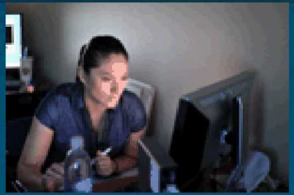
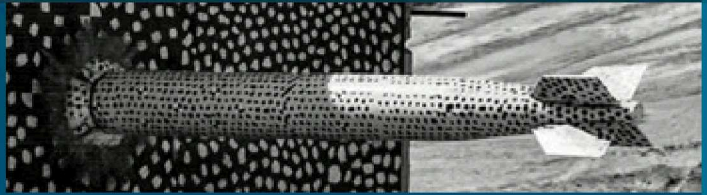
SAND2019-2168C

Spectral Clustering for Customer Phase Identification Using AMI Voltage Time Series





Logan Blakely
Sandia National Labs
Portland State University

Matthew J. Reno
Sandia National Labs

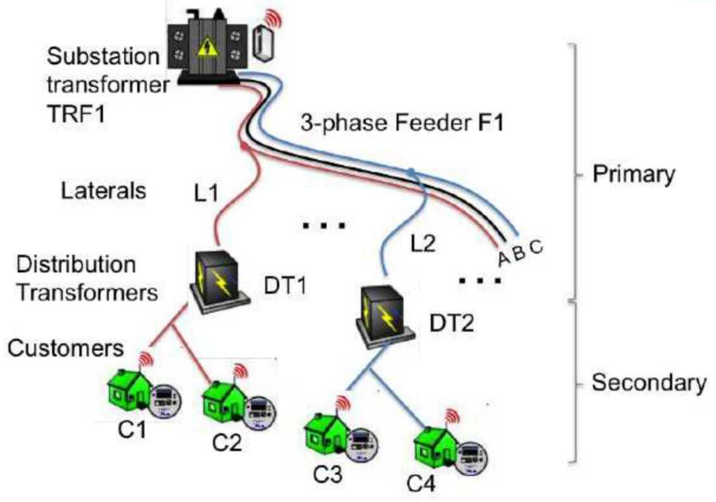
Wu-chi Feng
Portland State University

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Phase Identification

- **Motivation**: For distribution system planning and operations, especially with high penetrations of DER, accurate multi-phase distribution models are important, but utility models often have many errors
  - Customer transformer phase connections (top figure)
  - Single-phase laterals connected to different phases (bottom figure)
- **Problem**: Manually calibrating using PhaseTrakkers is time-consuming
- **Objective**: Use machine learning and big data from grid edge measurements to identify the phase of each customer
- **Hypothesis**: If customer voltage timeseries are correlated, they are likely to be on the same phase
- **Implementation**: Algorithm should only require AMI data
  - Only use AMI data, and should not require substation measurements, SCADA, or PMU data
  - Calibrate the distribution system model without assuming phasing, transformer connection, etc. are correct in the utility model

Customers Phase Connection



Feeder Laterals Colored by Phase

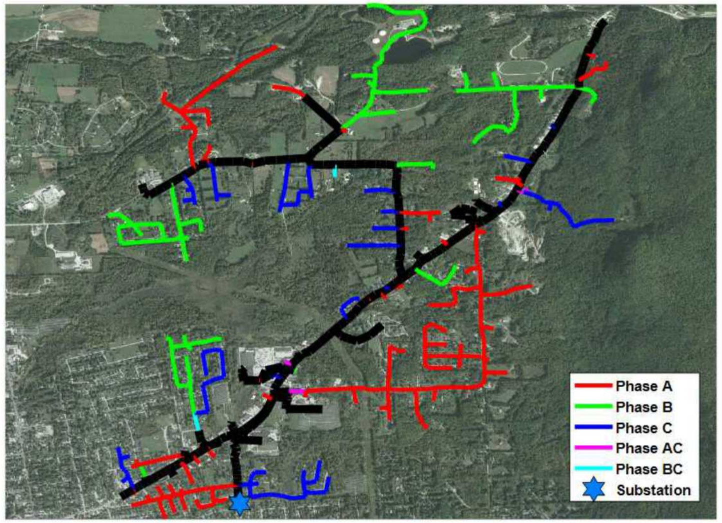


Figure Credit (top): R. Mitra *et al.*, "Voltage Correlations in Smart Meter Data," *ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1999–2008, 2015.

Spectral Clustering

- Unsupervised machine learning algorithm to group customers into a predefined number of clusters

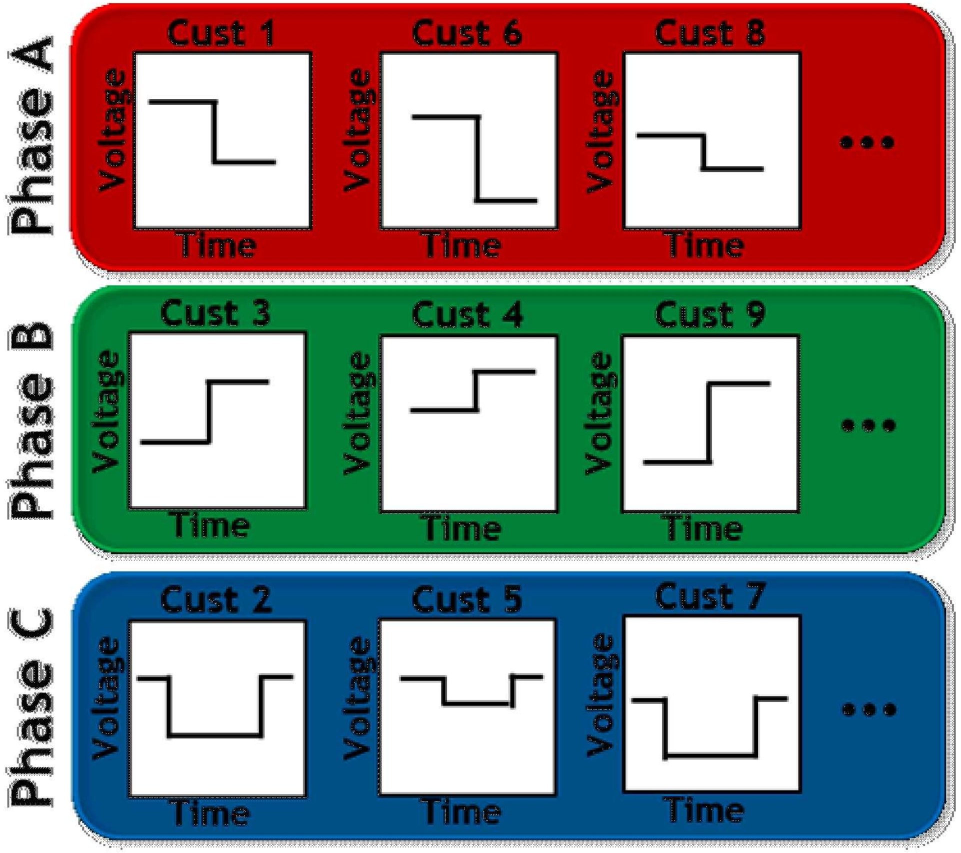
- Clustering timeseries of customers' voltage measurements based on their similarity

- Spectral clustering of each timeseries

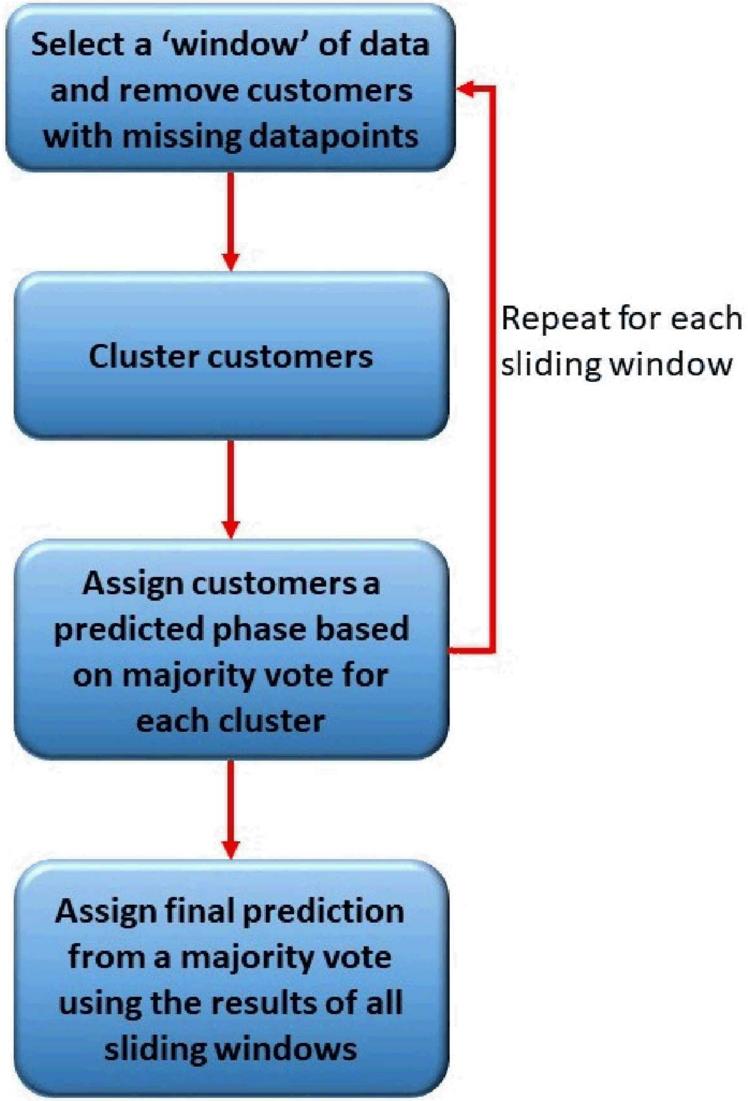
    - Uses affinity matrix to cluster similar inputs instead of distance from a centroid (basic K-Means clustering)

    - Eigenvalues of the Laplacian are used for clustering (nonlinear dimensionality reduction)

        - Scalable to datasets with high dimension
        - Transformation of raw voltage data



Phase Identification - Methodology

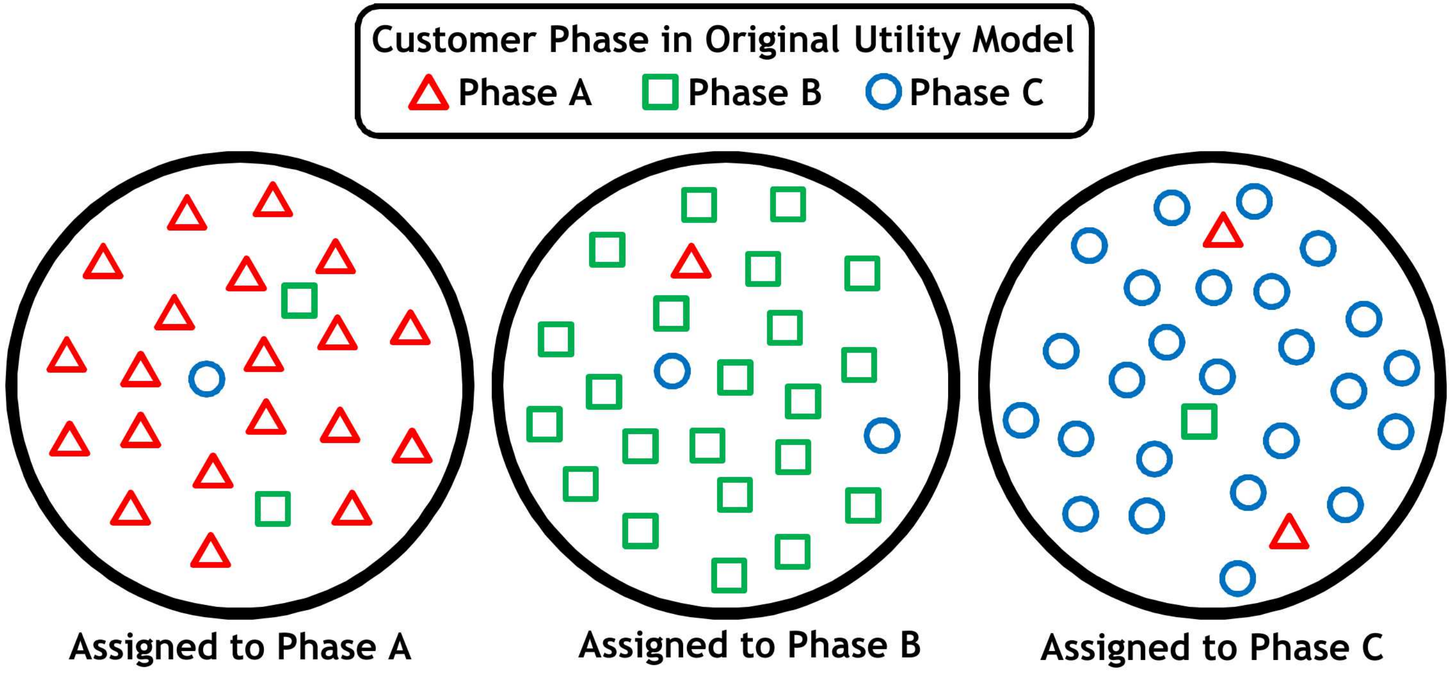


Overall Methodology

- The entire dataset is broken into timeseries windows (4-days)
- Spectral clustering of all customers in the window without missing data
- Assign predicted phase based on majority vote in each cluster
- Ensemble prediction determined by combining all windows

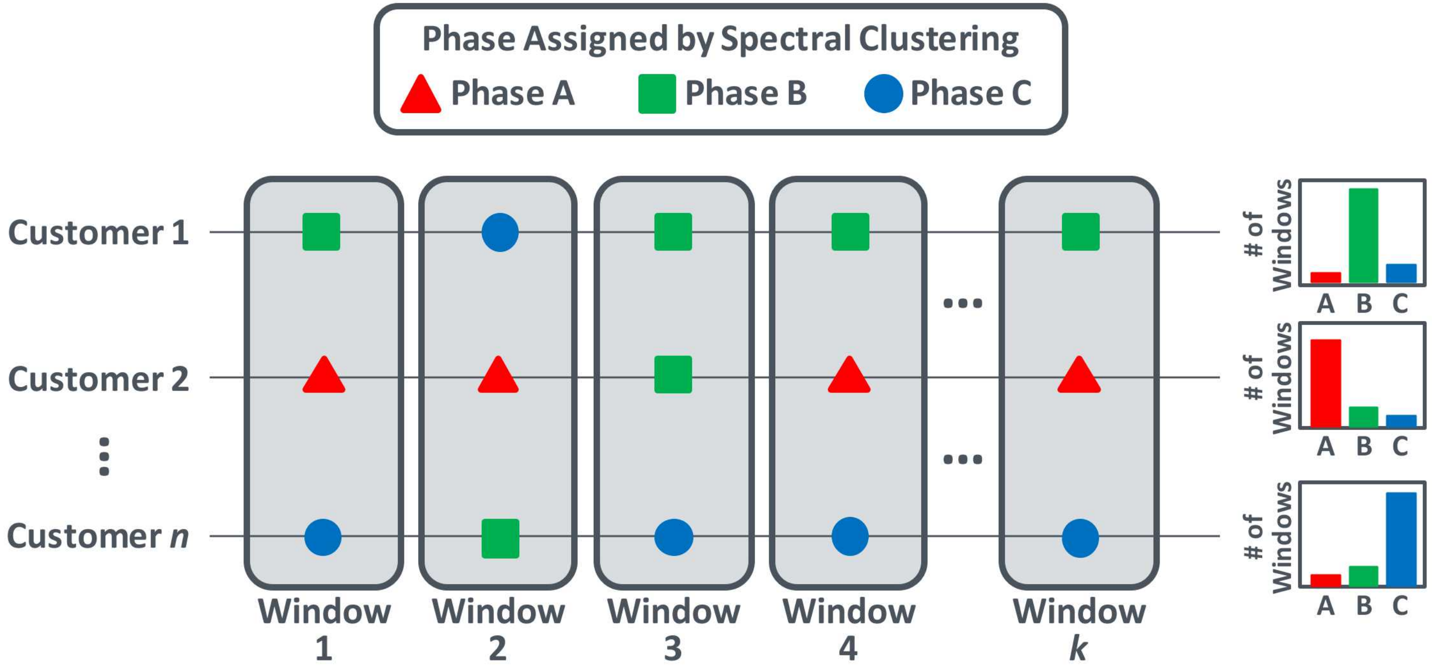
Determining the Phase of Each Cluster

- All customers are grouped into clusters, solely based on their voltage timeseries
- The phase of the cluster is determined by the majority of customers' phase labels in the original utility model
- Each customer is assigned a predicted phase based on the assigned cluster phase



Ensemble Machine Learning

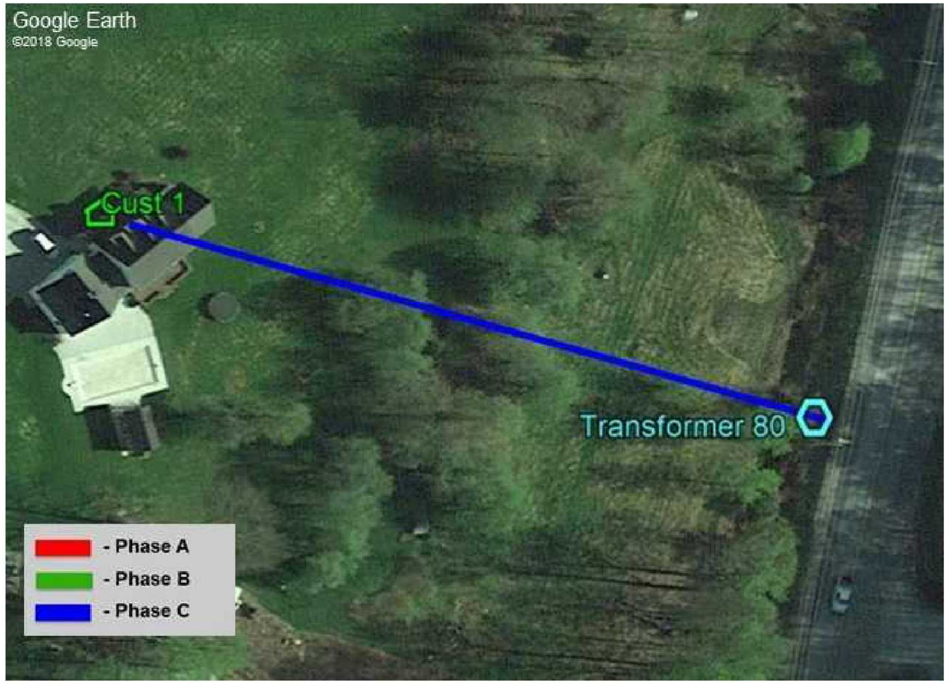
- Phase are calculated with voltage timeseries windows (for example 4 days), but often there is much more data available
- Large datasets can be processed by looking at many timeseries windows
- The ensemble prediction of each customer phase is determined using all available windows
- The confidence of the phase prediction for a customer is calculated using the percentage of windows in the ensemble that have the same phase prediction

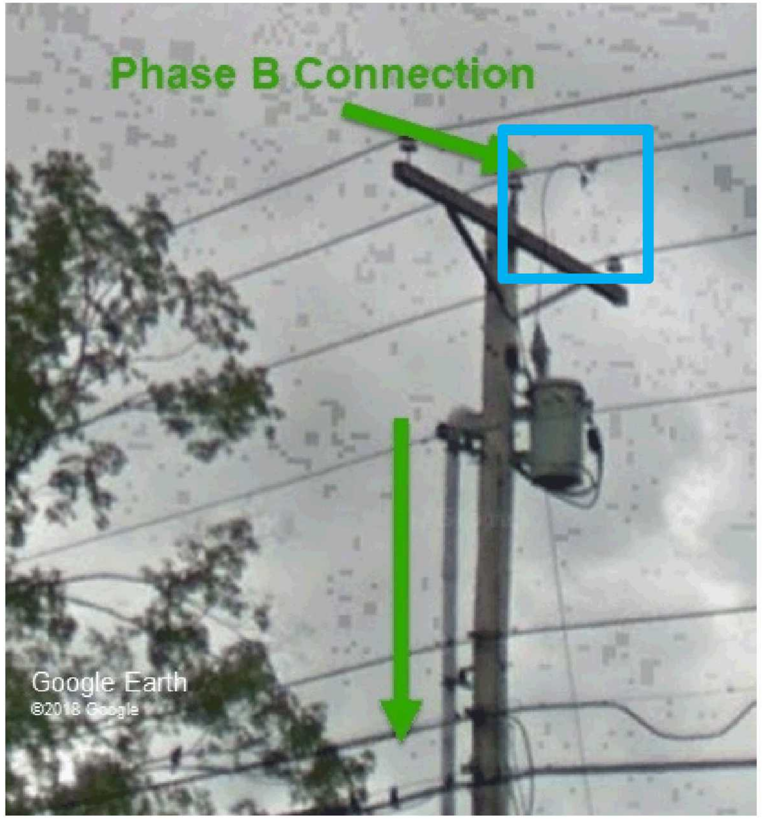


Phase Identification - Results

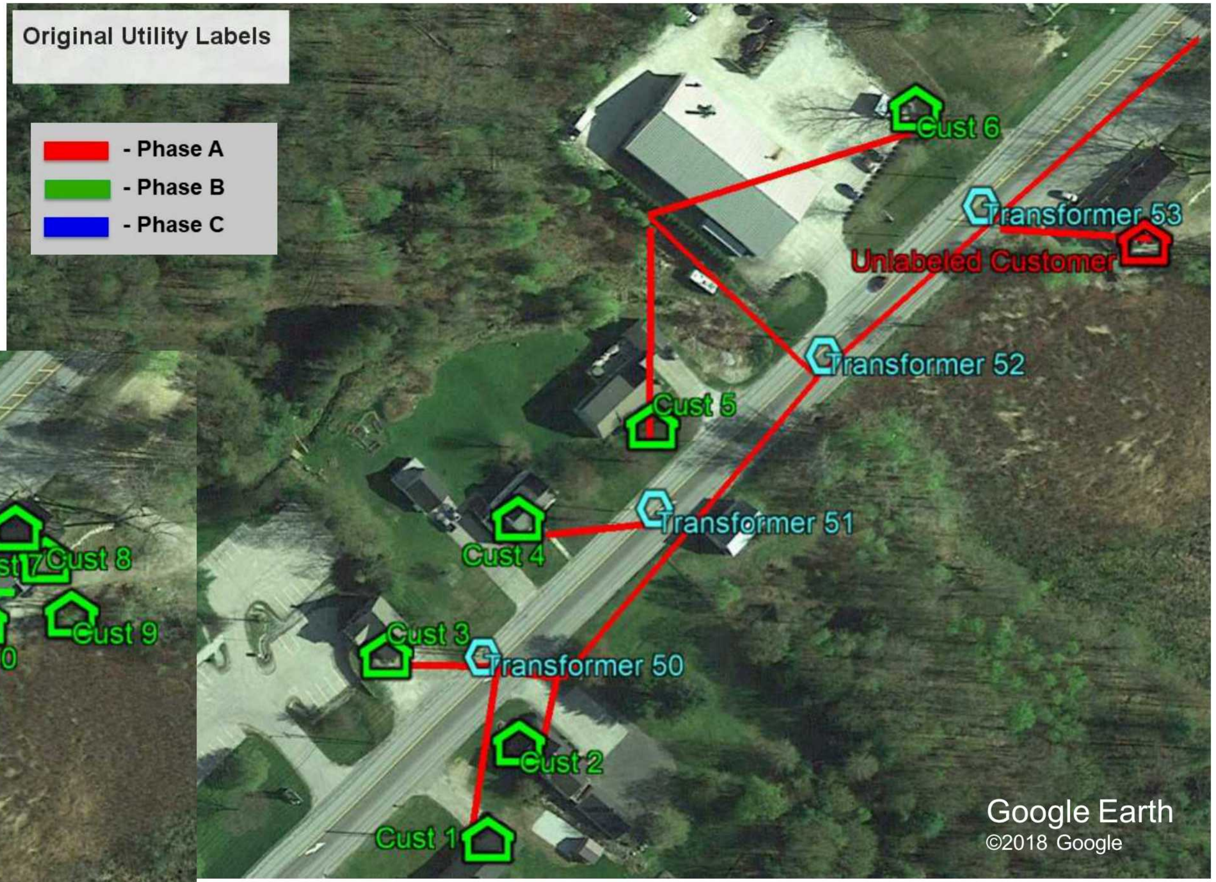
- Two-tier validation method
  - ‘Topology Validation’ – All customers connected to the same transformer predicted to be on the same phase. This is done for all customers on the feeder (1055 for Feeder 1)
  - Google Street View – Visually validate a subset of the total customers

- Example of an error in the utility phase labeling - Phase C (left figure)
- Predicted by the clustering as Phase B and verified in Google Street View (right figure)
- See the accompanying paper for a more detailed example validated with Google Street View

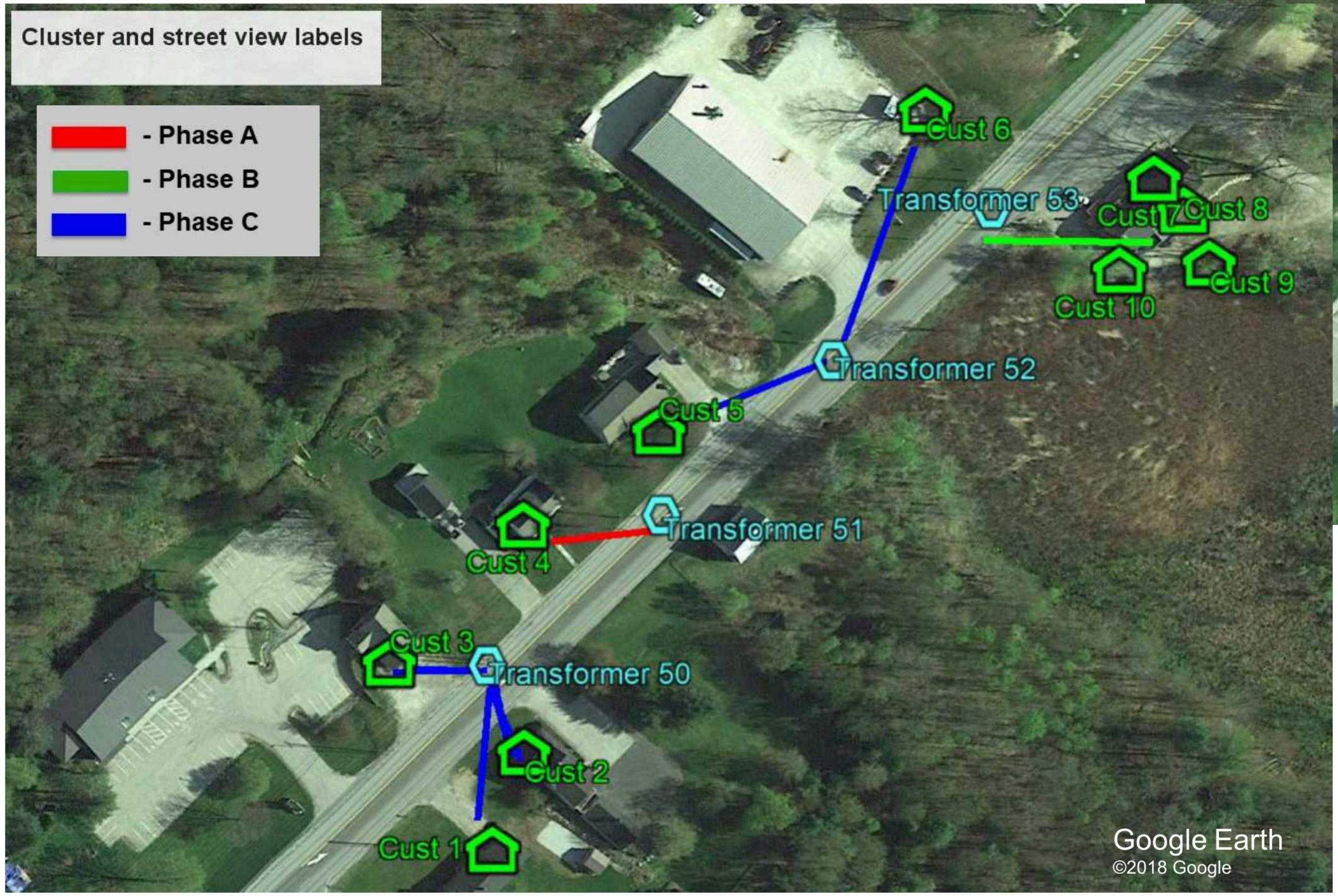




Phase Identification - Results



Predicted and Validated Labels



Original Utility Model Labels

Phase Identification – Feeder1Results

Feeder 1 Customers	Total Customers	Validated Utility Labels	Corrected Utility Labels	Remaining Customers
Customers	1055	957	92	6
Percentages	100%	~91%	~8.5%	~0.5%

99.5% of customers plausibly validated or corrected using this methodology

		Utility Label			
		A	**B**	**C**	**Total**
Clustering Predicted	A	506	24	8	538
	B	10	229	3	242
	C	48	5	222	275
	Total	564	258	233	

Phase Identification – Feeder Results Comparison

Feeders (12.47kV)	Total Customers	Percentage Predicted Errors	Peak Load (MW)	Feeder Length (KM)	Line Regulators	Capacitors
Feeder 1	1155	~9%	2.0	5.4	0	0
Feeder 2	1309	~16.2%	1.8	2.5	1 set	450 kVAr
Feeder 3	1188	~18%	1.4	3.2	0	300 kVar 300 kVar

- Results are similar when tested on two other nearby feeders – larger percentage of predicted errors compared to Feeder 1
- Feeder 2 and Feeder 3 are more complex than Feeder 1
- The utility notes that they have spent more time correcting Feeder 1 than Feeders 2 or 3

Conclusion

- The Spectral Clustering methodology with the sliding window ensemble successfully performed phase identification. Results were validated with Topology Validation and street view examples.
  - The results indicate that ~9% of the utility model for Feeder 1 contained errors.

- 99.5% of customers' phase labels in the utility model for Feeder 1 were either validated (~91%) or corrected (~8.5%).

- This methodology shows excellent promise for phase identification
  - Using only AMI data (no SCADA substation data)
  - Does not rely on accurate topology labeling by the utility

Future Work:

- Validate with field-verified labels

- Research is in progress to further test and validate this methodology using a synthetic dataset

- Use those findings to provide guidelines for AMI data collection methods

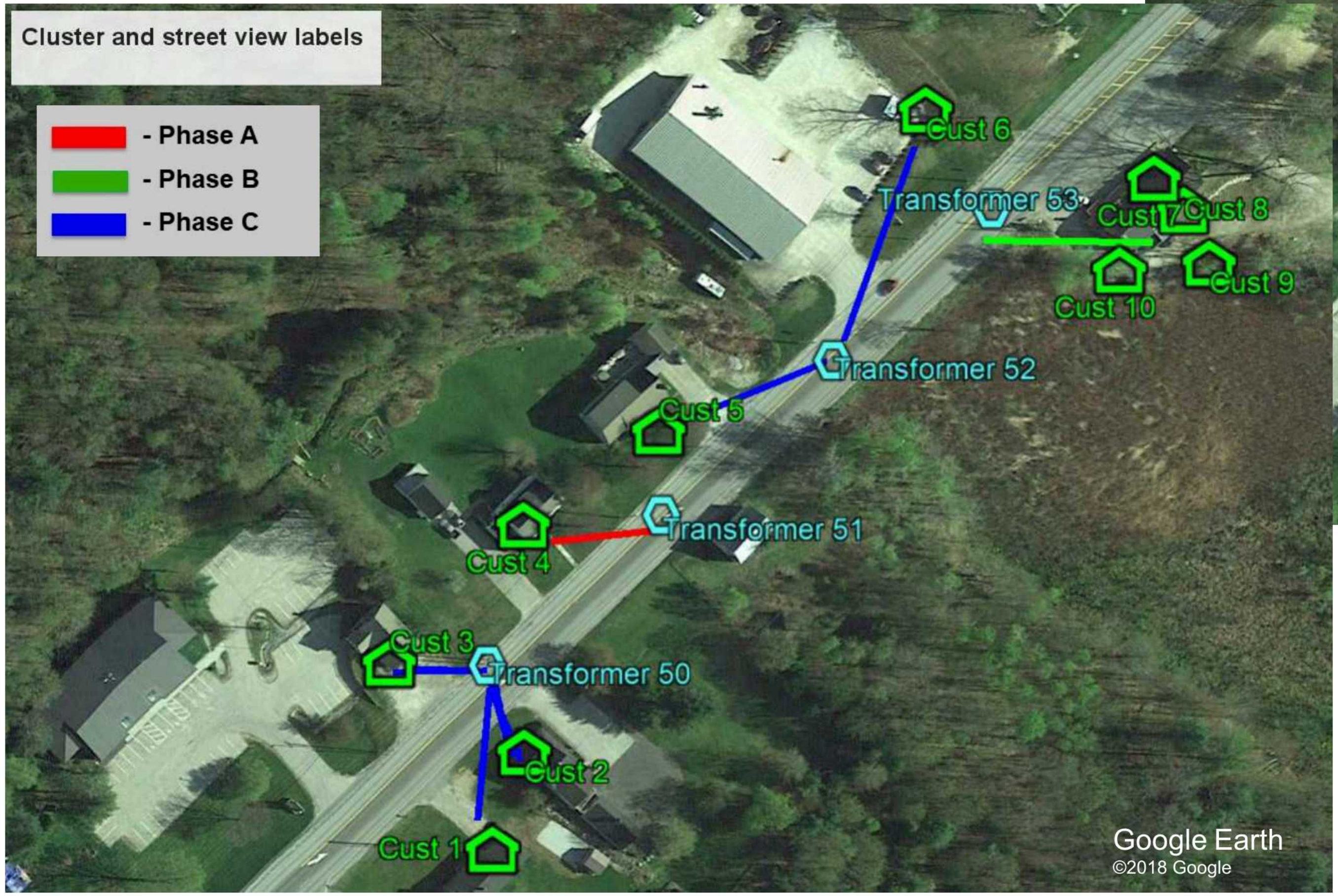
Questions?

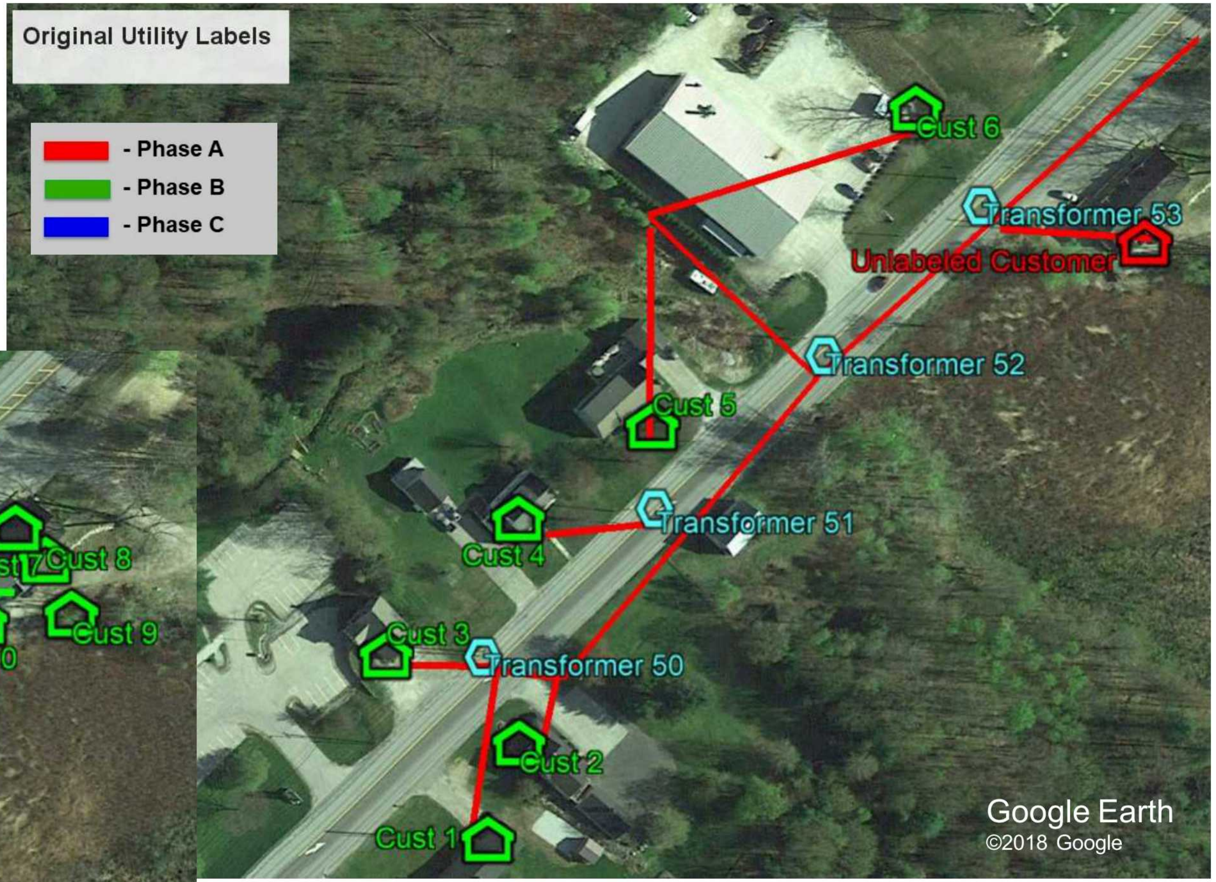
Logan Blakely

lblakel@sandia.gov

Supplemental Slides

Predicted and Validated Labels





Original Utility Model Labels

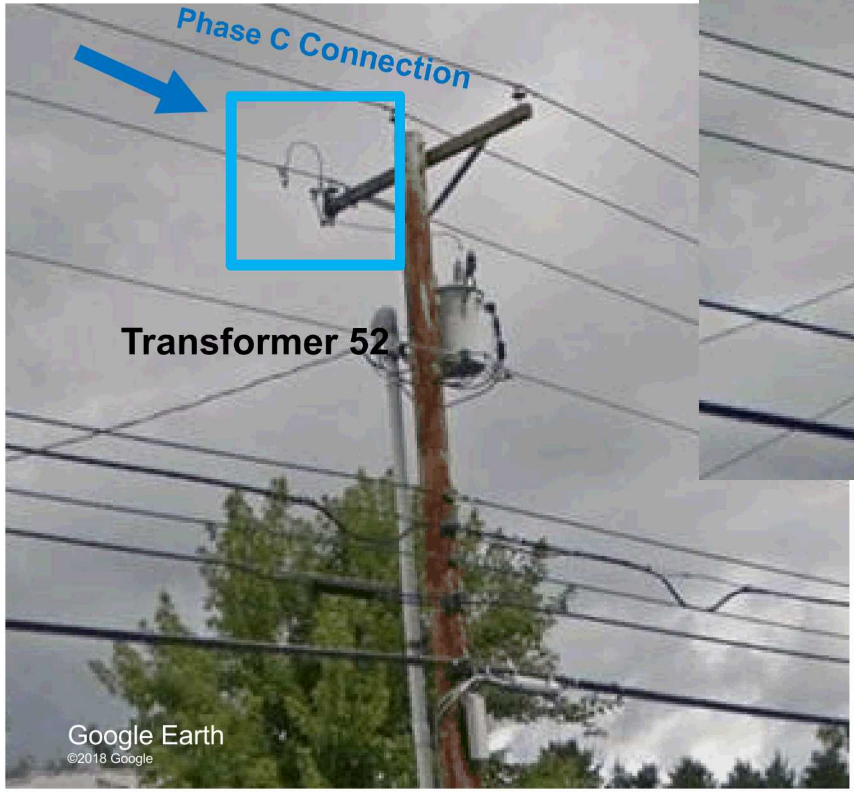
Transformer 53

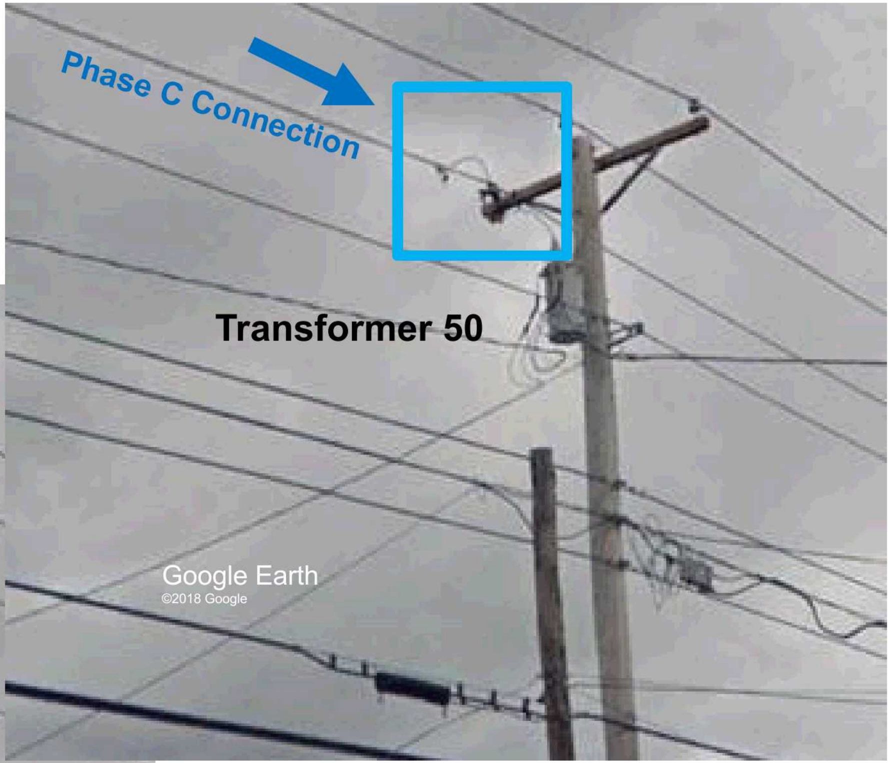
- Phase B interconnection
- 4 meters on the building

Unlabeled home (plotted in red) in the original utility model

Transformer 50, 51, & 52

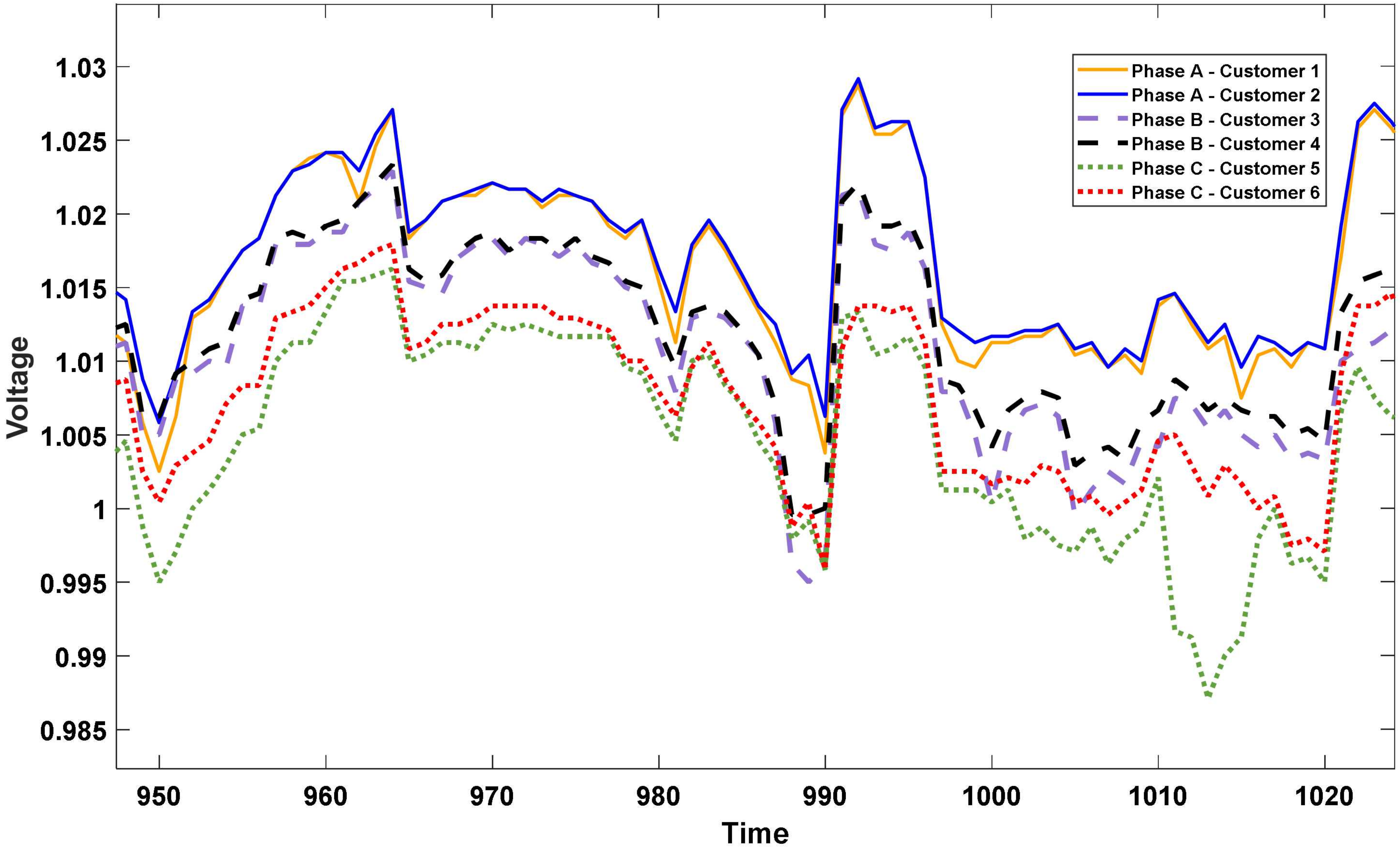
Labeled & predicted Phase A



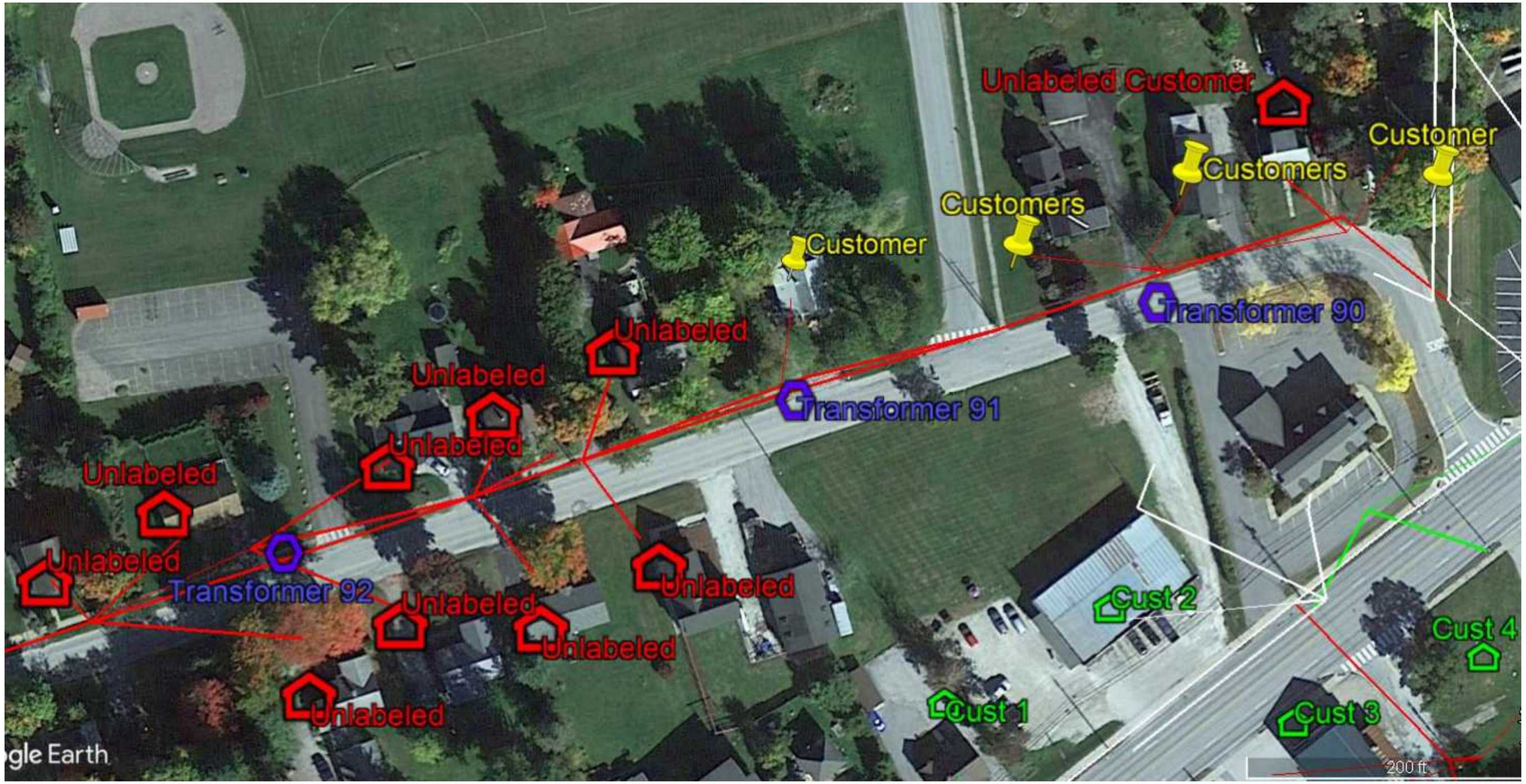


Labeled Phase A, predicted Phase C

Supplemental Slides



Supplemental Slides



Supplemental Slides