LEARNING TO INFER POWER GRID TOPOLOGIES: PERFORMANCE AND SCALABILITY

Yue Zhao[*]*, Jianshu Chen*[†] *and H. Vincent Poor*[‡]

[*]Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794
[†]Tencent AI Lab, Seattle, WA 98004
[‡]Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
Emails: yue.zhao.2@stonybrook.edu, jianshuchen@tencent.com, poor@princeton.edu

ABSTRACT

Identifying arbitrary topologies of power networks in real time is a computationally hard problem due to the number of hypotheses that grows exponentially with the network size. A "Learning-to-Infer" variational inference method is employed for efficient inference of every line status in the network. Optimizing the variational model is transformed to and solved as a discriminative learning problem based on Monte Carlo samples generated with power flow simulations. As the labeled data used for training can be generated in an arbitrarily large amount rapidly and at very little cost, the power of offline training is fully exploited to learn very complex classifiers for effective real-time topology identification. The Learning-to-Infer method is extensively evaluated in the IEEE 30, 118 and 300 bus systems. Excellent performance and scalability of the method in identifying arbitrary power network topologies in real time are demonstrated.

***Index Terms*—** power grid topology identification, line outage detection, machine learning, cascading failures

1. INTRODUCTION

Lack of situational awareness in abnormal system conditions is a major cause of blackouts in power networks [1]. Network component failures such as transmission line outages, if not rapidly identified and contained, can quickly escalate to cascading failures. In particular, when line failures happen, the power network topology changes instantly, newly stressed areas can unexpectedly emerge, and subsequent failures may be triggered that lead to increasingly complex network topology changes. While the power system is usually protected against the so called "$N-1$" failure scenarios, as failures accumulate, effective automatic protection is no longer guaranteed. Thus, when cascading failures start developing, smarter protective actions critically depend on correct and timely knowledge of the network status. Indeed, without knowledge of the network topology changes, protective control methods have been observed to further aggravate the failure scenarios [2]. Thus, real-time network topology identification is essential to all network control decisions for mitigating failures. In particular, since the first few line outages may have already been missed, the ability to identify in real time the network topology with an *arbitrary* number of line outages becomes critical to prevent system collapse.

Real-time topology identification is however a very challenging problem, especially when unknown line statuses in the network quickly accumulate as in scenarios that cause large-scale blackouts [1]. *The number of possible topologies grows exponentially with the number of unknown line statuses, making real time topology identification fundamentally hard.* Assuming a small number of line failures, exhaustive search methods have been developed in [3], [4], [5] and [6] based on hypothesis testing, and in [7] based on logistic regression. To overcome the prohibitive computational complexity of exhaustive search methods, [8] has developed sparsity exploiting outage identification methods with overcomplete observations to identify sparse multi-line outages. Without assuming sparsity of line outages, a graphical model based approach has been developed for identifying arbitrary network topologies [9]. On the other hand, *non-real-time* topology identification has also been extensively studied [10, 11, 12].

In this paper, we focus on *real-time* identification of *arbitrary* grid topologies based on instantly collected measurements in the power system. We formulate the topology identification problem in a Bayesian inference framework, where we aim to compute the posterior probabilities of the topologies given any instant measurements. To overcome the fundamental computational complexity due to the *exponentially large number of possible topologies*, we employ a "Learning-to-Infer" variational inference framework [13]. In particular, to find effective end-to-end variational models, we transform optimizing a variational model to a discriminative learning problem leveraging a Monte Carlo approach: a) based on *power flow equations*, data samples of network topology, network states, and sensor measurements in the network can be efficiently generated according to a *generative model* of these

This work was supported in part by the U.S. National Science Foundation under Grant DMS-1736417.

quantities; and b) with these simulated data, *discriminative models* are learned *offline*, which then offer *real-time* prediction of the network topology based on newly observed measurements from the real network.

We extensively evaluate the Learning-to-Infer method in the IEEE 30, 118, and 300 bus systems [14] for identifying topologies with an *arbitrary* number of line outages. It is demonstrated that, even with relatively simple variational models and a reasonably small amount of data, the performance is surprisingly good for this very challenging task. Moreover, the Learning-to-Infer method is demonstrated to scale efficiently as the size of the network increases.

2. PROBLEM FORMULATION

2.1. Power Flow and Observation Models

We consider a power system with N buses, and its *baseline topology* (i.e., the network topology when there is no line outage) with L lines. We denote the incidence matrix of the baseline topology by $M \in \{-1, 0, 1\}^{N \times L}$. We use a binary variable s_l to denote the status of a line l, with $s_l = 1$ for a connected line l, and 0 otherwise. The actual topology of the network can then be represented by $\boldsymbol{s} = [s_1, \ldots, s_L]^T$. We employ the DC power flow model, and denote the real power injections and the phase angles at all the buses by $\boldsymbol{P} \in \mathbb{R}^N$ and $\boldsymbol{\theta} \in \mathbb{R}^N$, respectively. The topology $\boldsymbol{s}$, nodal power injections and voltage phase angles satisfy the following [15]:

$$\boldsymbol{P} = M S \Gamma M^T \boldsymbol{\theta}, \tag{1}$$

where $S = \operatorname{diag}(s_1, \ldots, s_L)$, $\Gamma = \operatorname{diag}(\frac{1}{x_1}, \ldots, \frac{1}{x_L})$, and x_l is the reactance of line l.

To monitor the power system, we consider real time measurements taken by sensors measuring nodal voltage phase angles, and nodal real power injections. In particular, voltage phase angles measured by phasor measurement units (PMUs) located at a subset of the buses $\mathcal{M}$ can be modeled as

$$\boldsymbol{y} = \boldsymbol{\theta}_{\mathcal{M}} + \boldsymbol{v}, \tag{2}$$

where $\boldsymbol{\theta}_{\mathcal{M}}$ is formed by entries of $\boldsymbol{\theta}$ from buses in $\mathcal{M}$, and $\boldsymbol{v}$ captures the measurement noise.

2.2. Topology Identification as Bayesian Inference

We are interested in identifying the network topology $\boldsymbol{s}$ *in real time* based on instant measurements $\boldsymbol{y}$ collected in the power system. We formulate this as a *Bayesian inference* problem. Consider a joint probability distribution of $\boldsymbol{s}, \boldsymbol{P}$ and $\boldsymbol{y}$,

$$p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y}) = p(\boldsymbol{s}, \boldsymbol{P}) \cdot p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{P}). \tag{3}$$

We are interested in computing the posterior *conditional* probabilities: $\forall \boldsymbol{s}$,

$$p(\boldsymbol{s}|\boldsymbol{y}) = \frac{\int p(\boldsymbol{s}, \boldsymbol{P}) p(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{P}) d\boldsymbol{P}}{p(\boldsymbol{y})}, \tag{4}$$

and in particular, the *posterior marginal conditional probabilities* $p(s_l|\boldsymbol{y}), l = 1, \ldots, L$. However, computing the posterior marginals $p(s_l|\boldsymbol{y})$ is fundamentally intractable: even with $p(\boldsymbol{s}|\boldsymbol{y})$ given, summing out all $s_k, k \neq l$, to obtain $p(s_l|\boldsymbol{y})$ still requires exponential computational complexity [16].

3. A LEARNING-TO-INFER METHOD

To overcome the exponential computational complexity, we employ a "Learning-to-Infer" method for approximate inference of $p(s_l|\boldsymbol{y}), l = 1, \ldots, L$ [13]. The general idea is to find a variational conditional distribution $q(\boldsymbol{s}|\boldsymbol{y})$ that a) approximates the original $p(\boldsymbol{s}|\boldsymbol{y})$ very closely, and b) offers fast and accurate topology identification results. In particular, we assume that $q(\boldsymbol{s}|\boldsymbol{y})$ is modeled by some parametric form (e.g., neural networks), denoted by $\{q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})\}$, where $\boldsymbol{\beta}$ is a vector of model parameters. From a family of parametrized distributions $\{q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})\}$, we would like to choose a $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})$ that approximates $p(\boldsymbol{s}|\boldsymbol{y})$ as closely as possible, by minimizing the *expected Kullback-Leibler (KL) divergence* as follows:

$$\min_{\boldsymbol{\beta}} \mathbb{E}_{\boldsymbol{y}} \left[D(p \| q_{\boldsymbol{\beta}}) \right] \Leftrightarrow \min_{\boldsymbol{\beta}} \sum_{\boldsymbol{s}, \boldsymbol{y}} p(\boldsymbol{s}, \boldsymbol{y}) \log \frac{p(\boldsymbol{s}|\boldsymbol{y})}{q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})}$$
$$\Leftrightarrow \max_{\boldsymbol{\beta}} \mathbb{E}_{\boldsymbol{s}, \boldsymbol{y}} \left[\log q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}) \right], \tag{5}$$

where the expectation is taken with respect to the *true* distribution $p(\boldsymbol{s}, \boldsymbol{y})$. To evaluate $\mathbb{E}_{\boldsymbol{s}, \boldsymbol{y}} \left[\log q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y}) \right]$, we *approximate the expectation by the empirical mean* of $\log q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})$ over a large number of *Monte Carlo samples*, generated according to the *true* joint probability $p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y})$ (cf. (3)). We denote the relevant Monte Carlo samples by $\{\boldsymbol{s}^t, \boldsymbol{y}^t; t = 1, \ldots, T\}$. Accordingly, (5) is approximated by

$$\max_{\boldsymbol{\beta}} \frac{1}{T} \sum_{t=1}^{T} \log q_{\boldsymbol{\beta}}(\boldsymbol{s}^t|\boldsymbol{y}^t)]. \tag{6}$$

(6) can be viewed as an *empirical risk minimization* problem in machine learning [17], as it trains a *discriminative model* $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})$ with a data set $\{\boldsymbol{s}^t, \boldsymbol{y}^t\}$ generated from a *generative model* $p(\boldsymbol{s}, \boldsymbol{P}, \boldsymbol{y})$. As a result of this offline training process (6), an approximate posterior function $q_{\boldsymbol{\beta}^*}(\boldsymbol{s}|\boldsymbol{y})$ is obtained.

One great advantage of this Learning-to-Infer method is that we can generate *labeled data* very efficiently. As a result, we can obtain *an arbitrarily large set of data at very little cost* to train the discriminative model. This is quite different from the typical situations encountered in machine learning problems, where obtaining a large amount of labeled data is usually expensive as it requires extensive human annotation effort. Furthermore, once the approximate posterior distribution $q_{\boldsymbol{\beta}}(\boldsymbol{s}|\boldsymbol{y})$ is learned, it can be deployed to infer the power grid topology *in real-time* as the computational complexity of $q_{\boldsymbol{\beta}}(s_l|\boldsymbol{y})$ is very low by design.

Table 1: Data set size vs. the entire search space

The (reduced) IEEE 30 bus system with 38 lines	
Number of all topologies	$2^{38} = 2.75 \times 10^{11}$
Number of topologies with 8 disconnected lines	$\binom{38}{8} = 4.89 \times 10^7$
The generated data set	3×10^5

The (reduced) IEEE 118 bus system with 170 lines	
Number of all topologies	$2^{170} = 1.50 \times 10^{51}$
Number of topologies with 13 disconnected lines	$\binom{170}{13} = 9.94 \times 10^{18}$
The generated data set	8×10^5

The (reduced) IEEE 300 bus system with 322 lines	
Number of all topologies	$2^{322} = 8.54 \times 10^{96}$
Number of topologies with 12 disconnected lines	$\binom{322}{12} = 2.11 \times 10^{21}$
The generated data set	2.2×10^6

4. NUMERICAL EXPERIMENTS

We evaluate the Learning-to-Infer method for real-time grid topology identification with three benchmark systems of increasing sizes, the IEEE 30, 118, and 300 bus systems, as the baseline topologies. As opposed to considering just a small number of simultaneous line outages as in existing works, we allow *any number* of line outages, and investigate whether the learned classifiers can successfully recover the topologies.

4.1. Data Set Generation

In our experiments, we employ the DC power flow model (1) to generate the data sets. To generate a data set $\{\boldsymbol{s}^t, \boldsymbol{P}^t, \boldsymbol{y}^t, t = 1, \ldots, T\}$, we assume the prior distribution $p(\boldsymbol{s}, \boldsymbol{P})$ factors as $p(\boldsymbol{s})p(\boldsymbol{P})$. As such, we generate the network topologies $\boldsymbol{s}$ and the power injections $\boldsymbol{P}$ independently:

- We generate the line statuses $\{s_l\}$ using independent and identically distributed (IID) Bernoulli random variables, with $\mathbb{P}(s_l = 1) = 0.6, 0.9$ and 0.96 for the IEEE 30, 118, and 300 bus systems, respectively. We do not consider disconnected networks in this study, and exclude the line status samples if they lead to disconnected networks. As such, after some network reduction, the equivalent networks for the IEEE 30, 118, and 300 bus systems have $38, 170$, and 322 lines that can possibly be in outage, respectively.

- We would like our predictor to be able to identify the topology for *arbitrary values of power injections* as opposed to fixed ones. Accordingly, we generate $\boldsymbol{P}$ using the following procedure: For each data sample, we first generate bus voltage phase angles $\boldsymbol{\theta}$ as IID uniformly distributed random variables in $[0, 0.2\pi]$, and then compute $\boldsymbol{P}$ according to (1) under the baseline topologies.

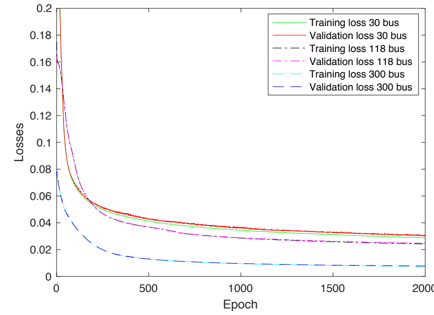


Fig. 1: Progressions of training and validation losses, IEEE 30, 118 and 300 bus systems.

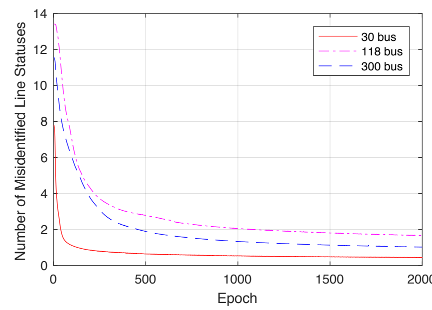


Fig. 2: Progressions of average numbers of misidentified line statuses, IEEE 30, 118 and 300 bus systems.

With each pair of generated $\boldsymbol{s}^t$ and $\boldsymbol{P}^t$, we consider two types of measurements that constitute $\boldsymbol{y}$: a) voltage phase angle measurements corrupted by IID Gaussian noises with a standard deviation of 0.01 degree, the state-of-the-art PMU accuracy [18], and b) power injections measured accurately.

In this study, we generate $300K$, $800K$, and $2.2M$ data samples for the IEEE 30, 118, and 300 bus systems, respectively. These $300K/800K/2.2M$ data are further divided into $200K/600K/1.8M$, $50K/100K/200K$, and $50K/100K/200K$ samples for training, validation, and testing, respectively. We note that *over* 99% of the generated $300K$ 30-bus topologies are distinct from each other, so are those of the generated $800K$ 118 bus topologies and those of the $2.2M$ 300 bus topologies. As a result, these generated data sets can very well evaluate the *generalizability* of the trained classifiers, as (almost) all data samples in the test sets have topologies *unseen* in the training sets.

Moreover, in the generated data sets, the *average numbers of disconnected lines* relative to the baseline topology are $7.8, 13.4$ and 11.6 for the IEEE 30, 118 and 300 bus systems, respectively. These numbers of simultaneous line outages are significantly higher than those typically assumed in sparse line outage studies. Furthermore, we would like to compare the *size of the generated data set* to the *total num-*

ber of possible topology hypotheses, as highlighted in Table 1. Clearly, a) it is computationally prohibitive to perform line status inference based on exhaustive search, and b) *the generated* $300K, 800K$ *and* $2.2M$ *data sets are only a tiny fraction of the entire space of all topologies*. Yet, we will show that the classifiers trained with the generated data sets exhibit excellent inference performance and generalizability.

4.2. Performance Evaluation

For the parametric model $q_{\beta}(\boldsymbol{s}|\boldsymbol{y})$, in this study, we employ two-layer (i.e., one hidden layer) fully connected neural networks. Rectified Linear Units (ReLUs) are employed as the activation functions in the hidden layer. In the output layer we employ hinge loss as the loss function. In training the classifiers, we use stochastic gradient descent (SGD) with momentum update and Nesterov's acceleration [19].

4.2.1. Performance of the Learning-to-Infer Method

We begin with $300, 1000$ and 3000 neurons in the hidden layer for the IEEE 30, 118 and 300 bus systems, respectively. For all the three systems, we plot in Figure 1 the achieved training and testing losses for every epoch. It is clear that the training and testing losses stay very close to each other for all the three systems, and thus no overfitting is observed. We observe very high testing accuracies, 0.989, 0.990 and 0.997 achieved for the IEEE 30, 118 and 300 bus systems, respectively. These can be equivalently understood by the *average numbers of misidentified line statuses*, plotted in Figure 2. We observe that, *at the beginning* of the training procedures, the average numbers of misidentified line statuses are 7.8, 13.4 and 11.6 for the IEEE 30, 118 and 300 bus systems, which are exactly the *average numbers of disconnected lines* in the respective generated data sets (cf. Section 4.1). Indeed, this coincides with the result from a naive identification decision rule of always claiming all the lines as connected (i.e., a trivial majority rule). As the training procedures progress, the average numbers of misidentified line statuses are drastically reduced to eventually 0.4, 1.7 and 1.0.

4.2.2. Model Size, Sample Complexity, and Scalability

In the proposed Learning-to-Infer method, obtaining labeled data is not an issue since data can be generated in an arbitrarily large amount using Monte Carlo simulations. This leads to two questions that are of particular interest: to learn a good classifier, a) what size of a neural network is needed? and b) how much data needs to be generated? To answer these questions, we vary the sizes of the hidden layer of the neural networks as well as the training data size, and evaluate the learned classifiers for the three benchmark systems. We plot the testing results for the IEEE 30, 118 and 300 bus systems in Figure 3(a), 3(b) and 3(c), respectively.

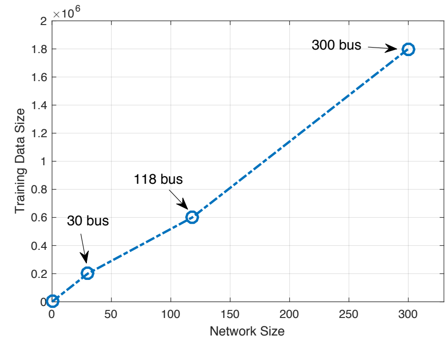


Fig. 4: Scalability of the Learning-to-Infer method, from the IEEE 30 bus system to the IEEE 300 bus system.

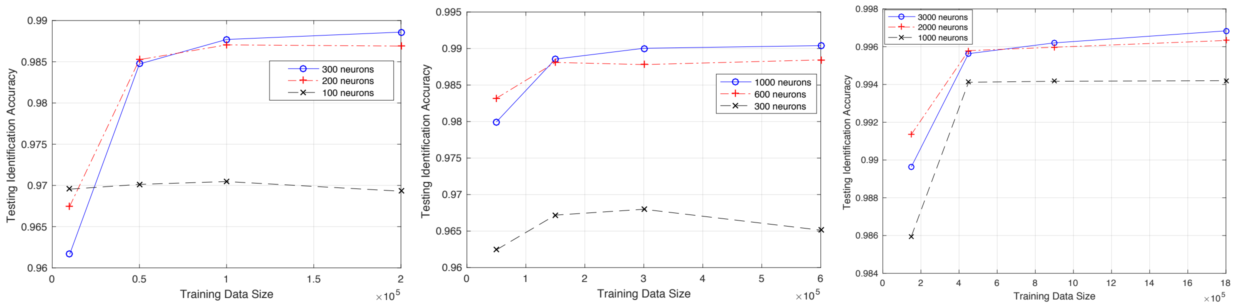
Based on all these experiments, we now examine the *scalability* of the proposed Learning-to-Infer method as the problem size increases. We observe that training data sizes of $200K, 600K$ and $1.8M$ and neural network models of sizes 300, 1000 and 3000 ensure very high and comparable performance with no overfitting for the IEEE 30, 118 and 300 bus systems, respectively. When these data sizes are reduced by a half, some levels of overfitting then appeared for these models in all the three systems. We plot the training data sizes compared to the problem sizes for the three systems in Figure 4. We observe that the required training data size increases approximately *linearly* with the problem size. This linear scaling behavior implies that the proposed Learning-to-Infer method can be effectively implemented for large-scale systems with reasonable computation resources.

5. CONCLUSION

We have employed a Learning-to-Infer variational inference method for real-time topology identification of power grids. The computational complexity due to the exponentially large number of topology hypotheses is overcome by efficient marginal inference with optimized variational models. Optimization of the variational model is transformed to and solved as a discriminative learning problem, based on Monte Carlo samples efficiently generated with power flow models. With the classifiers learned offline, their actual use is in real time, and topology identification decisions are made under a millisecond. We have extensively evaluated the Learning-to-Infer method with the IEEE 30, 118 and 300 bus systems. It has been demonstrated that arbitrary network topologies can be identified in real time with excellent performance using classifiers trained with a reasonably small amount of generated data. The method has been demonstrated to scale efficiently as the network size increases.

6. REFERENCES

[1] *US-Canada Power System Outage Task Force, Final Re-*



(a) (b) (c)

Fig. 3: Effect of model size and sample size, (a) IEEE 30 bus system, (b) IEEE 118 bus system, (c) IEEE 300 bus system.

port on the August 14, 2003 Blackout in the United States and Canada, 2004.

[2] *Arizona-Southern California Outages on September 8, 2011: Causes and Recommendations*, FERC, NERC, 2012.

[3] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1644 – 1652, Nov. 2008.

[4] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *Proceedings of the IEEE Power and Energy Society General Meeting*, July 2009.

[5] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor, "Identification of outages in power systems with uncertain states and optimal sensor locations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1140–1153, Dec. 2014.

[6] Y. Zhao, A. Goldsmith, and H. V. Poor, "On PMU location selection for line outage detection in wide-area transmission networks," in *Proceedings of the IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[7] M. Garcia, T. Catanach, S. Vander Wiel, R. Bent, and E. Lawrence, "Line outage localization using phasor measurement data in transient state," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3019–3027, 2016.

[8] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, Nov. 2012.

[9] J. Chen, Y. Zhao, A. Goldsmith, and H. V. Poor, "Line outage detection in power transmission networks via message passing algorithms," in *Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 350–354.

[10] Xiao Li, H Vincent Poor, and Anna Scaglione, "Blind topology identification for power systems," in *Proceedings of the IEEE International Conference on Smart Grid Communications*, 2013, pp. 91–96.

[11] Vassilis Kekatos, Georgios B Giannakis, and Ross Baldick, "Online energy price matrix factorization for power grid topology tracking," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1239–1248, 2016.

[12] Ye Yuan, Omid Ardakanian, Steven Low, and Claire Tomlin, "On the inverse power flow problem," *arXiv preprint arXiv:1610.06631*, 2016.

[13] Y. Zhao, J. Chen, and H. V. Poor, "Learning to infer: A new variational inference approach for power grid topology identification," in *Proceedings of the IEEE Workshop on Statistical Signal Processing*, June 2016, pp. 1–5.

[14] *Power Systems Test Case Archive*, University of Washington Electrical Engineering, https://www.ee.washington.edu/research/pstca/.

[15] J. D. Glover, M. Sarma, and T. Overbye, *Power System Analysis & Design*, Cengage Learning, 2011.

[16] Marc Mezard and Andrea Montanari, *Information, Physics, and Computation*, Oxford University Press, 2009.

[17] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.

[18] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Proceedings of the IEEE Innovative Smart Grid Technologies (ISGT)*, 2013.

[19] Yurii Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87, Springer Science & Business Media, 2013.