Big Data Analytics for Enhancing Distribution Grid Modeling

Zhaoyu Wang

Northrop Grumman Associate Professor

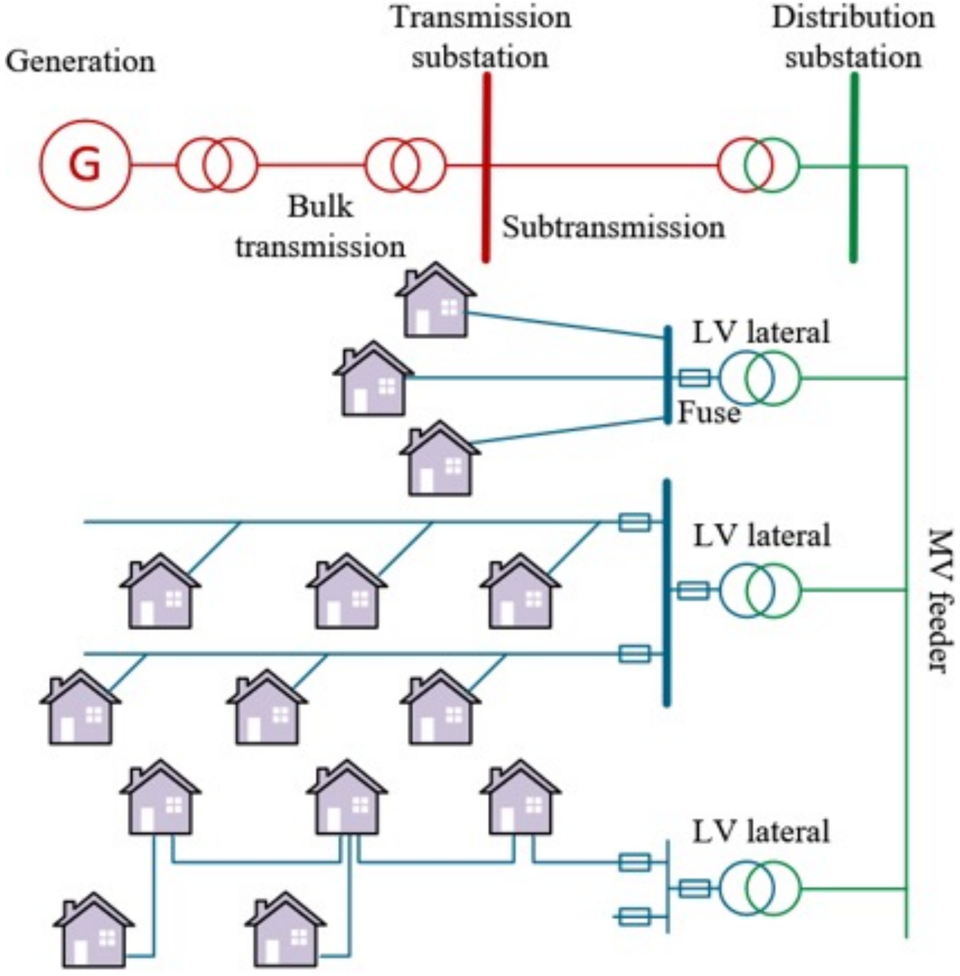
Iowa State University

(wzy@iastate.edu)

Presentation Outline

- Introduction to Smart Meter Data
- Distribution Grid Topology and Line Parameter Identification
  - Existing Work and Challenges
  - Robust Network Topology Identification via Weighted Laplacian Matrix
  - A Bottom-Up Sweep Approach for Line Impedance Estimation
- Conclusion and Future Work

Data in Power Distribution Grids



A Power distribution grid

Where Does The Data Come From?

SCADA
System for monitoring and controlling industrial processes. Typically installed at centralized control centers or remote monitoring stations.

Smart Meters
Digital devices for 15/30/60 mins energy and voltage measurement. Installed at consumer premises to measure individual usage.

Protection Devices
Electronics safeguarding electrical systems from faults and abnormal conditions.

Micro-PMUs
High-precision but expensive devices measuring power system characteristics for grid stability analysis.

Exemplary Real Data from Utilities

Time

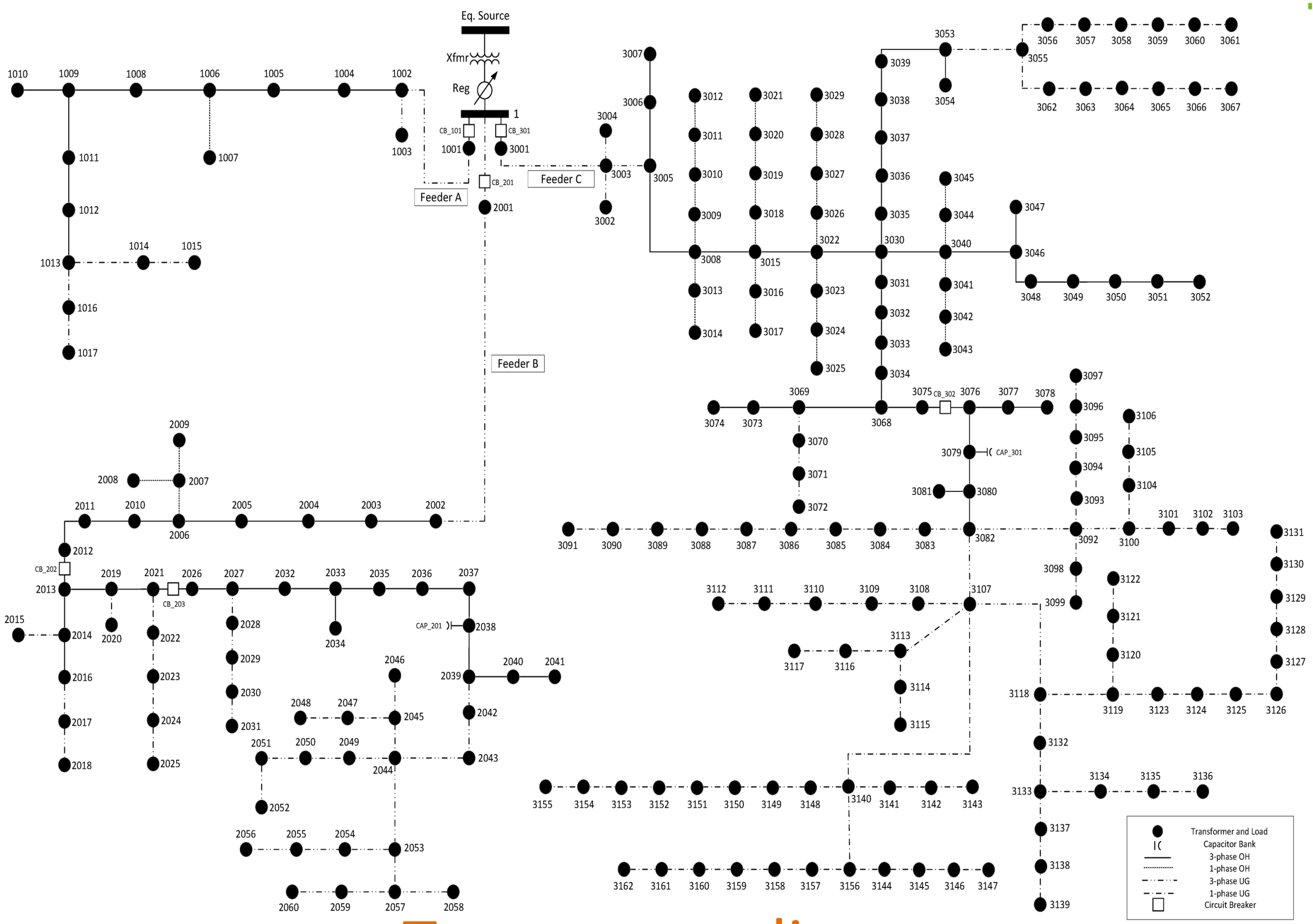
Hourly energy & instantaneous voltage

Account		time	kWH or V	time	kWH or V	time	kWH or V	time	kWH or V
100000001	KWH	201704010100	0.392	201704010200	0.257	201704010300	0.215	201704010400	0.239
100000001	VOLTS	201704010100	239	201704010200	239	201704010300	238	201704010400	240
100000002	KWH	201704010100	0.245	201704010200	0.204	201704010300	0.252	201704010400	0.342
100000002	VOLTS	201704010100	241	201704010200	240	201704010300	240	201704010400	240
100000003	KWH	201704010100	1.479	201704010200	0.417	201704010300	0.816	201704010400	0.414
100000003	VOLTS	201704010100	240	201704010200	239	201704010300	239	201704010400	240
100000004	KWH	201704010100	1.009	201704010200	0.555	201704010300	0.39	201704010400	0.382
100000004	VOLTS	201704010100	241	201704010200	237	201704010300	237	201704010400	239
100000005	KWH	201704010100	0.798	201704010200	0.809	201704010300	0.87	201704010400	0.692
100000005	VOLTS	201704010100	239	201704010200	238	201704010300	238	201704010400	240
100000006	KWH	201704010100	0.109	201704010200	0.188	201704010300	0.205	201704010400	0.148
100000006	VOLTS	201704010100	241	201704010200	240	201704010300	240	201704010400	242
100000007	KWH	201704010100	1.199	201704010200	1.512	201704010300	1.759	201704010400	1.474
100000007	VOLTS	201704010100	241	201704010200	240	201704010300	239	201704010400	241
100000008	KWH	201704010100	0.422	201704010200	0.419	201704010300	0.43	201704010400	0.537
100000008	VOLTS	201704010100	239	201704010200	239	201704010300	238	201704010400	240
100000009	KWH	201704010100	2.288	201704010200	2.278	201704010300	2.335	201704010400	2.297
100000009	VOLTS	201704010100	243	201704010200	242	201704010300	242	201704010400	242
100000010	KWH	201704010100	0.223	201704010200	0.257	201704010300	0.292	201704010400	0.25
100000010	VOLTS	201704010100	242	201704010200	241	201704010300	241	201704010400	241

Distribution System Data Sharing

With permission from our utility partner, we share a real distribution grid model with one-year smart meter measurements. This dataset provides an opportunity for researchers and engineers to perform validation and demonstration using real utility grid models and field measurements.

- The system consists of 3 feeders and 240 nodes and is located in Midwest U.S.
- The system has 1120 customers and all of them are equipped with smart meters. These smart meters measure hourly energy consumption (kWh). We share the one-year real smart meter measurements for 2017.
- The system has standard electric components such as overhead lines, underground cables, substation transformers with LTC, line switches, capacitor banks, and secondary distribution transformers. The real system topology and component parameters are included.



Test system diagram

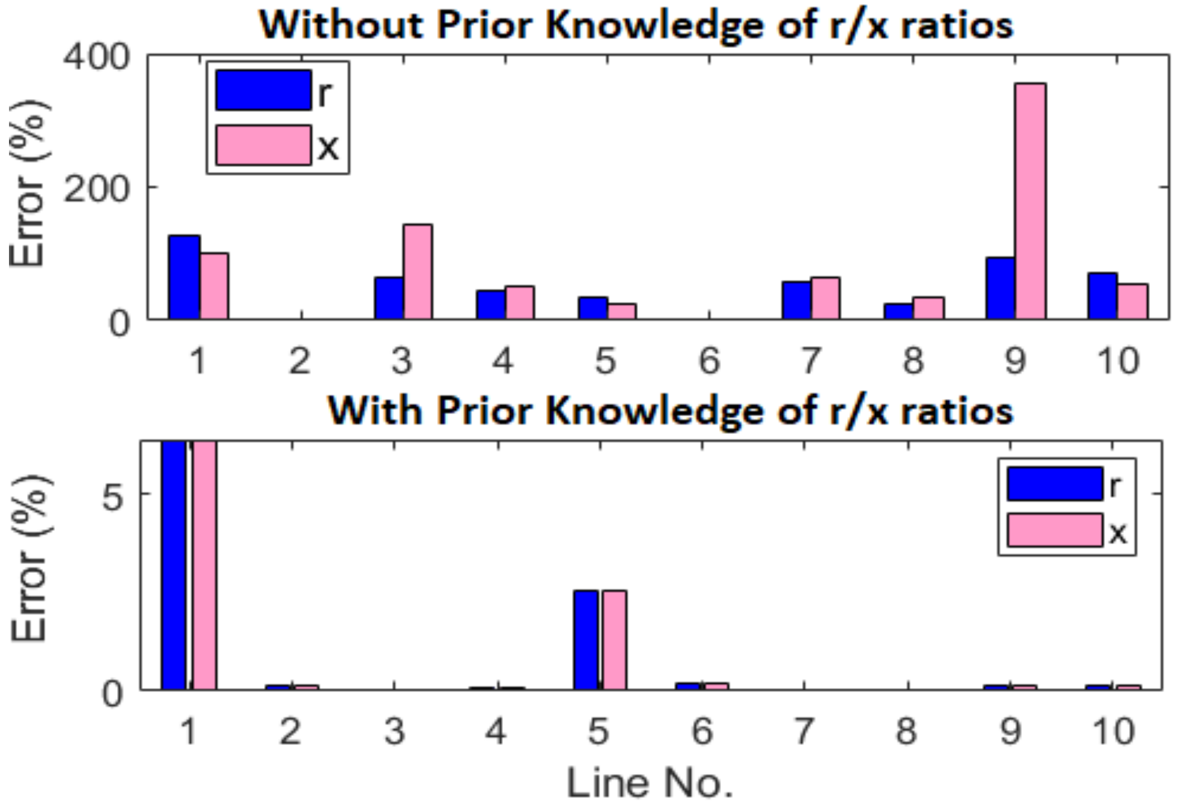
You may download the dataset at: http://wzy.ece.iastate.edu/Testsystem.html , including system description (in .doc and .xlsx), smart meter data (in .xlsx), OpenDSS model, and Matlab code for quasi-static time-series simulation.

Distribution Grid Topology & Parameter Identification

- Complete and accurate distribution grid models are essential to system monitoring and control.
- Many small and medium utilities only have simple one-line diagrams of their systems without any detailed information.
- System models are often incomplete or outdated due to the frequent system expansion and reconfiguration.
- Conventional field inspection is laborious, costly, and time-consuming, especially for large-scale systems.

Existing Work and Challenges

- Using Y-bus injection model and phasor information (J. Yu 19, O. Ardakanian 19, Y. Yuan 20)
  - **Limitation**: require full coverage of μPMUs (cost-prohibitive).
- Using Branch flow model and smart meter data (A. M. Prostejovsky 16, H. Xu 18, W. Wang 20)
  - **Limitation**: require prior knowledge (i.e., R/X ratios of ***all*** line sections and network connectivity).
  - **Reason for this requirement**: searching space of the optimization (***ill-conditioned***).
  - **Another challenge**: scalability and computational complexity.



How to perform real-time topology and parameter identification using very limited yet readily available SM data?

Our Solution

✓ **Topology Identification**: Modeling the distribution network as a graph and identify its *weighted Laplacian matrix* using SM data streams, where the matrix has a special structure that reveals the network connectivity.

- ✓High computational efficiency.
- ✓Robustness with respect to heterogeneous R/X ratios and model/measurement errors.

✓ **Parameter Estimation:** designing a *bottom-up sweep algorithm* to identify line impedances.

- ✓Based on the full nonlinear power flow, a least absolute deviations (LAD) with mixed-integer semidefinite programming (MISDP) model, and a least square (LS) model with mixed-integer second-order cone programming (MISOCP) model have been developed.
- ✓Adding a library of R/X ratios (rather than exact R/X of all line sections) as a constraint to narrow down the search space.
- ✓Dividing the network into multiple layers. Parameter identification and power flow calculations are performed layer-by-layer in an alternate manner from bottom to top layers.

Y. Guo, Y. Yuan, and Z. Wang, "Distribution Grid Modeling Using Smart Meter Data," arXiv preprint arXiv:2103.00660, 2021.

Distribution Grid Topology Identification

- Our topology identification approach builds on the linear approximation of the branch flow model.

$$\mathbf{v} \cong 2\mathbf{A}^{-T}\mathbf{R}\mathbf{A}^{-1}\mathbf{p} + 2\mathbf{A}^{-T}\mathbf{X}\mathbf{A}^{-1}\mathbf{q} - v_0\mathbf{A}^{-T}\mathbf{a}_0 \qquad (1)$$

  - For a radial distribution network, A is non-singular and $\mathbf{A}^{-T}\mathbf{a}_0 = 1_n$

$$\frac{1}{2}\mathbf{A}\mathbf{X}^{-1}\mathbf{A}^{T}(\mathbf{v} - v_0 1_n) = \mathbf{A}\mathbf{X}^{-1}\mathbf{R}\mathbf{A}^{-1}\mathbf{p} + \mathbf{q} \qquad (2)$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X}^{-1}\mathbf{A}^{T};\ \boldsymbol{\Phi} = \mathbf{A}\mathbf{X}^{-1}\mathbf{R}\mathbf{A}^{-1} \qquad (3)$$

where $\mathbf{v}$, $\mathbf{p}$, $\mathbf{q}$ denote the vectors collecting squared bus voltage magnitudes, real power, and reactive power injections; $[a_0, A^T]^T \in \{0, \pm 1\}^{(n+1)\times n}$ is the incidence matrix of the radial-topology graph; **R** and **X** are diagonal resistance and reactance matrices; **Y** is a *weighted Laplacian* matrix of the network with a sparse structure.

Weighted Laplacian Matrix of the Network

Proposition 1: $\mathbf{Y} \coloneqq \left[y_{ij}\right]_{n\times n}$ is a sparse symmetric matrix :

$$\begin{bmatrix} 3 & -1 & 0 & 0 & -2 & 0 \\ -1 & 5 & -4 & 0 & 0 & 0 \\ 0 & -4 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & -1 & -5 \\ -2 & 0 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -5 & 0 & 5 \end{bmatrix} \qquad (4)$$

- $\mathbf{Y}$ holds a salient feature: for any connected nodes *i* and *j*, $y_{ij} < 0$ and for any non-connected nodes $y_{ij} = 0$.

- If one can approximately identify **Y**, the topology can be extracted by observing the unique features of **Y**.

Weighted Laplacian Matrix Identification

- Assume the network has a homogeneous R/X ratio, $\frac{r_1}{x_1} = \cdots = \frac{r_n}{x_n} = \lambda$, $\boldsymbol{\Phi}$ reduces to $\boldsymbol{\Phi} = \mathbf{A}\mathrm{diag}\left(\frac{r_1}{x_1}, \ldots, \frac{r_n}{x_n}\right)\mathbf{A}^{-1} = \lambda 1_n$.
- For ***heterogeneous*** networks, our method still works because we do not require accurate estimation of $\mathbf{Y}$ and only need to distinguish zero and negative non-diagonal entries to identify connectivity, which will be proved in case study.
- The error vector regarding k-th measurement can be defined based on (2).

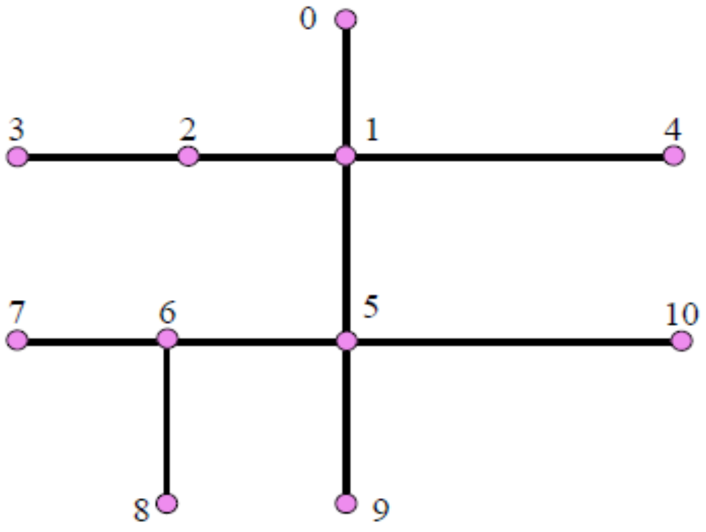
$$\frac{1}{2}\mathbf{A}\mathbf{X}^{-1}\mathbf{A}^{T}(\mathbf{v} - v_0 1_n) = \mathbf{A}\mathbf{X}^{-1}\mathbf{R}\mathbf{A}^{-1}\mathbf{p} + \mathbf{q} \Rightarrow \frac{1}{2}\boldsymbol{Y}(\mathbf{v} - v_0 1_n) = \lambda\mathbf{p} + \mathbf{q}$$

$$e^k := \mathbf{Y}\left(\mathbf{v}^k - v_0^k 1_n\right) - 2\lambda\mathbf{p}^k - 2\mathbf{q}^k \qquad (5)$$

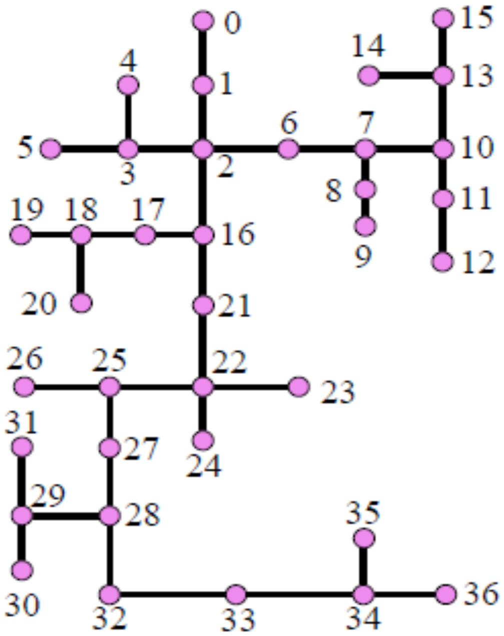
- **Our Model**: Based on (5) and a time window of length *K*, we develop a linear LS regression mode to estimate Y.

$$\min_{Y,\lambda} \|[e^1, \ldots, e^K]\|_2^2 \qquad (6)$$

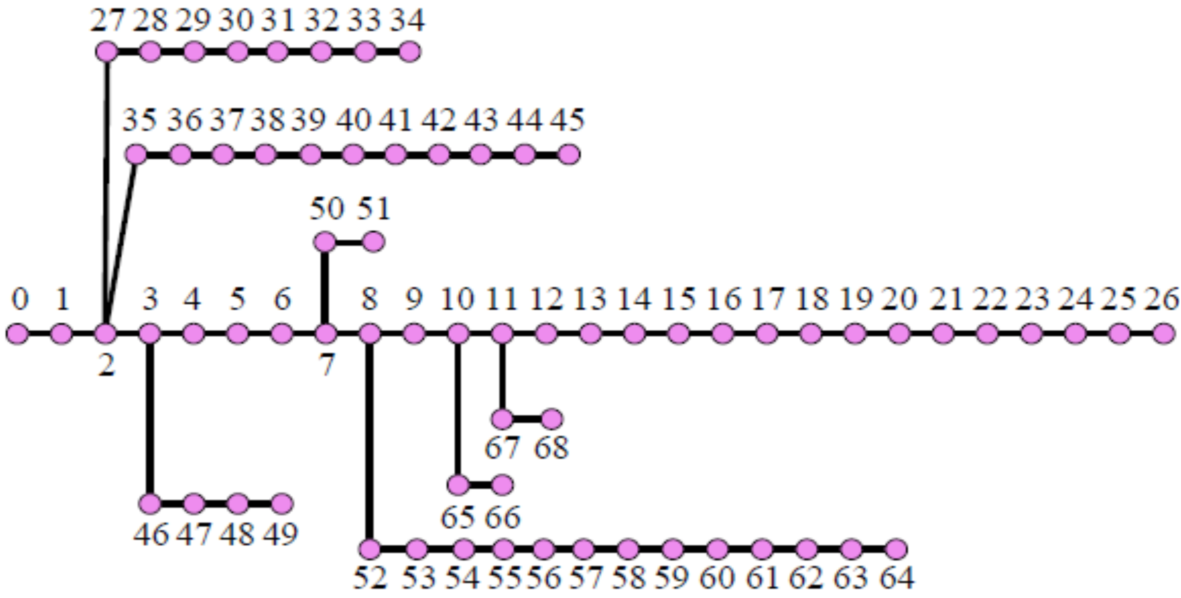
Estimated Weighted Laplacian Matrix for IEEE 13-, 37-, 69-Bus Test Feeder



13-Bus Test Feeder



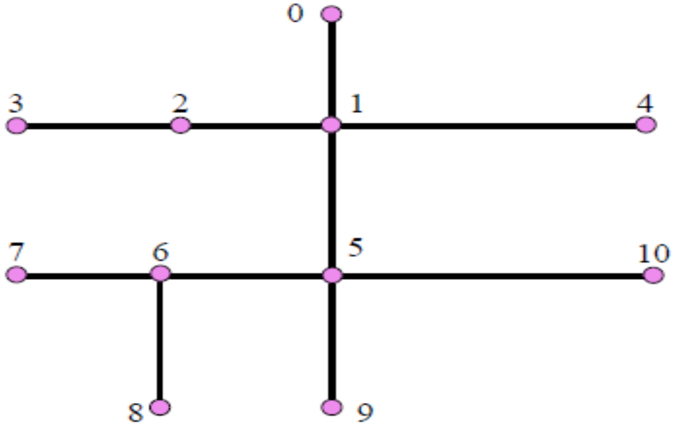
37-Bus Test Feeder



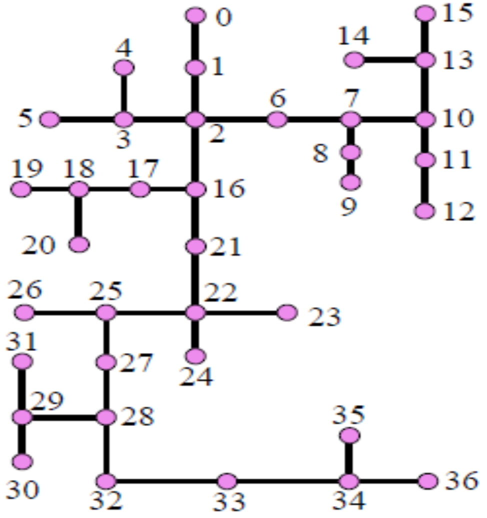
69-Bus Test Feeder

- The nodal load demand is calculated based on our real smart meter data with 1-h resolution. The length of window is selected as 200.
- Even though our method is derived on the assumption of a homogeneous R/X ratio, it shows the robustness to the systems with heterogeneous R/X ratios.
- The minimum and maximum R/X values of the three feeders are {0.5153, 2.0655}, {1.4536, 2.7482} and {0.4, 3.4}, respectively (***three heterogeneous feeders***).

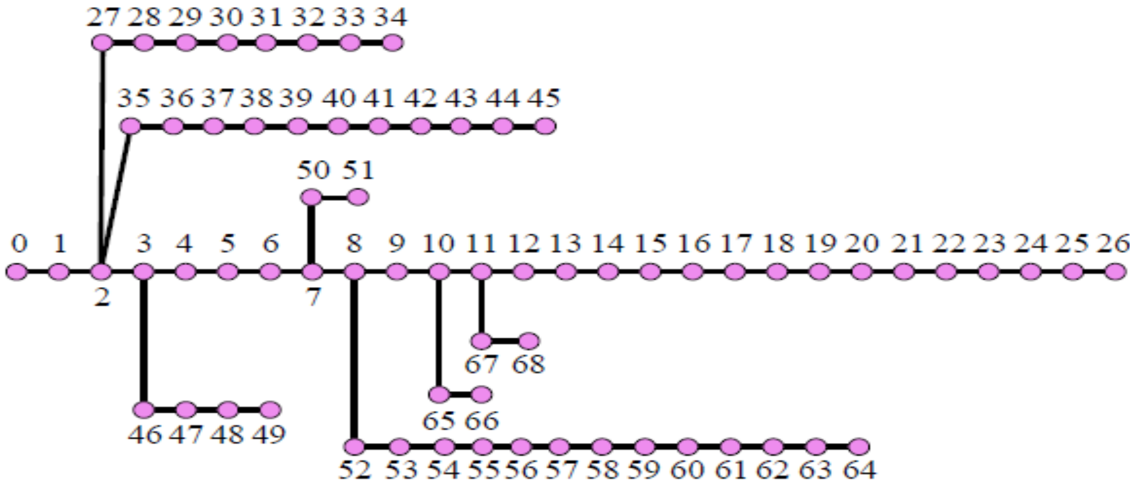
Estimated Weighted Laplacian Matrix for IEEE 13-, 37-, 69-Bus Test Feeder



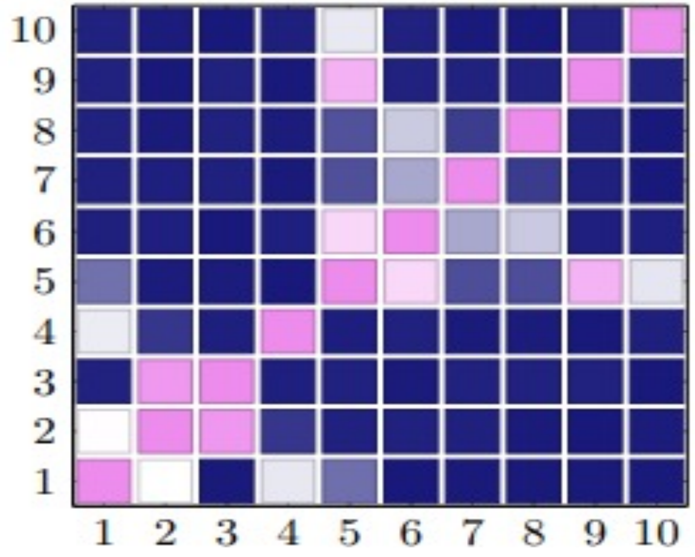
13-Bus Test Feeder

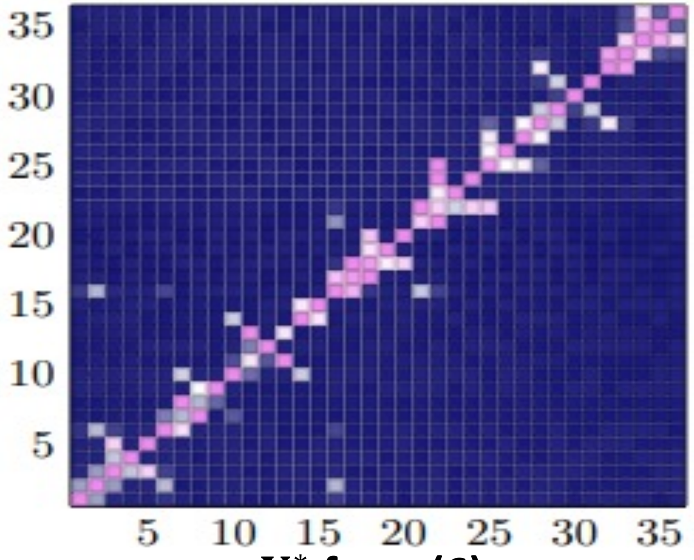


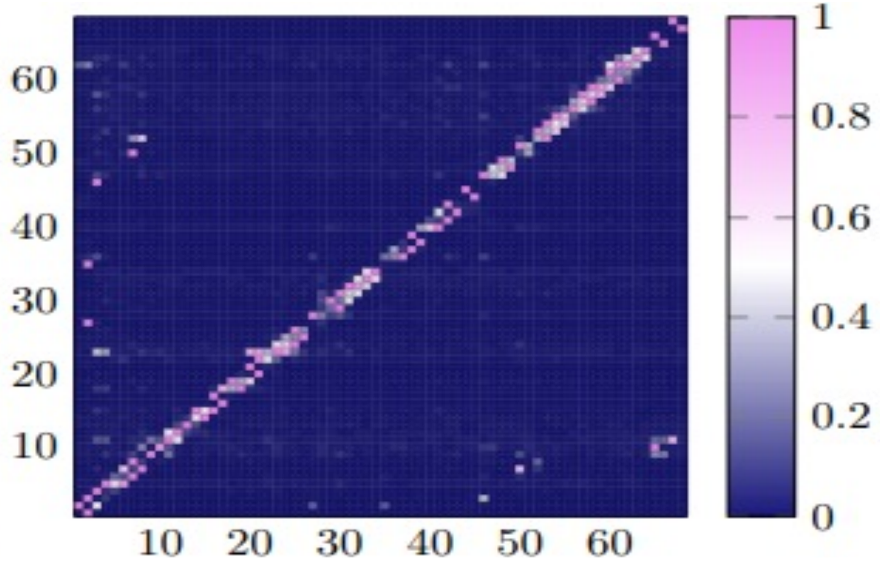
37-Bus Test Feeder



69-Bus Test Feeder



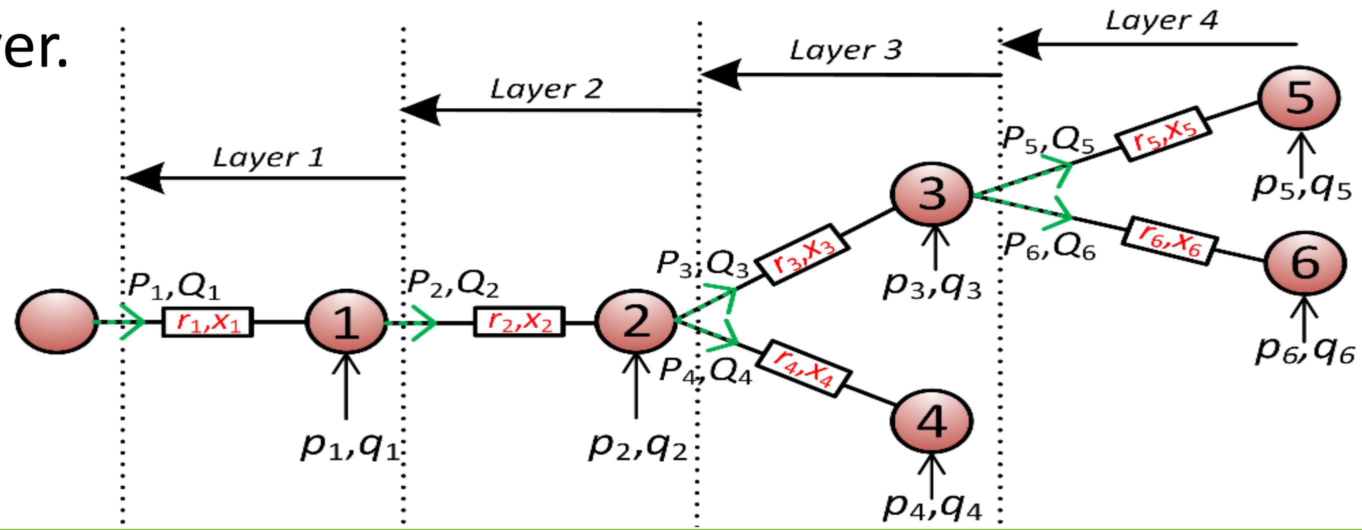


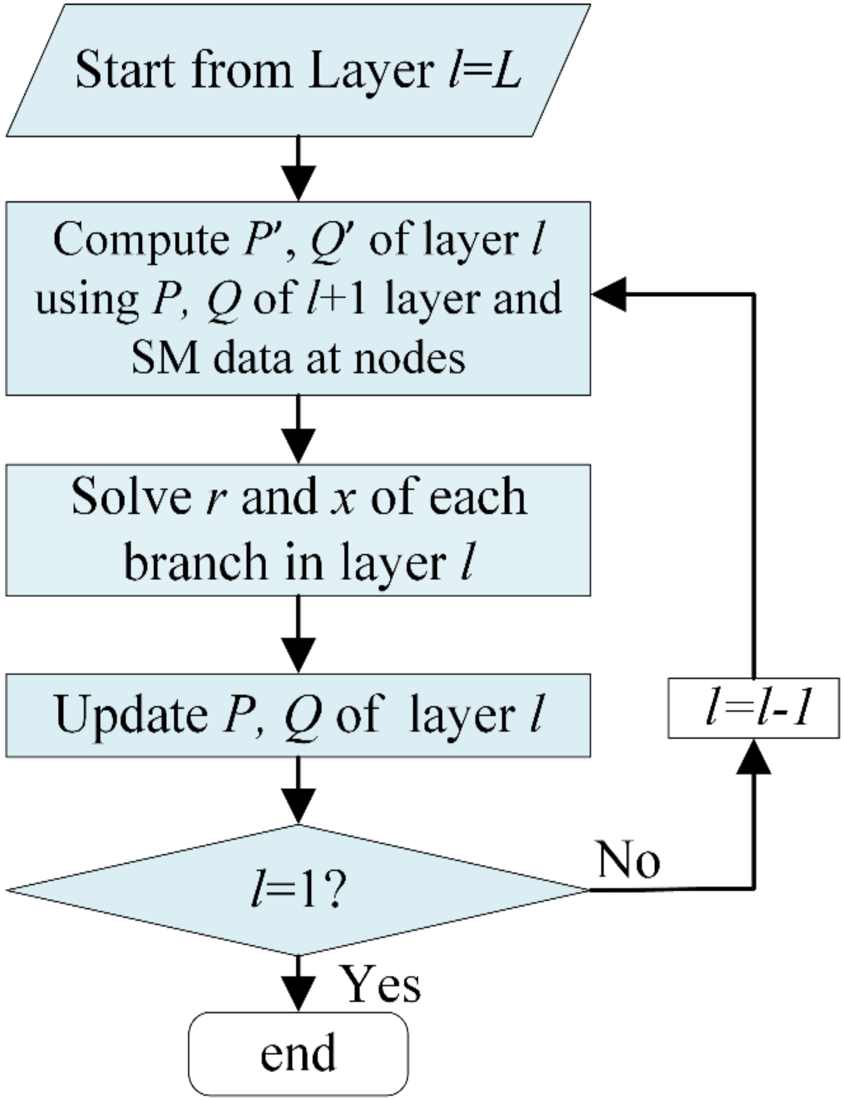


$\mathbf{Y}^*$ from (6)

Bottom-Up Sweep Parameter Identification

- Decomposing a radial distribution network into multiple layers labeled $1, \ldots, L$ (where L is the bottom layer).
- Our bottom-up sweep algorithm performs the line flow and line parameter estimation in an alternating way based on the layers of the network.
- Addressing the dimensionality issue and enabling parallel computation of all line sections within the same layer.





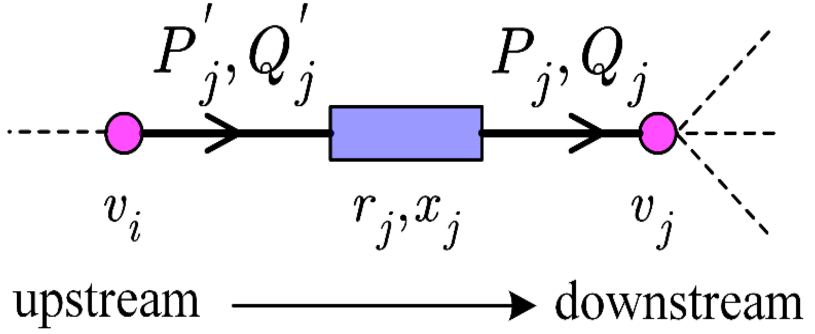
Branch Flow Model

- The proposed line impedance estimation establishes on the voltage drop relationship over a branch that can be modeled as:

Branch Flow: $P_j' = P_j + \frac{r_j\left(P_j^2+Q_j^2\right)}{v_j}$, $Q_j' = Q_j + \frac{x_j\left(P_j^2+Q_j^2\right)}{v_j}$

Voltage: $v_i - v_j = 2\left(r_j P_j + x_j Q_j\right) + \frac{\left(r_j^2+x_j^2\right)\left(P_j^2+Q_j^2\right)}{v_j}$

Branch model



- P_j' and Q_j' denote power flows out of the upstream node i; P_j and Q_j denote power flows into the downstream node j.

- "Upstream" and "downstream" represent the relative positions of the nodes and power could flow in either direction.

Network Parameter Estimation Model

- The parameter estimation establishes on the voltage drop over a branch that is defined as follows:

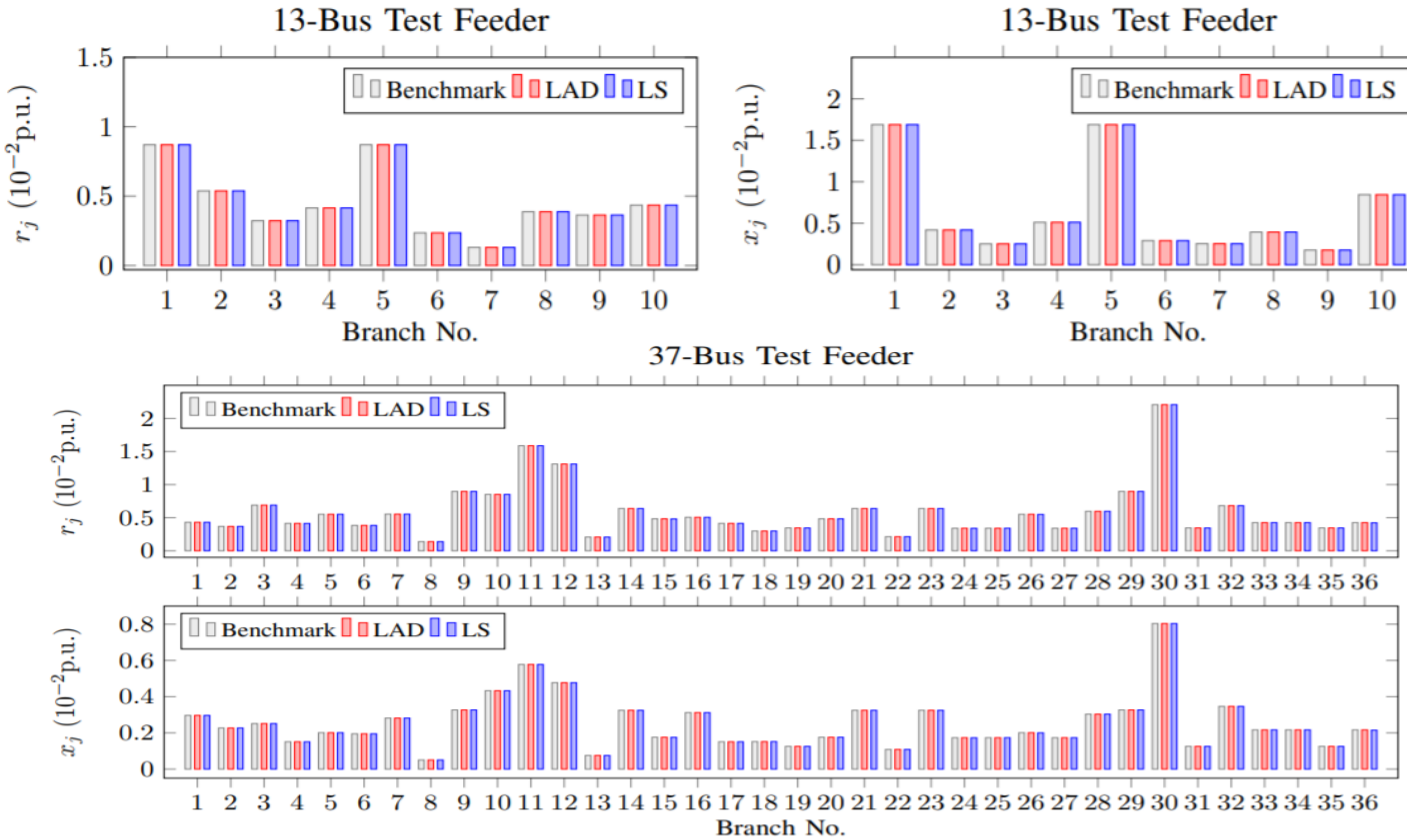
$$e_j^k := v_i^k - v_j^k - 2\left(r_i P_j^k + x_j Q_j^k\right) - \left(R_j + X_j\right) \cdot \frac{\left[\left(P_j^k\right)^2 + \left(Q_j^k\right)^2\right]}{v_j^k} \tag{7}$$

- Based on (7) and the R/X ratio library, the line parameter estimation is cast as a mixed-integer ***nonlinear*** programming model.

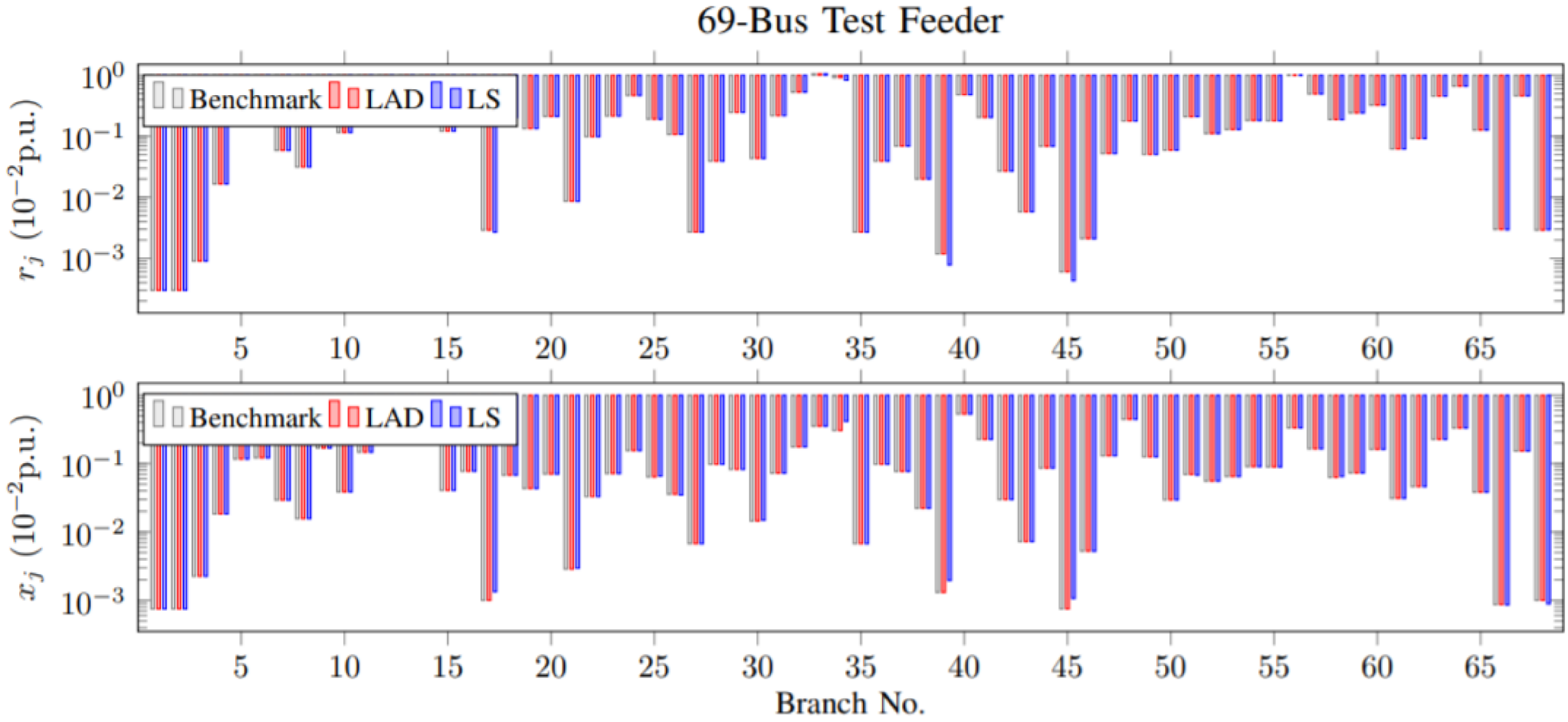
$$\min_{\alpha_z, r_j, x_j, R_j, X_j} \left|\left|e_j\right|\right|_1, \text{subject to } R_j = r_j^2, X_j = x_j^2, r_j = \sum_{z=1}^{Z} \lambda_z \alpha_z x_j, \tag{8}$$

$$\sum_{z=1}^{Z} \alpha_z = 1, \alpha_z \in \{0,1\}, \forall z.$$

Estimated Line Parameters for IEEE 13- and 37-Bus Test Feeder



Estimated Line Parameters for IEEE 69-Bus Test Feeder



Conclusion and Future Work

- Smart meter data, although may be of low resolution and limited measurement variables, can still be used to greatly help distribution system monitoring and operation. There are many applications such as network modeling, outage detection and behind-the-meter solar disaggregation.

- We demonstrated how to use smart meter data together with optimization and machine learning to estimate topology and line parameters in radial distribution systems.

- In the future, we will focus on using smart meter data to identify/calibrate network models in unbalanced mesh distribution systems.

Reference

[1] J. Yu, Y. Weng, and R. Rajagopal, “PaToPaEM: A data-driven parameter and topology joint estimation framework for time-varying system in distribution grids,” IEEE Transactions on Power Systems, vol. 34, no. 3, pp. 1682-1692, May 2019.

[2] O. Ardakanian, Y. Yuan, V. Wong, R. Dobbe, S. Low, A. von Meier, and C. J. Tomlin, “On identification of distribution grids,” IEEE Transactions on Control of Network Systems, vol. 6, no. 3, pp. 950-960, Sep. 2019.

[3] Y. Yuan, S. H. Low, O. Ardakanian, and C. J. Tomlin, “Inverse Power Flow Problem,” [Online]. Available: https://arxiv.org/pdf/1610.06631.pdf.

[4] A. M. Prostejovsky, O. Gehrke, A. M. Kosek, T. Strasser, and H. W. Bindner, “Distribution line parameter estimation under consideration of measurement tolerances,” IEEE Transactions Industrial Informatics, vol.12, no.2, pp. 726–735, 2016.

[5] H. Xu, A. D. Domínguez-García, and P. W. Sauer, “A data-driven voltage control framework for power distribution systems,” In Proceedings of IEEE Power and Energy Society General Meeting, pp. 1-5, Aug. 2018.

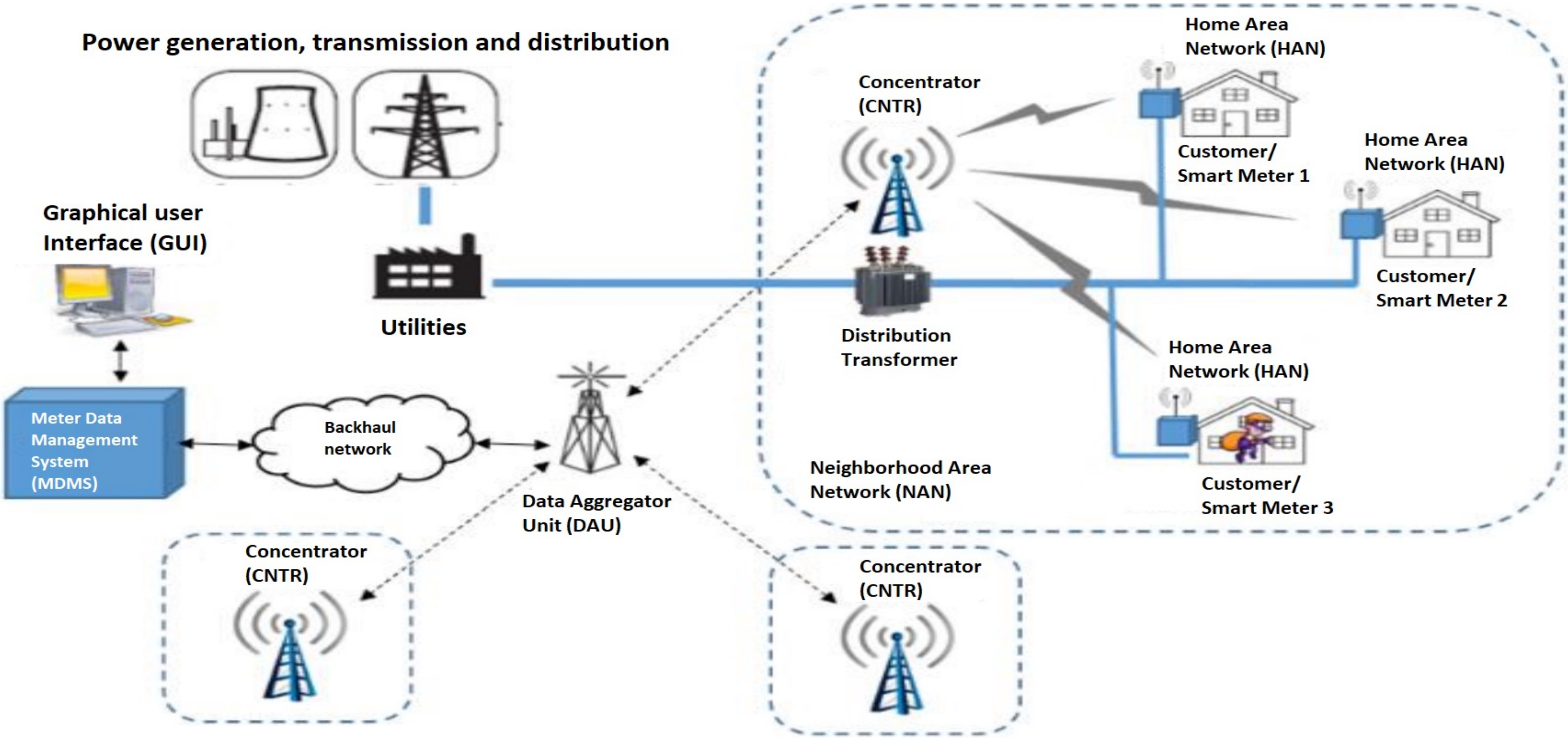
[6] W. Wang and N. Yu, “Parameter estimation in three-phase power distribution networks using smart meter data,”2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–6, 2020.

[7] Y. Guo, Y. Yuan, and Z. Wang, "Distribution Grid Modeling Using Smart Meter Data," arXiv preprint arXiv:2103.00660, 2021.

Thank You!
Q & A

Zhaoyu Wang
http://wzy.ece.iastate.edu

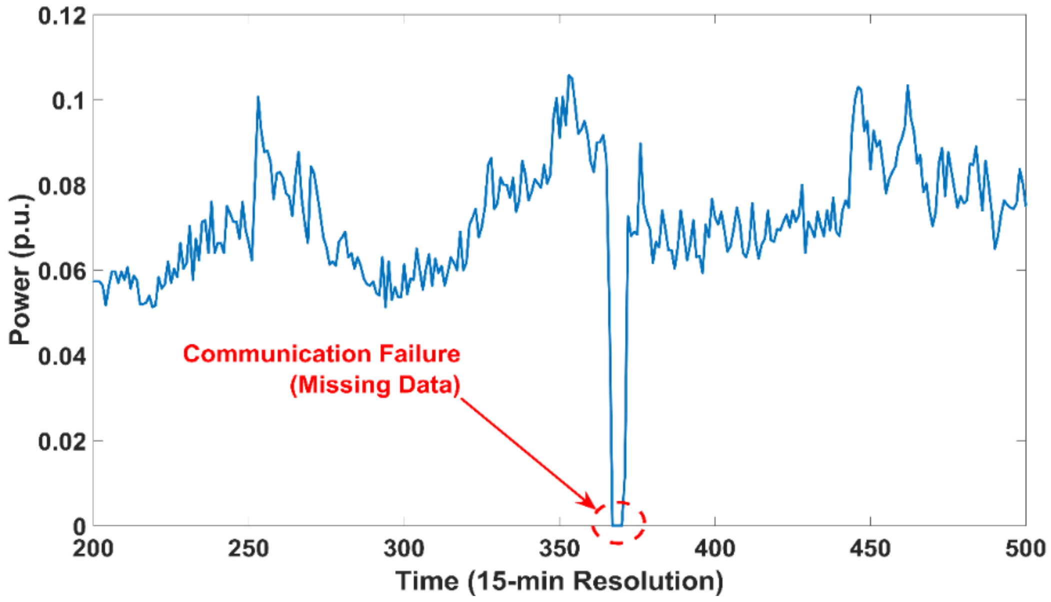
Smart Meter Data Collection

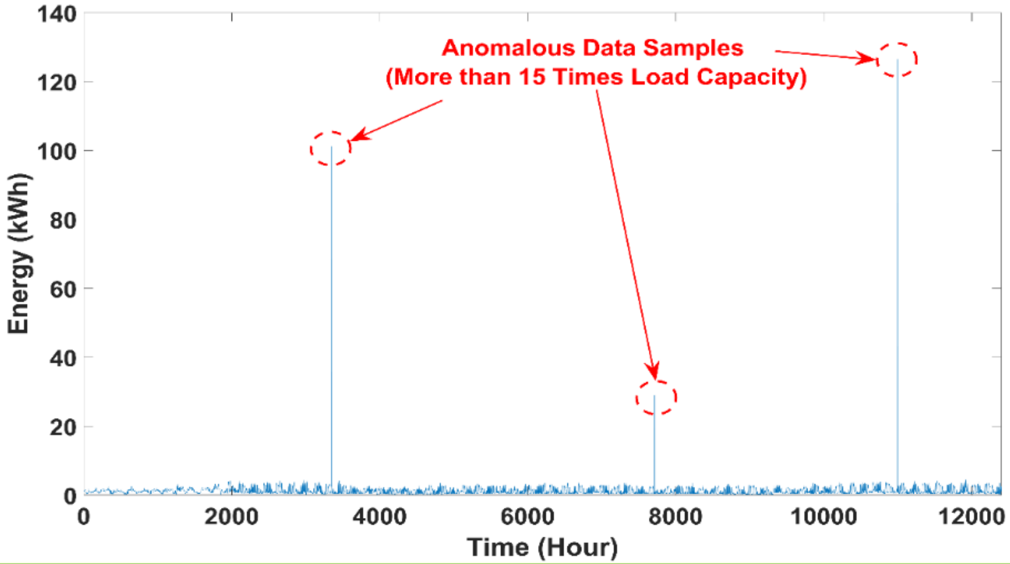


K. K. Kee, S. M. F. Shahab and C. J. Loh, "Design and development of an innovative smart metering system with GUI-based NTL detection platform"

Smart Meter Data Pre-Processing

- **Smart Meter Data Problems:**
  - ➢ Outliers/Bad Data
  - ➢ Communication Failure
  - ➢ Missing Data
- **Counter-Measures:**
  - ✓ Engineering intuition (data inconsistency)
  - ✓ Conventional Statistical Tools (e.g. Z-score)
  - ✓ Robust Computation (e.g. relevance vector machines)
  - ✓ Anomaly Detection Algorithms





Recovering Topology From Estimated $\mathbf{Y}^*$

- Recovering the topology from $\mathbf{Y}^*$ is cast as an *anomaly detection* problem.

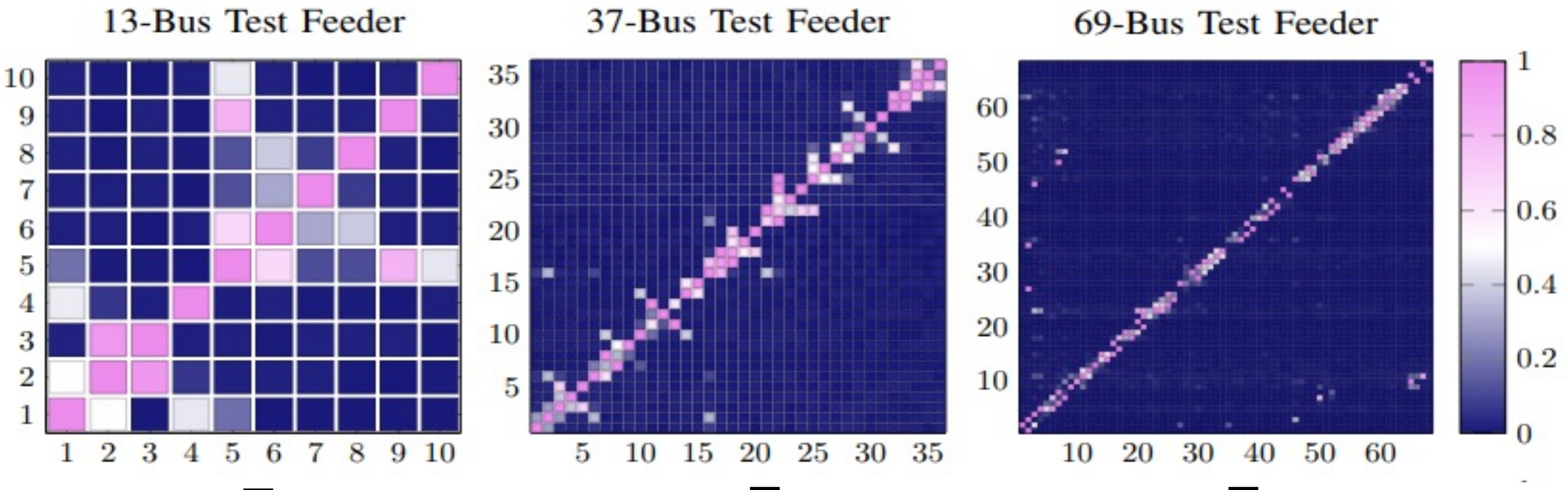
- **Our Solution**: We have utilized a density based spatial clustering of applications with noise (DBSCAN) method to extract the topology from $\mathbf{Y}^*$.
  - DBSCAN can marking as anomaly points that lie alone in low-density regions.
  - **Advantage**: DBSCAN can discover clusters with arbitrary shapes.
  - DBSCAN does not require a prior specification on the number of clusters.

Algorithm 1 Recovering Topology From $\mathbf{Y}^\star$ by Clustering

Initialization: Initialize $i \leftarrow 1, j \leftarrow 1, \gamma, \xi$
repeat
  [**S1**]: Select the ith row of $\mathbf{Y}^\star$.
  repeat
    [**S2**]: Pick $y^\star_{ij}$ and retrieve all direct density-reachable points using ξ.
    [**S3**]: Based on γ, if $y^\star_{ij}$ is a core point, a cluster is formed; otherwise, update $j \leftarrow j + 1$.
  until $j = n$ or no new point can be added to any cluster
  [**S4**]: Update $i \leftarrow i + 1$.
until $i = n$.

Estimated Topology for IEEE 13-, 37-, 69-Bus Test Feeder

Y^* from (6)



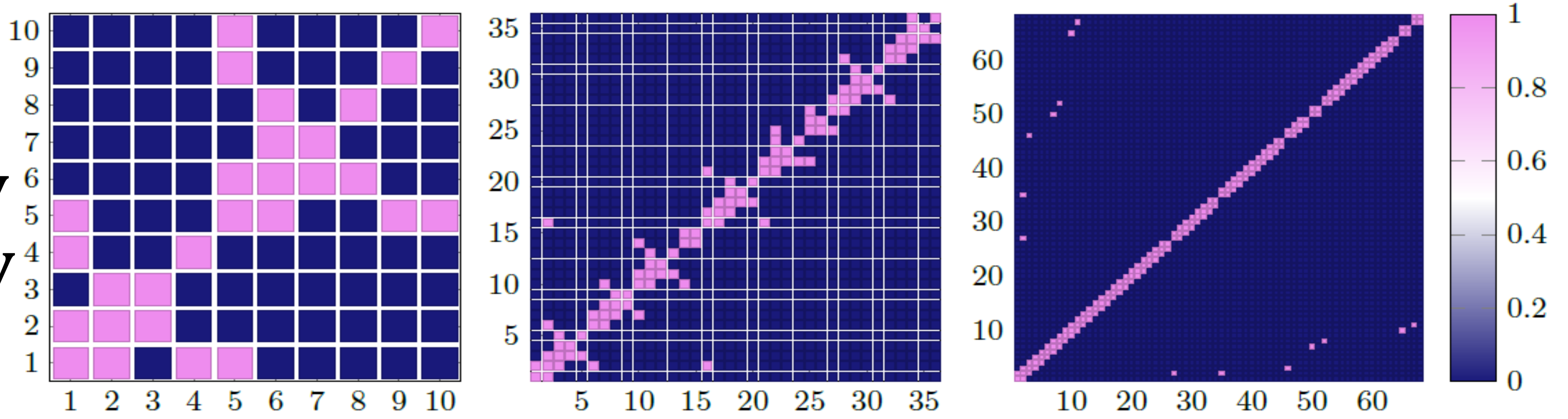
500 Simulations

100%

100%

98%

Topology Recovery



LAD Parameter Estimation Model

- The Big-M technique is exploited to linearize the bilinear term $\alpha_z x_j$.

- We rewrite (8) without L1-norm operator by introducing the auxiliary variables. The SPD relaxation is used to tackle the non-convex quadratic equalities.

$$\min_{\alpha_z, r_j, x_j, R_j, X_j, \theta_j^k} \sum_{k=1}^{K} \theta_j^k \qquad (9)$$

$$\text{subject to} \quad \theta_j^k \geq e_j^k, \forall k$$

$$-\theta_j^k \leq e_j^k, \forall k$$

$$\mathbf{W}_j^r = \begin{bmatrix} 1 & r_j \\ r_j & R_j \end{bmatrix} \succcurlyeq 0, \mathrm{rank}\{\mathbf{W}_j^r\} = 1, \forall j$$

$$\mathbf{W}_j^x = \begin{bmatrix} 1 & x_j \\ x_j & X_j \end{bmatrix} \succcurlyeq 0, \mathrm{rank}\{\mathbf{W}_j^x\} = 1, \forall j$$

$$\sum_{z=1}^{Z} \alpha_z = 1, \alpha_z \in \{0,1\}, \forall z.$$

$$-M_j(1-\alpha_z) \leq r_j - \lambda_z x_j \leq M_j(1-\alpha_z), \forall z$$

LS Parameter Estimation Model

- The Big-M technique is exploited to linearize the bilinear term $\alpha_z x_j$.
- We rewrite (8) by introducing the auxiliary variable μ_j, and additionally imposing the constraints.
- Relaxing the quadratic equalities, we obtain a MISOCP model.

$$\min_{\alpha_z, r_j, x_j, R_j, X_j, \mu_j} \quad \mu_j \tag{10}$$

$$\text{subject to} \quad \left\| \begin{matrix} \frac{\mu_j - 1}{2} \\ e_j \end{matrix} \right\|_2 \leq \frac{\mu_j + 1}{2}$$

$$\left\| \begin{matrix} \frac{R_j - 1}{2} \\ r_j \end{matrix} \right\|_2 \leq \frac{R_j + 1}{2}$$

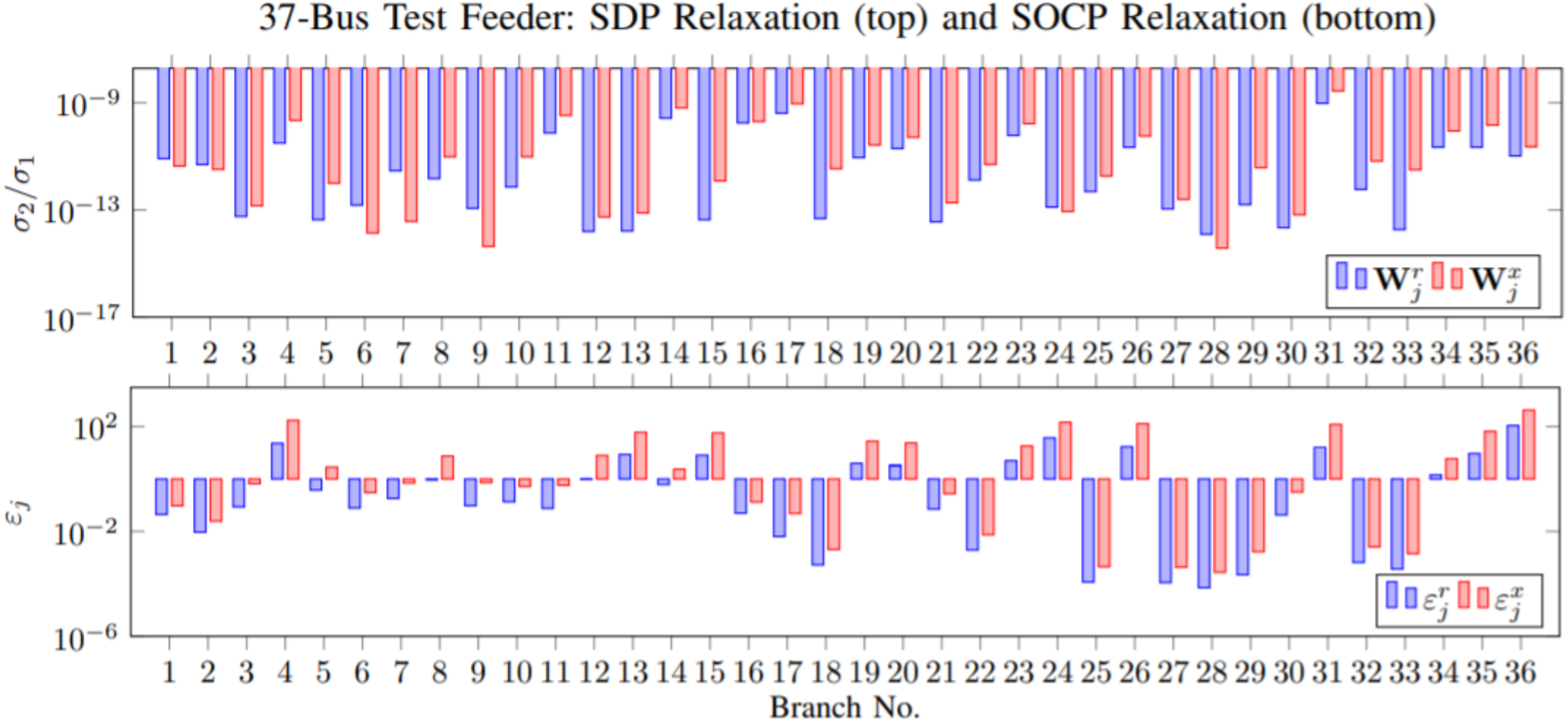
$$\left\| \begin{matrix} \frac{X_j - 1}{2} \\ x_j \end{matrix} \right\|_2 \leq \frac{X_j + 1}{2}$$

$$\sum_{z=1}^{Z} \alpha_z = 1, \alpha_z \in \{0,1\}, \forall z.$$

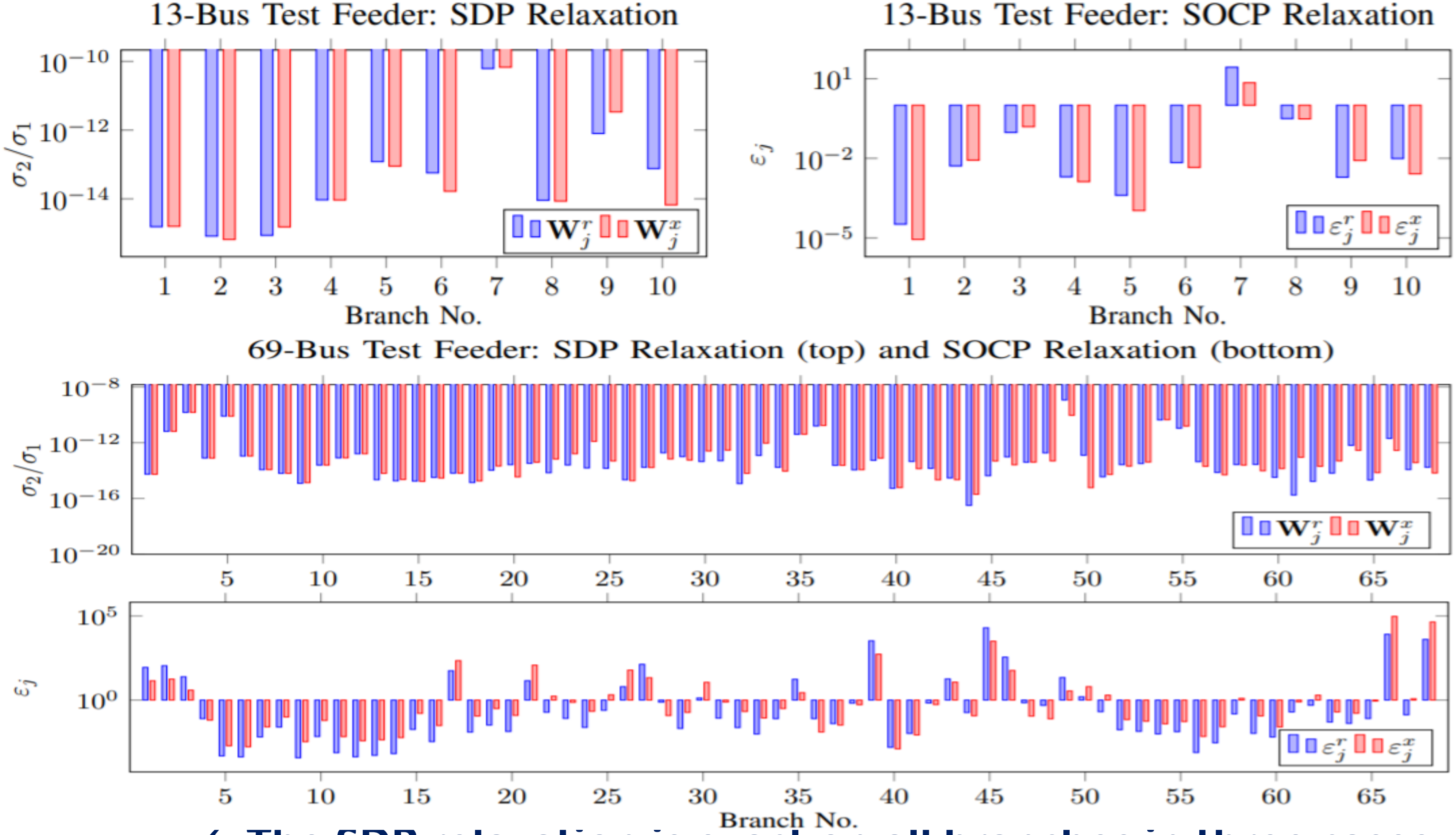
$$-M_j(1 - \alpha_z) \leq r_j - \lambda_z x_j \leq M_j(1 - \alpha_z), \forall z$$

Exactness of SDP and SOCP Relaxation of IEEE 37-Bus Test Feeders

- To quantify the exactness of SDP and SOCP relaxation in (9) and (10), we compute the ration between the largest two eigenvalues of $\mathbf{W}_j^r$ and $\mathbf{W}_j^x$ and the resultant errors ε_j^r and ε_i^x, respectively.



Exactness of SDP and SOCP Relaxation of IEEE 13- and 69-Bus Test Feeders



✓ **The SDP relaxation is exact on all branches in three cases.**

Estimated Line Parameters for IEEE 69-Bus Test Feeder - Summary

SDP-based LAD model:

- For the IEEE 13-bus feeder, the largest relative errors (among all branches) for r_j and x_j are $3.331\times10^{-5}\%$ and $3.335\times10^{-5}\%$.
- For the IEEE 37-bus feeder, the largest relative errors (among all branches) for r_j and x_j are $3.402\times10^{-4}\%$ and $3.403\times10^{-4}\%$.
- For the IEEE 69-bus feeder, the largest relative errors (among all branches) for r_j and x_j are $1.444\times10^{-4}\%$ and $7.061\times10^{-5}\%$.

SOCP-based LS model:

- For the IEEE 13-bus feeder, the largest relative errors (among all branches) for r_j and x_j are 0.256% and 0.952%.
- For the IEEE 37-bus feeder, the largest relative errors (among all branches) for r_j and x_j are 0.251% and 0.958%.
- For the IEEE 69-bus feeder, the largest relative errors (among all branches) for r_j and x_j are 33.95% and 46.81%. But these large errors ($\geq 5\%$) only occur in a few branches with high R/X ratios (17,34,39,45, and 68).