdots). \tag{1}$$

Then the estimator combines the prediction and the remaining partial system states to calculate a better estimation of the

same time instance:

$$\hat{S}_{\tau^*+1} = \text{Estimator}(S^+_{\tau^*+1}, S^-_{\tau^*+1}, S^-_{\tau^*+2}, \cdots). \quad (2)$$

This estimation $\hat{S}_{\tau^*+1}$ is regarded as the best approximation of S_{τ^*+1}, and replaces it in following computations. The process repeats until no partial system states are left for the estimator, and subsequent predictions S^+_t are considered as S_t.

In the previous work, a gated recurrent unit ensemble is constructed to implement the predictor [8]. While the case studies demonstrated satisfactory performance, its performance on large power systems remains unknown. Furthermore, the technique discarded the power network topology information, which plays a critical role in determining power flow and dynamics. In the next sections, we propose a new graph-based deep neural network technique to overcome these drawbacks.

III. Graph-convolutional Recurrent Adversarial Network (GRAN)

As discussed above, the previous synchrophasor recovery and predictor implementation suffers from the information under-utilization problem. In this section, we propose a new deep neural network architecture for the predictor based on recent developments of deep learning techniques. We first elaborate on the computation of the proposed graph-convolutional recurrent layer in the neural network, and discuss its advantage over existing models. Then we construct the proposed GRAN, and illustrate its training and prediction methodology. The proposed predictor collaborates with the original estimator in [8] to develop recovered and predicted power system synchrophasors.

A. Graph-convolutional Recurrent Layer

In this work, we follow the design principle of graph convolutional network (GCN) and long-short term memory (LSTM) to construct a graph-convolutional recurrent layer (GRL), which aims to extract both the temporal and graphical-spatial characteristics from the input data. In particular, the layer inherits the idea of convolutional filters in conventional convolutional neural networks (CNN). Such filters perform neighborhood information mixing on the input data, which typically refer to images or signals in Euclidean space, in order to propagate local information in the result [22]. By tailoring the filters, spatial data correlations can be learnt from the training data.

Nonetheless, conventional design of convolutional filters mix information within their receptive fields, which are commonly rectangular. This design cannot be easily adopted in handling data with graph structures due to the irregularity of node connections [23]. Instead of sharing information with geographical neighborhoods as for image pixels, graphs concentrate more on the data correlation over graph edges, which are the power transmission lines in our case. To overcome this drawback, a representative description of the graph topology, e.g., adjacency matrix, can be adopted to replace the convolutional filter in the information propagation process. This is the conceptual idea of GCN [23], which maps a $B \times M$ input data matrix to a new $B \times F$ output feature matrix, where B, M, F are the number of nodes in the graph, number of input and output features for each node, respectively. The alternative graph-based propagation is performed by a non-linear function:

$$H^{(l+1)} = f(H^{(l)}, A) = \sigma(\hat{D}^{-\frac{1}{2}} A \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} + b^{(l)}), \quad (3)$$

where the graph-convolutional layer is the l-th one in a neural network. $H^{(l)}$ is the input data matrix of the l-th layer, i.e., output of the $(l-1)$-th layer. $f(\cdot,\cdot)$ is the propagation rule, $\sigma(\cdot)$ is a non-linear activation function, e.g., $\tanh(\cdot)$. Finally, $\hat{D}$ is the diagonal node degree matrix of A, and $W \in \mathbb{R}^{M \times F}$, $b \in \mathbb{R}^{B \times F}$ are layer-specific network weight and bias parameters, respectively. This propagation is motivated by the first-order approximation of Chebyshev polynomials in the spectral domain [23], [24], leading to a fast computational speed. Additionally, the approximation is accurate enough to develop competitive graph-learning results [23].

Thanks to the introduction of matrix A in (3), graph topology information can be integrated in the learning system. However, GCN by itself is designed to focus on spatial characteristics of different graphs from the same distribution [23]. As system dynamics are time-series information, additional data processing is required to extract the temporal correlation among data in the time sequence. We follow the state-of-the-art time-series data processing technique in deep learning, i.e., LSTM, to design the proposed GRL based on (3). In particular, the propagation rule of GRL is designed to be recurrent such that information of previous time-steps in a time-series can be passed to subsequent ones. Similar to LSTM, we first formulate[1] a cell state for GRL to develop latent time-series properties from the current time-step input $X_t \in \mathbb{R}^{B \times M}$:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (4a)$$

where

$$f_t = \sigma[\hat{D}^{-\frac{1}{2}} A \hat{D}^{-\frac{1}{2}} (X_t W_f + H_{t-1} U_f) + b_f], \quad (4b)$$

$$i_t = \sigma[\hat{D}^{-\frac{1}{2}} A \hat{D}^{-\frac{1}{2}} (X_t W_i + H_{t-1} U_i) + b_i], \quad (4c)$$

$$\tilde{C}_t = \tanh[\hat{D}^{-\frac{1}{2}} A \hat{D}^{-\frac{1}{2}} (X_t W_c + H_{t-1} U_c) + b_c]. \quad (4d)$$

In (4), $C_t \in \mathbb{R}^{B \times F}$ is the cell state of current time step of GRL, $*$ denotes Hadamard product, and σ is the sigmoid function. Matrices $f_t \in \mathbb{R}^{B \times F}$ and $i_t \in \mathbb{R}^{B \times F}$ are the forget and input "gate" activators that control information flow from previous cell states and the new input data, respectively [25]. They investigate the output of the previous time-step $H_{t-1} \in \mathbb{R}^{B \times F'}$ and current input to determine whether respective information presented in (4a) is kept or discarded in the new cell state, where F' is the number of output features in the previous layer. In the equations, $W \in \mathbb{R}^{M \times F}$, $U \in \mathbb{R}^{F' \times F}$, and $b \in \mathbb{R}^{B \times F}$ matrices are tunable neural network parameters. Finally, the output of this GRL is determined based on the cell state:

$$H_t = o_t * \tanh(C_t), \quad (5a)$$

$$o_t = \sigma[\hat{D}^{-\frac{1}{2}} A \hat{D}^{-\frac{1}{2}} (X_t W_o + H_{t-1} U_o) + b_o], \quad (5b)$$

[1]Superscript "(l)" is omitted for simplicity.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2899395, IEEE Internet of Things Journal

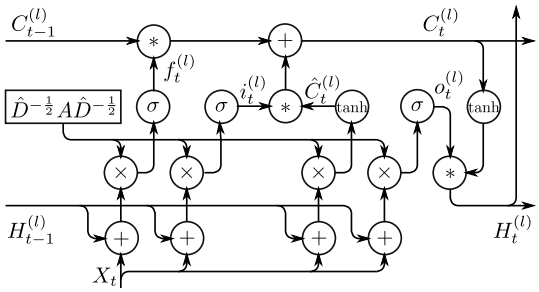


Fig. 1. Computation graph of the proposed graph convolutional recurrent layer.

where $o_t \in \mathbb{R}^{B\times F}$ is a final output gate activator which manipulates the output matrix based on the cell state. Consequently, the propagation rule of a GRL, i.e., (4) and (5), can be expressed as the following shorthand form:

$$H_t^{(l)} = \mathrm{GRL}(X_t, F, H_{t-1}^{(l)}, C_{t-1}^{(l)}, \Omega), \tag{6}$$

where Ω is the collection of network parameters. Fig. 1 presents an illustration of the data flow in GRL.

B. Network Architecture

In the previous sub-section, we introduced GRL to process time-series data with graphical topologies. Meanwhile, the neural network layer alone cannot enjoy the benefit of deep learning, i.e., outstanding feature extraction capability with adequately more layers in the same network. In this sub-section, we propose a GRL-enabled graph-convolutional recurrent adversarial model to fully address the challenge of learning dynamic power network states by machine.

Fig. 2 depicts the layered architecture of the proposed GRAN, in which "Full" refers to fully connected layers and numbers are the number of neurons (output features) in the respective layers. GRAN is composed of two sub-networks, namely, graph dynamics generator and discriminator. The former is composed of seven GRL with $M = 128, 256, 512, 512, 512, 256, 128$ and the respective F equals to the M value of the next layer, except that $F = 1$ for the last one. The latter is composed of a GRL with $M = 128$, and three fully connected layers with input dimensions equal to $512, 512, 64$, respectively. This design follows the principle of adversarial network training paradigm firstly proposed in generative adversarial network [26]. Both sub-networks form a two-player minimax game during the network training process. In the game, the generator tries to develop a sequence of new graph dynamics, which refer to system states in the investigated application, based on input historical dynamics. The discriminator aims to evaluate the authenticity of the newly generated dynamics. Its output is a real value indicating whether the generated ones are considered as real dynamics or not, which will be further elaborated in Section III-D. Within a well-trained network, the generator forges continuous system dynamics that can "fool" the discriminator. The architecture accepts time-series data in batches. For example, it first receives X_t and calculates the output using defined propagation rules (4), (5). In the process, latent information is also produced and stored as cell states C. Then the network receives a new batch X_{t+1} and do the calculation again. In this time, previous latent information is also involved according to (4a), and temporal correlation information is therefore propagated [22]. Fig. 3 presents the unrolled abstract computation graph of multi-layer GRL structure used in the architecture, which corresponds to any two adjacent blocks with recurrent links (up-left direction arrows) in Fig. 2. Compared with other generative models, e.g., Boltzmann machine and non-linear independent component analysis, this generative adversarial technique does not require Monte Carlo approximations or invertible generator structure to work well [27].

Let $\mathrm{G}(\cdot, \theta^{\mathrm{G}})$ and $\mathrm{D}(\cdot, \theta^{\mathrm{D}})$ be the mathematical formulation of the generator and discriminator, which are parameterized by θ^{G} and θ^{D}, respectively. Given time-series data $\mathbf{X}$ with steps $X_1, X_2, \cdots$ which follow a distribution $\mathbb{P}_{\mathbf{X}}$, the network tries to develop the subsequent dynamics of $\mathbf{X}$. The following objective function is used in the minimax game by the generator and discriminator:

$$\begin{aligned}\min_{\theta^{\mathrm{G}}} \max_{\theta^{\mathrm{D}}} V(\mathrm{G}, \mathrm{D}) = {} & \mathbb{E}_{\mathbb{P}_{\mathbf{X}}}[\log D(X_t, \theta^{\mathrm{D}})] \\ & + \mathbb{E}_{\mathbb{P}_{\mathbf{X}}}[\log(1 - \mathrm{D}(\mathrm{G}(X_{1,\cdots,t-1}, \theta^{\mathrm{G}}), \theta^{\mathrm{D}}))].\end{aligned} \tag{7}$$

During the network parameter training process, the generator optimizes θ^{G} to generate realistic system dynamics and fool the discriminator, leading to a small objective value $V(\mathrm{G}, \mathrm{D})$. At the same time, the discriminator adjusts θ^{D} to distinguish the artificially generated states from the real ones, rendering a large $V(\mathrm{G}, \mathrm{D})$. This process, which will be further elaborated in Section III-D, emulates the actions of the two players in the minimax game defined in (7).

C. Graph Dynamics Generator and Discriminator

In the proposed GRAN, GRLs constitute the generator and the first layer of the discriminator. At the beginning of the generator, available complete system states are first transformed into a series of graphs, in which each node corresponds to a bus in the power grid. Nodes are connected in accordance with the bus connectivity of the grid topology. The measured bus positive sequence voltage magnitudes and angles are assigned to the corresponding node in the graph as input node features[2]. Subsequently, the constructed graphs are input into the graph dynamics generator, in which seven GRLs are adopted to extract the temporal-graphical characteristics. Among these layers, the first four with 128, 256, 512, and 512 neurons, respectively, are used to encode the input graphs into latent feature graphs with 512 node features. The graphs are later decoded by the other three GRLs with 512, 256, and 128 neurons, respectively, to predict the next system state in the future. All layers are activated by Rectified Linear Units (ReLU) [28]. In the information diffusion process, the output prediction incorporates both temporal and graphical correlations of the measurements in the system. For instance,

[2]In this work, we consider positive sequence voltages. It is possible and simple to extend the model and adopt negative and zero sequence components of unbalanced measurements.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2899395, IEEE Internet of Things Journal

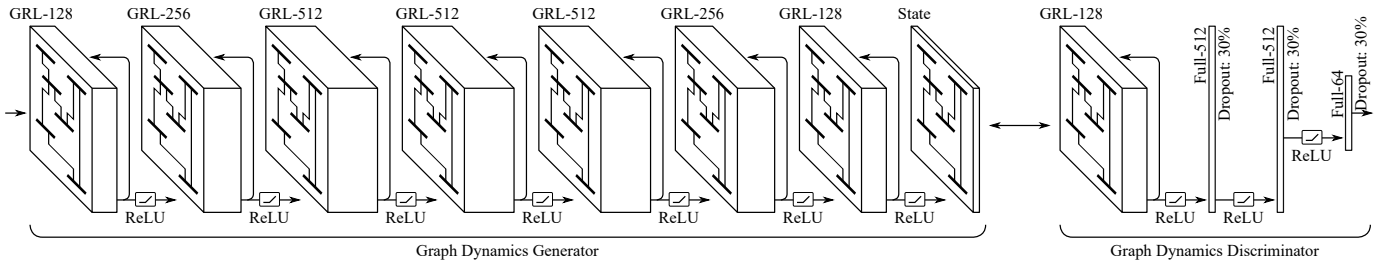


Fig. 2. Layered architecture of the proposed graph-convolutional recurrent adversarial network.

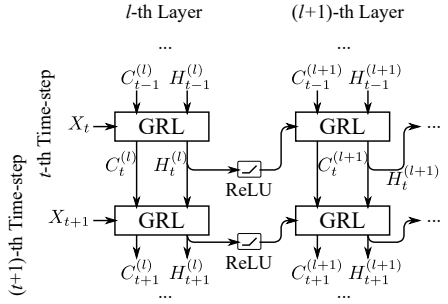


Fig. 3. Unrolled abstract computation graph of multi-layer GRL structure.

the generator outputs a prediction of X_{t+1} given X_t as the input. During the computation, both historical data features (C_{t-1}) and the graph features ($\hat{D}^{-\frac{1}{2}}A\hat{D}^{-\frac{1}{2}}$) are involved (cf. (4) and (5)).

Given the output of a system state prediction by the generator, the proposed graph dynamics discriminator determines whether the prediction is realistic based on the learnt power grid characteristics and immediately past states. This is achieved by a GRL with 128 neurons, which captures both the graphical and the temporal correlations of real consecutive system states. Subsequently, its output is flattened and input into three fully connected neuron layers, whose propagation rule is given as follows:

$$Y_t = \mathrm{ReLU}(W' X'_t + b'), \tag{8}$$

where X'_t is the flattened input data, Y_t is the output feature matrix, W' and b' are layer-specific network parameters. All layers cooperate to identify the authenticity of the input system state from the generator.

D. Training and Prediction

Note that in the proposed discriminator, only the first two fully connected layers with 512 neurons are activated with ReLU. We adopt linear activation for the last one with 64 neurons. This design facilitates an easier network training process. In traditional generative adversarial model, the discriminator output is activated with a sigmoid function to regulate the value in between zero and one [26]. While this setting makes the output human-readable (e.g., zero for fake and one for real), such networks suffer from several training issues such as model collapse where the generator collapses into a very narrow distribution, rendering the training unstable. In addition, there is no indication of convergence with the sigmoid activation function. Hence, we adopt the Wasserstein metric firstly proposed in [29] to construct smooth gradients for the discriminator, which utilizes the Earth-Mover distance as the cost function:

$$W(\mathbb{P}_{\mathbf{X}}, \mathbb{P}_{\hat{\mathbf{X}}}) = \inf_{\gamma \sim \prod(\mathbb{P}_{\mathbf{X}}, \mathbb{P}_{\hat{\mathbf{X}}})} \mathbb{E}_{(X,\hat{X})\sim\gamma} ||X - \hat{X}||, \tag{9}$$

where $\hat{X}$ is a prediction develop by the generator based on the learnt distribution $\mathbb{P}_{\hat{\mathbf{X}}}$, and $\prod(\mathbb{P}_{\mathbf{X}}, \mathbb{P}_{\hat{\mathbf{X}}})$ is the collection of all possible joint distributions of $\mathbb{P}_{\mathbf{X}}$ and $\mathbb{P}_{\hat{\mathbf{X}}}$. Due to the intractability of the infimum term, the distance is transformed to its equivalent form [30]:

$$W(\mathbb{P}_{\mathbf{X}}, \mathbb{P}_{\hat{\mathbf{X}}}) = \frac{1}{K} \sup_{||f||_L \le K} \mathbb{E}_{X\sim\mathbb{P}_{\mathbf{X}}} f(X) - \mathbb{E}_{\hat{X}\sim\mathbb{P}_{\hat{\mathbf{X}}}} f(\hat{X}), \tag{10}$$

where $f(\cdot)$ is an arbitrary function which refers to the aggregated propagation rule of the neural network, and K is the Lipschitz constant of $f(\cdot)$, such that

$$|f(x_1) - f(x_2)| \le K|x_1 - x_2|. \tag{11}$$

Specifically, we can use a neural network f_Ω parameterized by Ω to transform (10) into an approximated form

$$K \cdot W(\mathbb{P}_{\mathbf{X}}, \mathbb{P}_{\hat{\mathbf{X}}}) \approx \max_{\Omega:||f||_L \le K} \mathbb{E}_{X\sim\mathbb{P}_{\mathbf{X}}} f_\Omega(X) - \mathbb{E}_{\hat{X}\sim\mathbb{P}_{\hat{\mathbf{X}}}} f_\Omega(\hat{X}). \tag{12}$$

Finally, by clipping Ω into a finite range, e.g., $\Omega \in [-0.01, 0.01]$ as suggested in [29], the derivative $\partial f_\Omega / \partial X$ is restrained. This guarantees that there exists a constant K such that $||f||_L \le K$ [30]. The objective of the discriminator can be constructed accordingly:

$$\text{maximize } L_D = \mathbb{E}_{X\sim\mathbb{P}_{\mathbf{X}}} f_\Omega(X) - \mathbb{E}_{\hat{X}\sim\mathbb{P}_{\hat{\mathbf{X}}}} f_\Omega(\hat{X}). \tag{13}$$

In addition, since the first term in (13) is not involved in the generator, its objective function is designed as follows:

$$\text{maximize } L_G = -\mathbb{E}_{(X,\hat{X})\sim\gamma} ||X - \hat{X}||_2 - \mathbb{E}_{\hat{X}\sim\mathbb{P}_{\hat{\mathbf{X}}}} f_\Omega(\hat{X}), \tag{14}$$

where the first term refers to the expected distance between the generated and real system states.

Given a collection of historical or synthetic power system dynamics, the parameter training process is conducted in an offline manner. For each series of consecutive system states $\mathbf{X}$ with T steps, the parameters are updated progressively. In the i-th step ($i \le T$), $X_{1,\cdots,i}$ is input into the generator to develop a prediction $\hat{X}_{i+1}$ of X_{i+1}, which is adopted to

computed L_G. Subsequently, both the prediction and the real states are individually appended to $X_{1,\cdots,i}$, which are input into the discriminator to evaluate $f_\Omega(\hat{X}_{i+1})$ and $f_\Omega(X_{i+1})$ (cf. (13)). For every 32 dynamics, the objective values of both the generator and the discriminator are utilized to update the network parameters using RMSprop optimizer. This process repeats until the discriminator outputs converged L_D values. The resulting network parameters in the generator is employed in the online prediction of system states.

The training process aims to jointly optimize the numerous network parameters in the proposed neural network. Nonetheless, it is considered impractical to adjust their values without significant overfitting [22]. To overcome this issue, we adopt a technique called "dropout" to effectively reduce the correlation between the output and specific hidden layer neurons [31]. Specifically, this technique randomly sets the output data of randomly selected neurons[3] to zeros during the training process. Hence, the dropped neurons are temporarily removed from the computation graph, and more robust data features can be extracted. No neurons are removed from the graph during the prediction.

In the online prediction process, only the generator is involved. Upon finding the most recent time instance τ^* in the past with complete historical system states as introduced in Section II, the predictor inputs all the historical data into the generator to predict a $\hat{X}_{\tau^*+1}$, which corresponds to $S^+_{\tau^*+1}$ in (1). The prediction is combined with the remaining partial system states to develop a better estimation using the original estimator design in [8], which is an implementation of the XOR operation.

IV. Case Studies

To illustrate the efficacy of the proposed graph-based deep learning-driven predictor for synchrophasor recovery and prediction, we conduct a series of comprehensive simulations. We first introduce the employed testing systems and simulation configurations. Then, the detailed synchrophasor accuracy and system response time is demonstrated. Next, we investigate the impact of communication latency and measurement noise on the predictor. Finally, we study the capability of the predictor in adapting to power system changes. All case studies are conducted on computing servers with an Intel Core-i7 CPU at 4.7 GHz and 32 GB RAM. Additionally, nVidia GTX 1080Ti GPUs are employed for neural network computation acceleration. The testing code is developed in Python, and PyTorch [32] is used to construct the proposed neural network.

A. Test Systems and Configurations

In this work, we adopt three test systems to thoroughly investigate the system performance on different scales of power grids. In particular, the New England 10-machine system [33], Nordic system [34], and Iceland network system [35] are employed, each of which is equipped with $39/74/118$ buses, $10/20/35$ synchronous generators, and $34/102/206$ power lines, respectively. Classical PMU placement techniques, e.g., [36], are utilized to install PMUs to achieve $N-1$ observability.

We consider a variety of system dynamics of disturbances and faults. For each system, we first generate 100 random operating conditions by setting load level of each bus to 70% and 120% of the respective nominal value and use optimal power flow to establish the conditions with infeasible ones discarded. Subsequently, 20 000 one-, two-, and three-phase to ground short circuit contingencies are generated based on a random condition. The contingencies are cleared with a random fault clearance time from $0.1\,\mathrm{s}$ to $0.4\,\mathrm{s}$. PMU measurements are constructed using the respective voltages in the dynamics, which are developed by DSATools [37]. In practice, the synthetic data can be replaced or enriched by historical operational data of power grids. Furthermore, we adopt the real latency values in FNET/GridEye project [38]. For generated measurements in one system state instance, the corresponding latency values are randomly selected from those of all frequency disturbance recorders in the project at an arbitrary time. This configuration accords with the previous work [8]. Unless otherwise stated, we use the implementation proposed in [8] as the baseline performance.

To adjust the network parameters in the proposed neural network, 20 000 dynamics for each test system are divided into two data sets by the ratio of $3:1$, which accords with the common practice [39], [40]. The training set with 15000 dynamics is adopted to train the parameters as elaborated in Section III-D, while the other one, called testing set, is used to evaluate the system performance. This design is for cross-validation of the proposed model, and overfitting problem can be easily identified where the training and testing set performances deviate significantly.

B. Accuracy of Synchrophasors

We first investigate the accuracy of the generated synchrophasors by the proposed approach compared with the ground truth data. We adopt the total vector error (TVE) of voltage phasors in S_τ and $\hat{S}_\tau$ as the performance metric. Let $S_\tau = \{\mathrm{V}_{i,\tau}\angle\theta_{i,\tau}|\,\forall i\in\mathcal{V}\}$, $\hat{S}_\tau = \{\hat{\mathrm{V}}_{i,\tau}\angle\hat{\theta}_{i,\tau}|\,\forall i\in\mathcal{V}\}$ where $\mathcal{V}$ is the set of buses in the power system. TVE is defined by

$$\mathrm{TVE}\,(S_\tau,\hat{S}_\tau)=\frac{1}{|\mathcal{V}'_\tau|}\sum_{i\in\mathcal{V}'_\tau}\frac{|\hat{\mathrm{V}}_{i,\tau}\angle\hat{\theta}_{i,\tau}-\hat{\mathrm{V}}_{i,\tau}\angle\hat{\theta}_{i,\tau}|}{|\hat{\mathrm{V}}_{i,\tau}\angle\hat{\theta}_{i,\tau}|},\quad(15)$$

where $\mathcal{V}'_\tau\in\mathcal{V}$ is the set of buses whose measurements are not available at the control center, i.e., $S_\tau\setminus S^-_\tau$. This set is dynamic with respect to different time instances.

The system is trained according to the configurations introduced in Section III-D, and three models are trained for all three test systems. The TVEs for the proposed approach as well as the baseline performance are presented in Table I, and the 99.9% percentile TVE performance of New England system is demonstrated in Table II. In the table, we use $\tau=0$ to represent the current time. The accuracy of recovered synchrophasors in the past and predicted ones in the future are summarized, and the training set performance is also presented

[3]In this work, the probability is set to 30% in the last three fully-connected neuron layers.

TABLE I
TOTAL VECTOR ERROR OF PREDICTED SYNCHROPHASORS

τ	Proposed (Testing Set)			Baseline (SRPF)			Proposed (Training Set)			Unavailable Data Percentage
	New England	Nordic	Iceland	New England	Nordic	Iceland	New England	Nordic	Iceland	
-5	0.02%	0.03%	0.02%	0.02%	0.36%	0.44%	0.02%	0.03%	0.02%	2.49%
-4	0.03%	0.04%	0.04%	0.02%	0.41%	0.47%	0.03%	0.04%	0.04%	6.17%
-3	0.11%	0.12%	0.12%	0.11%	0.50%	0.54%	0.11%	0.12%	0.11%	50.43%
-2	0.56%	0.58%	0.59%	0.96%	1.14%	1.20%	0.53%	0.52%	0.54%	73.98%
-1	0.59%	0.60%	0.62%	1.03%	1.57%	1.63%	0.56%	0.59%	0.55%	94.50%
0	0.62%	0.65%	0.67%	1.04%	1.72%	1.79%	0.59%	0.63%	0.65%	99.86%
1	0.66%	0.70%	0.67%	1.07%	1.80%	1.96%	0.63%	0.67%	0.65%	100%
2	0.73%	0.75%	0.76%	1.10%	1.89%	2.13%	0.63%	0.70%	0.72%	
3	0.75%	0.75%	0.77%	1.11%	1.94%	2.21%	0.68%	0.73%	0.72%	
4	0.76%	0.78%	0.80%	1.12%	2.01%	2.37%	0.70%	0.73%	0.73%	
5	0.89%	0.94%	0.94%	1.14%	2.10%	2.44%	0.89%	0.87%	0.82%	
6 to 10	0.94%	0.99%	1.01%	1.34%	2.63%	2.91%	0.92%	0.95%	0.93%	
11 to 15	1.01%	1.07%	1.05%	2.08%	3.54%	4.07%	0.96%	1.04%	0.95%	

TABLE II
99.9% PERCENTILE TVE OF SYNCHROPHASOR RECOVERY ACCURACY ON NEW ENGLAND 10-MACHINE SYSTEM

τ	99.9% Percentile Total Vector Error		
	New England	Nordic	Iceland
-5	0.02%	0.03%	0.03%
-4	0.03%	0.04%	0.04%
-3	0.11%	0.13%	0.12%
-2	0.59%	0.64%	0.61%
-1	0.61%	0.71%	0.65%
0	0.68%	0.71%	0.68%
1	0.76%	0.71%	0.73%
2	0.74%	0.83%	0.79%
3	0.77%	0.88%	0.81%
4	0.83%	0.93%	0.83%
5	0.93%	0.98%	1.03%
6 to 10	1.00%	1.10%	1.06%
11 to 15	1.07%	1.17%	1.14%

TABLE III
COMPARISON OF SYNCHROPHASOR RECOVERY ACCURACY ON NEW ENGLAND 10-MACHINE SYSTEM

τ	Average Total Vector Error				Recoverable Cases
	Proposed	ICMC	SVT	Hankel St.	ICMC & SVT
-5	0.02%	0.01%	0.12%	0.05%	100.0%
-4	0.03%	0.14%	0.42%	0.09%	99.8%
-3	0.11%	1.46%	0.99%	0.47%	76.5%
-2	0.56%	3.43%	1.15%	0.91%	49.9%
-1	0.59%	4.69%	1.90%	1.36%	9.0%
0	0.62%	4.30%	1.73%	1.50%	1.8%
1 to 5	0.76%			1.58%	0.0%
6 to 10	0.94%	Un-recoverable		1.71%	
11 to 15	1.01%			1.86%	

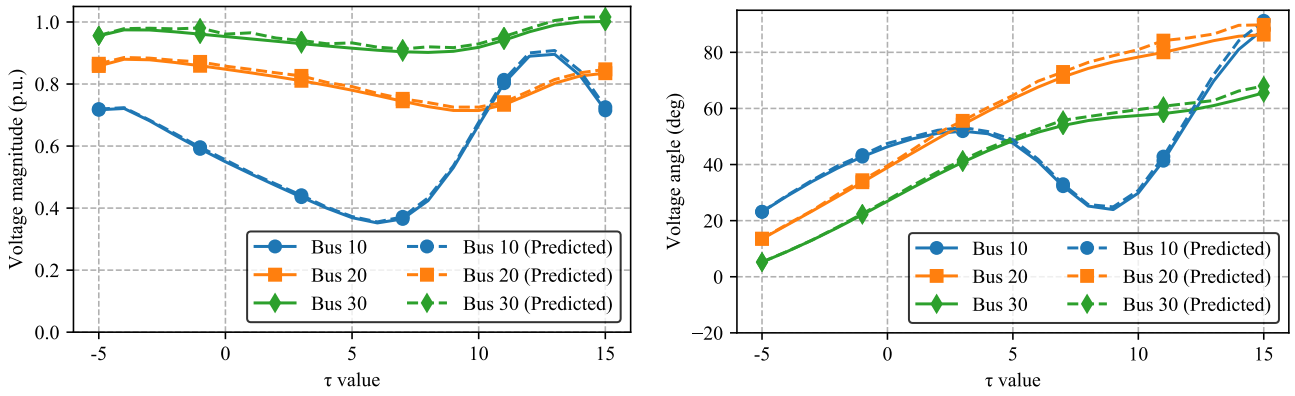
for reference. TVE values of $6 \leq \tau \leq 15$ are averaged for conciseness.

We first investigate the performance on the New England 10-machine system. From the simulation results, it can be concluded that the proposed graph-based approach significantly outperforms the baseline performance. For either compared approaches, TVE is closely related to the total number of unavailable data: the more unknown, the worse the accuracy. However, the proposed approach can develop much better synchrophasor predictions than the baseline for $\tau \geq -2$, where almost no measurements are received or sampled. This is contributed to by the incorporation of graph topology information into the learning process, which conventionally must be implicitly learnt from the training data. With this essential information, the system can greatly benefit from the small prediction accuracy improvement on each measurement. Since prediction errors in previous time steps are progressively strengthened in subsequent steps, the small improvements accumulate and constitute the significant performance gap on large τ values. Fig. 4 also demonstrates an illustration of the recovered and predicted synchrophasors compared with the ground truth data on a random contingency of the New England 10-machine system. The dynamics also support our previous discussion.

When comparing the other two test systems, the improvement of the proposed approach is more remarkable. Furthermore, it also possess a critical characteristic that can greatly enhance its practical implementation, i.e., system scalability up to more than 100 buses. For the baseline approach, the system performance deteriorates when the system size is increased. In both Nordic and Iceland systems, TVE ranges from 1.8% to more than 4% for synchrophasor predictions in the upcoming fifteen cycles. While the information can still demonstrate the general trend of the system dynamics, it can barely be used in accurate system analysis and control applications. Nonetheless, this drawback is addressed in the proposed approach. While the accuracy on Nordic and Iceland systems are also undermined by the notably increased system size, the performance gap is minuscule. This is due to the graph abstraction capability of the proposed GRAN. Recall the propagation rule of GRL (4) and (5), one may notice that the dimensions of weight parameters W and U are not related to the graph size B. This implies that other than the power system size, the system dynamics characteristics also plays a critical role in adjusting the weights. Therefore, moderately increasing the size of the system (up to few hundred buses) does not have noteworthy influence on the overall performance. We will further investigate this property of GRAN in Section IV-D.

We also present the TVE comparison of GRAN with matrix completion methods in the literature, namely, information cascading matrix completion (ICMC) [13], singular value thresholding (SVT) [13], and low-rank Hankel structure based approach (labeled as "Hankel St.") [16] in Table III. As illustrated in [13], ICMC and SVT can only recover system states with at least one available measurement, and therefore

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2899395, IEEE Internet of Things Journal



(a) Voltage magnitude (b) Voltage angle

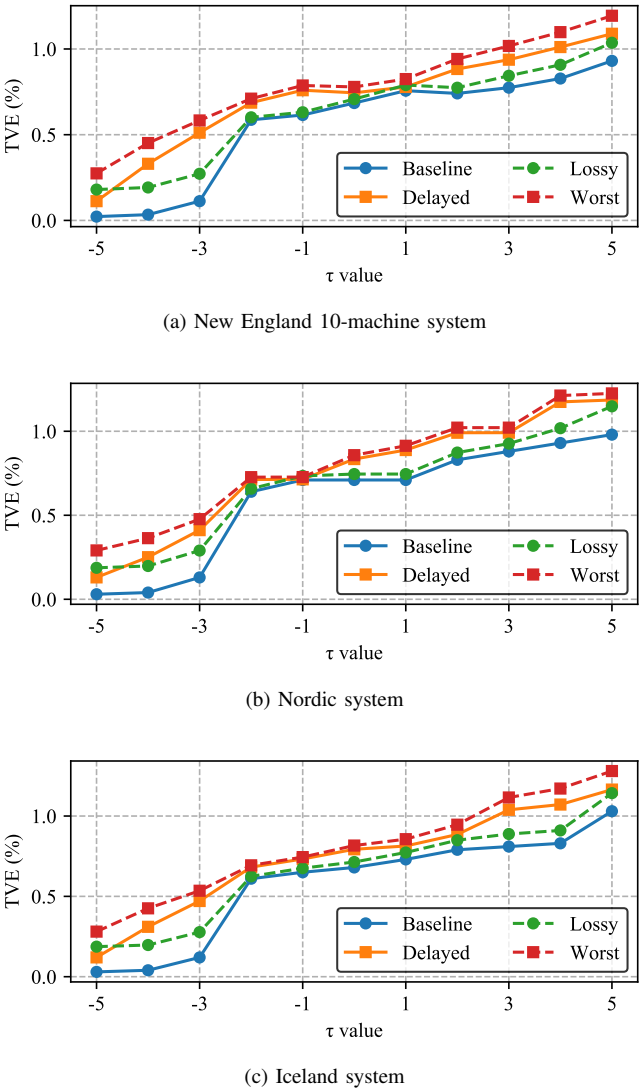
Fig. 4. An example of recovered and predicted voltage phasors of buses 10, 20, 30 in New England 10-machine system and their ground truths.

cannot be used to predict future states. Nonetheless, the approach proposed in [16] employs an iterative approach to gradually recover the missing data points in the Hankel matrix of synchrophasors, and thus can predict future data. The parameters are selected according to the recommendations in the literature. From the results it is obvious that GRAN outperforms all compared algorithms in most tasks, with the only exception at $\tau = -5$ where ICMC develops more accurate system states. However, considering that the TVE values are minuscule (approx. 0.02%), the improvement is not significant. Furthermore, as GRAN inherits the capability of recovering all past states and predicting future ones from SRPF, it is more capable of handling missing data caused by communication latencies in WAMS.

Finally, we summarize the computation time of training GRAN and making predictions. The training takes approximately $14\,\mathrm{h}$, $14\,\mathrm{h}$, and $15\,\mathrm{h}$ on the three test systems with $15\,000$ contingencies on eight testing GPUs, respectively. Considering that significant power network changes are typically well-planned by the system operators days earlier, the proposed GRAN can be re-trained with new training sets in sufficiently short time. We will further investigate the system performance in Section IV-D on small power network changes, which may not render enough time for a complete re-training. Furthermore, it takes less than $1\,\mathrm{ms}$ to recover one past system states or predict a future one with the proposed GRAN on all three test systems with one testing GPU. As modern WAMS develops $60\,\mathrm{Hz}$ samples to the control center, the small inference time required by GRAN makes it sufficient for real-time analysis.

C. Impact of Latency and Measurement Noise

In the previous case study, we employ the real world latency values from FNET/GridEye project, which represents the communication performance of a typical WAMS. Meanwhile, with the wide adoption of PMUs into modern power systems, the underlying communication infrastructure may have different latency properties. In this sub-section, we first assess the impact of bad communication links on the synchrophasor prediction accuracy. Similar to [8], we manipulate the real latency



(a) New England 10-machine system

(b) Nordic system

(c) Iceland system

Fig. 5. Total vector error on compared latency scenarios.

TABLE IV
TOTAL VECTOR ERROR WITH NOISY MEASUREMENTS ON NORDIC SYSTEM

τ	Testing Set	Baseline	Training Set	Noise TVE
-5 to -1	0.69%	1.18%	0.64%	0.59%
0	0.72%	1.96%	0.67%	
1 to 5	0.85%	2.14%	0.81%	-
6 to 10	1.10%	2.97%	1.03%	
11 to 15	1.21%	4.04%	1.15%	

TABLE V
TOTAL VECTOR ERROR OF TESTING DATA WITH VARIOUS NOISE LEVELS ON NORDIC SYSTEM

τ	Up to 0.5%	Up to 1.0%	Up to 1.5%	Up to 2.0%
-5 to -1	0.43%	0.69%	1.00%	1.43%
0	0.44%	0.72%	1.05%	1.50%
1 to 5	0.57%	0.85%	1.22%	1.68%
6 to 10	1.05%	1.10%	1.46%	1.88%
11 to 15	1.13%	1.21%	1.63%	2.06%

values adopted in the previous test (labeled by "baseline") and construct the following communication scenarios to emulate WAMS with bad data transmission conditions:

- *Delayed:* All latency values are increased by 50%.
- *Lossy:* Randomly 30% of all PMU measurements are lost due to communication problems.
- *Worst:* Both modifications in *delayed* and *lossy* scenarios.

The trained systems in Section IV-B is employed to make predictions based on the new latency scenarios. The results are depicted in Fig. 5.

From the results it can be observed that bad communication conditions have negative influences on the accuracy of the system, though not significant. The performance degradation compared with the baseline scenario is similar to that recorded in [8]. This is due to two factors. On the one hand, the modular sub-system design of SRPF can partially mitigate the influence of deteriorated communications since the predictor always looks back in the history to find the most recent complete system state. All subsequent predictions are based on valid and true information. Given that the prediction accuracy is satisfactory, high latency or data loss does not introduce critical challenges to the system. On the other hand, the proposed graph-based approach inherits the robustness of deep neural networks [22]. This is among the main reasons of adopting deep learning techniques in designing the system.

Besides communication latency, measurement noise is another uncertainty in WAMS [5]. According to IEEE Std. C37.118.2-2011 [41], measurements sampled by complied PMUs may contain up to 1% TVE noise. This imposes great challenges to accuracy-sensitive power system applications. In this test, we investigate the influence of PMU measurement noise on the synchrophasor prediction accuracy. In accordance with previous work [5], [11], [42], we sample a noise phasor from a truncated complex Gaussian distribution for each measurement developed in Section IV-A. The noise phasor is imposed on the corresponding ground truth one to formulate new training and testing cases, which are subsequently employed to train and assess the system performance on the Nordic system. The predicted synchrophasors are compared with the ground truth ones, and the accuracy is presented in Table IV. In the table, averaged TVE of the noise phasor is also listed for reference.

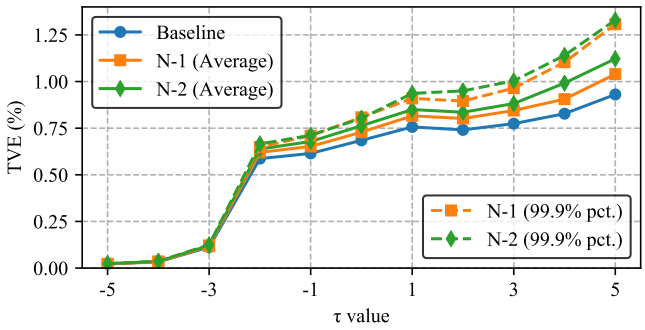
From the summarized results it can be concluded that the influence of noisy measurements is notable only when recovering synchrophasors in the past, i.e., $\tau \leq 0$. This can be observed by comparing Tables I and IV. Nonetheless, the performance degradation for $\tau \leq 0$ is mainly caused by the uncertainty of noise itself, which constitutes 0.59% TVE. To explore the reason insignificant noise-caused error occurs for $\tau > 0$, we conduct a series of simulations and change the maximum noise TVE to 0.5%, 1.5%, and 2.0%. The results are presented in Table V. From this table it is clear that when the noise is small (e.g., maximum 0.5% TVE), the error is mainly contributed to by the prediction error of GRAN. For large noises, the dominating factor becomes the noise error. Nonetheless, as standard PMUs can develop less than 1% TVE noise [41], the performance degradation is still tolerable. In addition, the recovery error is consistent with the noise level despite its changes. This shows the robustness of the proposed approach, which benefits from the extra graph information in the training data and the adversarial model.

D. Impact of Power Network Topology Changes

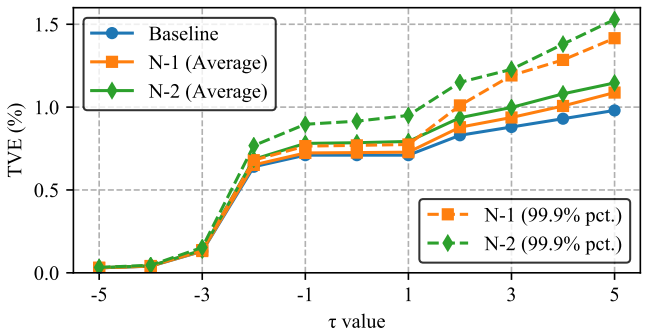
In Section IV-B we discussed the capability of handling power systems of different sizes with GRAN. In practice, the topology of real-world power systems is not static. It is quite common that power grids operate under $N-1$ or occasionally $N-2$ conditions. Due to the numerous types of $N-k$ contingencies, it is almost impossible to enumerate all system dynamics in these conditions to train the prediction system. Hence, it is of critical importance that the system adapts to power network changes automatically.

In this sub-section, we investigate the synchrophasor accuracy developed by the system when power grids undergo contingencies on random $N-1$ and $N-2$ topologies. Particularly, we select 25 $N-1$ and 25 $N-2$ stable operating conditions for each of the testing system. In each condition, 100 short circuit contingencies are randomly generated in accordance with the methodology introduced in Section IV-A, resulting in 5000 new $N-k$ contingency dynamics. We adopt the trained system in Section IV-B to predict the missing synchrophasors in these dynamics to evaluate the impact of power network topology changes. Note that since the new data is not used in training, the trained system does not have any knowledge on the network topology changes.

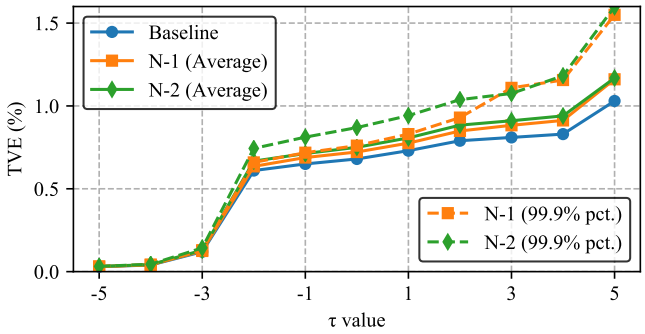
Fig. 6 depicts the average and 99.9% percentile TVE of the system on the compared operating conditions. In the figure, "baseline" refers to the original testing data performance as in Section IV-B. The simulation results imply that GRAN possesses great generalization capability. In all three test systems, $N-k$ topology changes do not have noteworthy performance impact. This result is mainly contributed to by the inference robustness of GRAN thanks to the graph-learning characteristics of the neural network architecture. Furthermore, a careful investigation in the system dynamics of the generated $N-k$ contingencies also suggests that the topology changes



(a) New England 10-machine system



(b) Nordic system



(c) Iceland system

Fig. 6. Average and 99.9% percentile total vector error on compared topology changes.

do not lead to dramatic changes on the dynamics. This is also a critical factor that leads to the satisfactory performance as shown in Fig. 6. On the one hand, $N-1$ and $N-2$ operating conditions do not impose drastic change on the physical characteristics of power networks. As a result, the system dynamics of an event on $N-k$ networks with small k values are similar to that on the nominal one. On the other hand, instead of storing the topology in layers of GRAN, the information is only used to help train the neural network which extracts and captures high-level characteristics of the topology. This means that after training, if a new system with similar topological characteristics is employed, the system can still yield satisfactory results. This is exactly the case in $N-1$ and $N-2$ operating conditions. Both factors contribute to the insignificant performance degradation of GRAN on changed power system topology.

To conclude, when undergoing $N-k$ operations, utilities can use the nominal system for synchrophasor prediction, which yields sub-optimal performance. At the same time, new systems can be trained in parallel to cope with the new topology if the outage is confirmed to be prolonged. No service outage will be experienced with this operating paradigm. In addition, the 99.9% percentile TVE values deviate from the average ones by less than 0.5%. This statistical indicator suggests that the proposed system can develop generally stable recovery and prediction accuracy under moderate topology changes.

We also adopt the baseline approach in [8] to handle the same topology changes. With more than 1% TVE increase for $\tau = 0$ in all three test systems, the performance are not as satisfactory. Consequently, we can conclude that the proposed system can better handle power network topology changes without re-training the system.

V. Conclusions

In this paper, we propose a new graph-based deep neural network for predicting synchrophasors in a synchrophasor recovery and prediction framework to address the data transmission latency issues in modern wide area measurement systems. The proposed system utilizes tailor-made neural network structures to learn from both the graphical (network) and temporal characteristics of the power system dynamics. We incorporate the design principle of recent advancements in deep learning techniques and propose a new graph-convolutional recurrent layer, which is capable of explicitly including graph topologies in the learning process. Furthermore, a graph-convolutional recurrent adversarial network is constructed based on the new neural network layer to make predictions on power system states based on existing network and dynamics information. To assess the efficacy of the proposed system, a series of comprehensive case studies are conducted on three test systems of difference sizes. The results indicate the outstanding accuracy of the proposed system over the existing implementation. Additionally, we investigate the robustness of the proposed system on bad communications and measurement noise, which do not influence the system notably. Finally, we use a case study to conclude that the proposed system can address topology changes well without re-training.

References

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[2] G. Bedi, G. K. Venayagamoorthy, R. Singh, R. R. Brooks, and K. Wang, "Review of internet of things (IoT) in electric power and energy systems," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 847–870, Apr. 2018.

[3] J. Bertsch, C. Carnal, D. Karlson, J. McDaniel, and K. Vu, "Wide-area protection and power system utilization," *Proc. IEEE*, vol. 93, no. 5, pp. 997–1003, May 2005.

[4] Q. Yang, L. Jiang, W. Hao, B. Zhou, P. Yang, and Z. Lv, "PMU placement in electric transmission networks for reliable state estimation against false data injection attacks," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1978–1986, Dec. 2017.

[5] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam, J. Gu, and V. O. K. Li, "Intelligent time-adaptive transient stability assessment system," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1049–1058, 2018.

[6] W. Liao, S. Salinas, M. Li, P. Li, and K. A. Loparo, "Cascading failure attacks in the power system: A stochastic game perspective," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2247–2259, Dec. 2017.

[7] N. Bui, A. P. Castellani, P. Casari, and M. Zorzi, "The internet of energy: a web-enabled smart grid system," *IEEE Netw.*, vol. 26, no. 4, pp. 39–45, Jul. 2012.

[8] J. J. Q. Yu, A. Y. S. Lam, D. J. Hill, Y. Hou, and V. O. K. Li, "Delay aware power system synchrophasor recovery and prediction framework," *IEEE Trans. Smart Grid*, in press.

[9] R. Quint, K. Thomas, A. Silverstein, and D. Kosterev, "Synchrophasor maturity model – extracting value from synchrophasor investments," North American Synchrophasor Initiative, Tech. Rep., Mar. 2015.

[10] L. E. Miller, A. Silverstein, D. Anand, A. Goldstein, Y. Makarov, F. Tuffner, and K. Jones, "PARTF: PMU data quality: a framework for the attributes of PMU data quality and a methodology for examining data quality impacts to synchrophasor applications," North American Synchrophasor Initiative, Tech. Rep. NASPI-2017-TR-002, Mar. 2017.

[11] J. J. Q. Yu, A. Y. S. Lam, D. J. Hill, and V. O. K. Li, "Delay aware intelligent transient stability assessment system," *IEEE Access*, vol. 5, pp. 17 230–17 239, 2017.

[12] B. P. Padhy, S. C. Srivastava, and N. K. Verma, "A wide-area damping controller considering network input and output delays and packet drop," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 166–176, Jan. 2017.

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

[14] Y. Seyedi, H. Karimi, and J. M. Guerrero, "Centralized disturbance detection in smart microgrids with noisy and intermittent synchrophasor data," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2775–2783, Nov. 2017.

[15] J. J. Q. Yu and A. Y. S. Lam, "Autonomous vehicle logistic system: joint routing and charging strategy," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2175–2187, Jul. 2018.

[16] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, "Modelless data quality improvement of streaming synchrophasor measurements by exploiting the low-rank Hankel structure," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6966–6977, Nov 2018.

[17] J. J. Q. Yu, D. J. Hill, and A. Y. S. Lam, "Delay aware transient stability assessment with synchrophasor recovery and prediction framework," *Neurocomputing*, vol. 322, pp. 187–194, 2018.

[18] R. Zhang, Y. Xu, Z. Y. Dong, and K. P. Wong, "Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system," *IET Generation, Transmission and Distribution*, vol. 9, no. 3, pp. 296–305, 2015.

[19] D. R. Gurusinghe and A. D. Rajapakse, "Post-disturbance transient stability status prediction using synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3656–3664, Sep. 2016.

[20] F. Aminifar, M. Fotuhi-Firuzabad, A. Safdarian, A. Davoudi, and M. Shahidehpour, "Synchrophasor measurement technology in power systems: panorama and state-of-the-art," *IEEE Access*, vol. 2, pp. 1607–1628, 2014.

[21] A. G. Phadke, B. Pickett, M. Adamiak, M. Begovic, G. Benmouyal, R. O. Burnett, T. W. Cease, J. Goossens, D. J. Hansen, M. Kezunovic, L. L. Mankoff, P. G. McLaren, G. Michel, R. J. Murphy, J. Nordstrom, M. S. Sachdev, H. S. Smith, J. S. Thorp, M. Trotignon, T. C. Wang, and M. A. Xavier, "Synchronized sampling and phasor measurements for relaying and control," *IEEE Trans. Power Del.*, vol. 9, no. 1, pp. 442–452, Jan. 1994.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, F. Bach, Ed. Cambridge, MA: MIT Press, Nov. 2016.

[23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. International Conference on Learning Representations*, Toulon, France, Apr. 2017.

[24] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Advances in Neural Information Processing Systems*, Barcelona, Spain, Dec. 2016.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[27] I. Goodfellow, "NIPS 2016 tutorial: generative adversarial networks," 2016, arXiv preprint arXiv:1701.00160.

[28] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. International Conference on Artificial Intelligence and Statistics*, Ft. Lauderdale, FL, Apr. 2011, pp. 315–323.

[29] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. International Conference on Machine Learning*, Sydney, Australia, Aug. 2017, pp. 214–223.

[30] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *Proc. International Conference on Learning Representations*, Toulon, France, Apr. 2017.

[31] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, Jul. 2012.

[32] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. Advances in Neural Information Processing Systems*, Long Beach, CA, Dec. 2017.

[33] M. A. Pai, *Energy function analysis for power system stability*. Springer Science & Business Media, 2012.

[34] "Test systems for voltage stability analysis and security assessment," IEEE Power & Energy Society, Power System Dynamic Performance Committee, Tech. Rep. PES-TR19, Aug. 2015.

[35] F. Milano, L. Vanfretti, and J. C. Morataya, "An open source power system virtual laboratory: the PSAT case and experience," *IEEE Trans. Educ.*, vol. 51, no. 1, pp. 17–23, Feb. 2008.

[36] J. J. Q. Yu, A. Y. S. Lam, D. J. Hill, and V. O. K. Li, "A unified framework for wide area measurement system planning," *International Journal of Electrical Power & Energy Systems*, vol. 96, pp. 43 – 51, 2018.

[37] "Dynamic security assessment software," accessed Jan. 2018. [Online]. Available: http://www.dsatools.com/

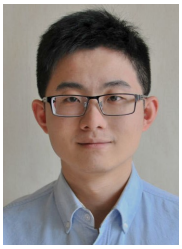
[38] "FNET server web display," accessed Apr. 2018. [Online]. Available: http://fnetpublic.utk.edu/

[39] J. J. Q. Yu, Y. Hou, A. Y. S. Lam, and V. O. K. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Trans. Smart Grid*, in press.

[40] J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Online false data injection attack detection with wavelet transform and deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 2371–2380, Jul. 2018.

[41] "IEEE Standard for Synchrophasor Data Transfer for Power Systems," IEEE Std C37.118.2-2011, Dec. 2011.

[42] T. Guo and J. Milanovic, "Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 377–385, Jan. 2014.



James J.Q. Yu (S'11–M'15) received the B.Eng. and Ph.D. degree in Electrical and Electronic Engineering from the University of Hong Kong, Pokfulam, Hong Kong, in 2011 and 2015, respectively. He was a post-doctoral fellow at the University of Hong Kong from 2015 to 2018. He is currently an assistant professor at the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China, and an honorary assistant professor at the Department of Electrical and Electronic Engineering, the University of Hong Kong. He is also the chief research consultant of GWGrid Inc. and external consultant of Fano Labs. His current research interests include smart city technologies, deep learning and big data, intelligent transportation systems, and energy systems.

David J. Hill (S'72-M'76-SM'91-F'93-LF'14) received the PhD degree in Electrical Engineering from the University of Newcastle, Australia, in 1976. He holds the Chair of Electrical Engineering in the Department of Electrical and Electronic Engineering at the University of Hong Kong. He is also a part-time Professor of Electrical Engineering at The University of Sydney, Australia. During 2005-2010, he was an Australian Research Council Federation Fellow at the Australian National University. Since 1994, he has held various positions at the University of Sydney, Australia, including the Chair of Electrical Engineering until 2002 and again during 2010-2013 along with an ARC Professorial Fellowship. He has also held academic and substantial visiting positions at the universities of Melbourne, California (Berkeley), Newcastle (Australia), Lund (Sweden), Munich and in Hong Kong (City and Polytechnic).

His general research interests are in control systems, complex networks, power systems and stability analysis. His work is now mainly on control and planning of future energy networks and basic stability and control questions for dynamic networks. Professor Hill is a Fellow of the Society for Industrial and Applied Mathematics, USA, the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering and the Hong Kong Academy of Engineering Sciences.. He is also a Foreign Member of the Royal Swedish Academy of Engineering Sciences.

Victor O.K. Li (S'80–M'81–F'92) received SB, SM, EE and ScD degrees in Electrical Engineering and Computer Science from MIT. Prof. Li is Chair of Information Engineering and Cheng Yu-Tung Professor in Sustainable Development at the Department of Electrical & Electronic Engineering (EEE) at the University of Hong Kong. He is the Director of the HKU-Cambridge Clean Energy and Environment Research Platform, an interdisciplinary collaboration with Cambridge. He was the Head of EEE, Assoc. Dean (Research) of Engineering and Managing Director of Versitech Ltd. He serves on the board of Sunevision Holdings Ltd., listed on the Hong Kong Stock Exchange and co-founded Fano Labs Ltd., an artificial intelligence (AI) company with his PhD student. Previously, he was Professor of Electrical Engineering at the University of Southern California (USC), Los Angeles, California, USA, and Director of the USC Communication Sciences Institute. His research interests include big data, AI, optimization techniques, and applications of AI to important societal problems. In Jan 2018, he was awarded a USD 6.4M RGC Theme-based Research Project to develop deep learning techniques for personalized and smart air pollution monitoring and health management. Sought by government, industry, and academic organizations, he has lectured and consulted extensively internationally. He has received numerous awards, including the PRC Ministry of Education Changjiang Chair Professorship at Tsinghua University, the UK Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of the HKSAR. He is a Fellow of the Hong Kong Academy of Engineering Sciences, the IEEE, the IAE, and the HKIE.

Yunhe Hou (M'08–SM'15) received the B.E. and Ph.D. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2005, respectively. He was a Post-Doctoral Research Fellow at Tsinghua University, Beijing, China, from 2005 to 2007, and a Post-Doctoral Researcher at Iowa State University, Ames, IA, USA, and the University College Dublin, Dublin, Ireland, from 2008 to 2009. He was also a Visiting Scientist at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2010. Since 2017, he has been a Guest Professor with Huazhong University of Science and Technology, China. He joined the faculty of the University of Hong Kong, Hong Kong, in 2009, where he is currently an Associate Professor with the Department of Electrical and Electronic Engineering. Dr. Hou is an Editor of the IEEE TRANSACTIONS ON SMART GRID and JOURNAL OF MODERN POWER SYSTEMS AND CLEAN ENERGY.