Machine Learning for Power Systems: From Pure Data-Driven to Physics-Informed Methods

Prof. Nanpeng Yu

Department of Electrical and Computer Engineering

Department of Computer Science

(cooperating faculty)

Director of Energy, Economics, and Environment Research Center

nyu@ece.ucr.edu

951.827.3688

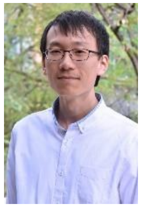
UNIVERSITY OF CALIFORNIA, RIVERSIDE

Team Members

- Energy, Economics, and Environment (E3) Research Center
  - Director: Dr. Nanpeng Yu.
  - Over 40 affiliated faculty in CSE, ECE, STAT and SPP
- Project Scientists and Postdoctoral Researchers
  - Dr. Koji Yamashita, Dr. Mikhail Bragin, Dr. Yinglun Li, and Dr. Zuzhao Ye
- PhD and MS Students
  - Yuanbin Cheng, Osten Anderson, Jingtao Qin, Joseph Brown, Maojie Tang, Zhixuan Tan, Shaorong Zhang, and Jeffrey Chan
- Alumni PhD Students

Dr. Brandon Foggo
Machine Learning Scientist at Habitat Energy

Dr. Wei Wang
Senior Data Scientist at Walmart Labs



Dr. Yuanqi Gao
Machine Learning Engineer at Lucid Motors

Dr. Jie Shi
Postdoctoral Researcher at Cornell

Dr. Wenyu Wang
Senior Engineer at Quanta Technology

Dr. Farzana Kabir
EV Grid Infrastructure Specialist at California Energy Commission

Dr. Yinglun Li
Senior ML Scientist at Alibaba International Digital Commercial Group

Dr. Zuzhao Ye
Founder of AmpTrans

Research Projects and Sponsors

- Over $16 Million of R&D Funding in the past 10 years
  - Federal government, state agencies, private companies
- Research Tracks
  - Physics-informed Machine Learning for Power Systems
  - Scalable Optimization in Critical Infrastructure Systems
  - Transportation and Building Electrification
  - Decarbonization Planning
  - Energy Efficient Data Center

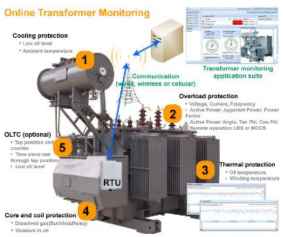
Outline

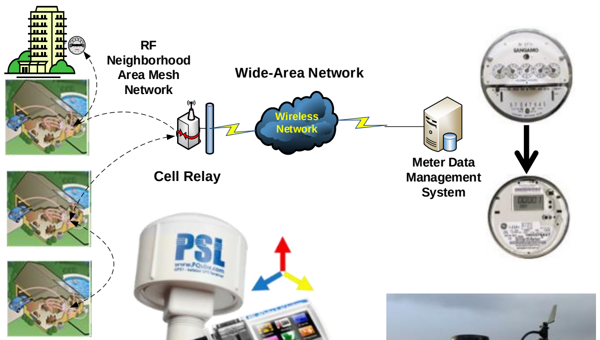
- Volume, Variety, Velocity & Value of Big Data in Power Systems
- Applications of Machine Learning in Power Systems
  - Transmission system, distribution system, end-use customers
- Motivation for Physics-informed Methods in Power Systems
- Leverage Unique Properties of Data from Power Systems
  - Low rank and sparsity, high and low entropy
- Integrate Machine Learning Models with Physical Models of Power Systems
  - Renewable resource model, power flow model
  - Topology of power network, system control model
  - Power system dynamic model
- Summary and Discussions

Volume and Variety

- Variety of Big Data in Power Systems
  - Supervisory control and data acquisition (SCADA) system
  - **Phasor Measurement Unit (PMU)**, Micro-PMU
  - Digital Fault Recorder, Equipment Monitors
  - Census Data, Advanced Metering Infrastructure (AMI)
  - Weather Station, Electricity Market
  - Geographical Information System
- Volume of Big Data in Power Systems
  - PMUs
    - Over 2,500 PMU (10,000 measurement channels) in the U.S.
    - Roughly 450 TB of data generated annually (60Hz)
  - AMI
    - Over 1.1 billion smart meter installations worldwide as of 2022.
    - Over 2 petabytes of smart meter data.





Velocity and Value

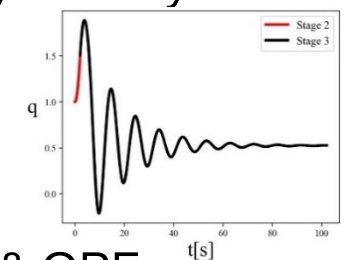
- Velocity of Data in Power Systems
  - Sampling Frequency
    - AMI Data: Sampling frequency increases from once a month to 1 reading every 15 mins
    - PMU: Mature technology (30 – 240 Hz), continuous point on wave (1920 – 61,400 Hz)
  - Bottleneck in Communication Systems of Distribution Network
    - Limited bandwidth for ZigBee network
    - Most of the utilities in the US receives smart meter data with ~24 hour delay
  - Edge Computing Trend
    - Itron and Landis+Gyr extend edge computing capability of smart meters
    - Centralized → distributed / decentralized monitoring, computing, control and decision making
- Value of Big Data Analytics in Power Systems
  - According to GTM Research, electric utilities around the world spent over **$3.8 billion** on data analytics solutions in 2020.
  - Analysis from Indigo Advisory group indicates that the market for AI in the energy sector could be worth **$13 billion** as of the end of 2023.

Applications of Machine Learning in Transmission Systems and Electricity Market

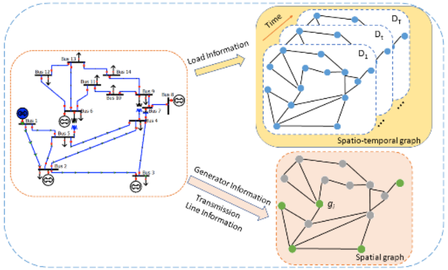
Electricity Market Applications

Price & Load Forecasting, Algorithmic Trading

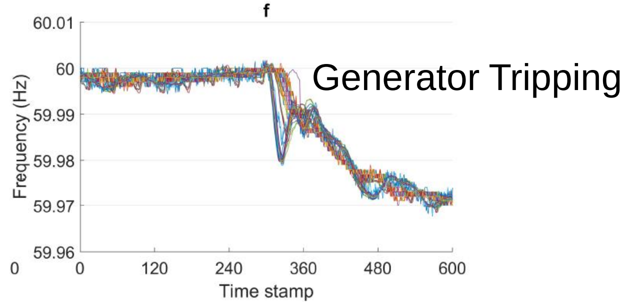
Learn Power System Dynamics

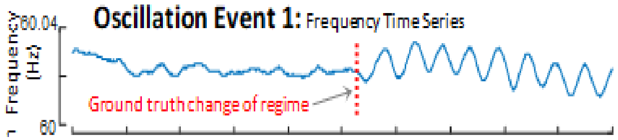


Accelerate UC & OPF



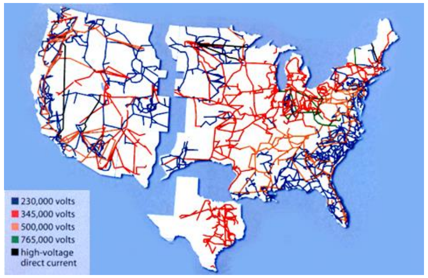
Event Detection & Classification





Equipment Monitoring

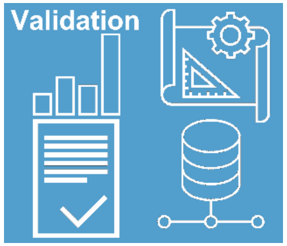
Identify Faulty Equipment in Substations



State Estimation

Linear State Estimation

Model Validation & Parameter Estimation



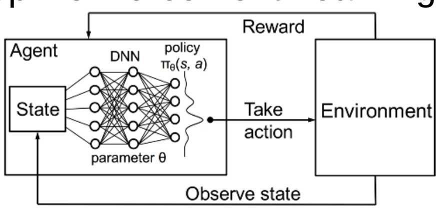
Applications of Machine Learning for Power Distribution Systems and End Use Customers

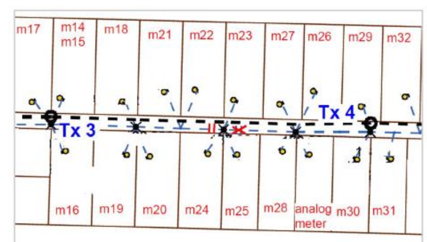
Spatio-temporal Forecasting
Electric Load / DERs – Short-Term / Long-Term

Anomaly Detection
Electricity Theft, Unauthorized Solar Interconnection

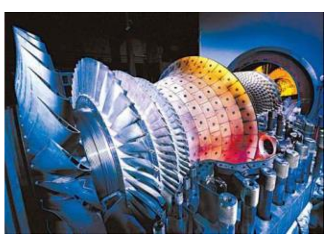
System Monitoring
State Estimation & Visualization

Distribution System Controls
Deep Reinforcement Learning





Network Topology and Parameter Identification
Transformer-to-customer, Phase connectivity, Impedance estimation



Equipment Monitoring
Predictive Maintenance
Online Diagnosis

Customer Behavior Analysis
Customer segmentation, nonintrusive load monitoring, demand response

Physics-Informed ML for Power Systems: Motivation

- Purely Data-driven ML algorithms: widely adopted by researchers and practitioners to solve a myriad of problems in power systems.
  - Big data: e.g. load & price forecasting, predictive maintenance of transformers
  - Struggle to deal with system monitoring, sequential decision-making, large-scale optimization, and control problems in power systems.
- Technical Challenges
  - Accuracy, generalization capability, sample efficiency, safety and interpretability
- Physics-informed ML Algorithms
  - Synergistic combination of machine learning & physical model or information
  - Embed domain knowledge, unique data property, system model in ML algorithms
  - Introduce inductive bias, improve explainability, and generalize to unforeseen scenarios from a finite training dataset

Low Rank and Sparsity Properties of Power System Measurement Data

Rank of the matrix: the number of linearly independent rows or columns in the matrix

<u>Low Rank and Sparsity</u>: Voltage Event Detection Using Optimization with Structured Sparsity Inducing Norms

- Key Observations
  - <u>Voltage related events</u> trigged by system faults are often regional events
  - The $X - L$ during voltage event periods have <u>row-sparse structure</u>
  - Rows of residual matrix correspond to PMUs highly impacted by the event
- Main Idea
  - Decompose the streaming PMU data matrix X into
    - A low-rank matrix L, a row-sparse event-pattern matrix S, and a noise matrix G
  - Extract anomaly features from L & S
  - Use clustering algorithm to identify power system voltage events

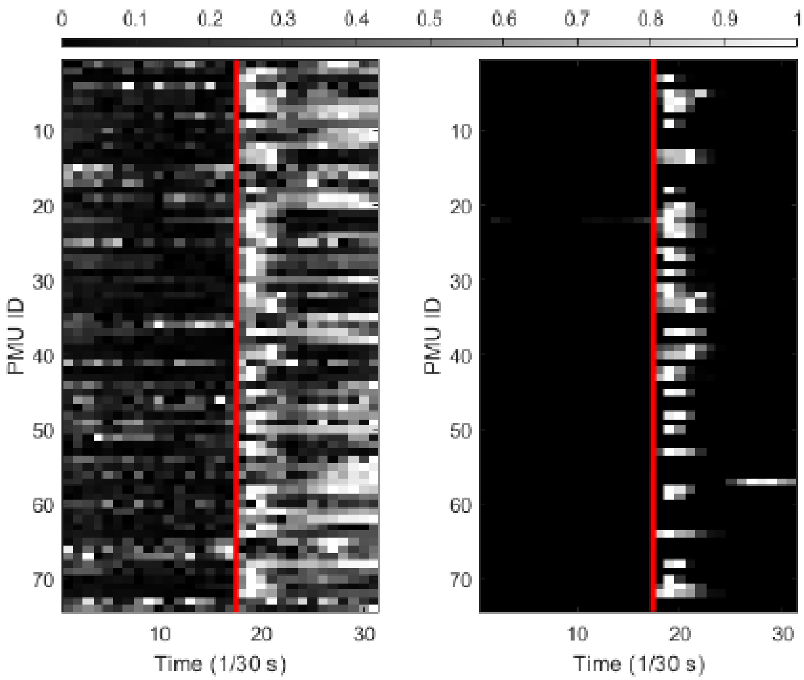


Fig. 2. The heatmap of "$X - L$" (left) and "$X - L - G$" (right) for normalized active power data (scaled from 0 to 1). The event happens approximately at the red line.

X. Kong, B, Foggo, and N. Yu, "Online Voltage Event Detection Using Optimization with Structured Sparsity-Inducing Norms," *IEEE Transactions on Power Systems*, vol. 37, no. 5, Sep. 2022.

Decompose PMU Data Matrix with Proximal Bilateral Random Projection (PBRP) to Detect Events

$$\min_{L,S} \quad \frac{1}{2}\|X - L - S\|_F^2 \qquad s.t. \quad \begin{cases} rank(L) = r, \\ S \text{ is row-sparse.} \end{cases}$$

$$\Rightarrow$$

$$\min_{L,S} \quad \frac{1}{2}\|X - L - S\|_F^2 + \lambda\|S\|_{21} \qquad s.t. \quad rank(L) = r,$$

Solution Approach: Coordinate Descent

$$\begin{cases} L^{(k)} = \arg \min_{rank(L)=r} \frac{1}{2}\|X - L - S^{(k-1)}\|_F^2 \\ S^{(k)} = \arg \min_{S} \frac{1}{2}\|X - L^{(k)} - S\|_F^2 + \lambda\|S\|_{21} \end{cases}$$

F SCORES OF THREE ALGORITHMS ON THE TESTING DATASET

Statistics	OLAP	HOLAP	P-BRP
Precision	**0.8889**	0.8824	0.8881
Recall	0.8955	0.8955	**0.9478**
F1 Score	0.8922	0.8889	**0.9170**
Precision	0.8089	**0.8571**	0.8000
Recall	0.9478	0.9403	**0.9851**
F2 Score	0.9163	0.9224	**0.9415**

AVERAGE COMPUTATION TIME OF EVENT DETECTION ALGORITHMS OVER THREE-MINUTE TIME PERIOD

Number of PMUs		50	100	150
Computation Time (s) (partial/total)	HOLAP	61.78/68.46	181.50/189.25	336.27/344.58
	OLAP	7.53/15.01	9.58/17.33	16.99/24.79
	P-BRP	2.18/8.46	3.13/9.40	4.29/10.53

- Residual PMU data matrices during voltage events have distinctive sparsity structure
- Computationally efficient PBRP algorithm is proposed to decompose PMU data matrices
- Online voltage event detection algo. shows better accuracy & scalability on PMU data (Eastern Interconnection)

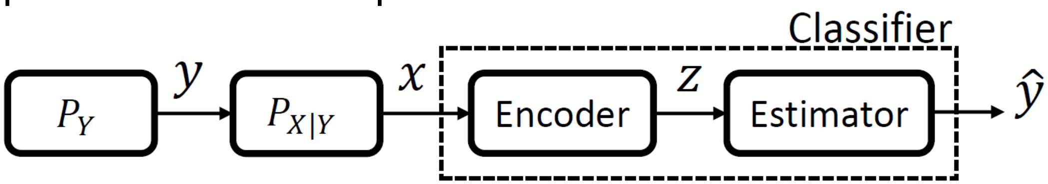
Low and High Entropy Power System Measurements

Information entropy of data: the average amount of information conveyed by an event, when considering all possible outcomes

$$H(x) \doteq -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

High & Low Entropy of Dataset: Information Loading-based Regularization

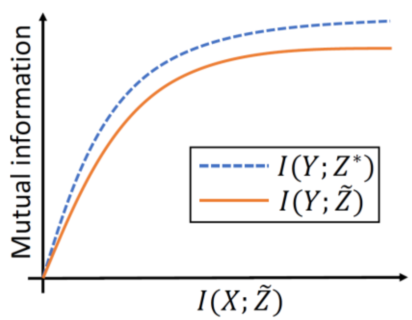
- Background
  - Abstract Representation of Deep Neural Network based Classifier

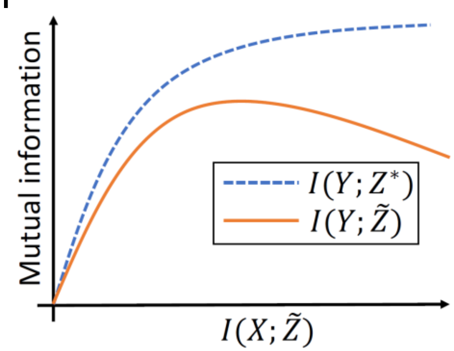


- Main Idea
  - Control the amount of information compression between the input layer and the last hidden layer of a deep neural network
  - Balance memorization and generalization

Low entropy input feature space

e.g. Vol. mag. in distribution system (4–11 bits)





High entropy input feature space

e.g. PMU data during power system events (> 60 bits)

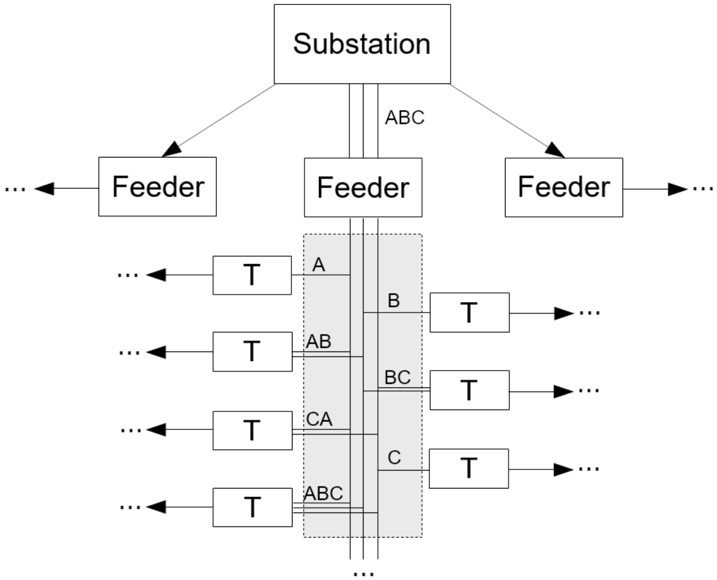
- Algorithm
  - Augment the typical cross-entropy loss function with estimated mutual information between the input layer and the hidden representation

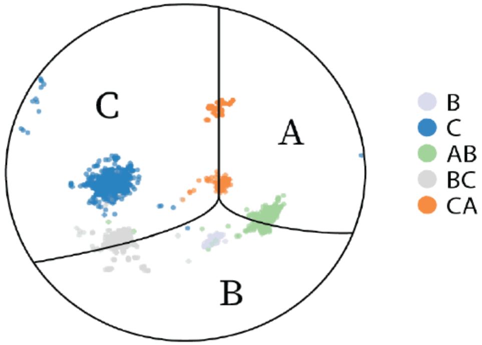
$$L_T = L_{CE} - \beta \hat{I}(X;Z)$$

B. Foggo, N. Yu, J. Shi and Y. Gao, "Information Losses in Neural Classifiers from Sampling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4073-4083, 2020. DOI: 10.1109/TNNLS.2019.2952029.

Phase Connectivity Identification

- Very few electric utility companies have completely accurate phase connectivity information in GIS!
- Validated using real-world distribution circuits data from SCE and PG&E.





Learned Representations on Circuit V

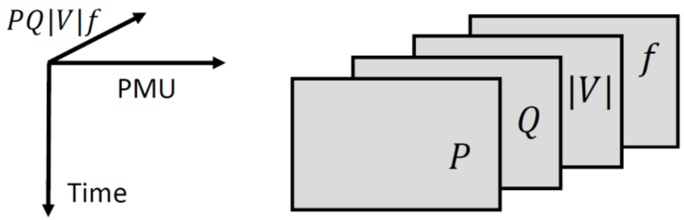
Phase Identification Accuracy

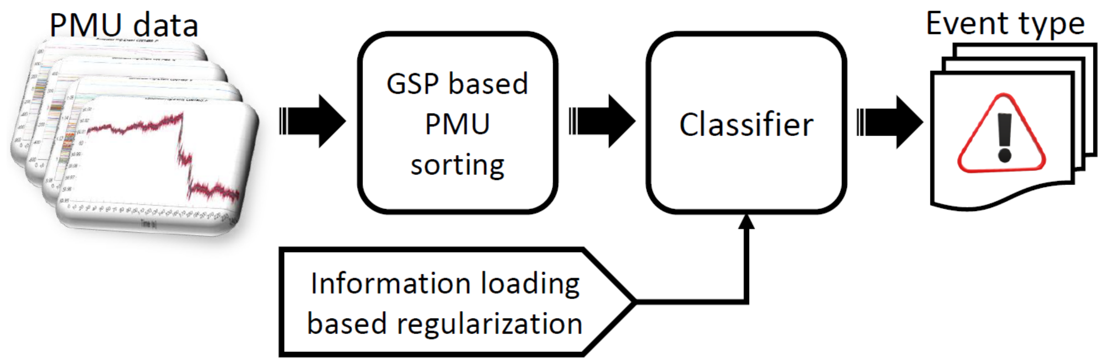
Circuit	Baseline	I-loading	Facility	Inverse	Inverse+I-loading
I	80.7%	81.5%	86.7%	89.9%	**91.0%**
II	64.7%	80.6%	90.5%	95.4%	**96.3%**
III	74.1%	75.2%	90.6%	91.5%	**93.1%**
IV	75.0%	78.0%	91.2%	94.6%	**98.8%**
V	51.7%	59.1%	94.2%	96.1%	**97.3%**

B. Foggo and N. Yu, "Improving Supervised Phase Identification Through the Theory of Information Losses," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2337-2346, 2020.

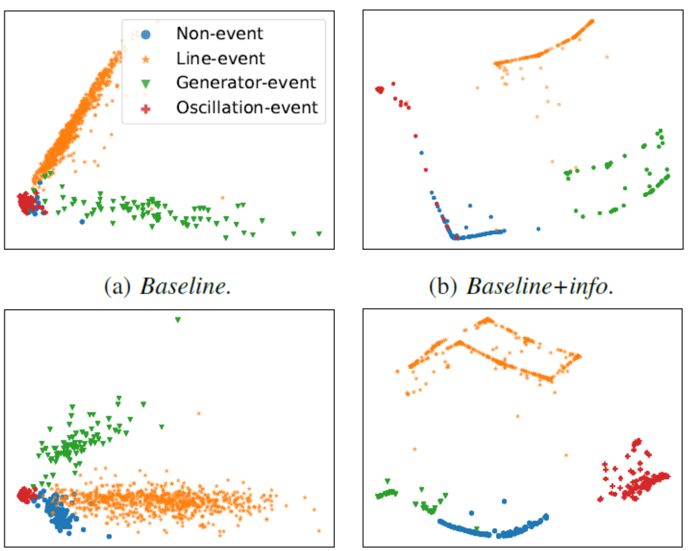
System Event Classification with PMU Data

- Formulated as a classification problem
  - Normal operation condition, line event, generator event, oscillation event





Learned Representation



(a) *Baseline.* (b) *Baseline+info.* (c) *Baseline+GSP.* (d) *Baseline+GSP+info.*

F1 Score on Testing Dataset from Eastern Interconnection

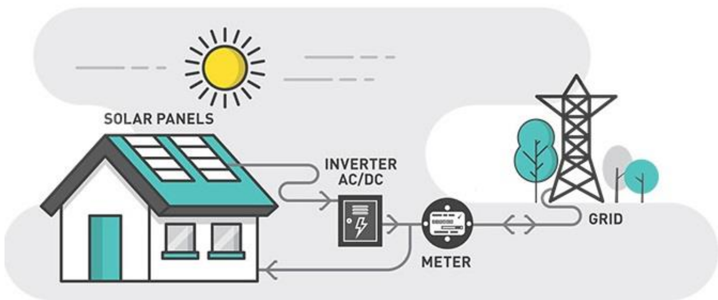
	Non-event	*Line-event*	*Generator event*	*Oscillation event*
Baseline	0.554	0.879	0.881	0.208
Baseline+info	0.596	0.928	0.924	0.205
Baseline+GSP	0.894	0.937	0.907	0.922
Baseline+GSP+info	0.973	0.971	0.962	0.986

J. Shi, B. Foggo, and N. Yu, "Power System Event Identification based on Deep Neural Network with Information Loading," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5622-5632, Nov. 2021.

Renewable Energy Resource Model

Solar PV System Performance Model: used to understand and predict energy or power output from PV systems under a wide range of environmental, design, and site.

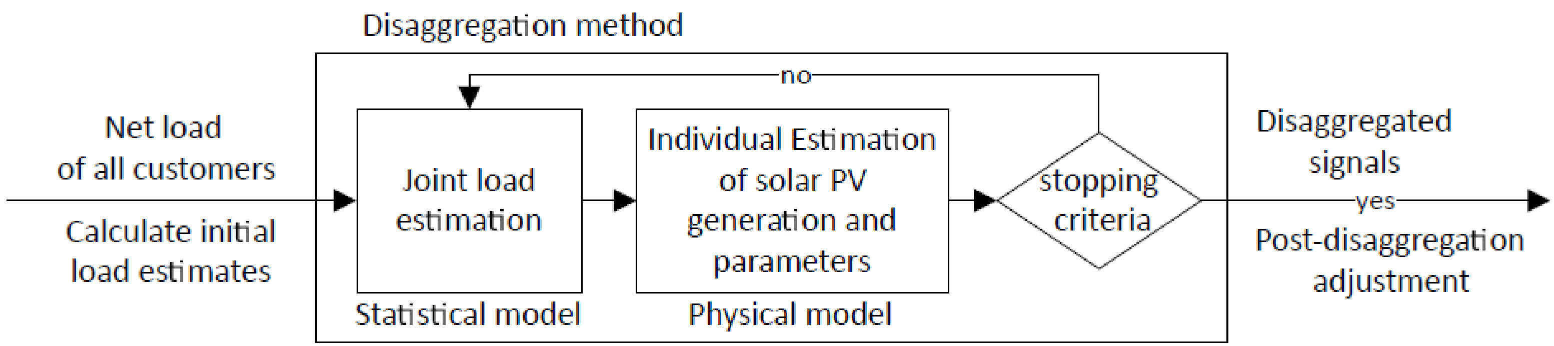
Physical Solar PV Performance Model
Estimation of Behind-the-Meter Solar Generation



Net Metering: $net\ load = load - solar\ generation$

Mixture Hidden Markov Model for Load Modeling

$$L_{nt} = a_{z_{nt}} + b_n + c_{n,z_{nt}} x_t + \epsilon_{nt},\ n = 1 \dots N,$$
$$t = 1 \dots T,\ b_n \sim N(0, \sigma^2),\ \epsilon_{nt} \sim N(0, \lambda_{z_{nt}}^2)$$



$$\underset{\boldsymbol{\theta}_S}{\operatorname{argmin}} \sum_{t=1}^{T} \left(S_t - g_t(\boldsymbol{\theta}_S)\right)^2$$
subject $\boldsymbol{\theta}_{S,min} \le \boldsymbol{\theta}_S \le \boldsymbol{\theta}_{S,max}$

$g_t(\boldsymbol{\theta}_S)$: solar PV generation at time t based on the **physical solar PV system performance model** with the technical parameters $\boldsymbol{\theta}_S$.

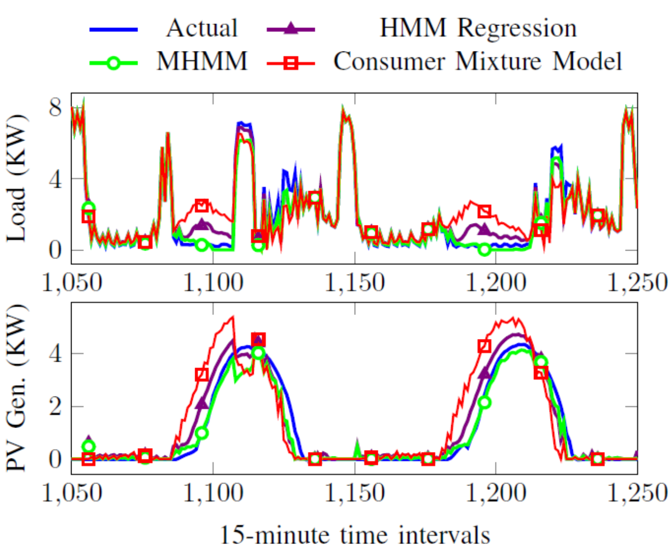
$$P_{ac} = g(\boldsymbol{\theta}_S) = \eta(\eta_{nom}, P_{dc}) P_{dc}$$

$$P_{dc} = g'(P_{dc0}, \theta_t, \theta_{az}, l) = (1 - l) \times \frac{E_{tr}(\theta_t, \theta_{az})}{E_0} P_{dc0}[1 + \gamma(T_c(\theta_t, \theta_{az}) - T_0)]$$

Testing Results

Error Metric	Variable	MHMM (solar panel scenario 1)	HMM reg. (solar panel scenario 1)	MHMM (solar panel scenario 2)	HMM reg. (solar panel scenario 2)	Consumer Mixture Model	SunDance Model
MSE	Solar	**0.13**	0.19	**0.12**	0.18	0.37	0.54
	Load	**0.13**	0.19	**0.12**	0.18	0.37	0.49
MASE	Solar	**2.13**	2.61	**2.11**	2.58	3.85	3.74
	Load	**0.43**	0.48	**0.42**	0.48	0.74	0.81
CV	Solar	**0.47**	0.58	**0.45**	0.57	0.77	0.85
	Load	**0.29**	0.33	**0.28**	0.32	0.46	0.57

- Validated using real-world smart meter and solar PV generation data from Austin, Texas.
- The MHMM follows the actual load much more closely than the benchmark algorithms during the low load periods.
- The synergistic combination of physical solar PV system performance model and the statistical community load model yields more accurate solar PV generation estimation.
- The MHMM allows sharing of info. across individual customers, which leads to more accurate load and solar PV generation estimates.



F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Joint Estimation of Behind-the-Meter Solar Generation in a Community," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 682-694, Jan. 2021.

Steady State Model of Power Systems

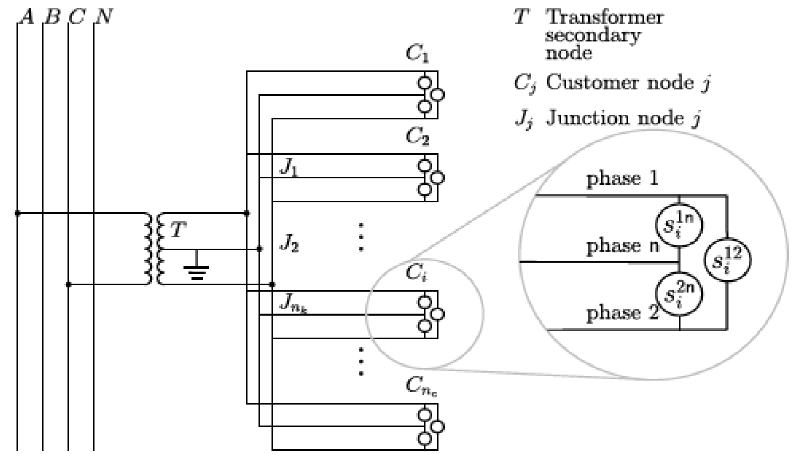
Power Flow Model: Linearized or Nonlinear Model of Power Flow

Power Flow Model:
Physics-informed Electricity Theft Detection

- Key Idea
  - There exists approximate linear models between power and voltage magnitudes to distribution secondaries.
  - A large discrepancy between estimated & measured power consumption data indicates potential theft.

$$y_i(t) = \left[\boldsymbol{x}(\boldsymbol{t})^T \quad \sum_{j=1}^{n_c} y_j(t)\right]\begin{bmatrix}\boldsymbol{\beta}_i^{\chi} \\ \beta_i^{y}\end{bmatrix} + \epsilon_i'(t)$$

Where $\boldsymbol{x}(\boldsymbol{t}) = [1, v_1(t), v_2(t), \cdots, v_{n_c}(t)]^T$, $y_i(t) = p_i(t)$



- Let $\tilde{y}_i^{(e)}$ and $\tilde{y}_i$ denote the out-of-sample residual time series for the energy thief

Lemma 1. $\tilde{y}_i^{(e)} - \tilde{y}_i = -\sum_{j \neq i} \beta_j^y y_i^s$

Lemma 1 & Lemma 3 combine to show that a thief's residuals will become negative once he or she begins to steal power.

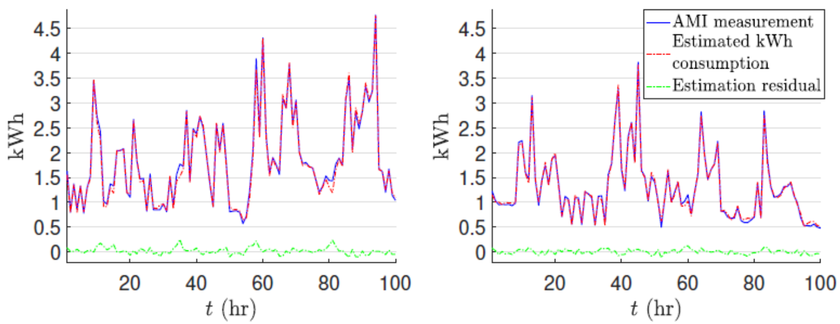
Lemma 2. $\sum_j \tilde{y}_j^{(e)} = \sum_j \tilde{y}_j = 0$

Lemma 2 shows that the residuals of the other customers will raise to balance their sum.

Lemma 3. $\forall \delta > 0$, there exists a training data window length $T > 0$ such that for each j: $\mathbb{P}\left(\beta_j^y \geq -\delta\right) > 1 - \delta$

Yuanqi Gao, Brandon Foggo, and Nanpeng Yu, "A Physically Inspired Data-driven Model for Electricity Theft Detection with Smart Meter Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, 2019.

Testing Results with Real-world Smart Meter Data

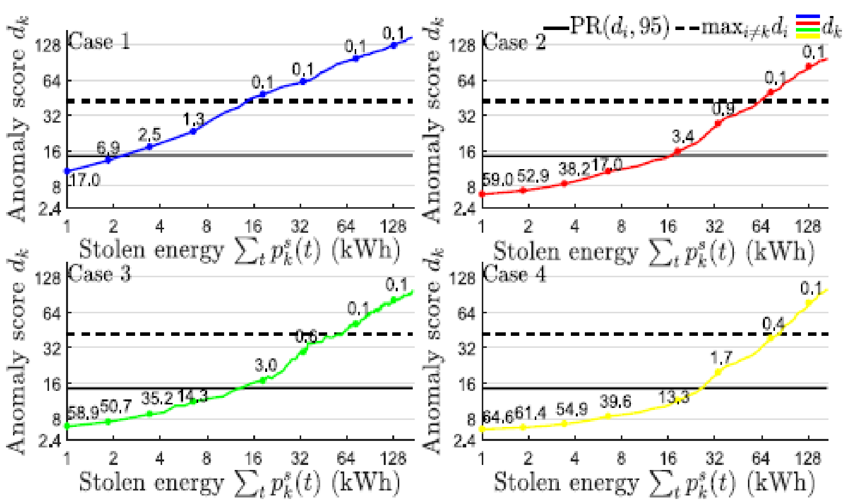


(a) In-sample ($t^{\mathcal{D}}$) (b) Out-of-sample ($t^{\mathcal{D}_a}$)

- 12 KV circuit from Southern California Edison, 6 months of smart meter data from 980 customers and 190 transformers.
- The average electric load consumed by the customer is 1.6 kWh.
- The mean of the estimation residual is -0.01 kWh and its standard deviation is 0.1 kWh.

d_k	Ranking	$\sum_i d_i/N$	PR(d_i,95)	$\max_{i \neq k} d_i$
79.4	1 (0.1%)	8.1	14.6	42.2

- Anomaly score for customer k is much higher than that of all other customers in the feeder



- Anomaly score increases monotonically with the amount of stolen electricity.
- In all cases, customer k's anomaly score will surpass the 95th percentile of all customers if it steals more than 32 kWh in two weeks or 0.19 kWh per hour.
- In cases 1-3, if customer k steals more than 0.38 kWh of power per hour, then its anomaly score will the largest of all customers.

Physics-informed Phase Connectivity Identification in Power Distribution System

- The time difference version of the physical model of distribution network

$$\widetilde{\boldsymbol{v}}(t) = X\widetilde{\boldsymbol{v}}^{ref}(t) + X\hat{K}X^T\widetilde{\boldsymbol{p}}(t) + X\hat{L}X^T\widetilde{\boldsymbol{q}}(t) + \boldsymbol{n}(t)$$

- The likelihood of observing $\{\widetilde{\boldsymbol{v}}(t)\}_{t=1}^{T}$, given $\boldsymbol{x}$, $\{\widetilde{\boldsymbol{p}}(t)\}_{t=1}^{T}$ and $\{\widetilde{\boldsymbol{q}}(t)\}_{t=1}^{T}$ is

$$Prob(\{\widetilde{\boldsymbol{v}}(t)\}_{t=1}^{T} \mid \{\widetilde{\boldsymbol{p}}(t)\}_{t=1}^{T}, \{\widetilde{\boldsymbol{q}}(t)\}_{t=1}^{T}; \boldsymbol{x})$$
$$= \frac{|\Sigma_N|^{-\frac{T}{2}}}{(2\pi)^{\frac{MT}{2}}} \times \exp\{-\frac{1}{2}\sum_{t=1}^{T}[\widetilde{\boldsymbol{v}}(t) - \widetilde{\boldsymbol{v}}(t,\boldsymbol{x})]^T \Sigma_N^{-1}[\widetilde{\boldsymbol{v}}(t) - \widetilde{\boldsymbol{v}}(t,\boldsymbol{x})]\}$$

Lemma 1. Let $\boldsymbol{x}^*$ be the correct phase connection. If the following two conditions are satisfied, then as $T \to \infty$, $\boldsymbol{x}^*$ is a global optimizer of $f(\boldsymbol{x})$.

1. $\boldsymbol{n}(t_k)$ is i.i.d. and independent of $\widetilde{\boldsymbol{v}}^{ref}(t_l), \widetilde{\boldsymbol{p}}(t_l), and\ \widetilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+$.

2. $\widetilde{\boldsymbol{v}}^{ref}(t_k), \widetilde{\boldsymbol{p}}(t_k), and\ \widetilde{\boldsymbol{q}}(t_k)$ are independent of $\widetilde{\boldsymbol{v}}^{ref}(t_l), \widetilde{\boldsymbol{p}}(t_l), and\ \widetilde{\boldsymbol{q}}(t_l)$, for $\forall t_k, t_l \in Z^+$, $t_k \neq t_l$.

- Directly minimizing $f(\boldsymbol{x})$ is difficult.
- **Key Idea**: phase identification problem → maximum marginal likelihood estimation (MMLE) problem.

W. Wang and N. Yu, "Maximum Marginal Likelihood Estimation of Phase Connections in Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp.3906-3917, Sep. 2020.

Numerical Results with Smart Meter Data

Number of Loads per Phase in the IEEE Test Circuits

Feeder	A	B	C	AB	BC	CA	ABC	Total
37-bus	5	5	6	3	2	2	2	25
123-bus	18	17	17	9	9	10	5	85
342-bus	30	38	31	35	31	33	10	208

- Smart Meter Data
  - Length: 90 days of hourly average real power consumption data
  - Provided by FortisBC
- Measurement noise ~ 0.1 and 0.2 accuracy class smart meters established in ANSI.
- Voltage measurements rounded to the nearest 1 V for the primary loads & 0.1 V for the secondary loads.

Phase Identification Accuracy of Different Methods with 90 days of Meter Data

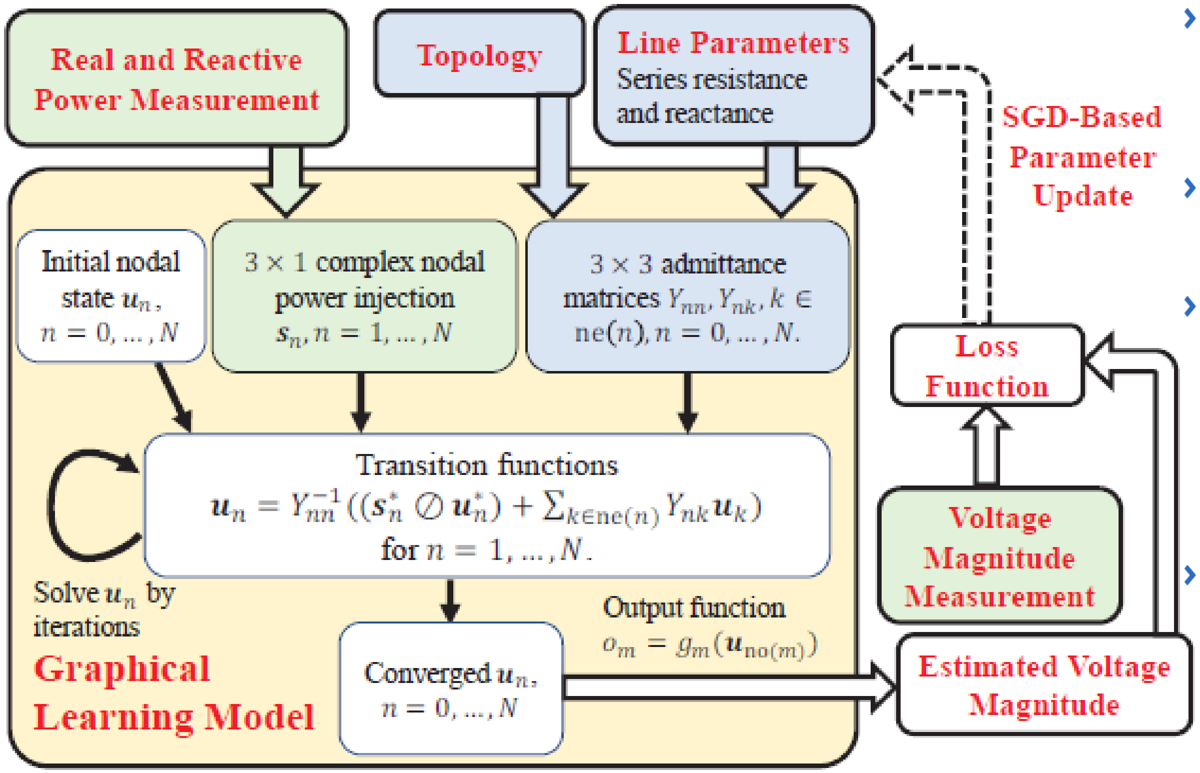
Method	Meter Class	37-Bus Feeder	123-Bus Feeder	342-Bus Feeder
Correlation-based Approach	0.1%	100%	98.75%	81.82%
	0.2%	100%	97.5%	81.31%
Clustering-based Approach	0.1%	100%	100%	93.43%
	0.2%	100%	98.75%	91.41%
MMLE-based Algorithm	0.1%	100%	100%	100%
	0.2%	100%	100%	100%

- The MMLE-based algorithm outperforms the correlation and clustering-based approaches.
- The improvement in accuracy increases as the complexity of the distribution feeder increases.

Graphical Model of Power Network

Graph Model for Power Systems: Nodes → Vertices, Power Lines → Branches

Graph Model: Physics-informed Graphical Learning for Distribution Line Parameter Estimation

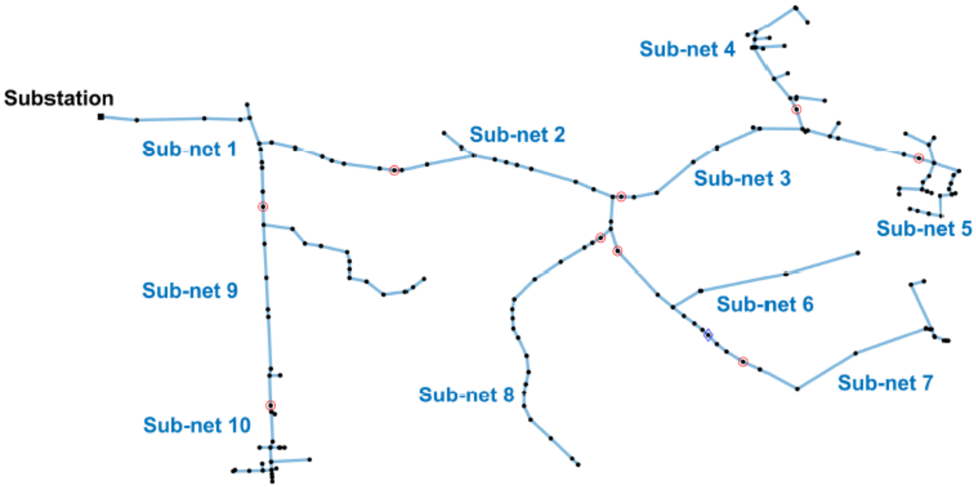


- Key Idea: **Embed physical equations of power flow in the graphical learning model**
- Inspired by graphical neural network (GNN)
- Difference between physics-informed GL and GNN
  - Leverage 3Φ power flow-based physical transition fcn. to replace the deep neural networks in GNN.
- Key Step: Derive the gradient of voltage magnitude loss function w.r.t. line segment's resistance and reactance parameters with an iterative method.

- Estimate distribution line parameters with SGD considering prior estimates of line parameters and physical constraints.
- Improve computation efficiency with grid partition scheme and fast forward/backward function.

Wenyu Wang and Nanpeng Yu, "Estimate Three-phase Distribution Line Parameters with Physics-informed Graphical Learning Method," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3577-3591, Sep. 2022,

Fast GL and Numerical Study Results



MADR Improvement of Parameter Estimation Methods in the Test Feeder (Avg. / Choose Optimal Value)

Network	LMLE	FGL	FGL+CON
Whole Network	10.8% / 13.5%	20.7% / 25.5%	29.4% / 30.9%
Sub-Net 1	10.3% / 9.1%	20.1% / 21.6%	23.1% / 27.1%
Sub-Net 2	7.3% / 9.2%	13.6% / 20.9%	26.7% / 29.3%
Sub-Net 3	9.9% / 12.2%	34.5% / 40.8%	41.6% / 43.7%
Sub-Net 4	4.5% / 4.7%	5.1% / 5.0%	12.4% / 13.2%
Sub-Net 5	11.6% / 14.8%	20.8% / 22.1%	21.0% / 22.3%
Sub-Net 6	12.0% / 24.3%	37.0% / 62.4%	61.8% / 63.5%
Sub-Net 7	9.3% / 9.3%	16.2% / 18.0%	31.6% / 32.9%
Sub-Net 8	13.9% / 16.5%	22.3% / 23.0%	24.9% / 25.5%
Sub-Net 9	17.3% / 21.3%	30.8% / 34.5%	37.9% / 38.2%
Sub-Net 10	7.6% / 9.6%	2.2% / 8.9%	19.0% / 20.4%

- 15 days of smart meter data from 13.2 kV distribution circuit in National Grid
- 177 line sections, 491 loads, 23 solar PV systems
- Feeder partitioned into 10-subnetworks (parallel estimation)
- Fast Graphical Learning has significantly higher MADR improvement (30% improvement in MADR)
- Approximately 10 times reduction in computation time

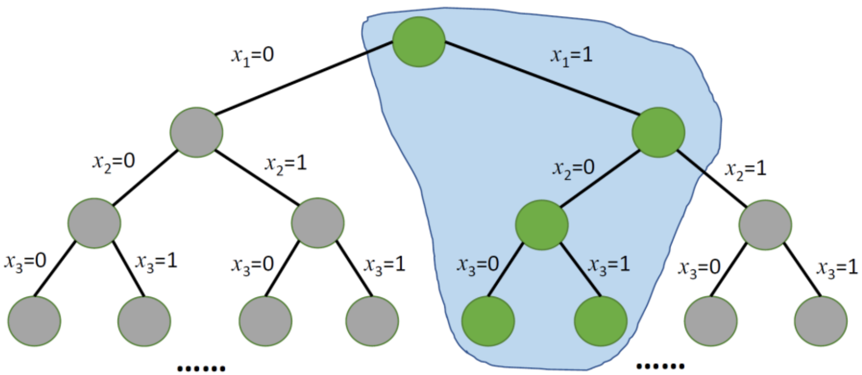
Avg. Runtime (Seconds) of Main Functions of Parameter Estimation Methods in Sub-networks of different sizes

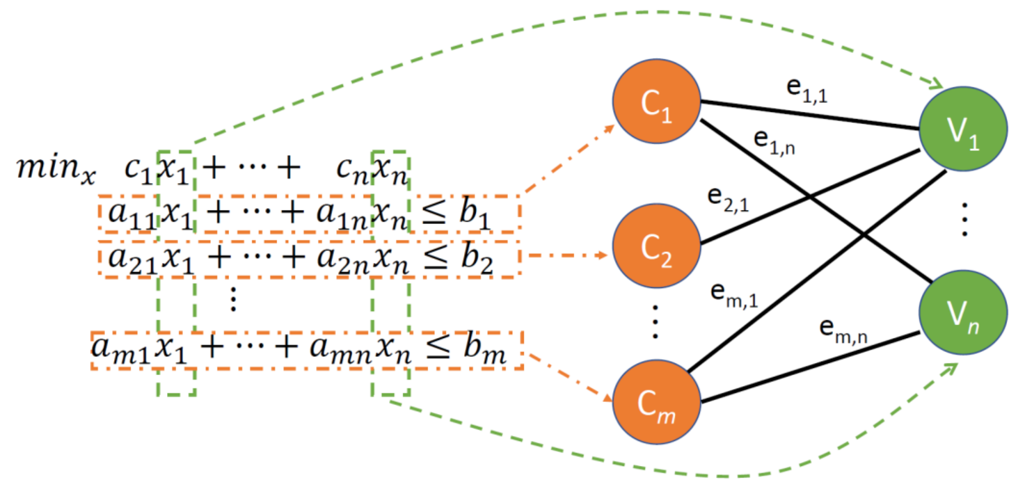
Method	Function	7-Bus	14-Bus	22-Bus
FGL	Fast-FORWARD	0.0142	0.0313	0.0519
	Fast-BACKWARD	0.068	0.1046	0.1761
GL	FORWARD	0.3028	1.1603	3.2839
	BACKWARD	0.1057	0.3325	1.0065
LMLE	Gradient Calculation	0.0079	0.1866	0.8531

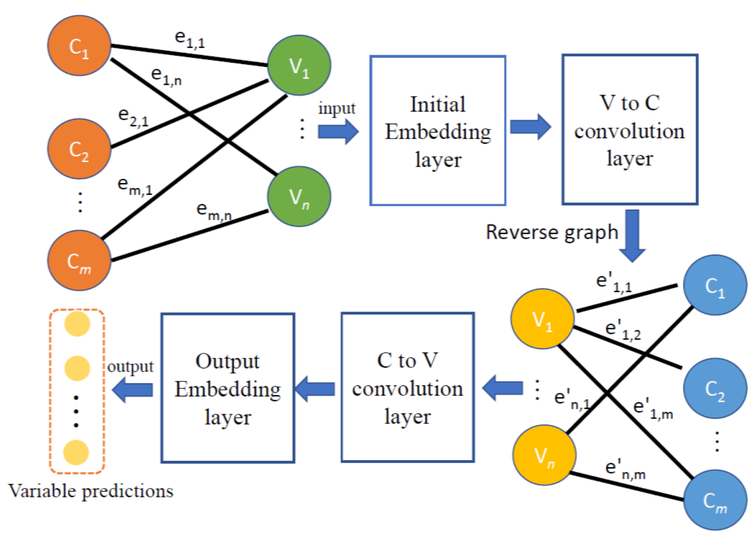
Wenyu Wang, Nanpeng Yu, and Yue Zhao, "Fast Graphical Learning Method for Parameter Estimation in Large-Scale Distribution Networks," *IEEE SmartGridComm*, 2022.

Solve Large-Scale MIP Problem with Generic GL

- Solving unit commitment (UC) or security constrained UC (SCUC) problems is crucial to market operations
- The UC or SCUC problem boils down to solving large-scale mixed integer programing (MIP) problems.
- State-of-the-art commercial solvers (e.g. Gurobi, CPLEX) use sophisticated branch and bound algorithms





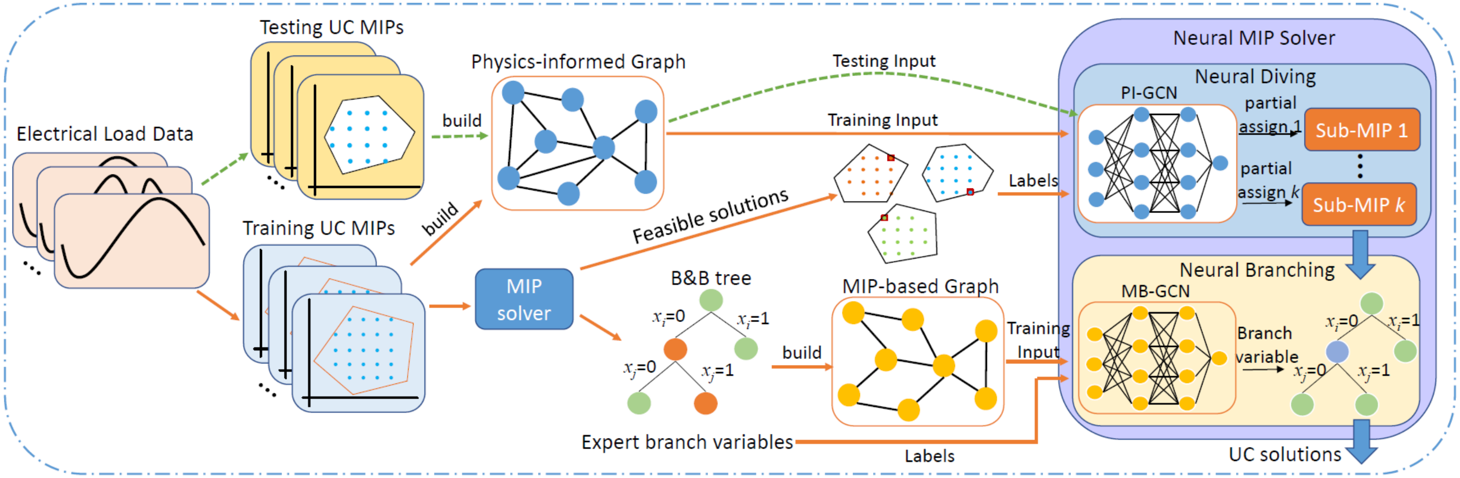


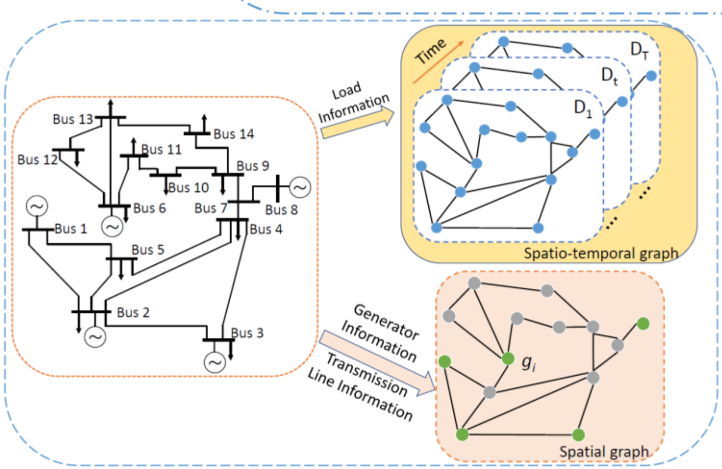
- Used to solve large-scale MIP problems (e.g. UC)
- Drawbacks
  - Scalability: the number of constraints in the bipartite graph increases exponentially with the number of nodes in a power system

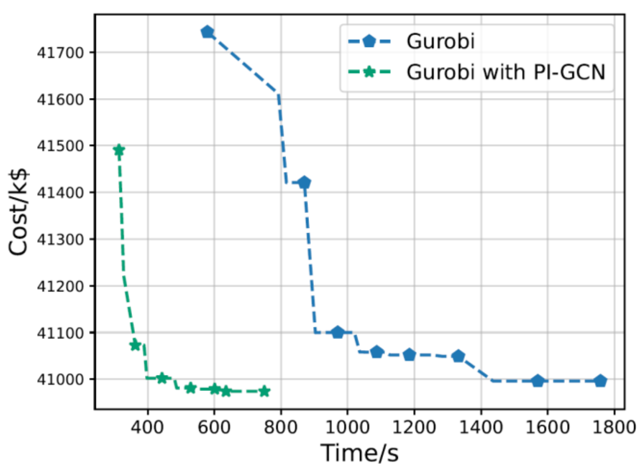
Nair, Vinod, et al. "Solving mixed integer programs using neural networks." arXiv preprint arXiv:2012.13349 (2020).
Google Deep Mind

Accelerate UC Solution with Physics-informed GL

- Learn to branch and dive for optimization problems on large-scale networks
- Instead of adopting bi-partite graph, leveraging power network to perform physics-informed GL







- Outperforms commercial MIP solver (e.g. Gurobi and CPLEX) on large-scale UCR problems (6515-bus)
- Better scalability, optimality and interpretability

J. Qin & N. Yu, "Solve Large-scale Unit Commitment Problems by Physics-informed Graph Learning" arXiv preprint arXiv:2311.15216

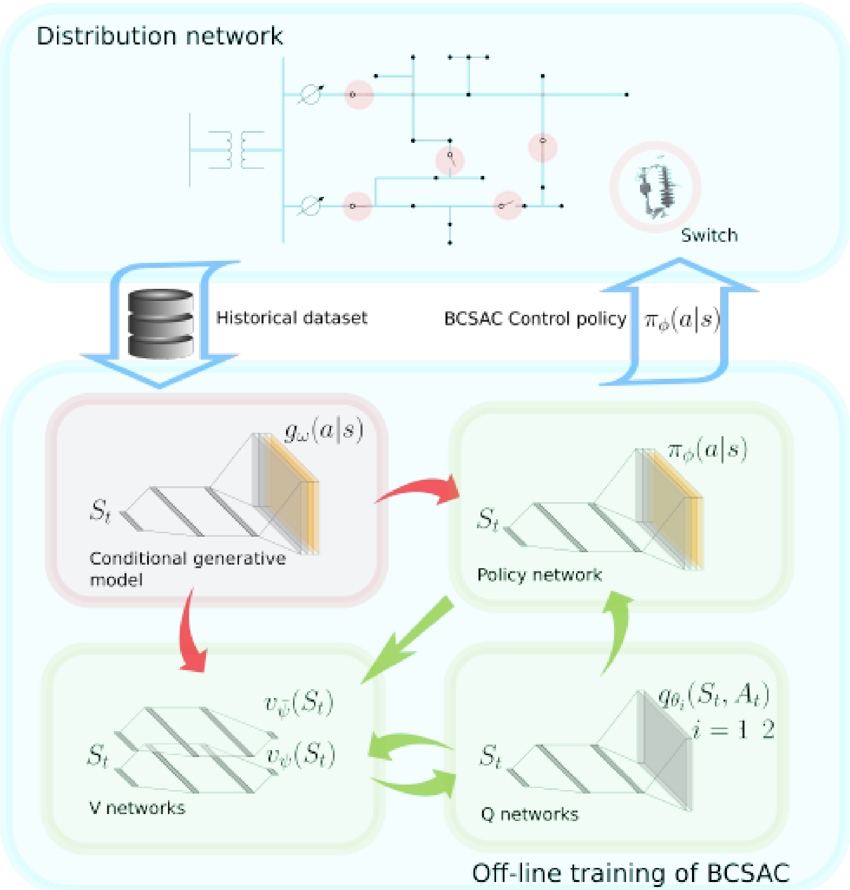
System Control Model

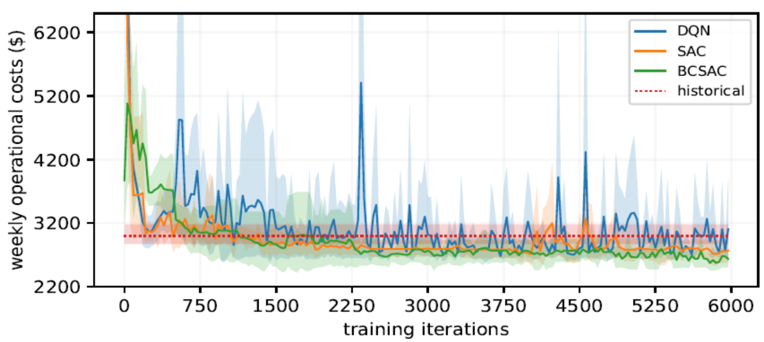
System Control Model: Model-based Controller or Heuristic Control Policy from Human Operator

Control Model: Batch Constrained Reinforcement Learning-based Distribution System Control

- Motivation
  - Costly & time consuming to learn optimal control strategy by directly interacting with physical network.
  - Learning from finite historical dataset lead to large extrapolation errors.
- Solution: Batch-constrained soft actor-critic algorithm
- Key Idea: train a control policy
  - Maximize the total discounted return
  - Minimize dissimilarity between *learned control policy* & *behavior policy* of the batch data

Overall Framework





Weekly Operational Cost of test feeder

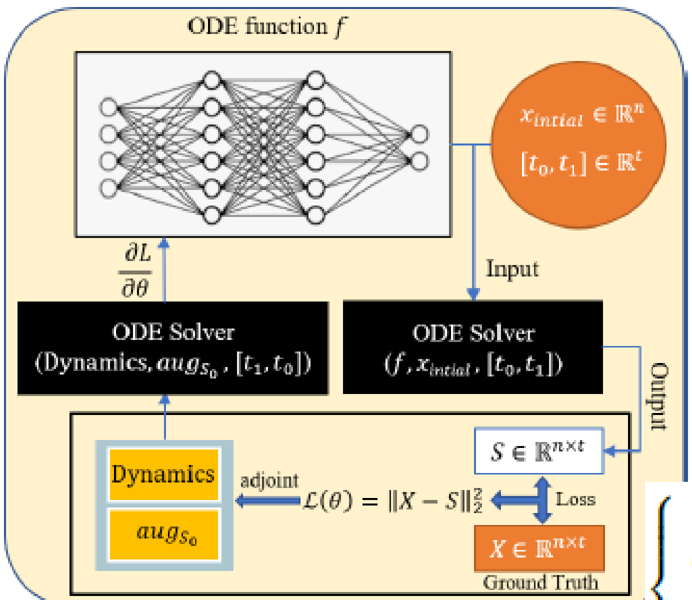
Y. Gao, W. Wang, J. Shi, and N. Yu, "Batch-constrained Reinforcement Learning for Dynamic Distribution Network Reconfiguration," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5357-5369, 2020.

Dynamic Model of Power Systems

Dynamic Model of Power System: Differential and Algebraic Equations (DAEs), Hamiltonian and Nearly Hamiltonian System

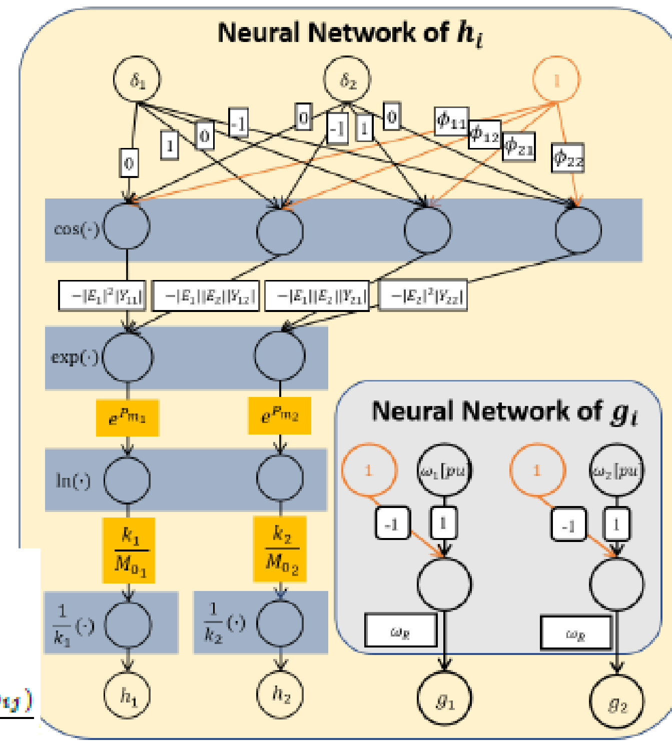
Dynamic Model: Generator Dynamic Parameter Estimation with Physics-based Neural Ordinary Differential Equations

- Key Ideas
  - Convert the forward solver of ODEs for power system dynamics into physics-informed neural networks.
  - Calculate the loss function based on the difference between dynamic simulation results from the neural networks and pseudo PMU measurements.



- Calculate the gradient of the loss function w.r.t. dynamic parameters based on the neural ODE technique and the adjoint method.
- Update the dynamic parameters iteratively with a quasi-Newton Method.

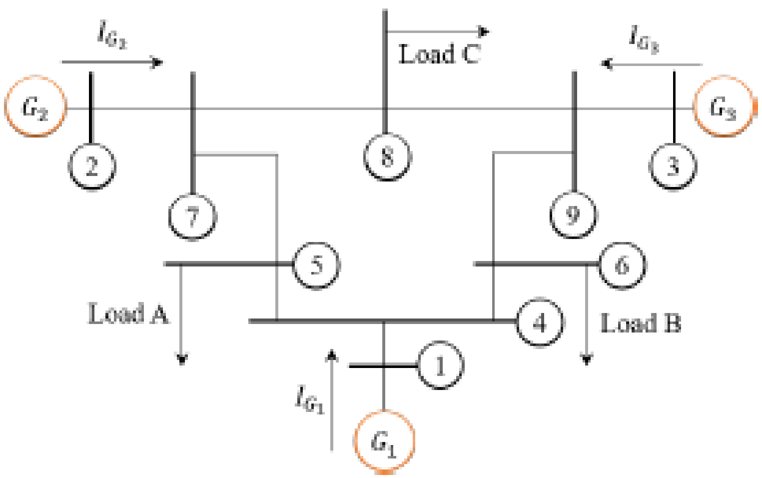
$$
\begin{cases}
\dot{\delta}_i = g_i(\omega_i[\text{pu}]) \\
\dot{\omega}_i[\text{pu}] = h_i(\delta_{i=1,\dots,n}) \\
g_i = \omega_R(\omega_i[\text{pu}] - 1) \\
h_i = \dfrac{P_{m_i} - \sum_{j=1}^{n} |E_i||E_j||Y_{ij}|cos(\delta_i - \delta_j - \phi_{ij})}{M_{0_i}}
\end{cases}
$$



X. Kong, K. Yamashita, B. Foggo, and N. Yu, "Dynamic Parameter Estimation with Physics-based Neural Ordinary Differential Equations" *IEEE PES GM*, 2022.

Numerical Study Results

WECC 3-machine 9-bus system



- Physics-based neural ODE algorithm has much shorter computation time than the baseline algorithm.

RELATIVE ESTIMATION ERROR (%) OF BASELINE AND NEURALODE-BASED METHOD

		Data Length 1s	3s	5s	Initial REE
Parameters	P_{m1}	2.36 / **1.50**	7.70 / **1.49**	7.47 / **1.26**	4.21
	P_{m2}	0.76 / **0.01**	5.05 / **0.12**	5.01 / **0.18**	5.77
	P_{m3}	1.05 / **0.41**	6.27 / **0.32**	6.52 / **0.46**	4.76
	M_{0_1}	5.80 / **2.86**	5.52 / **2.06**	6.31 / **3.09**	4.80
	M_{0_2}	4.36 / **3.19**	5.92 / **2.18**	5.88 / **3.67**	6.10
	M_{0_3}	**4.58** / 8.37	5.50 / **4.82**	**5.49** /10.51	5.33
Average		3.15 / **2.72**	5.99 / **1.83**	6.11 / **3.20**	5.16

Note: Baseline / Physics-based Neural ODE Algorithm.

- Generate noisy PMU measurements from dynamic simulation data.
- A single transmission line is disconnected at 5s. The simulation time is 10s.
- Two disturbance scenarios (between nodes 5 and 7, between nodes 8 and 9)
- When the PMU data length is 3s, our proposed algorithm achieves the lowest relative estimation error.
- The physics-based neural ODE algorithm outperforms the baseline algorithm in terms of estimation accuracy for most of the unknown parameters

RUNNING TIME (S) OF BASELINE AND NEURAL ODE-BASED METHOD.

Data Length	Running Time (second) Baseline	Neural ODE-based	Learning Rate
1s	8.38	3.78	0.5 / 0.5
3s	38.55	4.82	0.05 / 0.5
5s	100.25	4.67	0.05 / 0.5

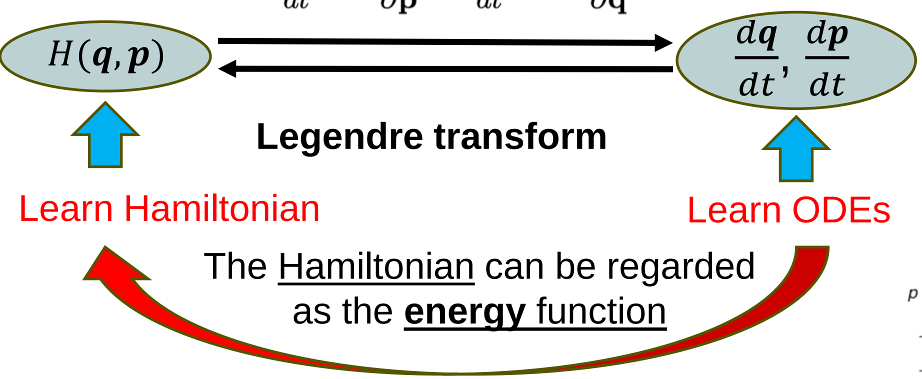
Note: Baseline / Physics-based Neural ODE Algorithm.

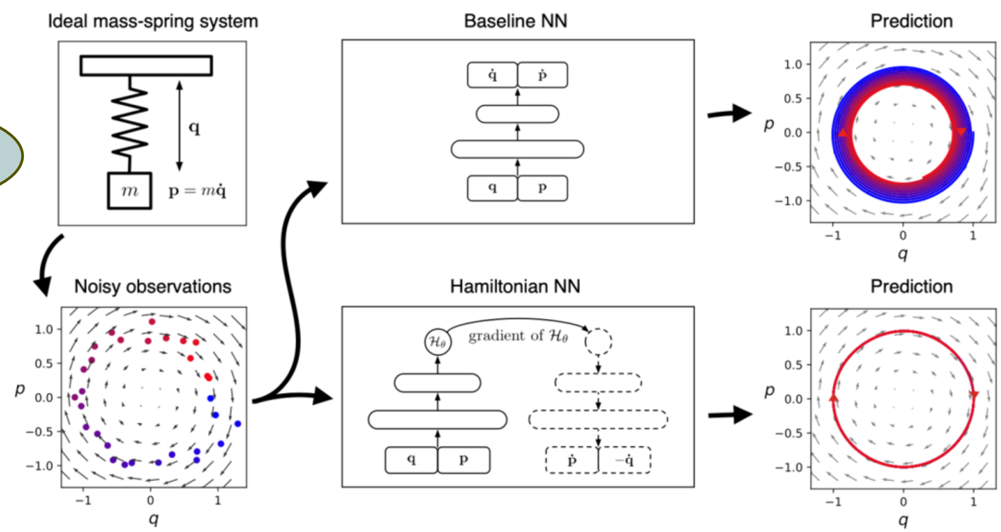
- When the data length is 3s, the running time of our model is just 4.82 seconds, which is nearly 8 times faster than the baseline model.
- Mini-batch scheme of neural network training shortens model running time.

Learning Dynamic System with Hamiltonian Neural Networks

- **Hamiltonian mechanics:** can predict the motion of a energy-conserved system.
- State variables: generalized position $\mathbf{q} = [q_1, q_2, \cdots, q_n]^T$ momentum $\mathbf{p} = [p_1, p_2, \cdots, p_n]^T$
- $\mathbf{q}$ and $\mathbf{p}$ correspond to voltage angle δ and angular speed ω in power system dynamic model
- Model single machine infinite bus system as a Hamiltonian: $m_1\ddot{\delta} + d_1\dot{\delta} + B_{12}V_1V_2\sin(\delta) - P_1 = 0$
- Can we learn the Hamiltonian or Lagrangian function (energy conservation law) directly?

$$\frac{d\mathbf{q}}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}}.$$





[1] Greydanus S, Dzamba M, Yosinski J. Hamiltonian neural networks. Advances in neural information processing systems, 2019.
[2] Cranmer M, Greydanus S, Hoyer S, et al. Lagrangian neural networks. arXiv preprint arXiv:2003.04630, 2020.

Learning Power System Dynamics: Nearly Hamiltonian NN

- Can we formulate the SMIB system as a Hamiltonian function?
- $m_1\ddot{\delta} + d_1\dot{\delta} + B_{12}V_1V_2\sin(\delta) - P_1 = 0$
- Unfortunately, the answer is **No**. If the damping coefficient d_1 is positive, then the SMIB system is a dissipative system instead of a conservative system.

Hamiltonian System

Hamiltonian differential equations

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}}$$

Total energy is conserved:

$$\frac{dH}{dt} = \frac{\partial^T H}{\partial \mathbf{q}}\frac{d\mathbf{q}}{dt} + \frac{\partial^T H}{\partial \mathbf{p}}\frac{d\mathbf{p}}{dt} = 0$$

Nearly Hamiltonian System

Hamiltonian differential equations

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}} \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}} - \mathbf{D}\frac{\partial H}{\partial \mathbf{p}}$$

Total energy decreases:

$$\frac{dH}{dt} = \frac{\partial^T H}{\partial \mathbf{q}}\frac{d\mathbf{q}}{dt} + \frac{\partial^T H}{\partial \mathbf{p}}\frac{d\mathbf{p}}{dt} = -\frac{\partial^T H}{\partial \mathbf{p}}\mathbf{D}\frac{\partial H}{\partial \mathbf{p}} \leq 0$$

$$\begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -d_1 \end{bmatrix} \begin{bmatrix} \frac{\partial H}{\partial q} \\ \frac{\partial H}{\partial p} \end{bmatrix}$$

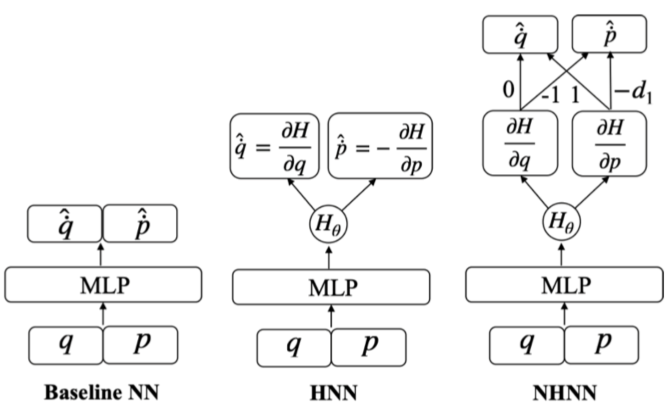
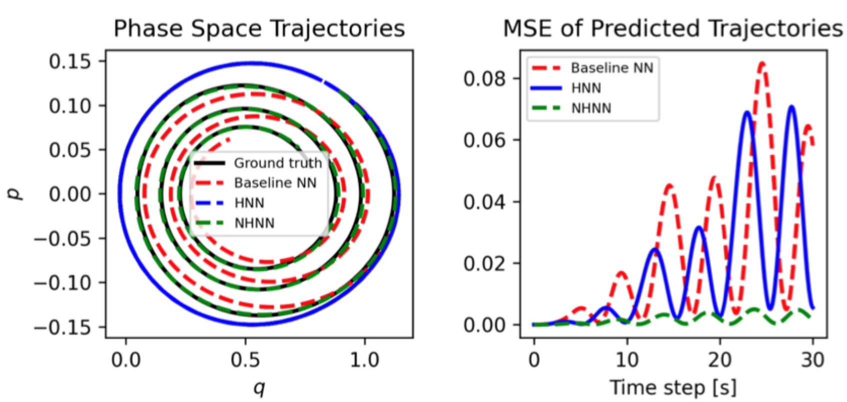


Fig. 2. Architecture of baseline neural network, HNN and NHNN



$(b)\ d_1 = 0.02$

Shaorong Zhang and Nanpeng Yu, "Learning Power System Dynamics with Nearly-Hamiltonian Neural Networks," *IEEE PES General Meeting*, 2023.

Summary

- Key: synergistic combine ML algorithms with physical domain knowledge
- Drastic improvement in accuracy, generalization capability, sample efficiency, safety and interpretability
- Types of power system domain knowledge or physical model
  - Unique data properties: low rank, high/low entropy
  - Steady state models: renewable resource model, topology, power flow model
  - Dynamic models: dynamic simulation and control models, Hamiltonian function
- How to combine physical model with machine learning model?
  - Iterative fitting the physical and data-driven models
  - Embed physics model in machine learning algorithms
    - Data preprocessing, loss function, neural architecture

Discussions

- Importance of Collaborating with Electric Utilities
  - Learn about current practice and realistic challenges
  - Work on solutions and datasets that could address real-world problems
    - Smart meter data for phase connectivity identification
    - PMU data from all three U.S. interconnections for system monitoring
- Real-world data is the best source for validating your ML algorithms
  - Simulation data does not possess the same property as real-world data
  - Algorithm trained with simulation data will struggle to solve real-world problems
- Reproducibility: Dataset and Open-Source Software
  - Learning Power System Dynamics with Neural ODE and NHNN
    - https://github.com/szhan311/Neural_ODE_Power_System_Dynamic, https://github.com/szhan311/Nearly-Hamiltonian-Neural-Network
  - RL-based Control in Distribution System
    - A RL-based VVC Dataset and Testing Environment. https://github.com/yg-smile/RL VVC dataset
  - Repository of synthetic PMU data generated from U.S. power grid data
    - pmuBAGE: https://github.com/NanpengYu/pmuBAGE

Recent Publications

- Theoretical Machine Learning
  - B. Foggo, N. Yu, J. Shi and Y. Gao, "Information Losses in Neural Classifiers from Sampling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4073-4083, 2020.
  - Brandon Foggo and Nanpeng Yu, "Analyzing Data Selection Techniques with Tools from the Theory of Information Losses," *IEEE International Conference on Big Data*, pp. 1-10, 2021.
  - Brandon Foggo and Nanpeng Yu, "On the Maximum Mutual Information Capacity of Neural Architectures," *ICML Workshop on Neural Compression: From Information Theory to Applications*, 2023.
- Machine Learning for Power Transmission Systems
  - Shaorong Zhang and Nanpeng Yu, "Learning Power System Dynamics with Nearly-Hamiltonian Neural Networks," *IEEE PES GM*, 2023.
  - X. Kong, B, Foggo, and N. Yu, “Online Voltage Event Detection Using Optimization with Structured Sparsity-Inducing Norms,“ *IEEE Transactions on Power Systems*, vol. 37, no. 5, Sep. 2022.
  - J. Shi, B. Foggo, and N. Yu, "Power System Event Identification based on Deep Neural Network with Information Loading," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5622-5632, Nov. 2021.
  - Shaorong Zhang, Koji Yamashita, and Nanpeng Yu, "Learning Power System Dynamics with Neural Ordinary Differential Equations," *IEEE PES General Meeting*, 2024.
  - Yinglun Li, Nanpeng Yu, and Wei Wang, "Machine Learning-Driven Virtual Bidding with Electricity Market Efficiency Analysis," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 354-364, Jan. 2022.
  - Yuanbin Cheng, Nanpeng Yu, Brandon Foggo, and Koji Yamashita, "Online Power System Event Detection via Bidirectional Generative Adversarial Networks," to appear in *IEEE Transactions on Power Systems*, 2022.
  - Brandon Foggo and Nanpeng Yu, "Online PMU Missing Value Replacement via Event-Participation Decomposition," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 488-496, Jan. 2022.
  - Xianghao Kong, Koji Yamashita, Brandon Foggo, and Nanpeng Yu, "Dynamic Parameter Estimation with Physics-based Neural Ordinary Differential Equations" *IEEE PES General Meeting*, 2022.
  - Jie Shi, Koji Yamashita, and Nanpeng Yu, "Power System Event Identification with Transfer Learning Using Large-scale Real-world Synchrophasor Data in the United States," *IEEE ISGT North America*, 2022.
  - Yuanbin Cheng, Koji Yamashita, and Nanpeng Yu, "Adversarial Attacks on Deep Neural Network-based Power System Event Classification Models," *IEEE ISGT Asia*, 2022.

Recent Publications

- Machine Learning for Power Distribution Systems and DERs
  - Farzana Kabir, Yuanqi Gao, Wenyu Wang and Nanpeng Yu, "Deep Reinforcement Learning-based Two-Timescale Volt-VAR Control with Degradation-aware Smart Inverters in Power Distribution Systems," *Applied Energy*, vol. 335, April 2023.
  - Wei Wang, Nanpeng Yu, Yuanqi Gao, and Jie Shi, "Safe Off-Policy Deep Reinforcement Learning Algorithm for Volt-VAR Control Problems in Power Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3008-3018, 2020.
  - Daner Hu, Zhenhui Ye, Yuanqi Gao, Yonggang Peng, and Nanpeng Yu, "Multi-agent Deep Reinforcement Learning for Voltage Control with Coordinated Active and Reactive Power Optimization," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4873-4866, Nov. 2022.
  - Zuzhao Ye, Yuanqi Gao, and Nanpeng Yu, "Learning to Operate an Electric Vehicle Charging Station Considering Vehicle-grid Integration," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3038-3048, 2022.
  - Wenyu Wang and Nanpeng Yu, "Maximum Marginal Likelihood Estimation of Phase Connections in Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp.3906-3917, Sep, 2020.
  - Yuanqi Gao, Brandon Foggo, and Nanpeng Yu "A Physically Inspired Data-Driven Model for Electricity Theft Detection with Smart Meter Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5076-5088, 2019.
  - Brandon Foggo and Nanpeng Yu, "Improving Supervised Phase Identification Through the Theory of Information Losses," *IEEE Transactions on Smart Grid*, vol. 11, pp. 2337-2346, 2020.
  - Jie Shi, Yuanqi Gao, Wei Wang, Nanpeng Yu, and Petros Ioannou, "Operating Electric Vehicle Fleet for Ride-Hailing Services with Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4822-4834, Nov. 2020.
  - Yuanqi Gao and Nanpeng Yu, "Model-Augmented Safe Reinforcement Learning for Volt-VAR Control in Power Distribution Networks," *Applied Energy*, vol. 313, May 2022.
  - Wenyu Wang and Nanpeng Yu, "Estimate Three-phase Distribution Line Parameters with Physics-informed Graphical Learning Method," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3577-3591, Sep. 2022.
  - Yuanqi Gao, Wei Wang, and Nanpeng Yu, "Consensus Multi-agent Reinforcement Learning for Volt-VAR Control in Power Distribution Networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3594-3604, July, 2021.
  - Yuanqi Gao, Wei Wang, Jie Shi, and Nanpeng Yu, "Batch-constrained Reinforcement Learning for Dynamic Distribution Network Reconfiguration," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5357-5369, 2020.
  - Farzana Kabir, Nanpeng Yu, Weixin Yao, Rui Yang, and Yingchen Zhang, "Joint Estimation of Behind-the-Meter Solar Generation in a Community," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 682-694, Jan. 2021.

Contact Information:
Nanpeng Yu, nyu@ece.ucr.edu
Lab Website: https://intra.ece.ucr.edu/~nyu/

Thank You

Questions?

Collaborating Companies: SCE, PG&E, National Grid, Fortis BC, Con Edison, NYPA, ComEd, Exelon, EPRI, RPU, ISO-NE and CAISO.

Funding Support: DOE, NSF, CEC, NYSERDA, EPRI, DEED, UCOP, NREL & Utilities.