SAND2022-4058C

Using Artificial Intelligence for Grid Data

Matthew J. Reno and Joseph A. Azzolini

Sandia National Laboratories

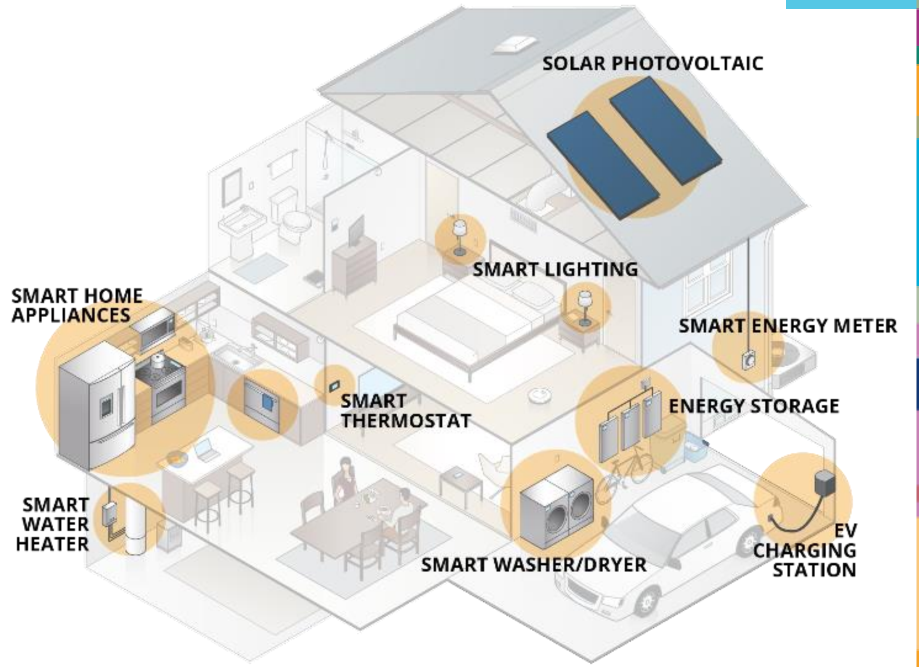
2022 Southeastern Energy Conference

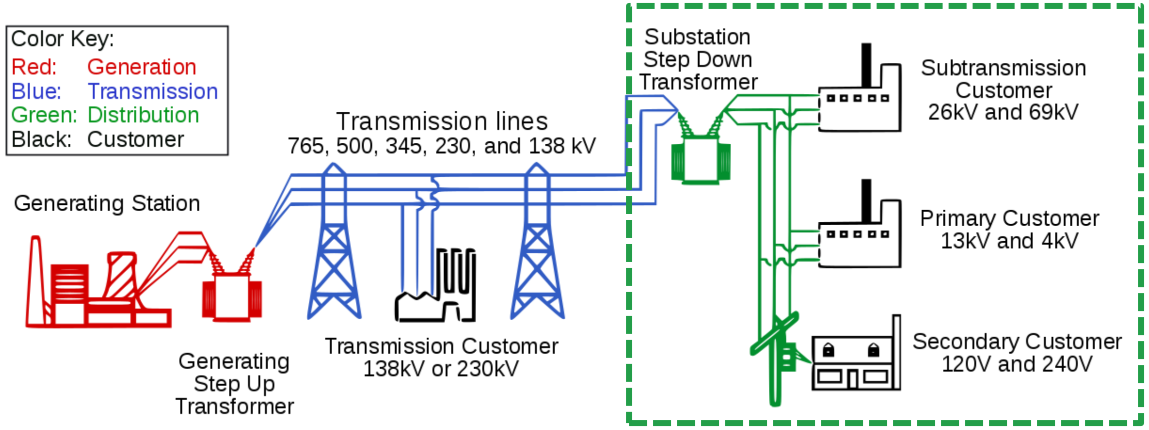
April 7, 2022

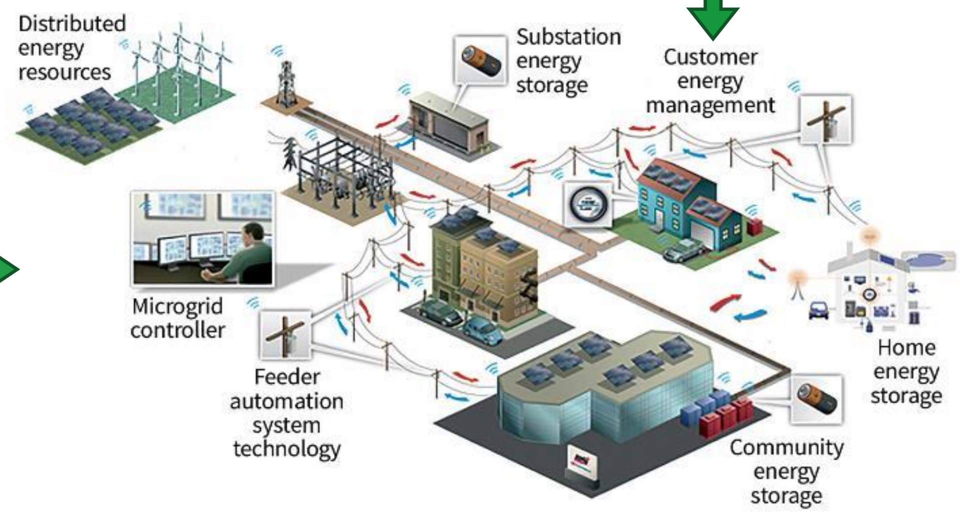
Electric Power Systems

Power Systems—and specifically distribution systems— are a perfect application for Machine Learning due to their complexity and large amounts of data

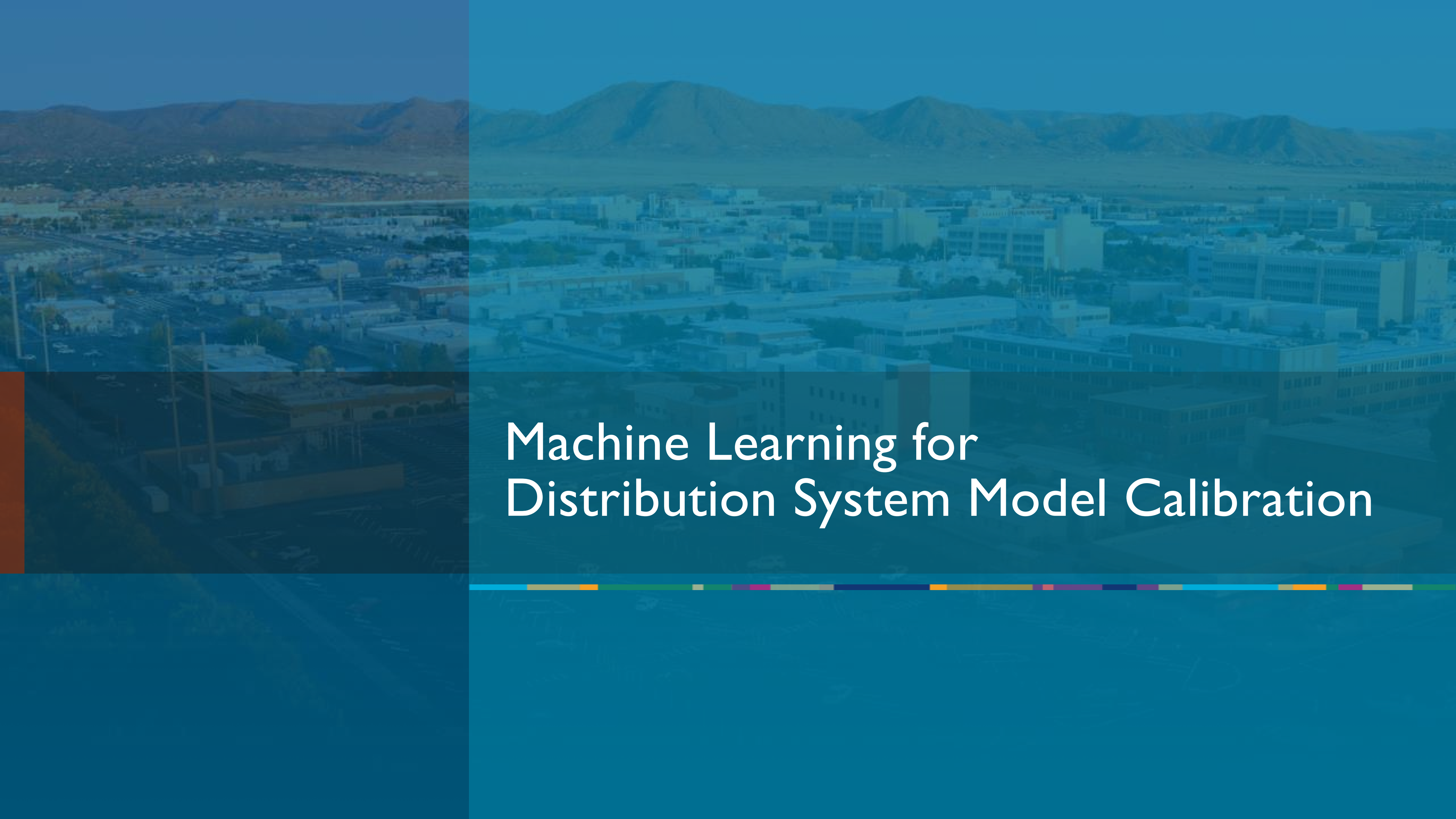
- About 300,000 miles of transmission lines vs. 6,000,000 miles of distribution lines
- About 20,000 substation transformers vs. 200,000,000 service transformers
- Additions of new sensing equipment such as smart meters and PMUs
- Advances in computing power for real-time learning/decision making
- New Artificial Intelligence algorithms to handle large datasets, transferable learning, and physics-based algorithms







http://www.ferc.gov/industries/electric/indus-act/reliability/blackout/ch1-3.pdf



Machine Learning for Distribution System Model Calibration

Distribution System Modeling

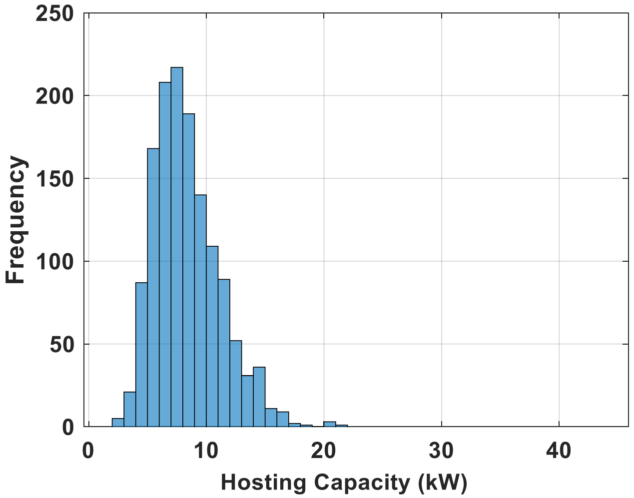
Power system models provide the foundation for nearly all analysis and planning tasks

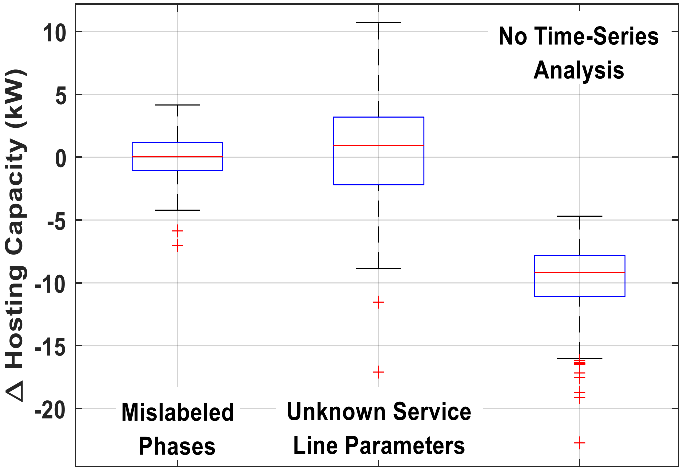
- Based on detailed knowledge of network topology, the physical and control parameters of various components, energy consumption data, etc.

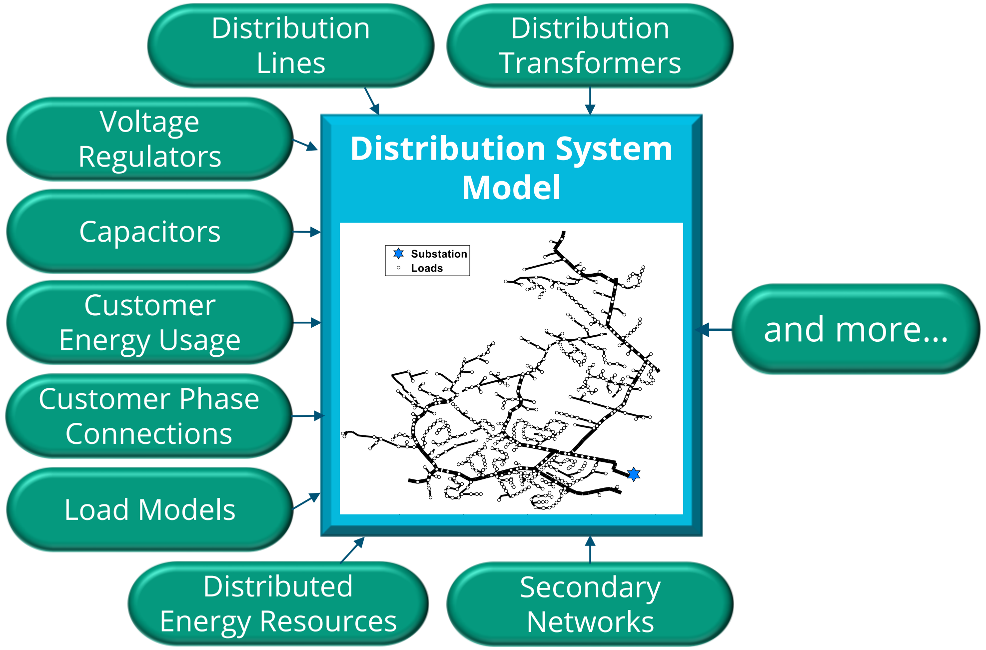
However for distribution systems…

- Creating and maintain models is **labor intensive**, and the process is **prone to errors**
- *A few volts* difference can result in PV hosting capacity varying by more than *200%*

Translates to high levels of uncertainty and conservative, worst-case analyses or rough rules of thumb







Distribution Lines

Distribution Transformers

Voltage Regulators

Capacitors

Customer Energy Usage

Customer Phase Connections

Load Models

Distribution System Model

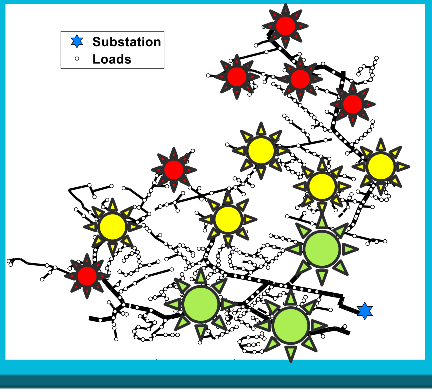
and more…

Distributed Energy Resources

Secondary Networks

Hosting Capacity Example:

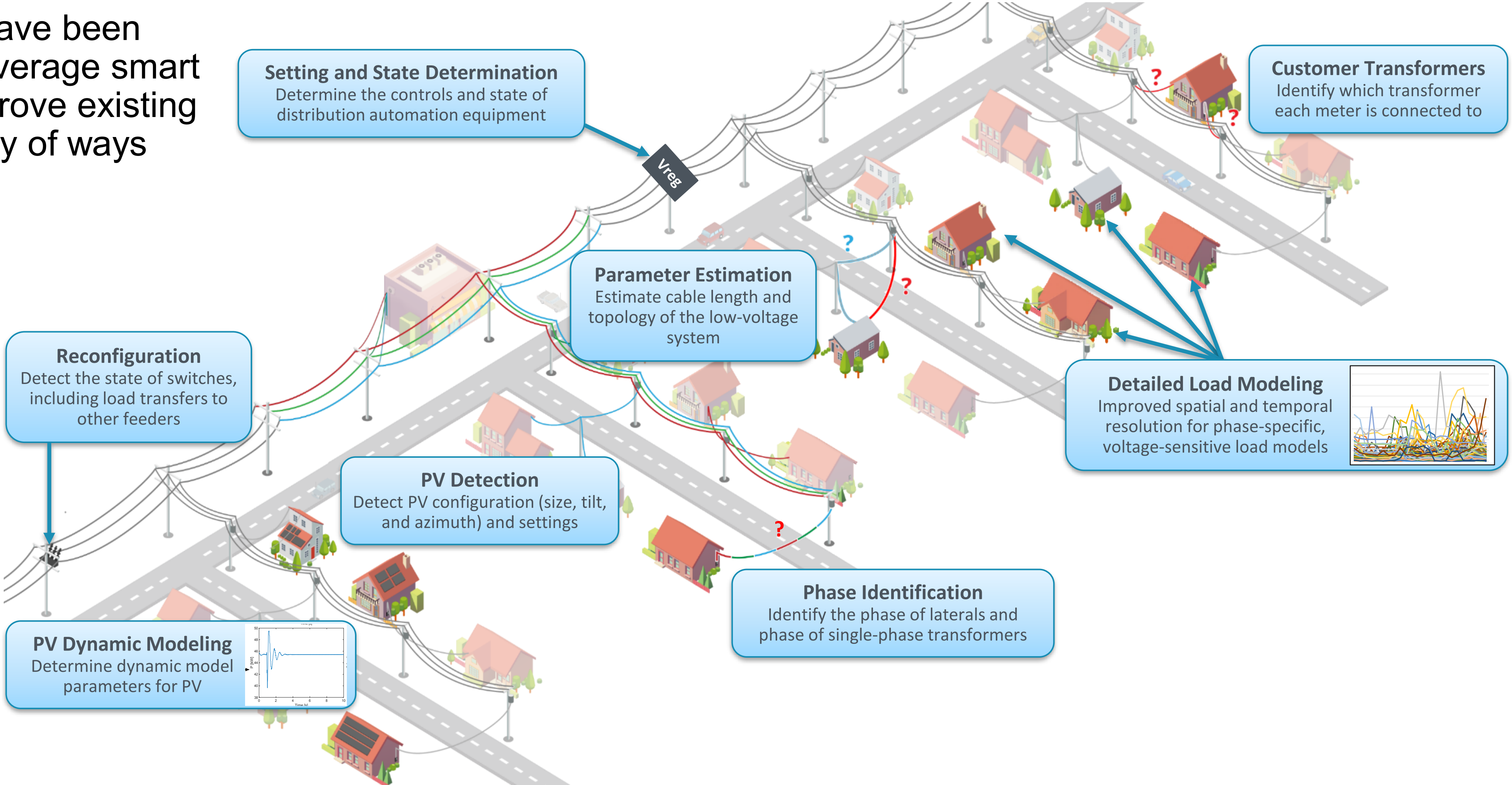
- Add PV, solve power flow
- Change PV, solve again



Improving Distribution Models with ML

Recent additions of Advanced Metering Infrastructure (AMI), or smart meters, provide measurements of each customer's power consumption, and possibly other quantities, such as voltage and reactive power

- ML approaches have been developed that leverage smart meter data to improve existing models in a variety of ways



Setting and State Determination
Determine the controls and state of distribution automation equipment

Customer Transformers
Identify which transformer each meter is connected to

Parameter Estimation
Estimate cable length and topology of the low-voltage system

Reconfiguration
Detect the state of switches, including load transfers to other feeders

Detailed Load Modeling
Improved spatial and temporal resolution for phase-specific, voltage-sensitive load models

PV Detection
Detect PV configuration (size, tilt, and azimuth) and settings

Phase Identification
Identify the phase of laterals and phase of single-phase transformers

PV Dynamic Modeling
Determine dynamic model parameters for PV

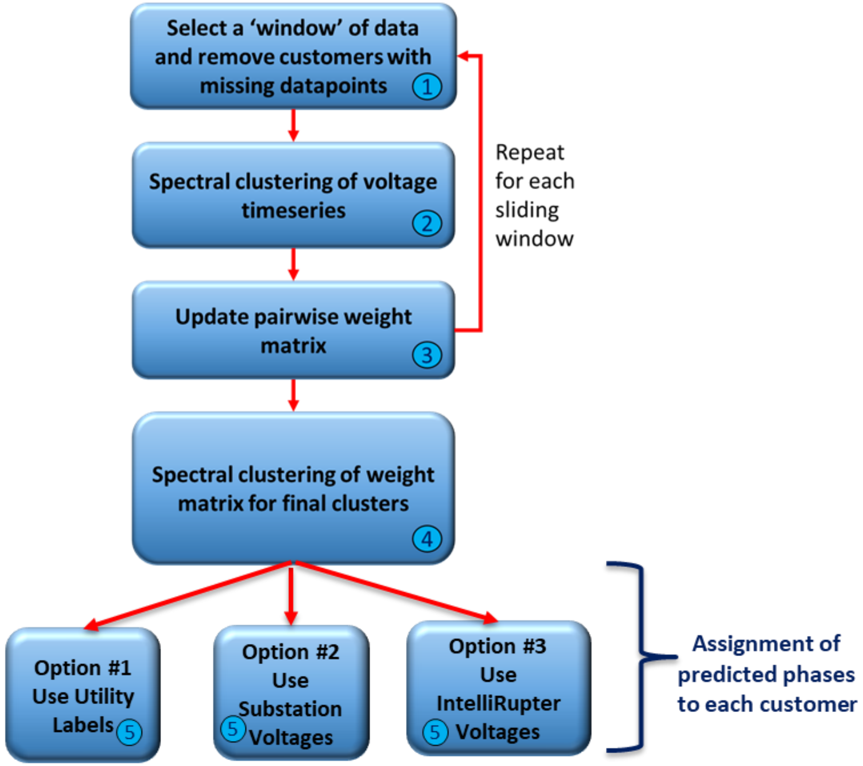
Phase Identification

Motivation: Distribution engineers must maintain customer voltages within a certain range and maintain a certain degree of load balance across the phases

Algorithms developed based on voltage correlation between customers on the same phase. For example…

Spectral Clustering Ensemble

(only AMI voltage)



Select a 'window' of data and remove customers with missing datapoints ①

Spectral clustering of voltage timeseries ②

Update pairwise weight matrix ③

Repeat for each sliding window

Spectral clustering of weight matrix for final clusters ④

Option #1 Use Utility Labels ⑤

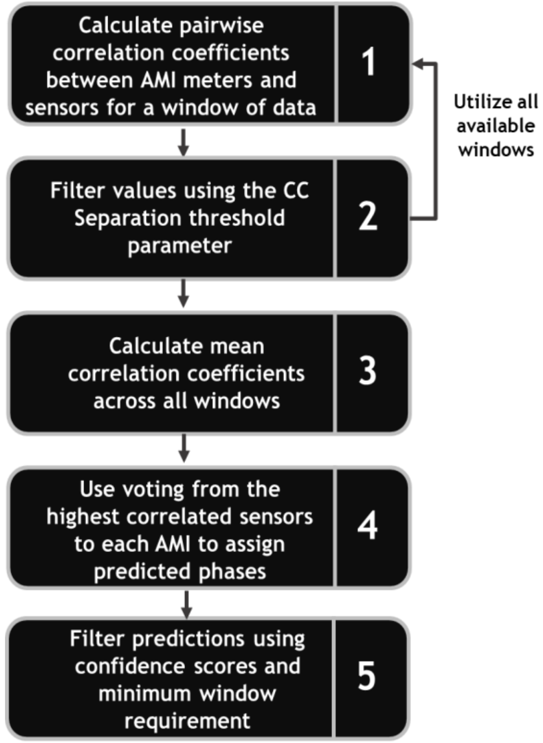
Option #2 Use Substation Voltages ⑤

Option #3 Use IntelliRupter Voltages ⑤

Assignment of predicted phases to each customer

Sensor-Based Method

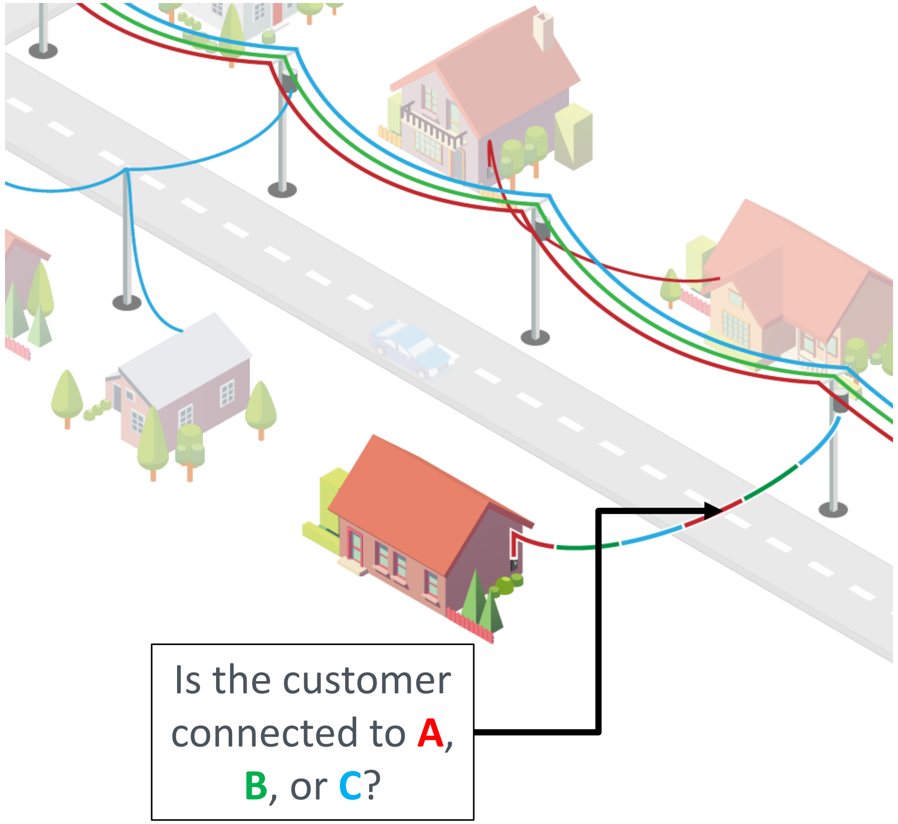
(AMI voltage and voltage from additional Sensors)



1. Calculate pairwise correlation coefficients between AMI meters and sensors for a window of data
2. Filter values using the CC Separation threshold parameter

Utilize all available windows

3. Calculate mean correlation coefficients across all windows
4. Use voting from the highest correlated sensors to each AMI to assign predicted phases
5. Filter predictions using confidence scores and minimum window requirement



Is the customer connected to A, B, or C?

Phase Identification

Results:

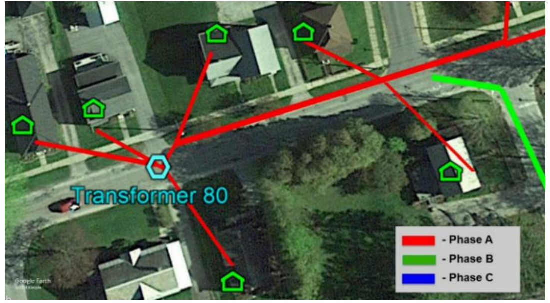
Algorithms were tested on both synthetic data and actual utility data

- Spectral clustering ensemble achieved 100% accuracy in correctly identifying customer phases in the presence of high levels of measurement noise and mislabeled customers on synthetic data
- Predictions on the actual utility datasets have also been 100% accurate thus far through field-verification of suspected mislabeled customers

Other comments:

- The algorithms can be integrated into existing AMI infrastructure and applied continuously
- This approach is beneficial to detect phase change events before problems arise

Original Utility Phase Labels



Predicted Phase Labels
Verified in Google Street View

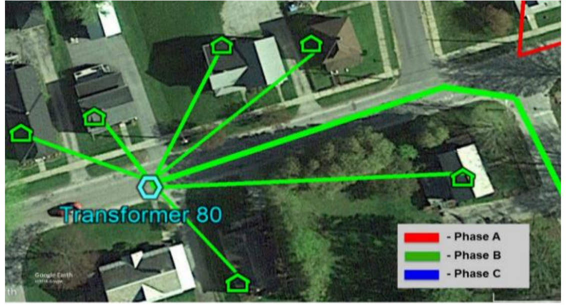
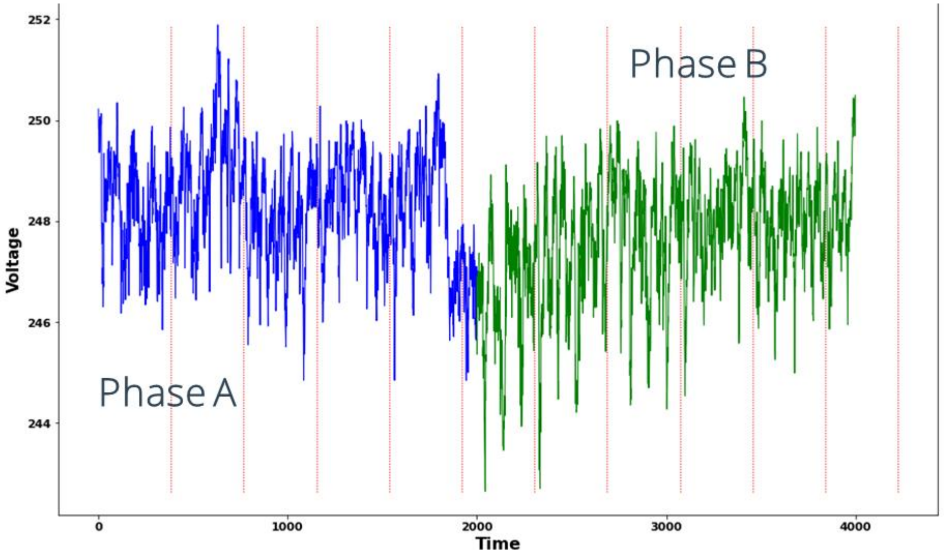


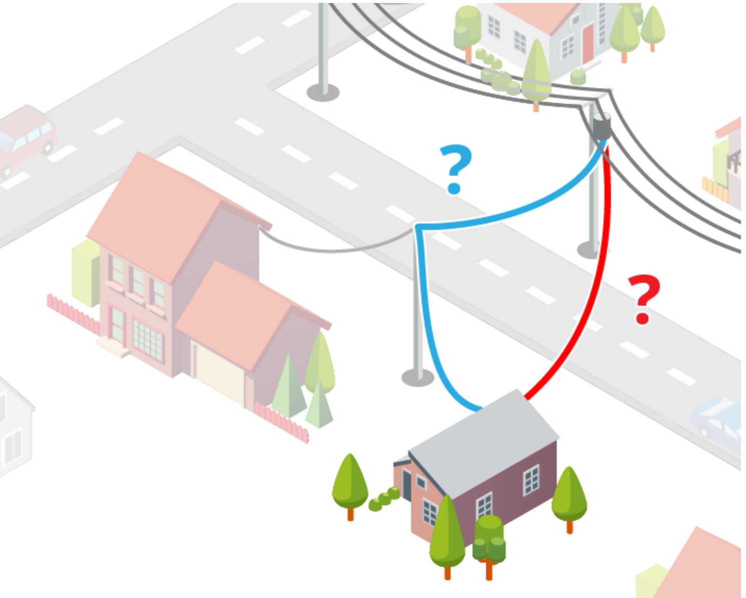
Illustration of a phase change event in AMI data



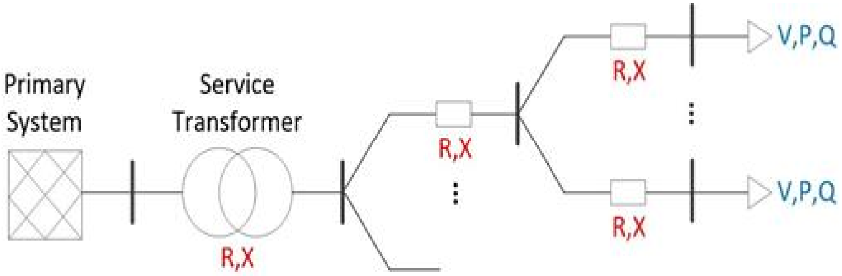
Secondary System Parameter and Topology Estimation

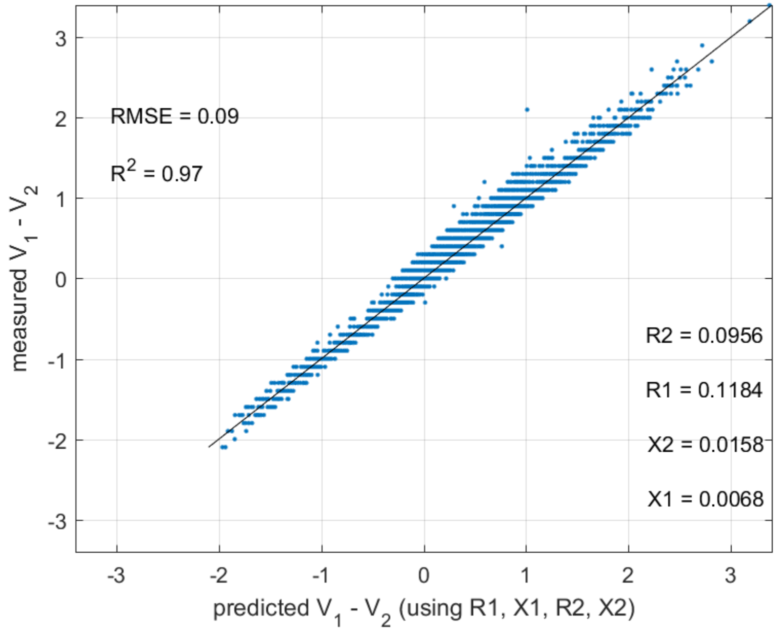
Motivation: Secondary circuits are typically not modeled or modeled with limited detail. Manual inspections require considerable man hours and may be hard to perform in urban areas with wiring underground

Objective: Use customer AMI voltage and power measurements to resolve secondary system parameters and topology



Parameter estimation is formulated as a pairwise regression problem





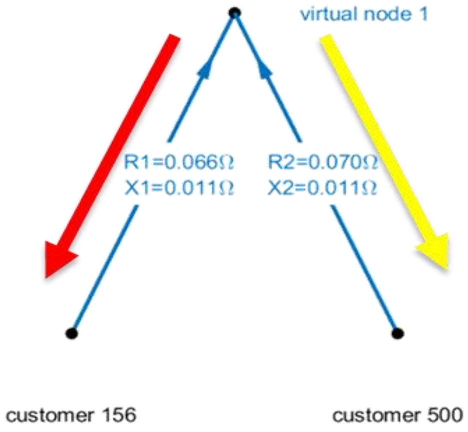
For all customers on a transformer find R1, R2, X1, X2
Which pair best fits the V1-V2 fluctuations?

Secondary System Parameter and Topology Estimation

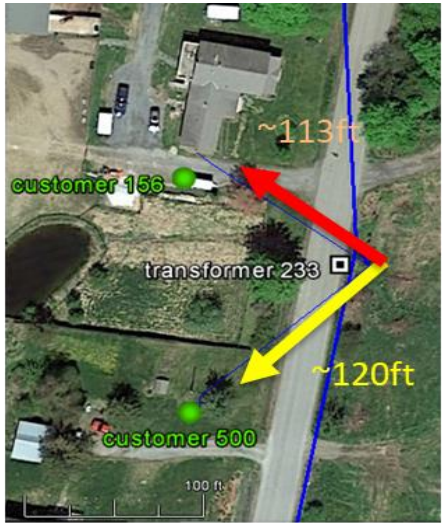
Results:

Extensively tested on synthetic data with data quality issues injected

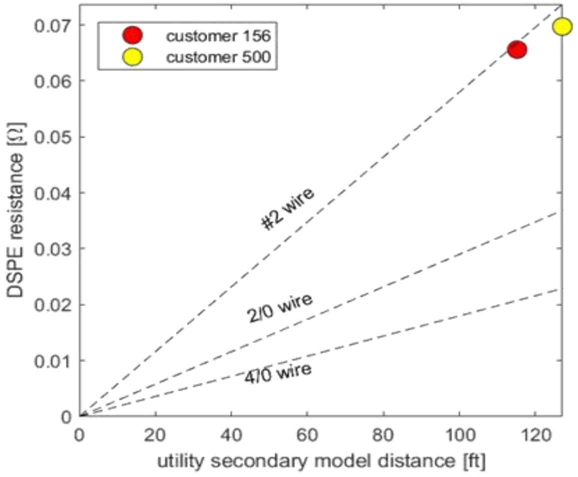
- Robust to meter bias, missing data, unknown power factor, and measurement interval
- Average vs Instantaneous measurement type does not affect this algorithm



(a)

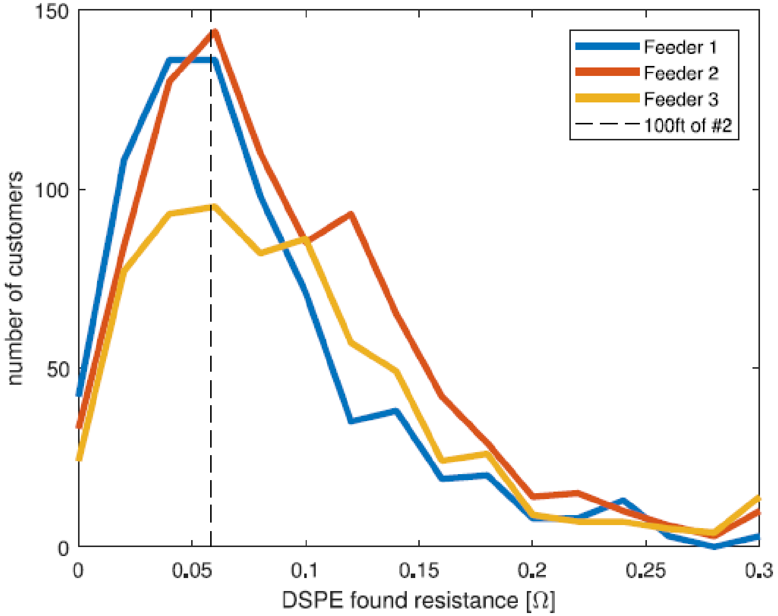


(b)

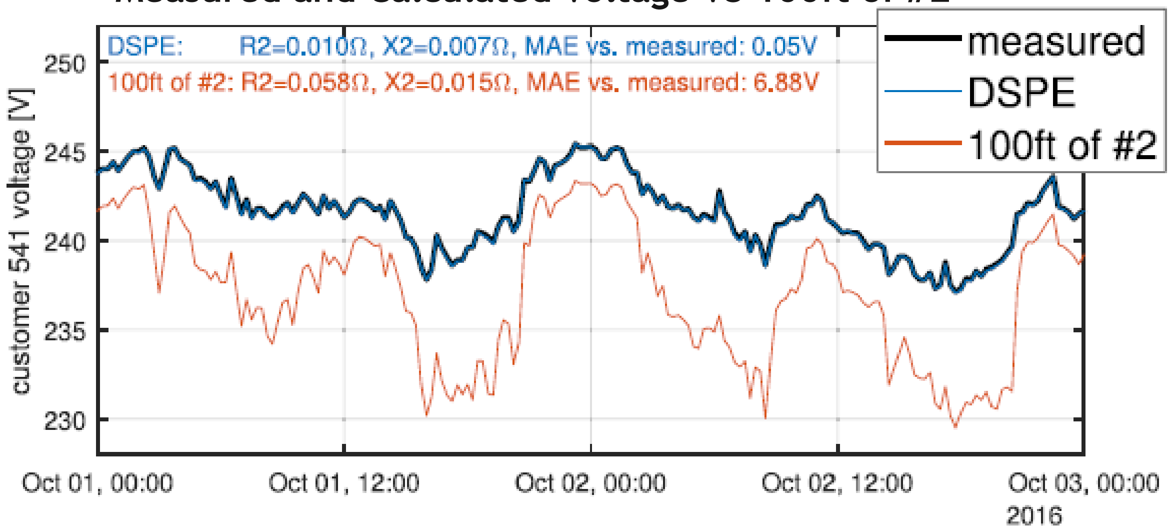


(c)

Distribution of Customers Resistances

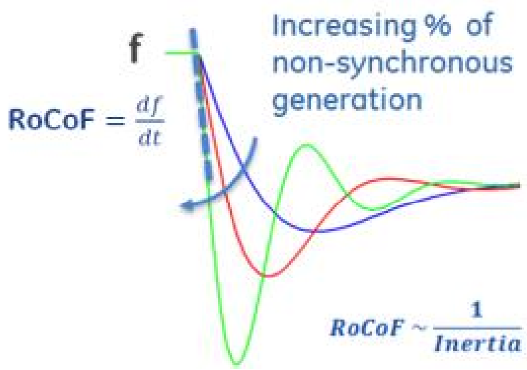


Measured and Calculated Voltage VS 100ft of #2



Machine Learning Based Fault Detection And Location In Electric Power Distribution Systems

Power System Protection



The protection system is designed to maintain safe operation and reliable service

- Rapidly remove the fault and minimize the disconnection of customers
- Relays measure voltage and current flowing through the line. Conventional protection uses logic pre-determined by a protection engineer to flag anomalous events (current is too high) for detecting faults and opening a breaker to isolate the part of the system with the fault
- Relays are coordinated and provide backup using time delays depending on the location of the fault.

Lost Communication from Ice Storm

Is it possible to detect faults fast enough to prevent arcing that starts wildfires?

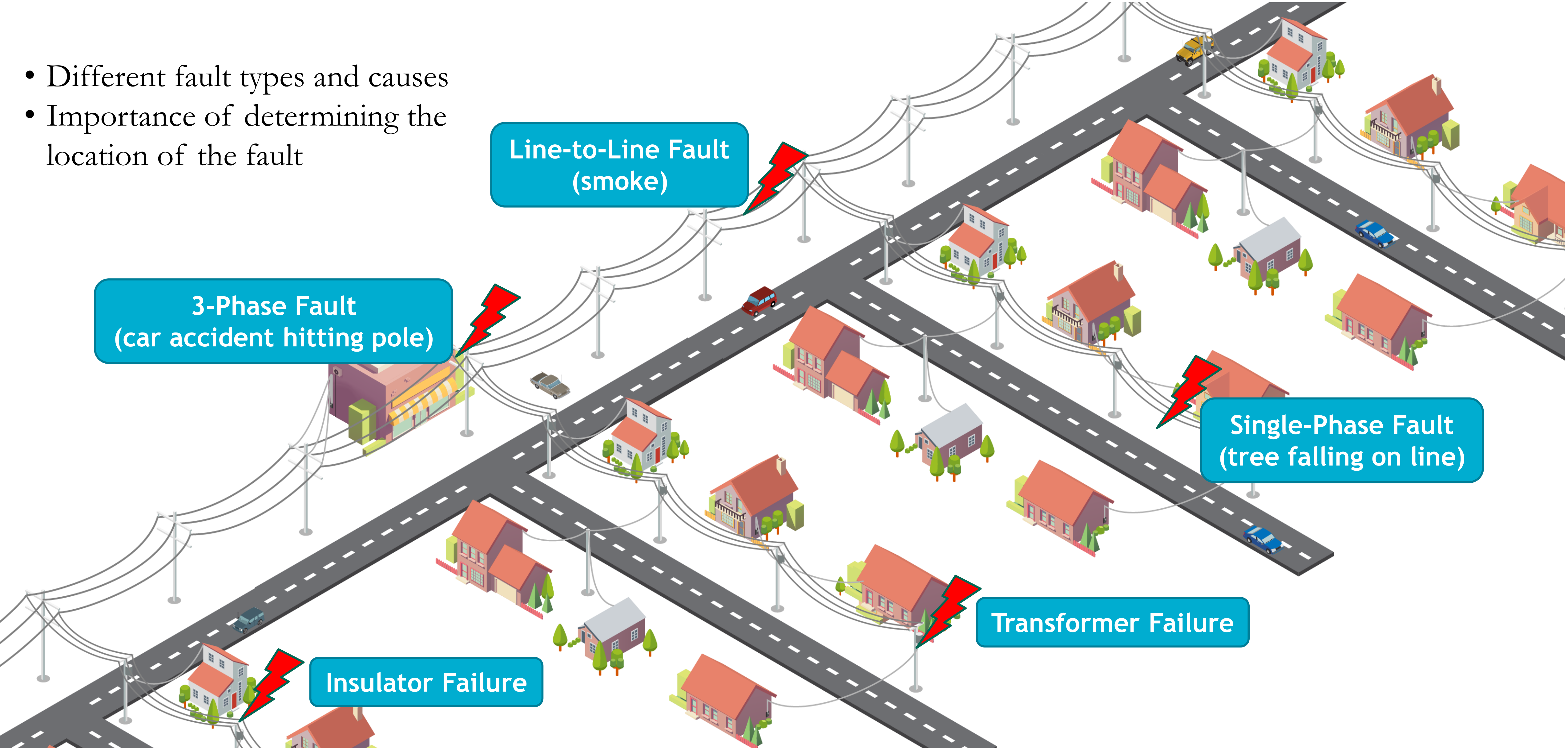
Power System Protection is getting more critical and more complicated:

- Inverter-based generation does not have the same fault characteristics
- Protection challenges in downtown networked systems
- Electric faults causing wildfires in California
- Fast-tripping protection schemes limited by communication networks
- Cyberattacks (e.g., 2016 Ukraine) on relays could cause damage or cascading outages

12/26/18 Transformer Fire in NYC from Failed Protection with Loss of Communication

Distribution System Faults

- Different fault types and causes
- Importance of determining the location of the fault



Line-to-Line Fault (smoke)

3-Phase Fault (car accident hitting pole)

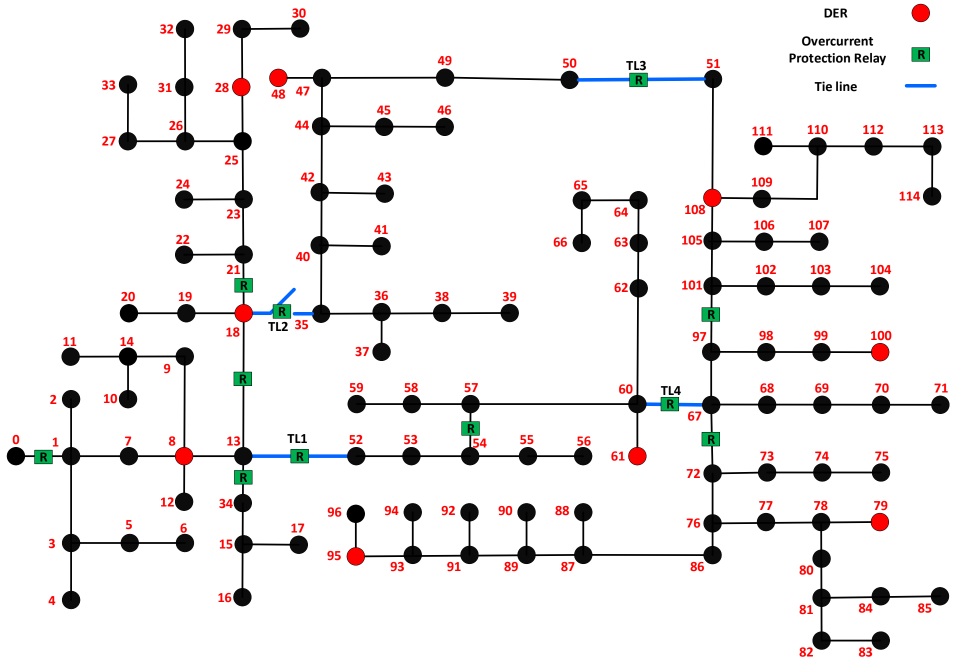
Single-Phase Fault (tree falling on line)

Transformer Failure

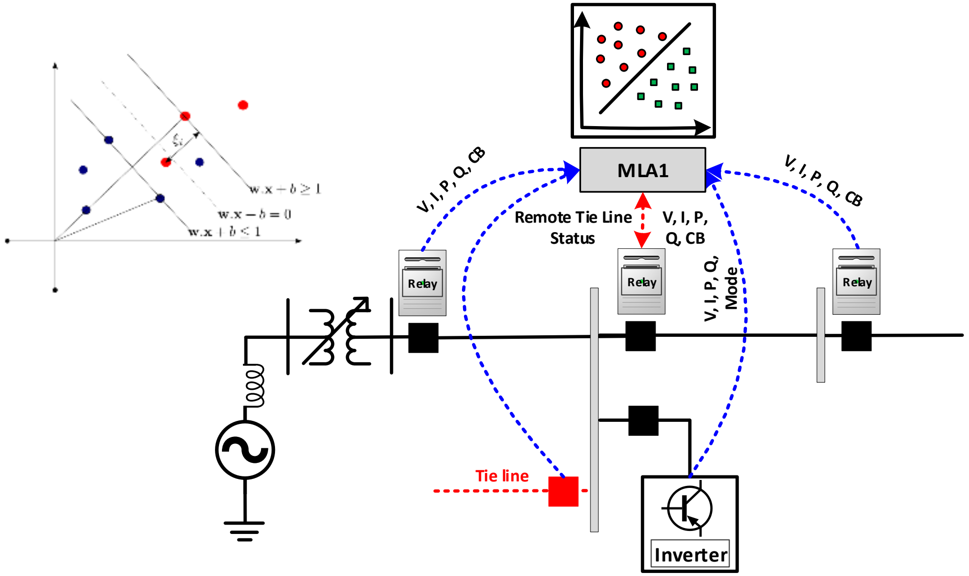
Insulator Failure

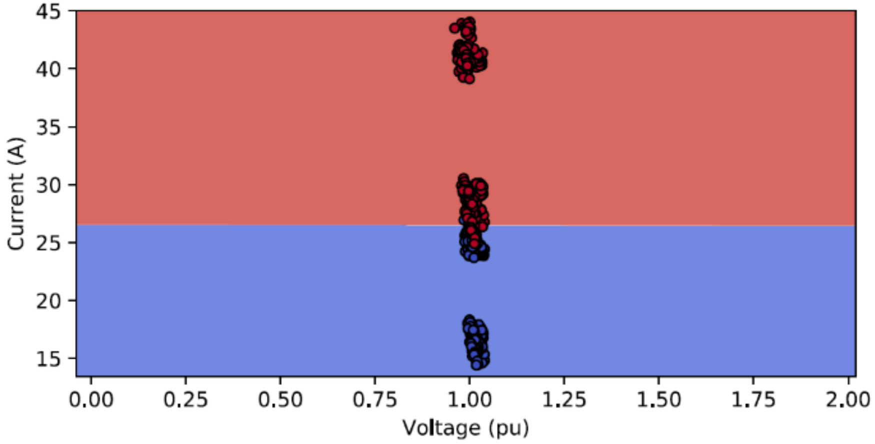
Machine Learning for Power System Operations

- During resilience events, there can be uncertainty in the status of the grid (breaker open/closed, microgrid connected or islanded) with possible loss of communication and many events happening at the same time
- Machine learning algorithms can learn correlations between local measurements and the configuration of the grid. This ensures reliable communication-free operations during resilience events
- Using Support Vector Machine (SVM) to learn and classify each switch in the network as open or closed



Tie Line / Breaker	Accuracy
TL1	98.2%
TL2	98.0%
CB2	99.5%
CB3	96.9%





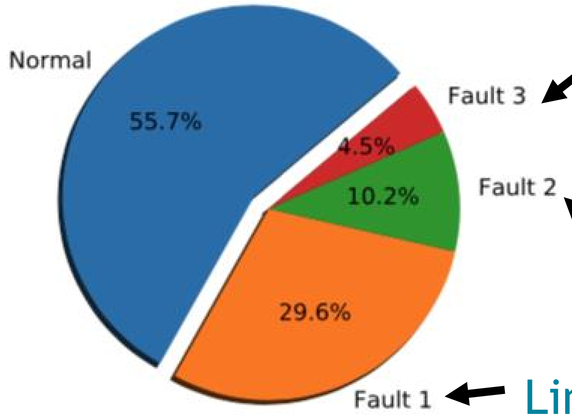
Machine Learning for Fault Detection and Location

Machine Learning (ML) Fault Analysis

- Using ML for power system protection instead of relays
- Test approach on IEEE 123 Model (Matlab Simulink)
- Simulate 3 fault types at 19 locations with varying resistances at different times of year

Intelligent Decision Making with Support Vector Machines

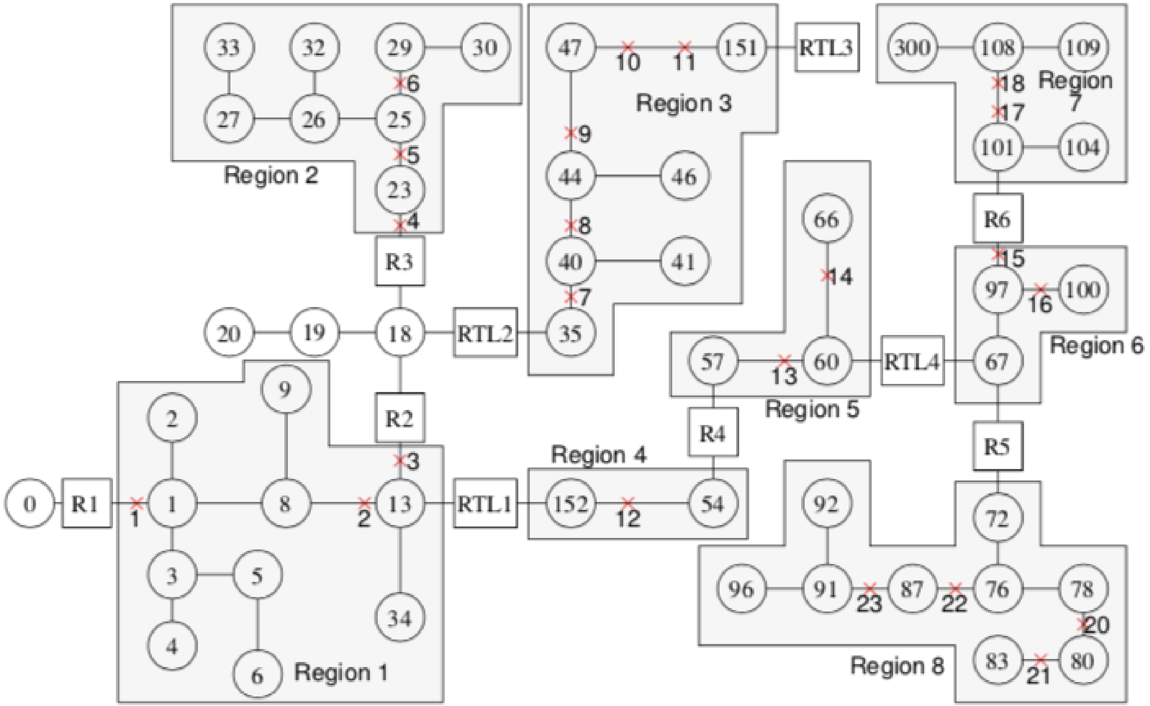
- Uses Sequence Current (I_0,I_1,I_2) and Voltages (V_0,V_1,V_2) as input features to SVM
- System specific learning that adapts
- ML at each breaker can distinguish faults inside its protective zone/region



Three Phase to Ground

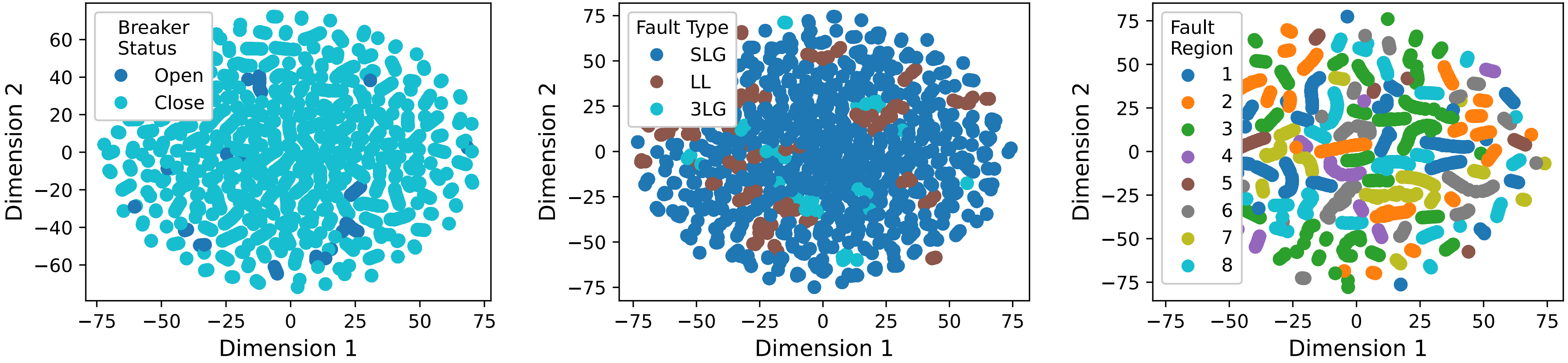
Line to Line

Line to Ground



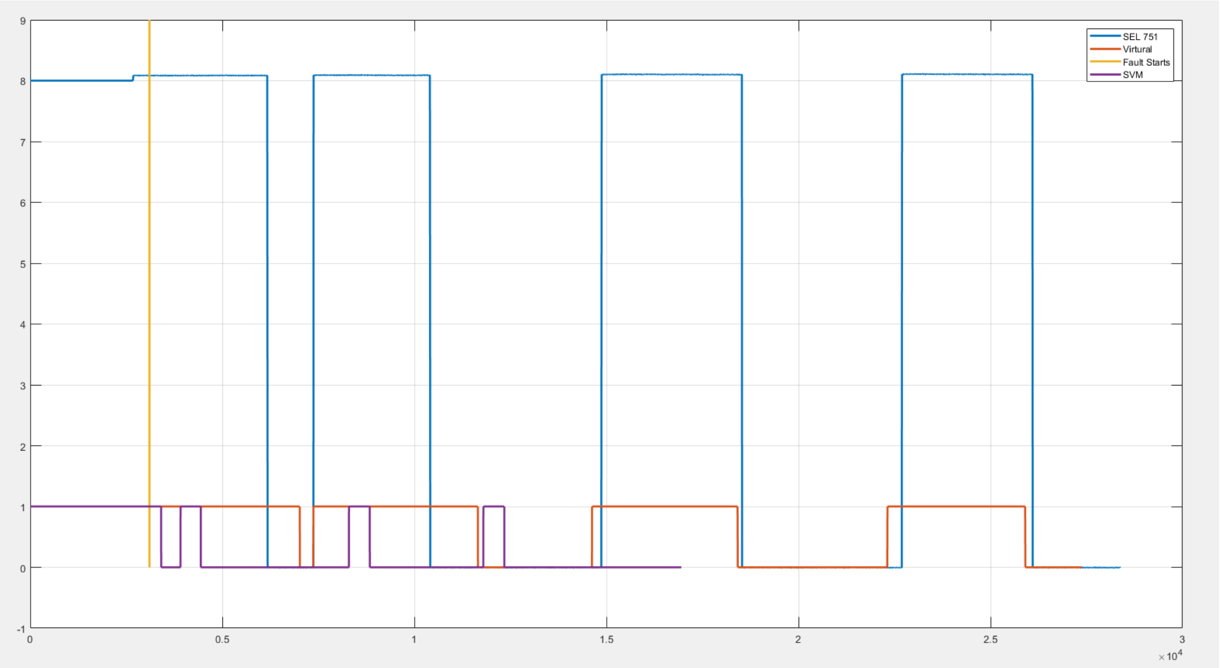
Machine Learning Embedded in Relays

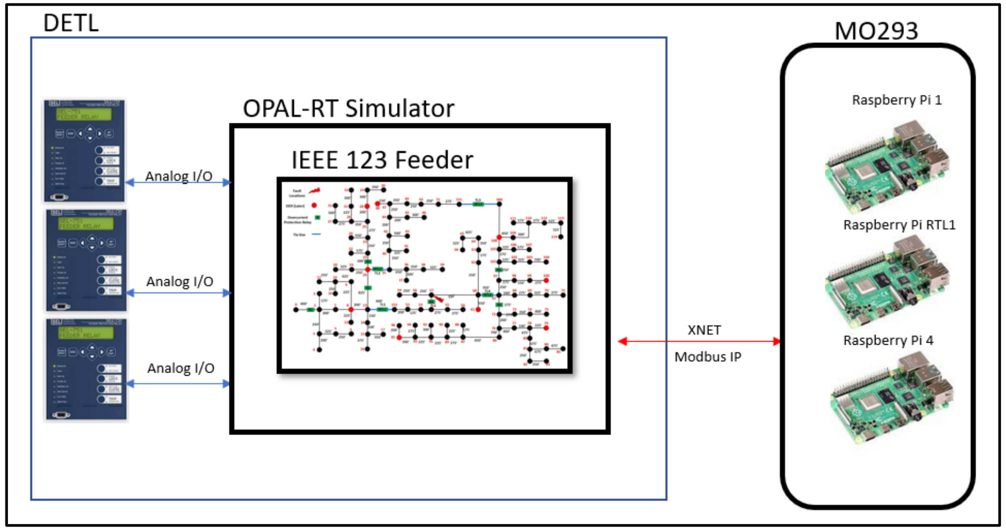
- SVM classified if there is a fault, fault type, and fault location. Tested on IEEE 123-node with 9 relays.
- No false trips in yearlong testing with varying and dynamic loads
- Correctly detects all fault events at different buses, resistances, etc.
- 100% accuracy for classifying the type of fault (SLG, LL, 3LG)
- Correct coordination (which relay trips first) for 99.6% of faults



Hardware In the Loop Testing

- The trained machine learning algorithm is placed into a Raspberry Pi for testing
- Using an Opal-RT real-time hardware-in-the-loop simulator, the real-time voltage and current signals from the system are fed into the analog inputs of the Raspberry Pi
- SEL-751 hardware relays are also included with a standard time overcurrent curve for comparison
- The SVM detects the faults in 0.17 seconds in comparison to the time overcurrrent of 1.94 seconds (11.4x faster)





Field Testing of Machine Learning for Fault Detection

Training must be done using simulations because there are only a few faults per year on a section of a distribution feeder, but there can be discrepancies between how the electric power system may operate in the real world vs. simulation training data

- Incorrect parameters/topology in the model, changes in the system
- Measurement noise, missing data, and differences in the fault

We trained an SVM classifier to identify faults and their type using data from a simulation of an actual system

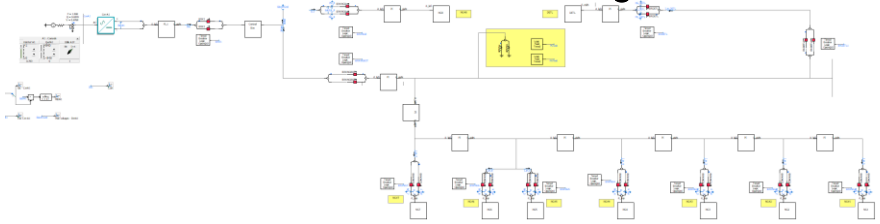
- The training data set (PSCAD simulation data) has a total of 160,000 points i.e. 40,000 points for each fault event

We then worked with Emera Technologies to install sensing and apply faults throughout their microgrid on Kirtland Air Force Base

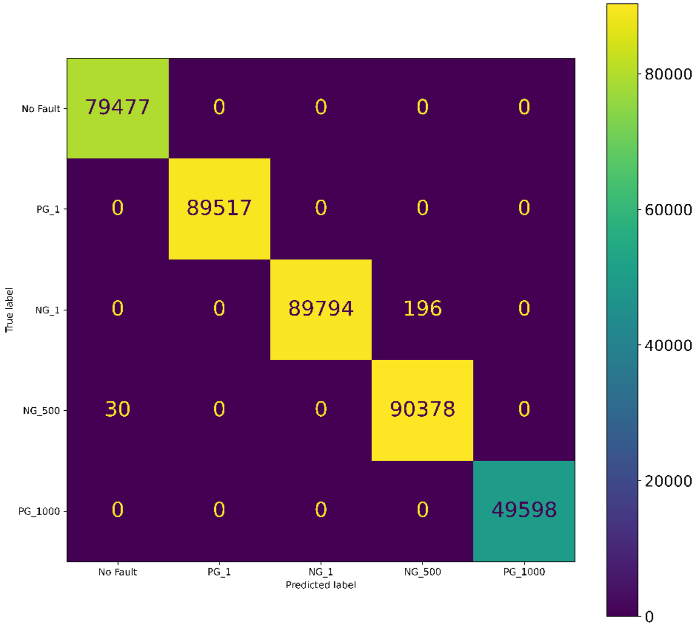
- The test data (captured from the actual system) has 400,000 data points, i.e. 100,000 data points for each fault event

The SVM multi-class algorithm embedded in the relay classified the faults with an accuracy of 99.943%.

PSCAD Model of the Microgrid



Actual Microgrid System



True label \ Predicted label	No Fault	PG_1	NG_1	NG_500	PG_1000
No Fault	79477	0	0	0	0
PG_1	0	89517	0	0	0
NG_1	0	0	89794	196	0
NG_500	30	0	0	90378	0
PG_1000	0	0	0	0	49598

Conclusions

- New sensors and machine learning provide more granular understanding and visibility into distribution systems
- Historical data combined with physics-based machine learning can detect model errors and calibrate system parameters
- Electric distribution system protective relays equipped with machine learning (ML) algorithms can improve power system reliability and resilience by performing an automated and self-learning monitoring and decision making analysis.
- ML algorithms can be trained offline using fault simulations of the system and location they are going to be installed. The trained algorithm is then embedded inside each relay to provide decision making support based on the grid measurements. Advanced, data-driven relays provide new value to power system resilience.

Questions?

Matthew Reno
mjreno@sandia.gov

Joseph Azzolini
jazzoli@sandia.gov

Sandia National Laboratories