# THE CREATION, VALIDATION, AND APPLICATION OF SYNTHETIC POWER GRIDS

A Dissertation

by

ADAM BARLOW BIRCHFIELD

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Thomas J. Overbye |
| Committee Members, | Katherine R. Davis |
| | Timothy A. Davis |
| | Alex Sprintson |
| Head of Department, | Miroslav M. Begovic |

December 2018

Major Subject: Electrical Engineering

# ABSTRACT

Public test cases representing large electric power systems at a high level of fidelity and quality are few to non-existent, despite the potential value such cases would have to the power systems research community. Legitimate concern for the security of large, high-voltage power grids has led to tight restrictions on accessing actual critical infrastructure data. To encourage and support innovation, synthetic electric grids are fictional, designed systems that mimic the complexity of actual electric grids but contain no confidential information.

Synthetic grid design is driven by the requirement to match wide variety of metrics derived from statistics of actual grids. The creation approach presented here is a four-stage process which mimics actual power system planning. First, substations are geo-located and internally configured from seed public data on generators and population. The substation placement uses a modified hierarchical clustering to match a realistic distribution of load and generation substations, and the same technique is also used to assign nominal voltage levels to the substations. With buses and transformers built, the next stage constructs a network of transmission lines at each nominal voltage level to connect the synthetic substations with a transmission grid. The transmission planning stage uses a heuristic inspired by simulated annealing to balance the objectives associated with both geographic constraints and contingency reliability, using a linearized dc power flow sensitivity. In order to scale these systems to tens of thousands of buses, robust reactive power planning is needed as a third stage, accounting for power flow convergence issues. The iterative algorithm presented here supplements a synthetic transmission network that has been validated

by a dc power flow with a realistic set of voltage control devices to meet a specified voltage profile, even with the constraints of difficult power flow convergence for large systems.

Validation of the created synthetic grids is crucial to establishing their legitimacy for engineering research. The statistical analysis presented in this dissertation is based on actual grid data obtained from the three major North American interconnects. Metrics are defined and examined for system proportions and structure, element parameters, and complex network graph theory properties.

Several example synthetic grids are shown as examples in this dissertation, up to 100,000 buses. These datasets are available online. The final part of this dissertation discusses these specific grid examples and extensions associated with synthetic grids, in applying them to geomagnetic disturbances, visualization, and engineering education.

# ACKNOWLEDGEMENTS

Dr. Ti Xu passed away just as this dissertation was coming together, and with him I lost a brilliant research partner and coauthor, a friend, and a mentor since my first graduate school days at the University of Illinois. He is greatly missed, and his advice and assistance were integral to making this work a success.

I thank my adviser, Prof. Thomas J. Overbye, for his vision and guidance of this work, and my other colleagues for their assistance. Building, validating, and applying the synthetic grids is a team effort. The collaborative research environment of the whole power engineering group at Texas A&M University has made my graduate school experience enjoyable and productive.

I am grateful for friendship and support from those at Westminster Presbyterian Church in Bryan, Texas, my home for these good years of graduate school. And I am glad to have shared the Pasler Street house with Philip Barnett, Ryan Haberman, Andrew N. Fearing, and J. Michael Martin.

I am thankful to my father and mother, Ken and Denise Birchfield, who love me and encourage me, along with my sister Audrey, brother Jack, and sister Carrie, and grandparents Jerry and Jackie Birchfield and Gary and Caroline Sayre. My family are my best friends.

I am thankful, most of all, to the Lord Jesus Christ, who made me, redeemed me, and sustains me by his grace. As an engineer, my faith gives motivation to my work, as I labor to use my gifts for the good of the world and the glory of God.

# CONTRIBUTORS AND FUNDING SOURCES

This work was supported by a dissertation committee consisting of adviser Prof. Thomas J. Overbye, Prof. Katherine R. Davis, and Prof. Alex M. Sprintson, of the Department of Electrical and Computer Engineering, and Prof. Timothy A. Davis of the Department of Computer Science. Prof. Srinivas Shakkottai of the Department of Electrical and Computer Engineering substituted for Prof. Sprintson during the defense of this thesis. All work conducted for the dissertation was completed by the student independently.

The work was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000714. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

The work was funded in part by the Thomas W. Powell '62 and Powell Industries Inc. Fellowship from the Department of Electrical and Computer Engineering at Texas A&M University.

The work was funded in part by a Graduate Teaching Fellowship from the College of Engineering at Texas A&M University.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# 1. INTRODUCTION

"Reproducible science is good science," wrote a recent NSF and IEEE joint workshop report. "Increasing the trustworthiness of published research reduces effort wasted in building on faulty science" [1]. With digital technology having largely displaced print for research dissemination and curation, the amount of scientific work produced is expanding rapidly even as more questions arise about the validity and usefulness of some results that are published. Thus the practices for ensuring the quality of research are having to keep up, particularly towards ensuring reproducibility. Funding and publishing agencies in nearly all disciplines are strengthening their requirements so that peer researchers can replicate each others' results.

Data availability is at the core of this push for research reproducibility. The same NSF and IEEE report wrote that, "In the eyes of many, research reproducibility and open science are two sides of the same coin. The premise is that if everyone has access to the data that underlies a research undertaking, then the results and conclusions are more likely to be reproducible.…While sharing data and code by no means guarantees reproducibility, it certainly opens new channels for peer validation" [1]. In step with this assessment of open data, NSF and other funding agencies now require data management plans as part of proposals. The National Institutes of Health (NIH) is investing heavily in cloud-based tools to promote the archival and accessibility of vast amounts of biomedical data, and in standardized test cases and datasets for their research community [2]. The United States Department of Agriculture, in a recent report from its Science Advisory Council, stressed

reproducibility and replicability through publication of datasets as key to maintaining rigor for research in agriculture and nutrition [3].

In the field of power engineering, this need for data availability and reproducibility is pronounced. The electric power grid is a critical infrastructure system that is undergoing rapid technological changes, and high-quality research is essential to meeting the world's future energy needs in a way that is efficient, safe, affordable, and environmentally responsible. With concerns over the security of power grids against malicious cyber or physical attacks, particularly since the 9/11 terrorist attacks, the need for open data is in tension with restrictions on who can access power grid data. The security needs of large, high-voltage power grids prohibit free dissemination of their detailed models, and most such datasets are considered in the United States to be Critical Energy Infrastructure Information (CEII), accessible only to regulators, utilities, and some researchers under strict non-disclosure agreements (NDAs). While these security concerns are legitimate, they hinder innovation in this field because the researchers who can access certain datasets cannot publish or share them. The National Academies of Sciences, Engineering, and Medicine have recognized this issue in a recent report on enhancing the electric grid's resilience, and have called for the development of more public datasets and test cases [4].

Some public test cases have existed in the power systems research community for many decades. *IEEE Transactions on Power Systems* and other research journals are full of papers that demonstrate an innovation or analysis on the IEEE 14 bus case, IEEE 118 bus case, or another of the standard cases. Often research papers present results on larger, actual datasets, but little data sharing is allowed for these due to the

associated NDAs. What have been missing in the power engineering community are large-scale, complex, high-quality test case datasets that are fully public. The largest IEEE test case has 300 buses, and most of these cases were developed in the 1960s and 1970s. The grid has changed and grown since then. The most recent models of the North American Eastern Interconnect have about 70,000 buses, with remote generator regulation, phase-shifting transformers, control systems, impedance correction tables, and other complexities.

The concept of synthetic power grids refers to a systematic way of building fully public test cases for the research community. These cases' size, structure, and features are anchored in a robust statistical and structural analysis of the actual grid. Synthetic grid models are situated on a real geographic footprint, with ties to existing public energy data; however, the transmission networks themselves are entirely fictitious, with no direct correspondence whatsoever to any actual grid. The grids look and feel real, and are solidly effective for a variety of research studies; but since they are synthetic they are able to be freely published.

The problem of building synthetic grids is to create a power system dataset: all the substations, buses, loads, generators, and branches, with all associated parameters. The system must be fully public and thus cannot use any actual power system information as an input; however, public data can be used. But the system must be realistic, matching characteristics of actual grids in size, complexity, structure, and parameter statistics. This is the validation: studying actual power systems to pick out the key characteristics that, when met by synthetic grids, quantifies their realism. Many

challenges arise as the size of the systems increases, up to 70,000 buses, and complexities are added in voltage and power control devices.

The contribution of this dissertation is a new, unified suite of methodologies for creating synthetic grids that extend and refine existing preliminary considerations, suitable for building large and complex realistic grids; a systematic validation of synthetic grids anchored in analysis of actual grid properties; applications of synthetic grids to research in the fields of geomagnetically induced current, visualization, and engineering education; and a set of nine new synthetic power grid datasets, sized 200 to 100,000 buses, which are made and validated with the new framework.

# 2. BACKGROUND

## 2.1 The grid as a complex network

There has been a lot of interest since the late 1990s in various "complex networks" or graphs and how their properties are important to the underlying systems' function and reliability. A complex network is a graph with many nodes (vertices) connected by links (branches or edges). Examples come from nearly all fields: biology, such as neuron connections; ecology, such as food chains; social networks, like Facebook or Linked-In; professional connections such as author citations or movie co-stars; as well as countless technical systems such as the internet, the world wide web, airline routes, roads, microchip circuits, and power grids. This last example, of course, is the focus of this dissertation, and the complex network properties observed on these types of networks in particular are fundamental to the synthetic approach.

Watts and Strogatz [5] first introduced the "small-world" concept, and cited the power grid as an example. Partway between a regular lattice and fully random graph, small-world networks have both the property of high clustering, like a regular lattice, and short average path length, like a random graph. As Figure 1 shows, a small perturbation from the original regular lattice, through random rewiring, can synthesize the highly-clustered but short-path properties through these cross-network links that are created. The paper [5] observes these characteristics in power grids along with neurons of a worm and the graph of movie star collaboration.

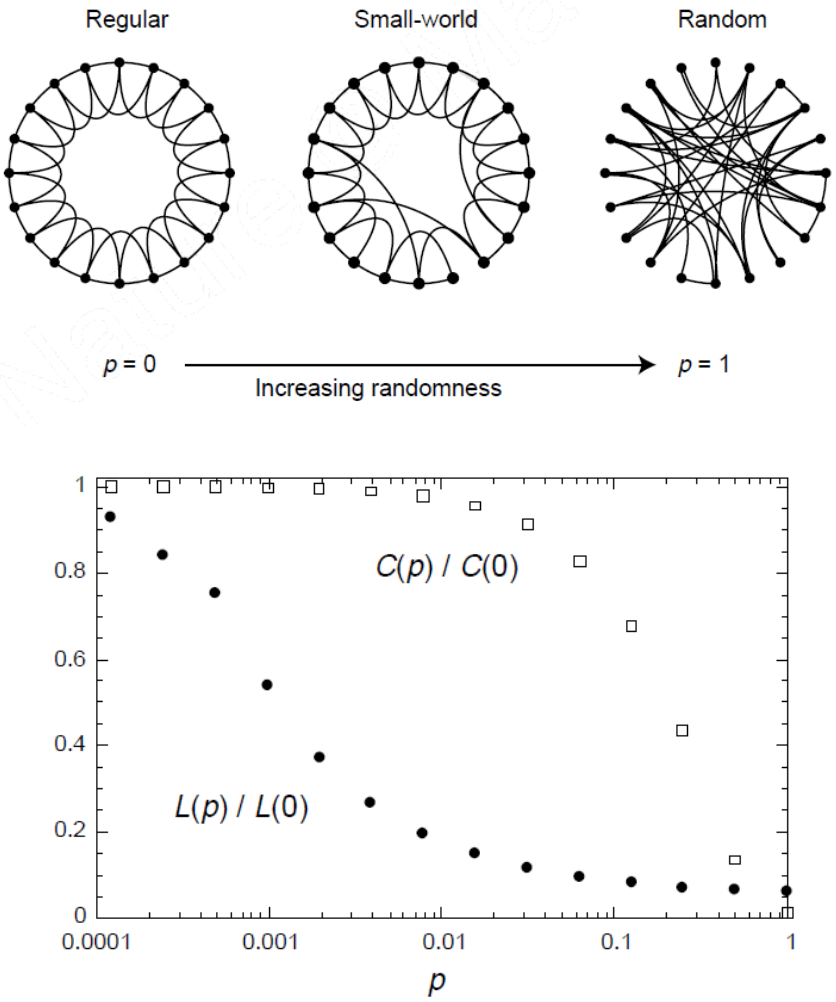


Figure 1. The small-world concept (reprinted from [5]). Above, a regular lattice becomes small world with tiny amounts of random rewiring. Below, the shortest path length $L$ decreases more rapidly than the clustering coefficient $C$ with this random rewiring, leaving the small-world sweet spot at about $p = 0.01$ for this model.

Many other analyses of power grid complex network properties exist. While some have touted this small-world model as representing grid structure well [6], at least one other study has suggested a modified scale-free model [7], more like the internet, where there are some large-degree nodes and the network appears the same at different zoom levels. Still others have rejected both models [8].

The purposes of studying the complex network properties of power grids, for many, is the insights that they show for research into system vulnerability and opportunities to improve system robustness [9]-[14]. Particularly for the study of cascading failure in propagating blackouts, these properties are used to generalize finding system vulnerability.

The challenge of accessing and sharing high-quality datasets for large power systems, due to legitimate security concerns through non-disclosure agreements, has contributed to a lack of consensus in the research community as to what the complex network properties of these mammoth machines actually are. A recent survey on the power system as a complex network showed that the reported average nodal degree in various studies of transmission systems ranged from 2.12 to 4.38, for example [15]. In addition to the influence of data non-availability, there is diversity among power grid models, stemming from historical engineering design decisions, particular needs of various locations, and different levels of modelling detail. So part of the purpose of this dissertation is to present some original analysis of power grid complex networks properties on actual system datasets, and to evaluate how they are matched in the synthetic grids produced.

### 2.2 Other test cases and synthesis methods

Some existing public test systems, such as the IEEE cases [16] (Figure 2, for example), are commonly used in power systems research, but their small size and limited complexity does not fully meet the needs of the research community today. Newly-developed grid data sets that are larger in size and have characteristics and

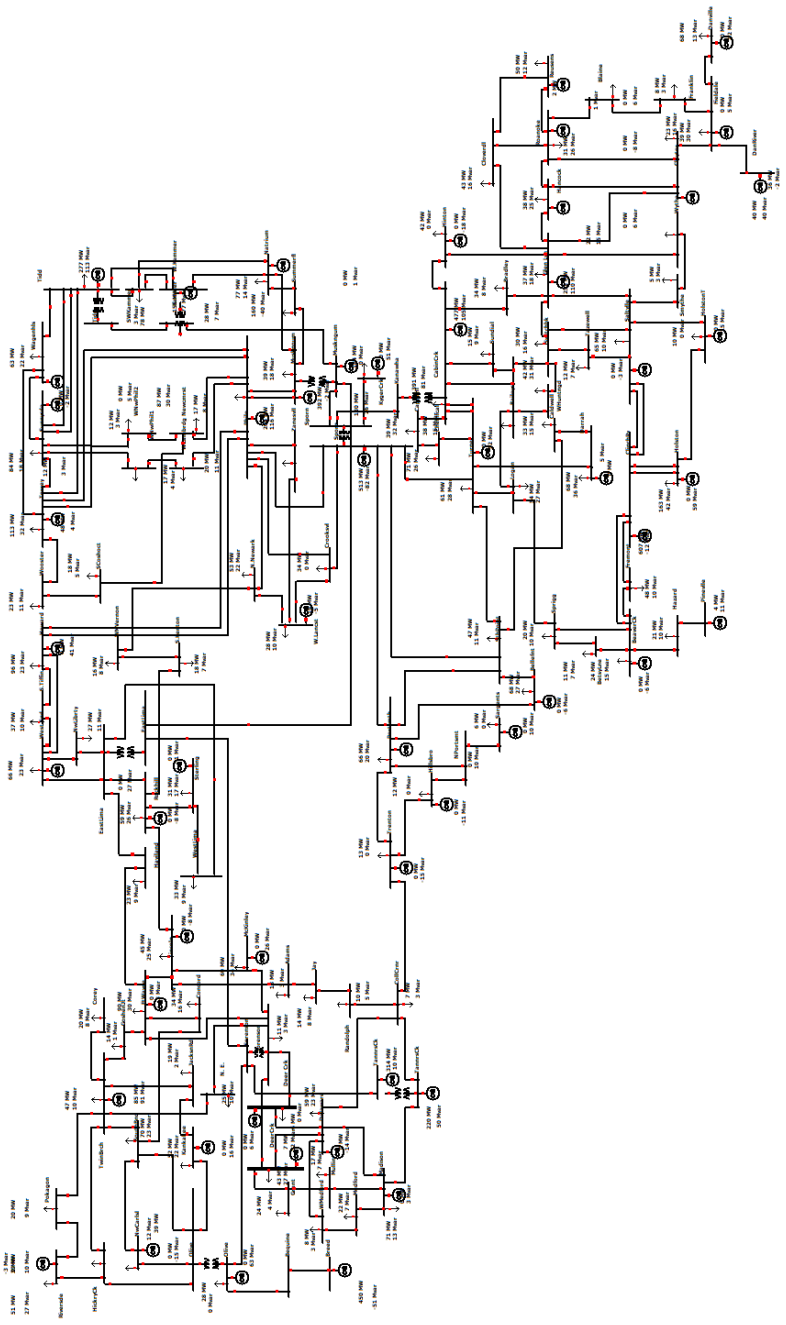

Figure 2. The IEEE 118-bus case. This is one of the largest public test cases prior to recent developments.

properties similar to actual grids, while maintaining the ability to be shared publicly, have many benefits for power system innovation. One example [17] discusses making

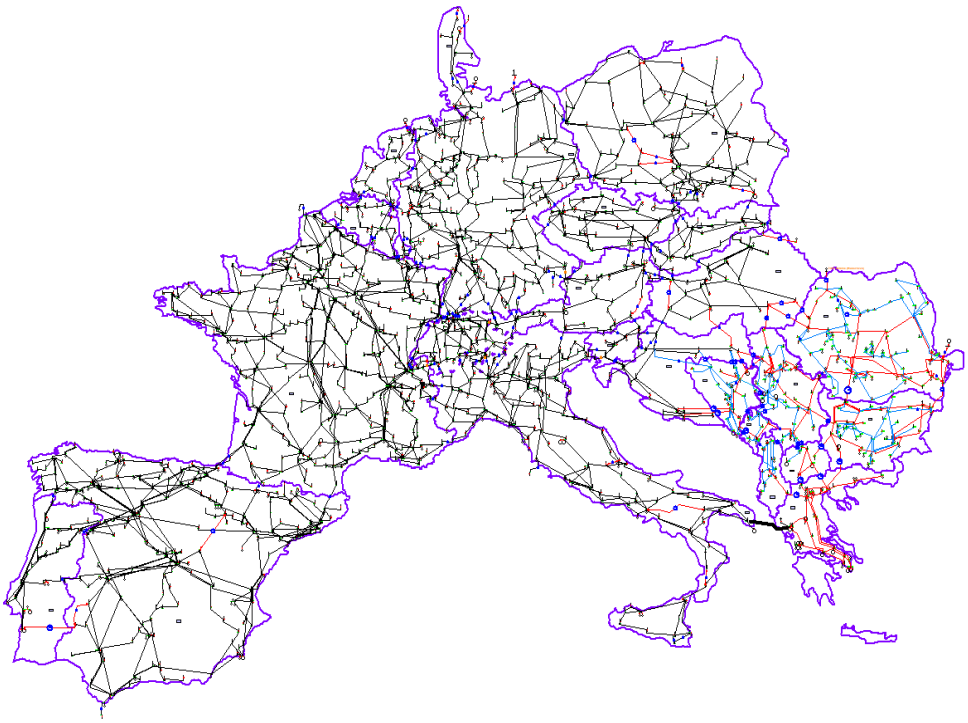

Figure 3. The Europe dc market test case. This case is introduced in [17].

a large transmission grid for dc power flow studies, based on public data for continental Europe, Figure 3. This approach mimics the existing grid rather than using a synthetic method. A recent report [18] points out both the usefulness of having more public power data available and several approaches that are being used to make more use of the data that is available, including building transmission networks.

Various initial work to generate realistic power grid network topologies are given in [6] and [19]-[21]. Reference [6] focuses on generating the topologies themselves, using a small-world model without any consideration of location except a notion of one-dimensional order. References [19]-[20] take geographically-based approaches and [21] uses a clustering-based method. Reference [22] points out the

challenges associated with modeling the variety of power system topologies in different places.

Other efforts to build synthetic grids have been recently published. In [23] and [24], the framework from [6] is extended to consider nominal voltage level and bus types. In [7], the framework of [20] is further expanded to build a large geographic grid. Other very recent developments on this topic include [25] and [26].

What is unique about the approach of this dissertation is that it addresses the full spectrum of data needed to use these synthetic grids as test cases. The cases have been released publicly in several industry data formats, and are being used for a variety of engineering studies. The process is geographic from the start and brings together the engineering design principles from transmission planning with complex network analysis. And the scalability of the methodology means the cases produced are by far the largest published, up to 100,000 buses, thoroughly validated, N-1 reliable and geographically mapped to maximize the usefulness to the research community.

### 2.3 Review of preliminary considerations

The author's previous work (MS Thesis [27]) in [28]-[30] integrates spatial, topological, and electrical requirements to make full power flow cases. This subsection briefly reviews the approach taken, which is greatly revised and extended for the work of this dissertation.

This preliminary method starts with geographic data to place loads, generators and substations. The underlying source datasets are the U.S. Census Bureau database [31] and the U.S. Energy Information Administration (EIA) Form 860 generator

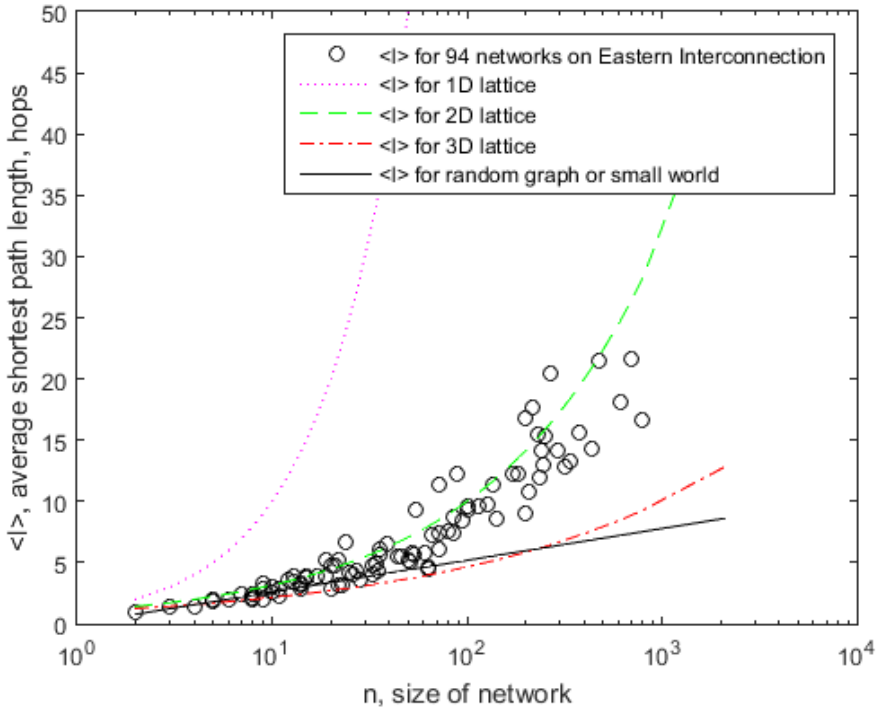


Figure 4. Average shortest path length for single-voltage networks. These scale better with 2D lattice than with random graph.

database [32]. These form the seed for the load and generation profiles, respectively, to which a clustering methodology is applied. Basic parameters are applied to each element, such as a fixed load power factor. The choice of which generators and loads will share a substation and which substations will be high-voltage is made by random assignment proportional to MW capacity.

The Delaunay triangulation, with a focus on networks of a single nominal voltage level, were introduced in the preliminary considerations as well [33]-[35]. Figure 4 shows sample results from that work, where the single voltage level shortest path analysis scales more like a 2D lattice than a random graph. This work [27] showed

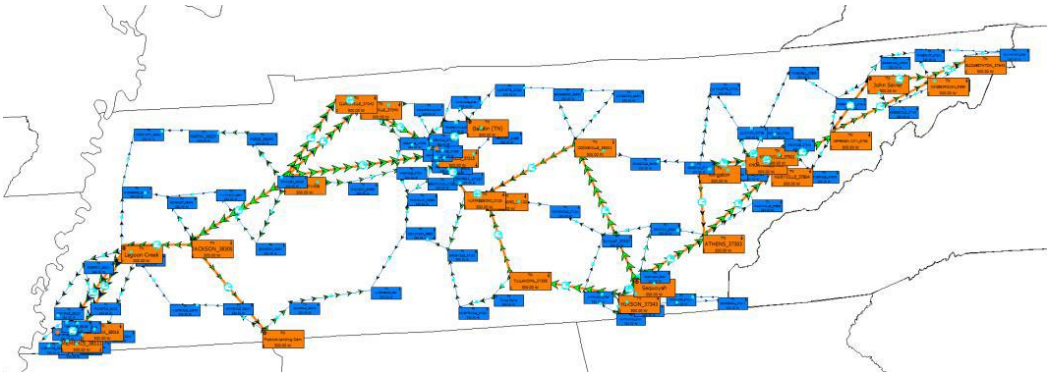

Figure 5. Diagram of the 150-bus test case.

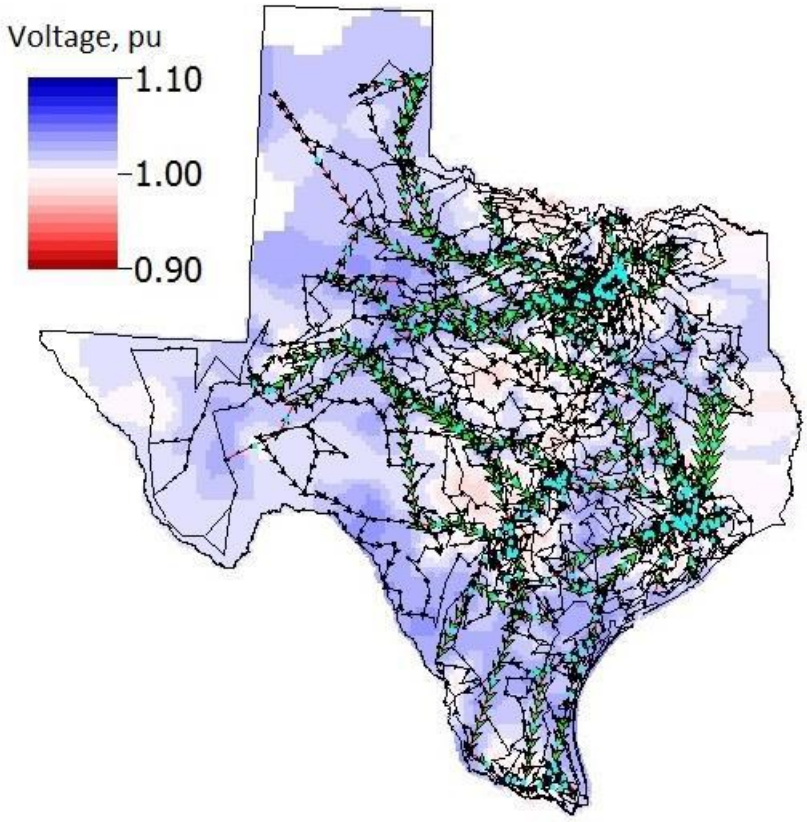


Figure 6. Diagram of the 2000-bus test case.

that many desired complex network properties of actual power grids can be met by using edges from the Delaunay triangulation.

The transmission grid for [27] is built with dc power flow assumptions by adding lines one at a time at each voltage level. Considerations are made for both the Delaunay triangulation and the predicted power flow, looking at voltage angle gradients. The resulting cases, Figures 5-6, are the 150-bus case on the footprint of Tennessee, and an initial 2000-bus case on the footprint of Texas. (Note that the 2000-bus case in this dissertation is also geo-located in Texas, but it is a completely new, rebuilt case that improves on the formulation from [27].)

This dissertation builds on and refines this basic framework in several ways. The substation planning and transmission planning are both generalized to work for cases with more than one area, including multiple overlapping nominal voltage levels, a critical piece of building very large synthetic grids. As part of this, the clustering process was extended for use in assigning voltage levels to substations. The initial transmission planning algorithm was replaced with a new approach that adds and removes lines at each iteration, considering N-1 contingency conditions with a dc power flow sensitivity to pick out the most helpful line to add at each iteration. The whole substation and transmission planning process was revamped to allow for building large, realistic cases.

In addition, a third stage is added in this dissertation for reactive power planning. While before only the dc power flow was considered, and then hopefully the ac power flow would converge (usually safe for small grids), that approach will not work for large grids; thus it is addressed in this dissertation. Furthermore, each parameter and aspect of these datasets is covered with a detailed study of actual power grid datasets, so that the created datasets from this dissertation can be validated.

Beyond the improvements discussed in this dissertation, other direct extensions of this framework have been made for different types of engineering studies. These extensions are not the primary work of this author, so they are not part of this dissertation, however they are interesting and an important part of making this work useful for a variety of engineering applications. This includes methods in [36] for economic studies, [37]-[38] for transient stability dynamics, and [39] for time series simulations.

### 2.4 Publications associated with this dissertation

Much of the material in this dissertation is reprinted from published works, accepted works, or works under preparation, with permission from the publishing organizations and the other authors.

Reference [40] focuses on the extension of the substation and transmission planning stages to scale to large systems, with multiple areas and overlapping voltage levels. Sections 3.1 and 3.2 are largely taken from here.

Reference [41] introduces the reactive power planning algorithm, a necessary third step for very large systems due to power flow convergence issues. Section 3.3 reproduces that material.

Reference [42] outlines the validation method and validation metrics, which are reproduced and expanded in section 4.

References [43] and [44] give applications of these grids to visualization and education, which are the focus of section 5.

# 3. CREATION OF SYNTHETIC GRIDS*

This section presents the methodology for building synthetic power grids. The overall synthesis methodology mimics actual system planning, by ordering first load and generation planning, then substations, followed by the transmission grid and real power with a dc approximation, finishing with the reactive power planning stage to manage voltage with an ac power flow. Creation and validation go together, and process starts with a statistical analysis of actual grid datasets. Thus the validation metrics which will be presented in section 4 are integrated into each stage of the synthesis process.

Geography, power flow feasibility, and scaling to large systems are key aspects of this synthetic grid approach. The substations are geo-located and transmission line parameters are dependent upon their geographic length. Geographic features are considered both in placing substations and in adding the lines. N-1 contingency reliability is a key consideration in planning the transmission network, which has not been addressed by any prior work. And each step is designed with computational considerations in mind, to be tractable for systems up to 100,000 buses.

### 3.1 Substation planning

The first two steps in the framework for building synthetic grids are substation and transmission planning. While the preliminary approach has been presented in the author's previous work [29]–[30], the methodology is being greatly expanded and refined in this dissertation [40]. The key additional contributions are network synthesis considering multiple nominal voltage levels and multiple geographic areas, handling lower transmission voltage levels down to 69 kV, and computational considerations needed to scale systems up to 70,000 or more buses, plus a sensitivity metric to target N-1 reliability in the transmission grid. Whereas for grid models with a size below 1000 buses it is acceptably realistic to have only two voltage levels (such as 115 kV and 345 kV) that fully cover a single area, for larger cases realistic grids will have more voltage levels, some or all of which will only span part of the entire system. Multiple areas and voltage levels, which are essential to building cases the size of real continental interconnects, present several challenges in that voltage levels can only be connected to each other through transformers when they coincide at substations, and thus the overall branch-bus topology must be considered in conjunction with the topology of the substation voltage networks individually. Given a set of substations geographically placed across a system, with each area having its designated set of voltage levels, the first problem is to assign these voltage levels to specific substations appropriate to real statistics and system needs.

By way of example, a new public test case is presented by this section. Similar to the case of [30], the geographic footprint of Texas is used to build a 2000 bus case.

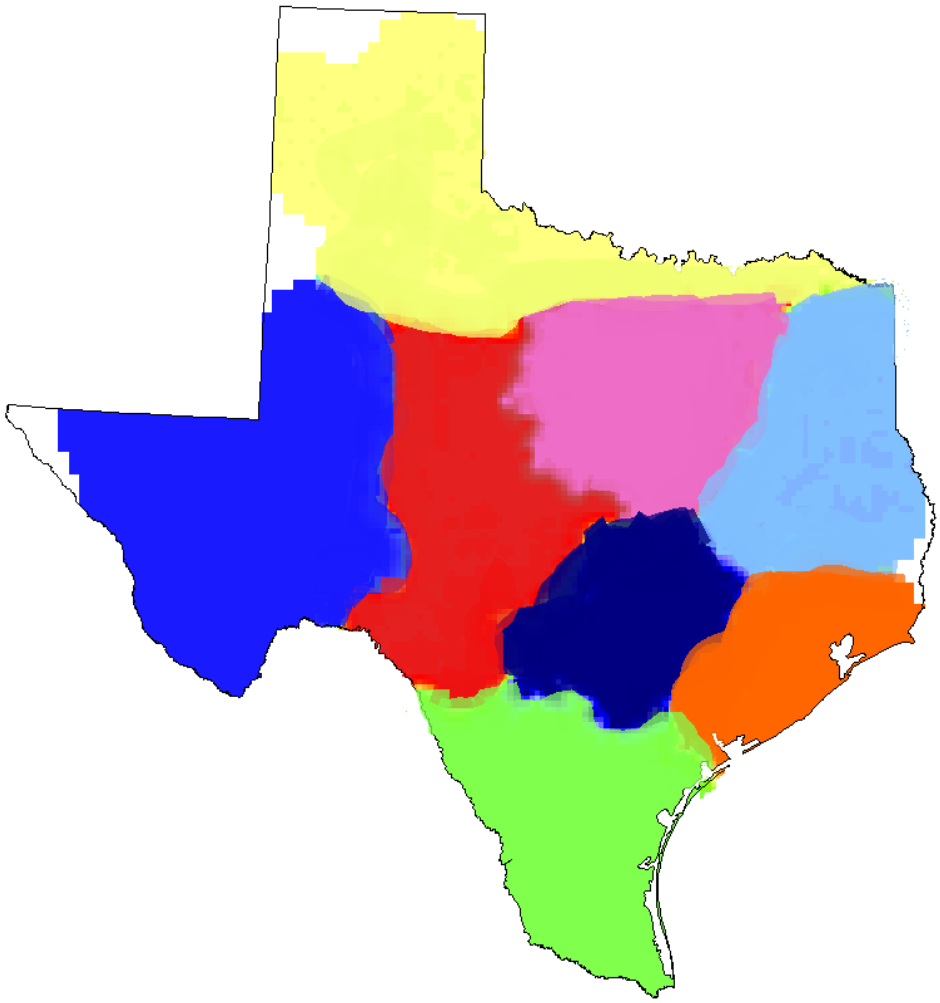

Figure 7. Eight areas of Texas used for the 2000-bus case. The areas are: Far West (blue), West (red), North (yellow), North Central (pink), South Central (indigo), South (green), East (cyan), and Coast (orange).

This case is entirely new and includes more voltage levels as well as an updated algorithm, building upon [30]. It is used as an example through this and the next subsections.

In [30], two voltage levels were used, 115 kV and 345 kV. Each substation was given a 115 kV bus, and 15% were selected at random to be given a 345 kV bus. The random selection was not uniform, but larger generators and loads were given a higher

| Area | Percent of substations containing | | | |
|---|---|---|---|---|
| | 500 kV | 230 kV | 161 kV | 115 kV |
| Far West | | 22% | | 100% |
| West | 8% | 18% | | 100% |
| North | 10% | | 100% | |
| North Central | 12% | | 100% | |
| South Central | 8% | 18% | | 100% |
| South | 8% | 18% | | 100% |
| East | 13% | 22% | | 100% |
| Coast | 13% | 22% | | 100% |

Table 1. Voltage level specifications for new 2000 bus case.

probability. This served as an excellent simplified approximation. To improve realism and prepare for larger and more complex cases, the analysis here rebuilds the case with four voltage levels, assigned to areas in Figure 7 according to Table 1. Though these areas were used before, here they are integral to the formation of the network.

The values in Table 1 are designed to create four overlapping voltage networks that cover the whole case with sufficient coverage for each area and sufficient diversity across the case. Many areas have three voltage levels, while some have only two. The 500 kV network ties most of the system together, but the underlying networks vary considerably. The percentages given are based on the principle, observed on actual grids, that nearly all substations will have a connection to an area's lowest voltage level, while about 10-20% will have a connection to the higher level. If there is a third level, 5-15% of substations will be contained.

With substations selected based on public population and energy data, similar to the method of [30], the set of substations are geographically assigned to the eight areas. Then, voltage level assignments must be made, satisfying the designed percentages such as from Table 1 and also the needs of the network.

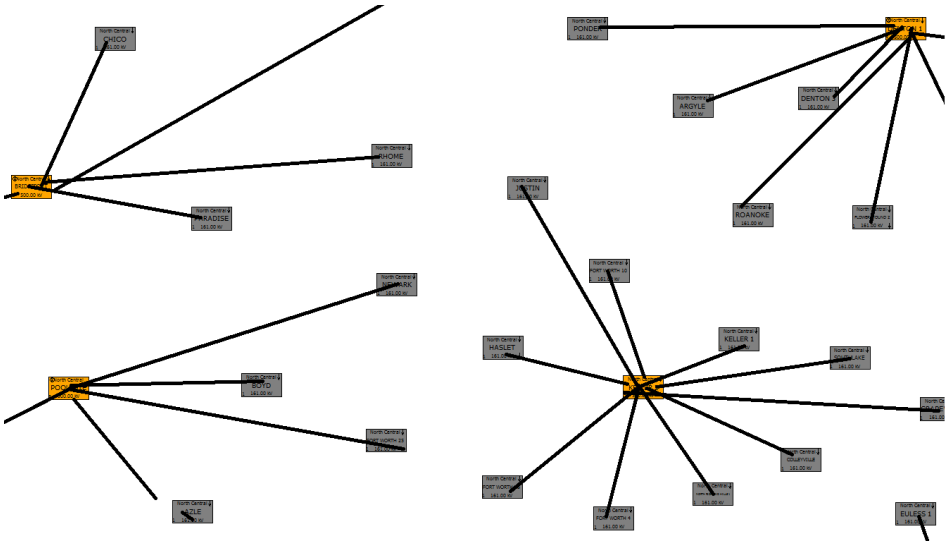


Figure 8. Example of the parent-child substation hierarchy. These are for the 161-500 kV North Central area near Fort Worth. The boxes represent substations. The 500 kV substations are selected to minimize the product of distance with the total load and generation of each child 161 kV substation. Note that this does not show the transmission lines, only the hierarchical structure used to assign voltage levels.

**Upgrades for connections within areas.** Since the lowest voltage level covers the most substations, all of them are initialized to that level; then, some are upgraded to higher levels. If desired, some of the higher-voltage substations could subsequently have the lower voltage levels removed.

In deciding which voltage levels to add to which substations, two considerations are made. The first is that there must be a sufficient high-voltage network available to meet the demand of the substation and its nearest neighbors' load. Second, the voltage levels must line up across area borders so that the areas can interconnect at all levels.

Computational complexity benefits from the fact that this first concern only considers the needs of a single area. No matter how large a system grows, the time to

complete this step will only grow linearly with the number of areas, with the size of each area being the primary determining factor.

The approach taken is to cluster the substations into groups, where one substation in the group is considered the "parent" substation and the others are "children." Ultimately, the parent substations will be upgraded to the higher voltage level, though it will not necessarily need to connect to its children in the transmission network. The parent-child hierarchy is illustrated in Figure 8. For networks with three voltage levels, two iterations of the clustering are done, first to cluster the lower voltage substations into medium-voltage substations, then to cluster the medium-voltage substations into the highest-voltage substations, with two levels of resulting hierarchy.

The algorithm assigns each substation a base weight, which is defined as the sum of the loads and the generator capacities at that substation.

$$W_i = \sum P_{g,i} + \sum P_{l,i} \tag{1}$$

For the two-level hierarchy, the weight in the second level is the same as the first level plus sum of the weights of all children from the first level.

The objective of the clustering is to minimize the weight of $N$ clusters, where $N$ is found from the total number of substations and Table 1. The weight of a cluster is a function of the weight of its members and their distance from the parent.

$$W_c = \sum_{i}^{i \in c} W_i d_{ic} + \alpha_p \sum_{i}^{i \in c} W_i^2 + \alpha_d \sum_{i}^{i \in c} d_{ic}^2 \tag{2}$$

where $W_i$ for each child is defined in (1), $d_{ic}$ is the geographic distance of the child to the cluster parent, and $\alpha_P$ and $\alpha_d$ are parameters, which can be used to tune the algorithm to focus more on limiting total power handled by a substation or total

distance. For this implementation, $\alpha_p = \alpha_d = 0$ seemed to give an acceptable solution.

Perfect satisfaction of the optimality formulation

$$\min \sum_{c}^{N} W_c \tag{3}$$

is not needed, but the definition in (2) provides a sufficient gradient to do the clustering. A greedy, steepest-descent approach is used, where each substation begins in its own cluster, with the total cost function equal to zero. At each iteration, one substation is selected to become the child of its nearest neighbor, and the cost function is updated. The substation is selected based on the smallest incremental change to the total cost function. The process continues until the number of clusters has been reduced to $N$. If at any iteration the selected substation already contains children, each child is assigned a new parent substation to which it is nearest. These changes are considered in the gradient analysis. To speed up this process, at the beginning each substation is assigned a parent priority list, which is the list of all other substations ordered by distance from it. Therefore, each time it needs to check for a new parent, it only needs to check at most a few substations on its list. The result is that the algorithm is relatively quick for up to a few hundred substations.

With this algorithm, each substation that is upgraded to a higher voltage level serves a purpose: it provides a network connection to the higher voltage level for a number of substations in its geographic neighborhood. The more load and generation in a neighborhood, and the more distance between substations in it, the more high-voltage substations it will have. Remote substations with very low loads may not need

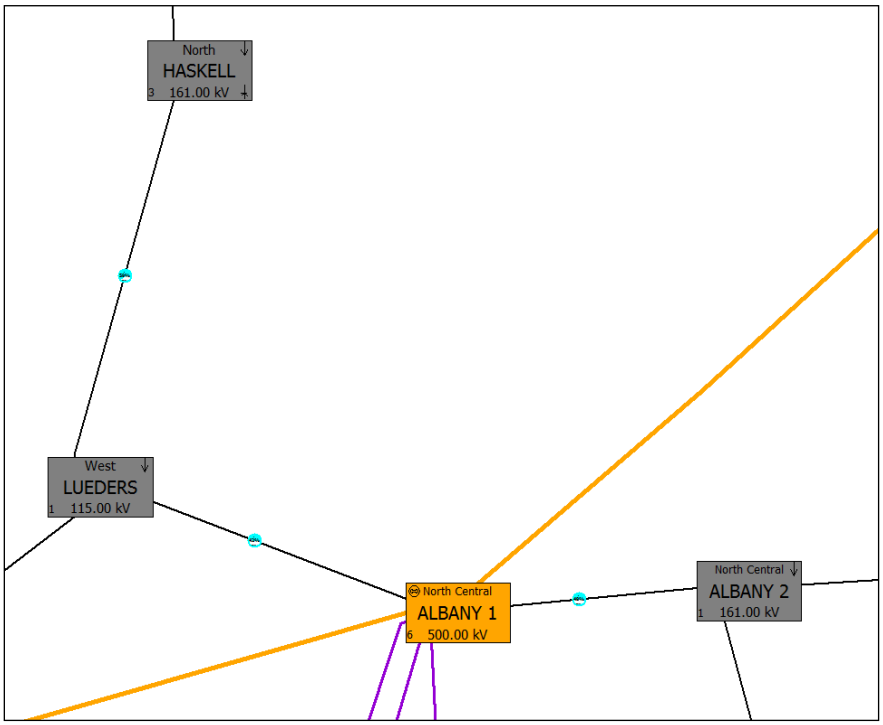


Figure 9. Example cross-area connection. The 161 kV substation (gray, top) and the 500 kV substation (orange, center) each have a 115 kV bus added, which is not part of their area's network, so that they can connect to the 115 kV substation in another area (gray, bottom left). Transmission network is shown.

the full high-voltage transmission out there; similarly, a large urban area will not need excessive high-voltage substations clumped together. This algorithm establishes a balance that covers the area in proportion to the load density and geographic density. At the same time, the largest generators are almost guaranteed to be upgraded to higher voltage levels, because making them children of another substation would increase the cost function too much.

**Upgrades for cross-area connections.** Focusing the voltage assignment problem on area needs individually is useful, not only to reduce the computation time, but also to avoid the confusion of voltage levels which span some but not all of the

system. However, at the boundaries of the areas there must be ways for differing voltage levels to connect. Appropriately selecting these substations is the subject of this subsection.

The first issue is finding out which areas border each other and how strong that connection is. A simple metric employed here is to use the Delaunay triangulation of all the system substations. The Delaunay triangulation is quick to calculate and provides a good approximation of neighborhood for geo-located points. For any two areas that are connected to each other by at least 10 Delaunay segments, there ought to be connections at high, medium, and low voltage levels. (Note that throughout this document, the terms high, medium, and low voltage levels refer to relative magnitude within the high-voltage transmission range of 69+kV).

For any boundary where the two areas share voltage levels, no additional connections need to be made. They will be connected using these shared levels. For any boundary with mismatched voltage levels, additional buses are added to some substations along with transformers that connect them. These substations are selected as the shortest ones in a separate Delaunay triangulation using only the substations involved. Figure 9 shows an example where additional buses have allowed lower-voltage connections between the areas that would not otherwise have been possible. So there will be 161:115 kV transformers at these boundary points but not anywhere else, similar to the actual grid. These inter-area connections are essential to the interconnected nature of the grid.

### 3.2 Transmission planning

The transmission planning stage is the second part of the synthetic grid creation process. The methodology developed in this section is a major revision and improvement on the preliminary framework proposed in [30]. The approach uses an add-and-remove iterative process which ranks lines to be added and removed, continuing until all desired criteria are met. The analysis done at each iteration includes a dc power flow along with a Delaunay triangulation comparison and depth first search to measure biconnectivity and protect bridges.

The algorithms to build the network topology must be adjusted to ensure that the designated characteristics are met, with multiple voltage level networks interacting. While the various algorithms used in this work do not consider all the complex factors that are used in actual power system planning, the resulting substation bus assignments and network properties are shown to match important features of the real grid, and the resulting cases can be validated for realism. These techniques have allowed scaling network synthesis up to 10,000 buses.

This section reformulates the network synthesis algorithm to make several changes. These modifications are related to the multiple voltage levels and multiple areas, but also contribute more generally to the quality of the cases, allowing flexibility to add additional constraints and objectives. Existing constraints fall in two categories, described below: the ones which consider power flow analysis and those which consider topology. The problem has some similarity to transmission expansion planning, but the starting point is an empty graph and the objective is realism rather than optimal performance. This new method begins with an initial set of lines and

| Nominal kV | Characteristic p.u. X/mile | Characteristic MW |
|---|---|---|
| 115 kV | 0.0054 | 160 |
| 138 kV | 0.0040 | 223 |
| 161 kV | 0.0029 | 265 |
| 230 kV | 0.0015 | 541 |
| 345 kV | 0.00058 | 1195 |
| 500 kV | 0.00025 | 2598 |
| 765 kV | 0.00012 | 4100 |

Table 2. Characteristic per-distance X and MW for voltage levels.

both adds and removes lines at each iteration, rather than simply adding them from scratch without removal. This means that analyses can affect the reward and penalty structures differently in the removal and add stages.

**Initial dispatch and power flow considerations.** The method described in [30] runs an iterative dc power flow solution that estimates the power that would flow through each potential line and contributes a reward for that line proportional to the potential power. This reward competes with other rewards and penalties from other analyses to determine whether the line will be added. A similar structure is used here, with the power flow results affecting points for both removal (using calculated power flow in existing lines) and addition (using potential power flow as before). However, four important adjustments are made. First, the power flow part of the analysis is heavily dependent on the assumptions made with respect to the generation dispatch. Second, the power flow must be normalized so that it affects multiple voltage levels equally. Third, the power flow ought to affect addition much more than removal, as often lines with low power flow exist for other reasons. Fourth, near the end of the

development, power flow considerations should not be allowed to swamp meeting the targets in other areas.

For the generator dispatch, several options are available. The solution proposed before is reasonable, which pre-specifies inter-area exchanges and then dispatches all generators proportional to net load. An improvement upon this method is to set non-dispatchable load such as wind, hydro, and solar to some fixed proportion such as 25%, then dispatch the rest according to an equal-lambda economic dispatch. Creating synthetic cost curves was described in [36], and the local cost of coal and gas can be adjusted to tweak the system inter-area power flows as design goals require. Making these assumptions for the initial dispatch will aid in making the cases more applicable to optimal power flow (OPF) analysis.

With multiple voltage levels across many areas, the amount of MW flow that would be considered large through a transmission line varies greatly. With this in mind, it is important to ensure that the reward given for large power flow is not overly dominant at the high voltage levels and insignificant in the low voltage levels. To address this concern, a characteristic MW value is assigned to each voltage level, as shown in Table 2. The actual or potential power value is divided by the characteristic value and capped at 1 to get a normalized value. This normalized value is then multiplied by some reward that can be adjusted, but will have a similar effect on all voltage levels. The reward used is 500 for addition and 50 for removal. These rewards are added for the first 80% of iterations, after which other considerations are allowed to dominate.

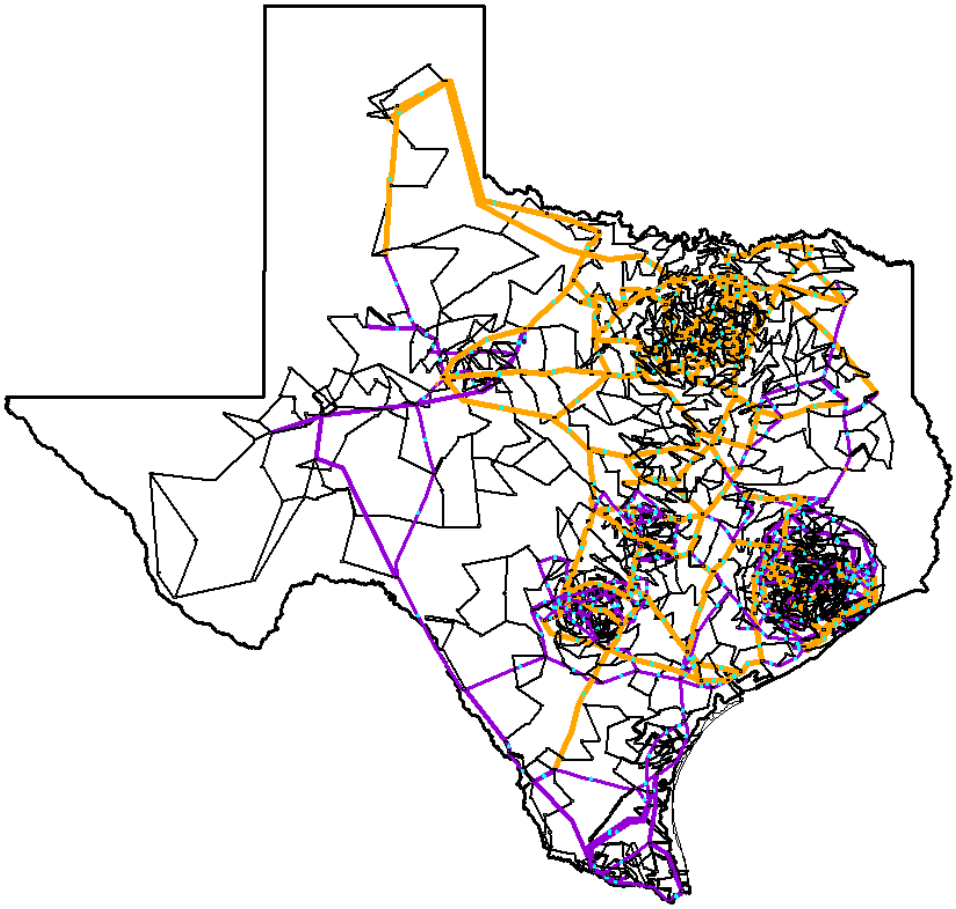

Figure 10. The transmission network for the 2000-bus case. The 500 kV network (orange) covers all of the system except the far west. The 230 kV network (violet) is not present in the northern part of the grid. The 161 kV and 115 kV networks (black) split the grid between north and south.

**Delaunay triangulation and topological considerations.** In [30], the Delaunay triangulation is noted for its usefulness in dramatically weeding out potential transmission lines as well as being a proxy for a variety of topological and geographic characteristics that are observed in actual grids. By restricting the search space to sections within three hops on the Delaunay triangulation, the computation time is dramatically reduced and over 99% of actual lines are considered. Moreover, by matching the percentages of first, second, and third neighbor Delaunay segments, the

clustering coefficients, shortest path length, and node degree distribution naturally turn out similar to an actual grid. Meeting these objectives remains a key priority in building synthetic networks.

Additional topological requirements are heuristic, as bi-connectivity is enforced, along with connectivity of each voltage network and the forbidding of any radial loads. Each of these can be checked with a linear-time depth first search analysis. These are given rather high penalties or rewards, to ensure that the requirements are met. For example, all candidate lines that connect to a radial sub are given a reward of 200, and any line which is one of only two connections to a substation is given a removal penalty since removing it would create a radial substation.

**Structure and implementation of topology algorithm.** The network generation algorithm is an iterative process, and at each iteration there are several steps. A few analyses are performed for the network as a whole, including the dc power flow and depth first search for bi-connectivity, as well as the identification of radial substations. Then each voltage level is considered in turn. Analysis is made of the proportions of Delaunay first, second, and third neighbors matched, respectively, and the connectivity of the voltage level. Then 0 or 1 lines are removed and 1 or 2 lines are added. Bridges (which would disconnect the network) can never be removed. Since the network is initialized to the minimum spanning tree, at the beginning all lines are bridges and therefore no lines can be removed. Until the designated number of lines is added, the number of lines added in an iteration is one more than the number removed. Each voltage network runs its own removal and addition at each iteration,

and the iteration continues for enough rounds to meet all the criteria. Figure 10 shows the resulting line topology for this case.

**Summary of advanced transmission planning considering N-1 contingency sensitivity.** In further refining the methodology just discussed, we employ the following procedure for building the transmission network of a synthetic power grid. While there are $n^2$ possible transmission lines connecting n substations at a voltage level, previous work has shown that nearly all practical transmission lines are in the tractable set of about $23n$ lines that come from the geographic Delaunay triangulation and its second and third neighbors [30]. These are the candidate lines, and $1.15n$ to $1.25n$ of them are picked arbitrarily to form an initial network. At each iteration, remove one branch at random from each subnet, and then pick a candidate line that best contributes to the goals for the transmission system, and add it back to the network. Inspired by simulated annealing, these repetitive steps of random removal, smart addition, produce a network which balances both key objectives in actual transmission system planning: geographic feasibility and electric reliability.

The geographic feasibility goal considers, primarily, line length. Cost-effective planning prefers shorter lines, so our method gives candidate lines a piecewise-linear penalty for length, gently encouraging shorter lines within the reasonable range for a nominal voltage level and sharply disallowing lines longer than are actually seen. The candidate line's length is calculated from the straight-line path between substations, scaled by any geographic features it crosses. In addition to length, the synthesis process works to match the distribution of Delaunay neighbor count: there should be the right proportion of lines in the system that are first, second, and third neighbors along the

Delaunay graph, matching actual grids [30]. To enforce this, uniform penalties are given to candidate lines in a category which is already overrepresented in the system.

The reliability analysis for synthesis focuses on the key planning criteria of N-1 security, that is, electric service should not be interrupted, nor should lines be overloaded, by the outage of any single element. To achieve this, there is both a topological analysis and a power flow analysis. The topological analysis performs a depth-first graph search, returning the connected components, bridges, articulation points, and biconnected components. The graph should be fully connected, and remain so with the loss of a single edge or vertex. With negative penalties, candidate transmission lines between two disconnected components and those between two biconnected components are strongly encouraged. In later iterations, bridges are protected against removal to ensure connectivity.

The power flow analysis uses linearized (or "dc") power flow modeling, which ignores the flow of reactive power and assumes voltage magnitudes are equal to the nominal value [45]. For each synthesis iteration, a full N-1 contingency set is run, and for each subnet a critical contingency is chosen which causes the most branch overloads. To encourage system N-1 reliability, transmission lines are favored which contribute to easing these critical contingency overloads. For a few critical contingency overloads, the sensitivity of all $21n$ candidate lines to the power flowing in the overloaded line can be calculated quickly [46]. Then the candidate lines are given a penalty negatively proportional to their sensitivity values, which encourages adding lines to mitigate these critical contingency overloads.

This paragraph reviews the planning sensitivity formulation from [46] which is used in the transmission planning process. Using the $\boldsymbol{B}$ matrix dc power flow formulation and differentiating it to a small change in system impedance:

$$\bar{P} = \boldsymbol{B} \cdot \bar{\theta} \tag{4}$$

$$d\bar{\theta} = -\boldsymbol{B^{-1}}(\boldsymbol{dB})\bar{\theta} = -\boldsymbol{B}^{-1}\bar{e}_i\bar{e}_i^T\bar{\theta}dB_i \tag{5}$$

where we are modifying the $\boldsymbol{B}$ matrix on right-of-way $i$, where $\bar{e_i}$ defines the right-of-way path by being zero at all buses except a 1 at the from bus and -1 at the to bus of the right-of-way. The $\boldsymbol{B}$ bus matrix, which would reflect the system configuration in a given contingency, would be modified with only four values by the differential change in admittance $dB_i$. Now (4) gives the sensitivity of system angles to that admittance change. To get the change in power across right-of-way $k$,

$$P_k = X_k\phi_k = X_k\bar{e}_k^T\bar{\theta} \tag{6}$$

$$d\phi_k = \bar{e}_k^T d\bar{\theta} \tag{7}$$

Thus the sensitivity formulation becomes

$$\frac{\partial\phi_k}{\partial B_i} = -\phi_i\bar{e}_k^T\boldsymbol{B}^{-1}\bar{e}_i \tag{8}$$

Since the $\boldsymbol{B}$ matrix is symmetric, we can then rearrange the sensitivity to be

$$\frac{\partial\phi_k}{\partial B_i} = -\phi_i\bar{e}_i^T\boldsymbol{B}^{-1}\bar{e}_k \tag{9}$$

So with one LU factorization of $\boldsymbol{B}$ and backward substituting for $\bar{e}_k$, the sensitivity of the power through that right-of-way $k$ can be determined with respect to adding admittance to any other right-of-way $i$ by only three floating-point operations each, as follows:

$$\bar{s}_k = B^{-1}\bar{e}_k \tag{10}$$

$$\frac{\partial \phi_k}{\partial B_i} = -\phi_i \bar{e}_i^T \bar{s}_k = -(\bar{\theta}[i_{\text{from}}] - \bar{\theta}[i_{\text{to}}]) \cdot (\bar{s}_k[i_{\text{from}}] - \bar{s}_k[i_{to}]) \tag{11}$$

Geographic, single-line circuit diagrams of the resulting 70,000 bus synthetic system on the footprint of the U.S. portion of the EI can be seen in Figures 11 and

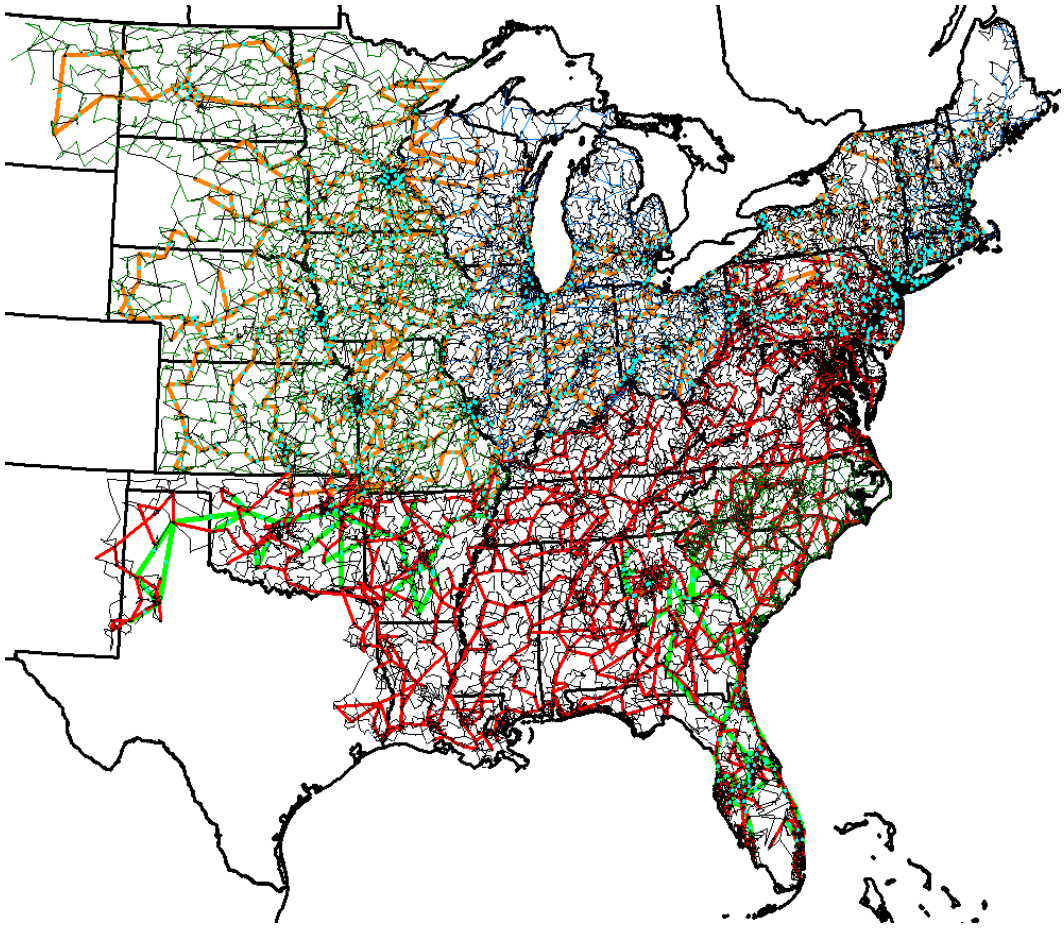

Figure 11. Single line diagram of the 70,000 bus synthetic power grid. This system is fictitious and does not represent any actual power grid. The transmission lines are shown, colored by voltage level. The geographic footprint for this case is the United States portion of the EI, with states as areas and geographic features including coastlines, mountain ranges, and urban centers considered. Arrows indicate the magnitude and direction of power flow along the lines.

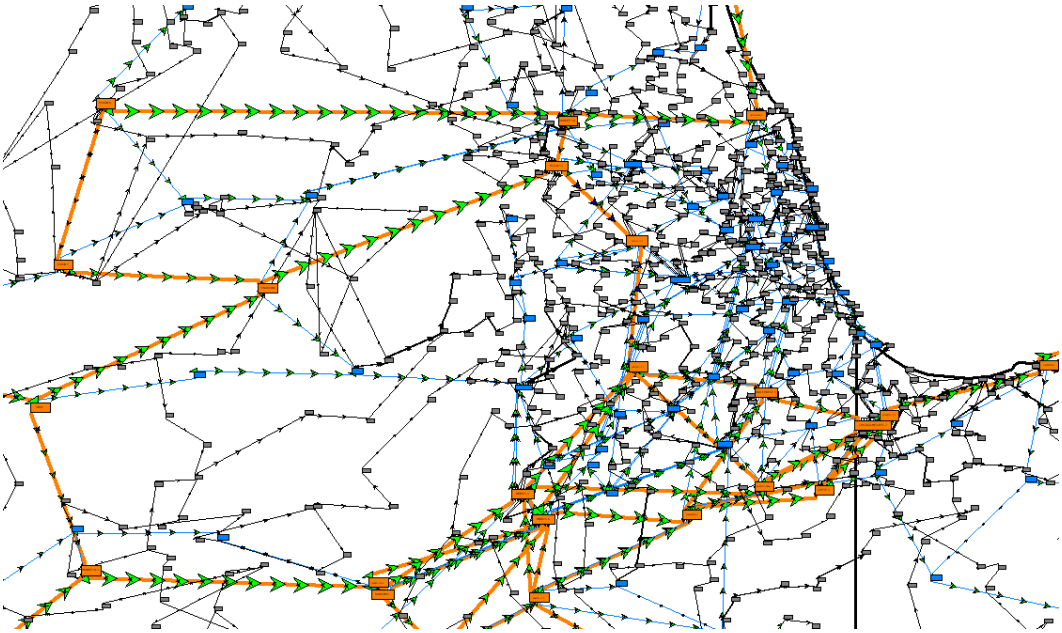

Figure 12. Zoomed single line diagram of the 70,000 bus synthetic power grid. The zoomed-in view of the Chicago metropolitan area is shown.

12. More information about the cases built through the methodologies of this section can be found in section 6.l.

### 3.3 Reactive power planning

This section builds on the network synthesis algorithm, extending the substation and transmission line placement process with a completely new third step to consider the reactive power planning requirements and other complexities related to system power and voltage control. The new developments lead to full and realistic ac power flow solutions, meeting the challenges introduced as the system scale becomes large. This section begins with addressing the challenges in finding an initial ac power flow solution on a large scale synthetic grid. Then there is a detailed algorithm

for reactive power planning in large synthetic grids, for which the 10,000 bus test case is used as an example.

The method, in short, is to move incrementally from a dc power flow, for which the system has already been optimized with a good solution, to a full ac power flow solution with a reasonable set of reactive power support devices. This is begun by initializing the system to have a very large number of devices controlling the voltage magnitudes of most system buses to a common, flat voltage. Then, iteratively, some of the temporary devices are removed at each step, adjusting the remaining ones by repeated ac power flow solutions.

**Background on power flow convergence.** The standard non-linear power flow equations for ac power system solutions are solved iteratively. The conventional Newton-Raphson method is described in [47]. When these methods were first being applied to computer simulations, and as the size of the systems studied increased, it was noted that the Newton-Raphson power flow solution, being non-linear and iterative, is not guaranteed convergence [48]-[49]. This is still an active area of research, and several factors can affect convergence such as the problem conditioning, voltage stability characteristics, and the choice of initial values for the variables [50]. Fractal domains of attraction make predicting or guaranteeing convergence to a specific solution difficult [51]-[52]. Reference [53] describes an application of homotopy solution methods, which have been proposed to improve convergence for cases that are perturbed further from the solution. Power flow convergence and analyzing the solvability of cases is applicable to voltage stability and dynamics studies, and has been studied in these contexts [54]-[57]. For many studies, the key to good convergence (if

a solution exists at all) is a good initial guess; this presents a problem for synthetic systems which have no previous solution and do not have reactive compensation in place, a problem which is addressed here.

**Background on reactive power planning.** Reactive power planning has many conventional optimization methods to add capacitors and other devices to an existing transmission system [58]-[61]. Methods are available to optimize over many objectives: installation cost, real power losses, fuel cost, voltage profile, and voltage stability [60]. Most of these methods, however, require a convergent initial power flow solution, and sometimes also require the initial voltage profile or other constraint to be met. The main purpose of these methods is incrementally adding reactive power support devices to an existing grid. Furthermore, there are computational limitations in many of the methods that prevents scaling to many thousands of buses, especially when a large number of devices must be placed. Building on this work, the problem addressed by this section involves adding resources to a case without any existing devices or solution, with the objectives being to meet statistical metrics of realism rather than minimizing cost (although synthetic generator cost curves are considered in the initial dispatch, leading to good starting solutions for ac optimal power flow solutions).

**Challenges in obtaining initial ac power flow solutions for large-scale synthetic grids.** Modeling power system loads as constant real and reactive power means that even the smallest systems might have no ac power flow solution, or multiple solutions. A bus with fixed real and reactive power will have a maximum loading, above which there is no solution, based on the impedances of the connecting

lines and voltage of remote buses. Within this loading constraint, there will often be at least one low-voltage solution in addition to the expected solution closer to nominal, due to the nonlinearities of the power flow equations. For cases where multiple solutions exist, the initial guess of a Newton-Raphson power flow will determine which solution, if any, is reached by this method [54].

In large systems, these constant power situations combine across interconnected buses to make a correct solution even more difficult to find, if one even exists. Often the non-existence of solutions can be interpreted as the inability of the required amount of real or reactive power to be transferred from available sources. In the case of real power, the typical formulation of the power flow problem requires that each generator's MW set point be specified beforehand except one slack bus. The slack bus must pick up whatever real power load and losses are not provided by the other generators. Hence the dispatching of the non-slack generators to meet the assumed load and losses has a significant impact the solution. For small cases, the slack bus can correct a higher relative error in the loss assumptions present in the dispatch, but for large systems even 1% error in the assumed losses could be far more than the slack bus is able to produce.

Synthetic transmission grids, created according to the previous subsection, are built initially using an iterative dc power flow solution, which models real power flows only, in a lossless approximation of the system. The generator dispatch used in this step, which is based on an economic dispatch with synthetic cost curves, must be adjusted to account for assumed losses, or the ac power flow will fail because the slack bus will be incapable of supplying all of the losses. Though generator participation

factors are used as an outer-loop adjustment to account for losses, there must still be an initial inner-loop solution (using a single slack bus for real power mismatch, as in [47]).

Large system reactive power flows are even more likely to produce an unsolvable case, since the high $X/R$ ratio of most transmission system branches prevents reactive power from traveling far, meaning reactive power must be supplied within a nearby region. If some region of buses does not have sufficient reactive power available from generators or other devices, it cannot be brought from very far away, causing the case to be unsolvable. Thus while for real power the main concerns are system-level (ensuring that the losses are distributed among all dispatched generators), for reactive power the concerns are largely localized, making sure that the net static reactive power in some neighborhood, summing the loads, losses, and shunt capacitors, is sufficiently supplied by nearby voltage-controlled devices like generators, subject to those devices' reactive power limits. In synthetic grids, no modeling of reactive power is done in the dc power flow, so there are no additional support devices and potentially many areas without sufficient support.

For large system cases which have a solution, getting the Newton-Raphson method to converge to that point depends heavily upon the initialization. A flat start, which is the most basic initialization option, refers to setting all bus voltage magnitudes to 1.00 per-unit and angles to 0°. There are additional assumptions, as to whether generators are at reactive power limits, which shunts are switched in, and where transformer taps are set, which would all also be initialized to some default in a full flat start. Basically, the assumption is that nothing from prior solutions is known. For

smaller cases, reaching a good solution from flat start initialization is possible, but the large interconnect cases (e.g., EI or WECC) often do not flat start in many commercial software packages ([62], page 48). These solutions diverge, even when the original settings are maintained for shunts, taps, and generator limits. In practice, flat starts are not used for large interconnects; instead, new solutions are obtained by small modifications to an existing solution.

A solution technique that requires previous good solutions will not work for synthetic cases, which are being solved for the first time. Synthetic grids, created with the dc-based method of the last section, not only lack a previous ac power flow solution, but have no initial set of voltage control devices, including reactive power resources. A key problem in building large-scale synthetic cases is adding a realistic set of reactive power devices, coordinating them in a way that they have a reasonable solution that fits a desired voltage schedule, and then actually finding this solution, which will likely not be available from a flat start.

Power flow solution convergence is also affected by additional complexities, which, while not inherently unique to large-scale systems, are often found in them. Three-winding transformers are typically modeled as three equivalent branches radiating from a fictitious star bus; the equivalencing can produce negative reactances and reactances close to zero, which can complicate sensitivities in voltage regulation. A related issue is remote bus voltage regulation, where a generator, shunt device, or load tap-changing (LTC) transformer regulates a bus other than its own terminal, sometimes with multiple remote buses regulating the same bus. It is quite common practice, for example, for a generator to be modeled in the power flow as regulating

the high-side bus of its step-up transformer (GSU) [63]-[65]. These devices can fight one another and lead to convergence issues. Phase angle regulating transformers (PARs, also known as phase-shifters), can further complicate the power flow as they regulate real power flow and integrate impedance correction tables. Each of these issues must also be addressed in building synthetic systems that match the complexity of the actual grid.

**Case study: ten thousand bus system.** To demonstrate the issues raised by the previous discussion, and to serve as an example for the proposed algorithm to follow, this portion introduces a 10,000 bus synthetic power grid. The geographic footprint selected corresponds to the U.S. portion of the North American western interconnect (WECC). This region has a population of over 70 million, with census and generator data publicly available. The region was divided into areas along state lines, with California, as the most populous state, being subdivided into five areas. Seven voltage levels were selected: 765, 500, 345, 230, 161, 138, and 115 kV, and each area was designed 2-3 voltage levels. The system has its generation and load substations placed from public information as well as a clustering method. Buses and transformers are assigned to each substation using the modified hierarchical clustering described in the previous subsection, and an initial generator economic dispatch is set. Then the transmission network is built through an iterative dc power flow solution, gradually adjusting to meet the statistics and metrics observed in actual cases.

Included in this case are 300 three-winding transformers, divided into three main types: GSUs, load step-down, and modified network transformer with a low-voltage tertiary. Parameters are calculated analogously to two-winding transformers, as

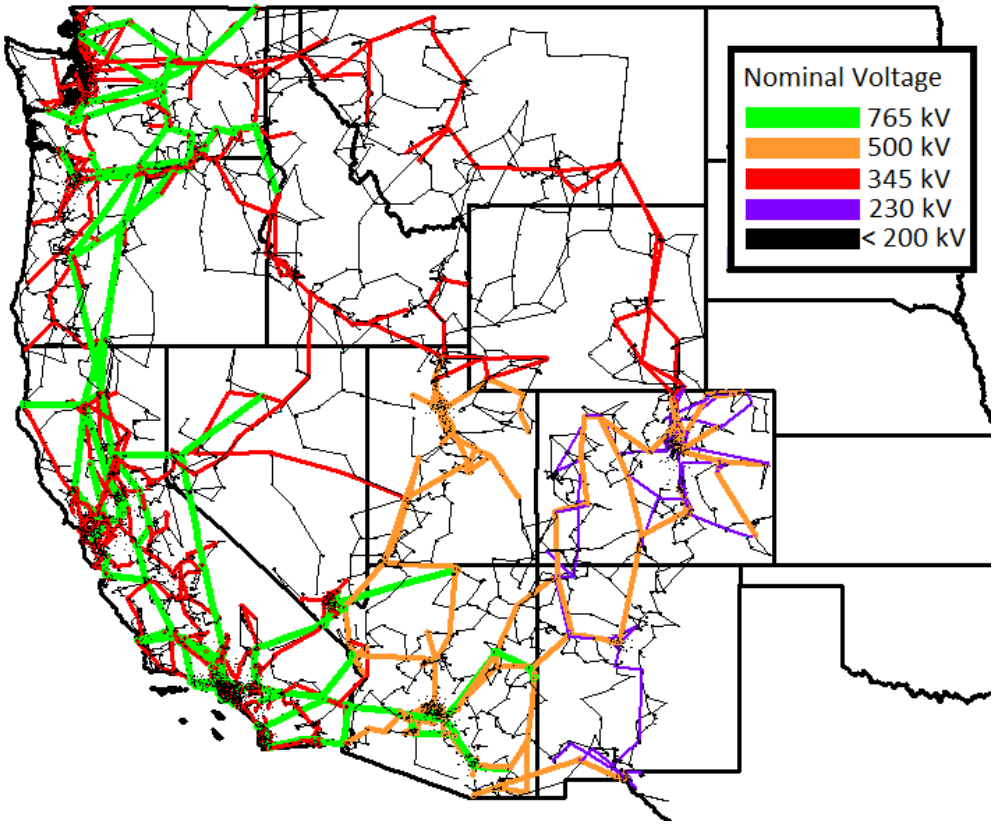


Figure 13. Oneline diagram for the synthetic 10k grid. The diagram shows the transmission line voltages. This case is totally fictitious, built from public information with a synthetic methodology, and does not represent the actual grid in this location.

in the metrics of section 4, then converted to an equivalent set of three branches with a star bus, as they are typically modeled in commercial power flow solvers. The result is a transmission system that has been validated with a dc power flow, but does not yet have an ac power flow solution. The oneline diagram can be seen in Figure 13, and the basic statistics are given in Table 3.

Unsurprisingly, the case did not converge in an ac power flow initially, despite having a reasonable dc solution. Various contributing factors to this are those mentioned in the previous section, including the wide spread of voltage angles, limited reactive power resources, and three-winding transformers. Next, an implementation

| System Element | Number |
|---|---|
| Buses | 10000 |
| Substations | 4762 |
| Areas | 16 |
| Voltage levels | 7 |
| Loads | 4899 |
| Generators | 2485 |
| Transformers | 2981 |
| Transmission lines | 9726 |
| Three-winding transformers | 300 |
| Tap-changing transformers | 294 |
| Phase-shifting transformers | 5 |
| Switched shunts | 387 |

Table 3. Statistics of the 10k case.

of the algorithm described below was applied to the system, placing shunt capacitors and reactors, and setting the control points and tap settings for voltage control devices.

**Overview of algorithm for synthetic reactive power planning.** The approach used here is to move incrementally from a dc power flow, for which the system has already been optimized with a good solution, to a full ac power flow solution with a reasonable set of reactive power support devices, as shown in the flow chart of Figure 14. This is begun by initializing the system to have a very large number of devices controlling the voltage magnitudes of most system buses to a common, flat voltage. Then, iteratively, some of the temporary devices are removed at each step, adjusting the remaining ones by repeated ac power flow solutions.

**Initial power flow solution.** Thus the first step was to take the 10,000 bus system transmission network and augment it with a large number of temporary voltage control devices. Most of these will eventually be removed, with the remnant becoming shunts. The temporary devices initially added to the system are generators, set to 0

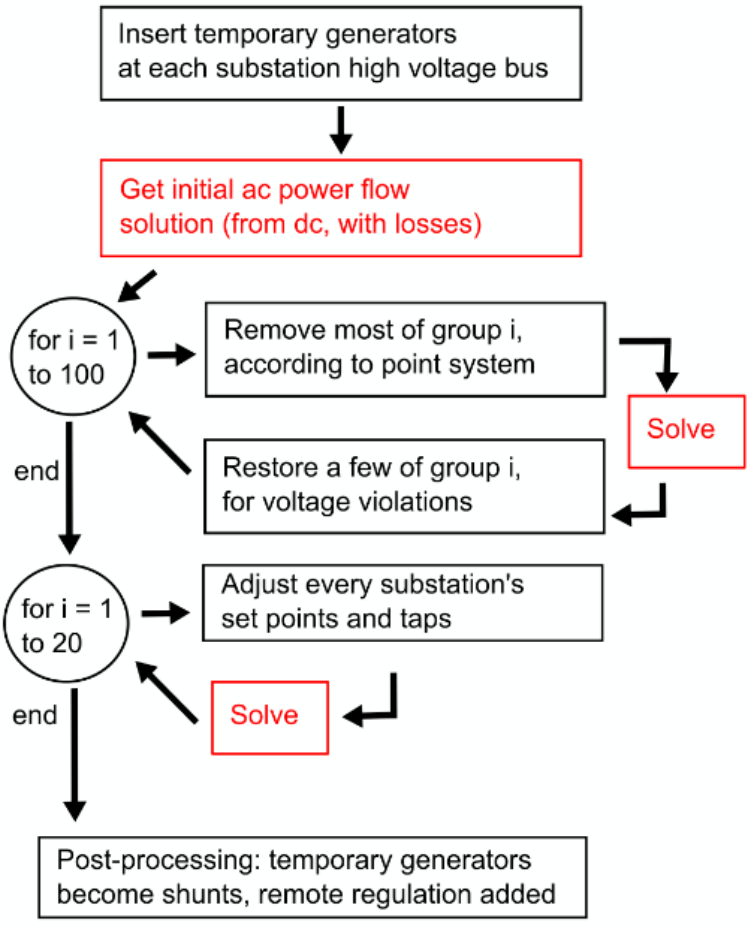


Figure 14. Flow chart of synthetic reactive power planning algorithm.

MW active power output, with reactive power limits of ±300 Mvar up to ±800 Mvar, depending on the nominal voltage level. In the ac power flow, these will be modeled as PV buses, which at first will all be set to regulate their own bus to 1.04 p.u., to avoid generators fighting one another. Also for this reason, the temporary generators cannot be added too densely, as they tend to converge to undesirable solutions (for example, one generator producing 250 Mvar, with its neighbor on the other side of a low-impedance branch absorbing 230 Mvar). A compromise that works well in the analysis

here is to initialize one temporary generator at the highest nominal kV bus in each substation, that is, 4762 in the 10,000 bus case. Of course, this is far more than a realistic number of shunts; however, the approach of this section is to work backwards from a feasible power flow solution to a realistic set of devices.

The first ac power flow solution is initialized with all the temporary generators in place. The voltage angles are initialized to the solution of a dc power flow, including some compensation for an assumed real power loss percentage, such as 1-3%. The actual generators must be dispatched accordingly. Sometimes this loss percentage must be adjusted over a few iterations to get a convergent initial solution, but, thanks to the large number of temporary 0 MW generators, reactive power is not a concern in this first solution. The voltage profile is largely flat. For the synthetic 10k case 1.5% assumed losses worked initially, with a convergent, flat solution.

Once the ac power flow has converged, the rest of the process is just a matter of making small changes to the system and analyzing at each step the impacts of the modifications on the resulting solution. The approach of this section does this in two stages. The first stage removes a large number of the temporary generators, to match the expected number of shunt compensating devices. Then the second stage modifies set points and transformer taps to fit the voltage magnitude distribution desired.

**First stage iterations: removing most temporary devices.** For the first stage, consideration is made of the principle that reactive power effects are mostly localized. Thus many temporary generators can be removed at once, provided they are not clumped too tightly together. This leads to the strategy of selecting 100 groups of temporary generators, uniformly at random, and addressing each group in turn. In the

10,000 bus case, the 4762 temporary generators will be divided into groups with 47 or 48 devices each, dispersed across the system. Most of the temporary generators will be far geographically from any other in its group. The approach of considering one group at a time strikes a balance between an intractable number of solutions and making too large of a localized change that could cause the solution to diverge.

Thus, the first stage will have exactly 100 iterations, with the following steps at each iteration (1) remove almost all of the temporary generators in the specified group, keeping a few a priori, (2) perform the ac power flow solution – restoring all generators in this group if it happens to diverge, and (3) restoring a few more of the original temporary generators as necessary for voltage support, based on the outcome of the solution.

The decision of whether a temporary generator should be removed is based on both factors related to the statistical observations from the actual grid and the specific localized reactive power needs of a neighborhood in the system. To integrate these notions, a points system is devised, with points corresponding to various heuristic factors that increase the likelihood that a temporary generator should be removed. The point system is specified in Table 4, with considerations for voltage level, additional resources around, and the topology of the system. Before the solution, all temporary generators in a group except a few with the lowest points are removed. For the 10,000 bus case only 4 were kept before the solution from each group. After the solution has been reached, all buses above a nominal voltage of 50 kV are analyzed for voltage violations outside the range [0.96, 1.06]. For any bus that violate this constraint, which had a temporary generator removed this iteration, that generator is

| Condition | Points |
|---|---|
| Nominal kV < 200 | 2 |
| Nominal kV < 400 | 2 |
| Substation generation > 100 Mvar | 1 |
| Substation generation > 10 Mvar | 1 |
| No. tie lines = 2 | 1 |
| At least one tie line is sending MW | 1 |
| Nearest Q resource is 1 hop away | 3 |
| Nearest Q resource is 2 hops away | 2 |
| Nearest Q resource is 3 hops away | 1 |
| Second-nearest Q resource is 1 hop away | 5 |
| Second-nearest Q resource is 2 hops away | 4 |
| Second-nearest Q resource is 3 hops away | 3 |
| Second-nearest Q resource is 4 hops away | 2 |

Table 4. Point system for removing temporary generators.

restored, up to a specified fraction. These fractions are chosen to match a maximum of 16% of total substations, which means in the synthetic 10k case up to 7 out of the 44 removed temporary generators could be restored at each iteration to fix voltage violations. Many iterations did not need all of these restorations.

If, by chance, the power flow solution fails to converge at some point, that group of temporary generators is considered critical to the system, and fully restored. Since this is a rare occurrence and will only be 1% of the total devices, this will not significantly impact the statistics or the solution. For the example case here, no iteration failed to converge.

The result of the first stage for the 10,000 bus case is that 387 out of the original 4762 temporary generators remain, providing voltage control and reactive power support to various needed regions of the system. The voltage profile remains high and flat.

**Second stage iterations: voltage schedule adjustment.** Next, the second stage of the reactive power planning algorithm adjusts all the generator set points, actual and temporary, which were initialized to 1.04, as well as transformer tap settings. Before beginning these iterations, a realistic fraction of transformers is selected to be allowed to adjust their taps to control the voltage, with distinction made between network and generator step-up (GSU) transformers. These selections are made probabilistically. Unlike the previous iterations, the system admittance matrix must be reformed at each step, since the changing of taps affects it. While each iteration is therefore slower, the overall time of this stage is low since the parallel adjustment of each substation's set points only requires about twenty iterated solutions.

With each substation largely independent during the voltage scheduling iterations, this step of the algorithm can be decoupled. During initialization, the substation is assigned a target voltage, which will be selected uniformly from the range [1.035, 1.045]. Buses nominally in the range [200 kV, 400 kV] will be regulated to 0.002 above that, and those above 400 kV will be regulated to 0.003 above the substation schedule. These are of course small modifications to the general principle that system voltages should be made flat and high, recognizing that in real systems this ideal cannot be perfectly matched.

For any substation with a temporary generator or an actual generator attached directly to the network voltage buses, the voltage set points will be adjusted in the second stage iterations to reduce the difference between the actual bus voltages and their schedule. The changes made are small and discrete, so that each power flow

solution is only a slight modification of the previous one, to avoid fighting between devices.

For transformers with an allowed tap-changing mechanism, these taps are also adjusted, in increments of 0.00625, to work on matching substation voltage schedules. For network transformers, the transformer is tapped as needed, maximum once per iteration, for the lower voltage bus voltage magnitude. For GSUs, the approach is slightly more complicated. If the high-voltage bus of the GSU already has another device controlling its voltage, the GSU and generator behind it work towards adjusting the generator's Mvar output to about 30% of the generator's maximum Mvar output, so that there will be plenty of reserve reactive power in the system. Otherwise, the GSU and associated generator must work to control the high-voltage bus's voltage magnitude.

Finally, the slight adjustments of the voltage scheduling iterations are complete, and the temporary generators are converted to shunt capacitors and reactors with the given reactive power set points. Many of the actual generators which are connected through a GSU are set to regulate the bus voltage of the high side of the transformer, according to [63] and [64]. For the synthetic 10k case, it was then exported to a commercial power flow solver, and the solution converged as expected.

### 3.4 Additional complexities

In addition to the grid size, complexities in synthetic power systems are what distinguish them from existing test cases. Phase-shifting transformers, remote bus

voltage regulation, tap-changing transformers, switched shunt reactors and capacitors, and impedance correction tables are added to synthetic grids as part of this work.

**Adding phase-shifters to 10k case.** Phase angle regulating transformers (PARs) are a small minority among transformers in interconnect cases, however, they are important to specialized purposes such as balancing active power between parallel paths and reducing loop flow through an area [65]-[67]. Often associated with PARs are impedance correction tables, which change the branch impedance according to the off-nominal phase shift value of the PAR. As the PAR taps from 0° phase shift to a significant shift such as 30° or more, the branch impedance increase significantly. Impedance correction tables are also used in some tap-changing transformers, but the focus is on ones associated with PARs. These tables have the following observed characteristics (1) They are approximately symmetric around a phase shift of 0° (2) the center impedance is the smallest (3) a good fit is quadratic.

# 4. VALIDATION OF SYNTHETIC GRIDS*

Validating full power system models, that is, determining how accurately their features match what is found in the actual grid, is key to ensuring the quality of new synthetic power grids for their use in research and development. This section introduces a systematic approach to validation, which contributes many new validation metrics and their defined criteria to match. These metrics are designed to help quantify the realism of a synthetic grid. Because of the variety in engineering design and modeling practices, actual grids are quite diverse; the interesting challenge in this work is to capture the distribution of network characteristics, in a way that synthetic grids can be adequately evaluated. In addition, the size of a network can affect its statistical properties, since large networks have averaging effects. Each of these issues is addressed in this work by studying a high-quality, diverse, large set of North American power system models. The initial suite of validation metrics defined here contributes a benchmark for developed cases.

Every aspect of the synthetic grid validation is anchored in a thorough analysis of high-quality real power grid models. The actual power system data for which statistics are given in this work comes from observations of the major North American power grid interconnections, as obtained from the FERC form No. 715 dataset, as

well as twelve subset cases created by extracting 84 areas along geographic and utility lines from the full interconnects. From these, statistics are gathered 85 on cases ranging from 400 to 5000 buses, in addition to the 70,000 and 16,000 buses in full eastern (EI) 86 and western (WECC) interconnect cases, respectively.

The framework of this validation process is broad in application, since collecting statistics on system properties and identifying benchmarks is appropriate for many aspects of the power system which may be synthesized. The focus of the metrics selected, however, is in three categories: the metrics of system proportions, those of system network structure, and those of power flow solvability with voltage control. Together, these categories cover much of what is needed for a base case power flow solution. The idea in picking metrics is to obtain wide coverage of parameters. Except transmission lines, everything in power system models are contained in substations, so these aggregations are the orientation of the questions answered by selected metrics – How many substations are there? What voltage levels do they contain? How much load and generation do they have? Then more detailed metrics are studied that set the power flow parameters of loads and generators. Covering the branch topology is the objective of the second set of metrics. Here, substation transformers are studied in their impedance and limit parameters. The same is studied for transmission lines, followed by topological observations, which likewise are focused on substations and voltage levels. At each stage, coupling is considered among metrics; clearly nominal voltage level will significantly impact transmission line impedance, for example.

For each metric selected, a quantitative threshold standard is decided, with the expectation that no realistic power system will violate that standard, unless there is an

exception that has a justification in engineering design choice. In other words, this validation is a screening process that looks at almost all parts of the grid model and picks out any unusual data for further scrutiny. Exceptions of this type are part of the diversity of engineering practices among many grids. The case size must also be considered when looking at exceptions, as large cases are bound to have a few outliers, but will have much more consistent trends than smaller cases, which are more sensitive to the peculiarities of location.

### 4.1 Metrics of structure, proportion, and parameters

**Number of buses per substation.** Substation aggregation of buses indicates how buses are related to a specific geographic location. While substation grouping and geographic location are not strictly necessary for power flow solutions, they are integral to an understanding of grid topology, since geography is a major driving factor in system design**.**

The EI averages 2.3 buses at each substation, and the WECC averages 2.5. The subset cases considered vary from 1.7 buses per substation to 4.5. The number of buses represented in each substation can be affected by modeling decisions about how much detail is represented, including generator step-up transformers and sub-transmission network equivalents. Figure 15 shows the distribution of substation size. There are many substations with 1-3 buses, much fewer with 4-10 buses, and fewer still with 10-25. The larger the case is, the longer the tail of this distribution, as Figure 15 shows. For cases on the order of 100 buses, the tail could end at about the 1% threshold, which would make a largest substation of about 8 buses acceptable. The EI

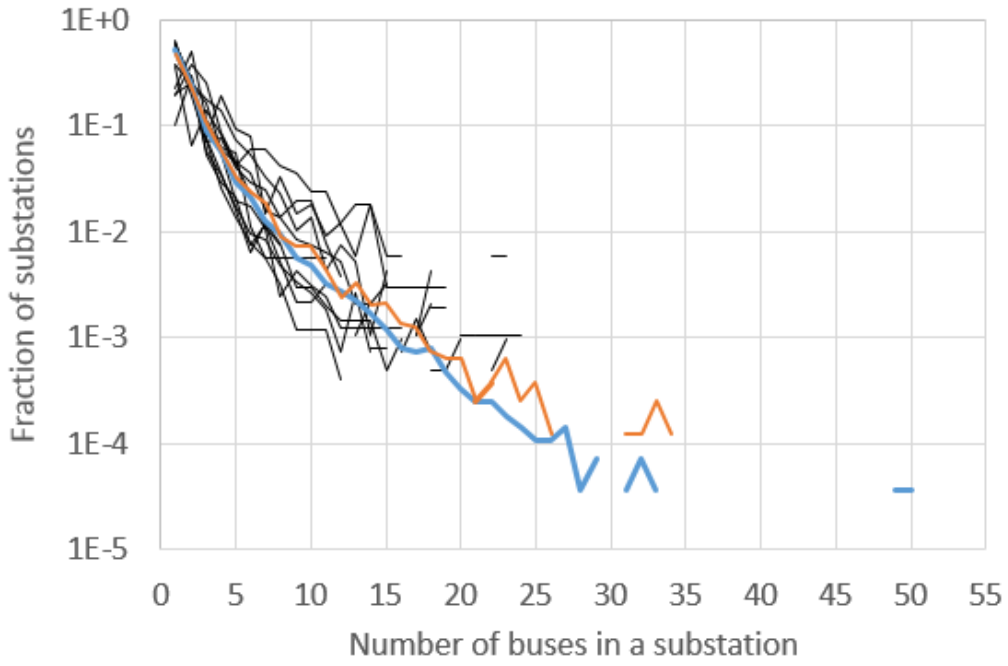


Figure 15. Number of buses in substations. Probability mass function, for EI (blue), WECC (orange), and 12 subset cases (black).

and WECC cases (orange and blue in Figure 15) have on the order of 10000 buses; their tail extends to the 1e-4 threshold at about 27 buses.

**Substation voltage levels.** The synthetic networks will focus on transmission nominal voltage levels of 69+ kV. Table 5 shows the percentages of such substations with buses in the 69-200 kV range and the 200+ kV range, for each of the fourteen cases. The majority cluster of areas in Table 5 indicates synthetic networks should have a 69-200 kV bus at 85-100% of substations, and 7-25% of transmission substations should have a bus in the range 201+ kV. The two exceptions to this rule, areas 8 and 10, use 230 kV as a system-wide voltage, while the rest of the areas use a voltage below 200 kV for a system-wide network. Synthetic networks could be designed in this way, in which case substations with 230 kV would fall in the lower category rather than the upper one.

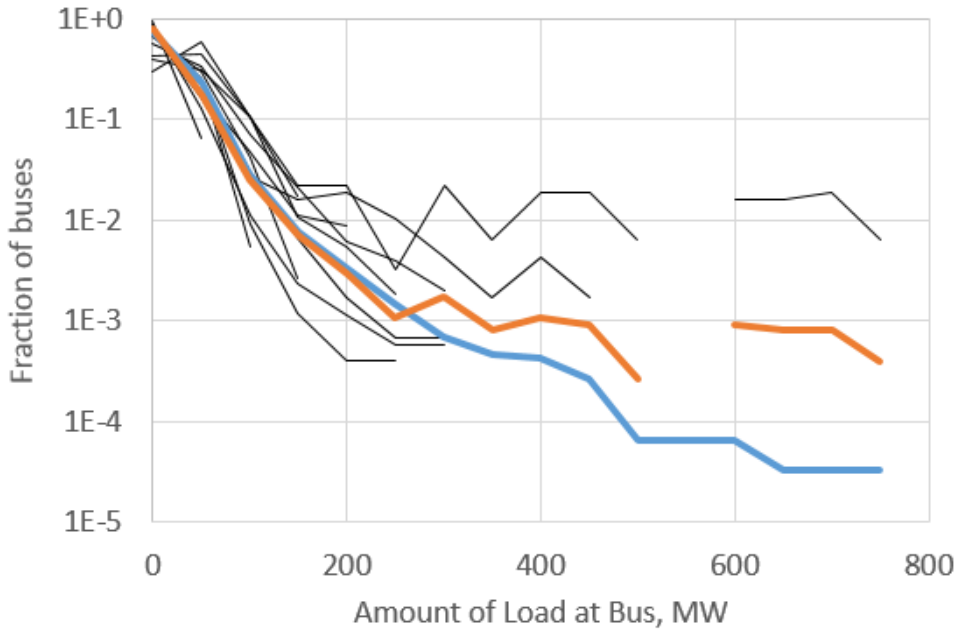


Figure 16. Amount of load at load buses. Probability mass function, for EI (blue), WECC (orange), and 12 subset cases (black).

**Percentage of substations containing load.** Categorizing buses or substations as load, generating, or neither plays an important role in synthesis methods and relates to the core energy delivery purpose of power systems. Except for two cases, areas 5 and 9, all of those studied show 75-90% of substations containing load, as shown in Table 5. Load, of course, is an aggregation of sub-transmission, distribution, and customer-level circuits, which for these exceptions appears to be grouped at a higher level than for typical grid cases.

**Load at each bus.** The selected cases vary from about 6-18 MW of load per bus on average. This excludes a couple of cases, which, because of their large net import or export of power, are outliers. Synthetic networks are often designed as self-contained systems. This average metric is important because it indicates the

| Case | No. Buses | Percent of transmission substations with bus at 69-200 kV | Percent of transmission substations with bus at 201+ kV | Percent of sub-stations with load | Percent of substations with generation | Ratio of generation capacity to load |
|---|---|---|---|---|---|---|
| EI | 62,605 | 93% | 15% | 87% | 11% | 1.35 |
| WECC | 20,131 | 89% | 22% | 76% | 17% | 1.56 |
| Area 1 | 4,939 | 99% | 7% | 88% | 4% | 1.19 |
| Area 2 | 1,505 | 93% | 21% | 79% | 14% | 1.28 |
| Area 3 | 3,363 | 97% | 13% | 81% | 28% | 1.37 |
| Area 4 | 693 | 97% | 8% | 90% | 8% | 1.54 |
| Area 5 | 4,013 | 94% | 15% | 79% | 10% | 2.04 |
| Area 6 | 434 | 98% | 13% | 89% | 18% | 1.33 |
| Area 7 | 2,762 | 96% | 12% | 83% | 29% | 1.49 |
| Area 8 | 768 | 56% | 67% | 88% | 15% | 0.87 |
| Area 9 | 3,266 | 87% | 21% | 67% | 22% | 1.45 |
| Area 10 | 1,453 | 73% | 38% | 59% | 39% | 1.28 |
| Area 11 | 4,322 | 90% | 19% | 90% | 4% | 1.33 |
| Area 12 | 1,885 | 98% | 7% | 90% | 9% | 1.25 |

Table 5. Statistics on case substations: voltage levels, load, and generation.

relationship between the size of a network in buses and the amount of peak load it serves.

Figure 16 shows the distribution of bus loads, for buses which have at least one load. The distribution varies widely, depending on the aggregation decisions used to model the loads at each bus. However, all cases show a large number of smaller loads, with a smaller percentage of larger ones. This distribution should be met in synthetic cases.

**Ratio of total generation capacity to total load.** The EI and WECC and their sub-regions generally have 20-60% more generation capacity than the peak load, as shown in the rightmost column of Table 5. There are two exceptions, one which imports lots of power and has 12% less generation capacity than total load, and one

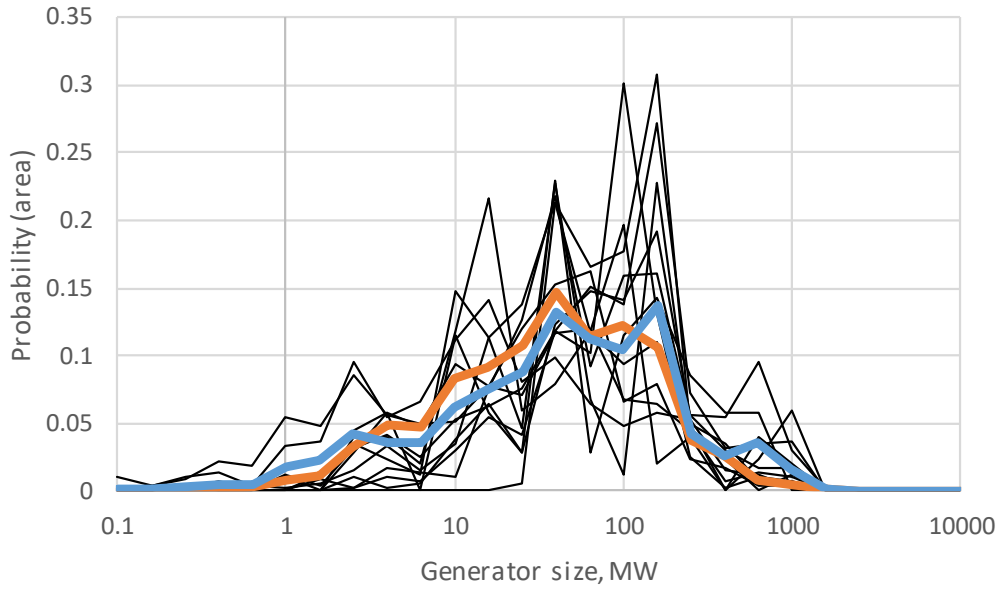


Figure 17. Generator capacities. Probability density, for EI (blue), WECC (orange), and 12 subset cases (black).

which has 104% more capacity as it exports a lot of power. The other cases fall within the realistic range of 20-60% capacity surplus. For any self-contained system, this metric should be almost inviolable.

**Percent of substations containing generation.** In the EI, 11% of substations contain generation, and in the WECC, the proportion is 16%. The values are shown in the fifth column of Table 5, which suggest that synthetic cases should contain generation in 5-25% of substations.

**Capacities of generators.** The selected cases consistently contain a wide variety of generator MW capacities, and it is important for synthetic cases to contain not only the correct total and average generation, but the spectrum of generator sizes real cases contain. Figure 17 shows these cases, with the range of 25 MW to 200 MW being the most common range for all cases, and most cases containing a few generators

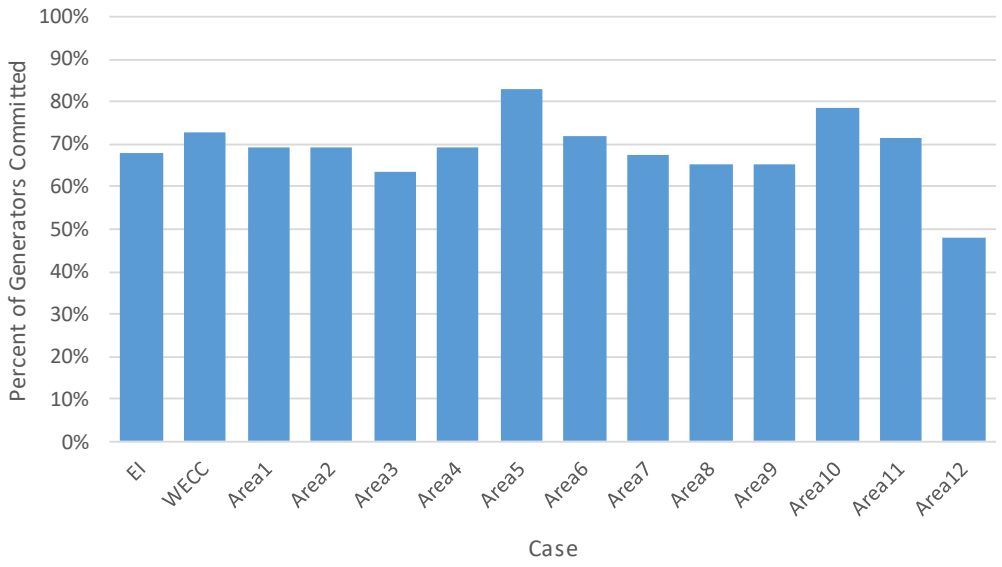


Figure 18. Fraction of committed generators for cases and sub-cases studied**.**

larger than 200 MW. Below 25 MW, the modeling varies. Some cases include a sizeable set of small generators, while at least a few areas largely ignore or aggregate them.

**Percent of generators committed.** The percent of generating units which are committed, that is, connected to the grid and generating power positive active power, is an important metric of the reserves and economics of the generation fleet. As shown in Figure 18, this value is 60-80% for most of the real cases considered.

**Generator dispatch percentage.** Most committed generators are operated close to their maximum MW capacity. This is especially true in certain cases in the EI. One defined criterion is that at least 50% of generators should be dispatched above 80%. The wide variation in this parameter is illustrated by Figure 19, where the EI and WECC are shown to have diverging distributions. Nevertheless, they share the characteristics that the majority of generators are operated close to their MW limit.

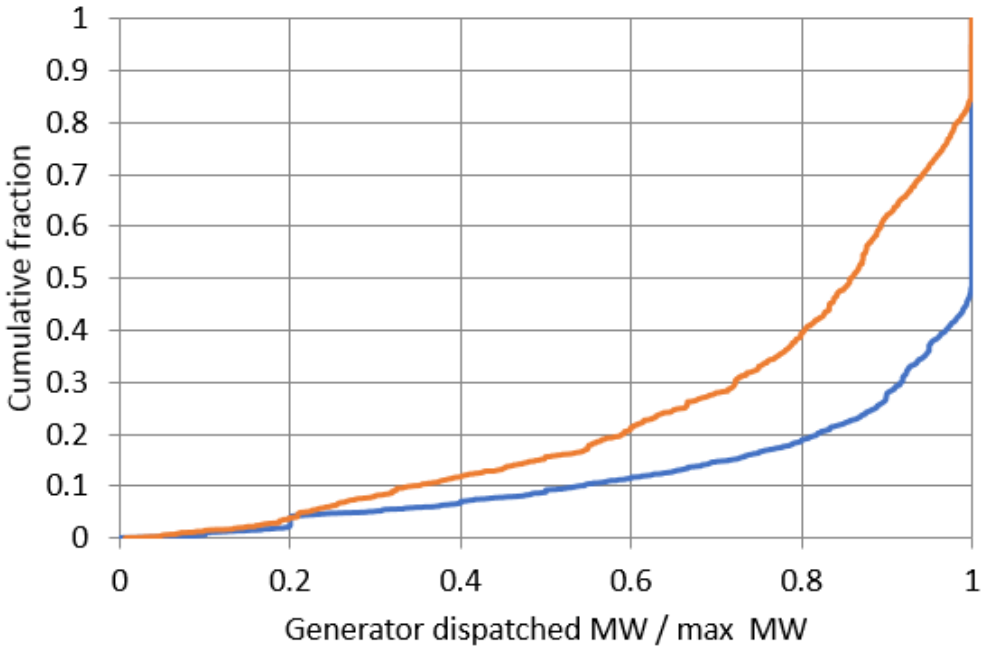


Figure 19. Cumulative fraction plot of generator dispatch percentage. Results shown for EI (blue) and WECC (orange).

This is an operational parameter that will change over time, but the focus is on the planning case values.

**Generator reactive power limits.** Generators' ratio between maximum reactive power limit and maximum active power limit, MaxQ/MaxP, shows the relationship between the size of a generator and how much voltage support it can give. This parameter also has a wide range of variety, since in actuality these are approximations for the capability curves, since reactive power limits are not the same at each active power operating point. As a basic qualification that seems to meet the data in real cases, for at least 70% of generators, that ratio of maximum reactive power to maximum active power should be between 0.40 and 0.55.

**Metrics of System Network: Transformers and Transmission Lines.**

**Transformer per-unit reactance.** Transformer reactance X is evaluated on the

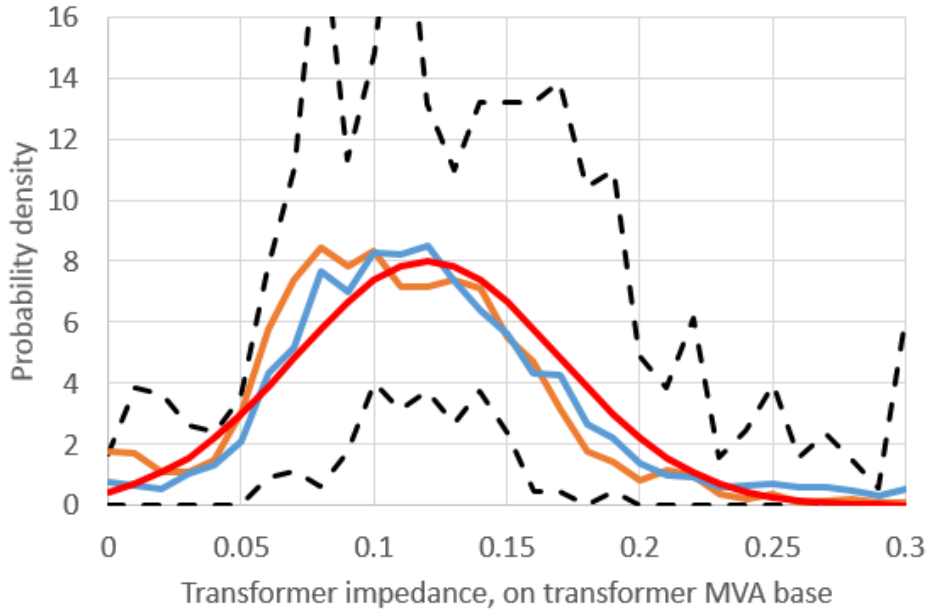


Figure 20. Transformer reactance. Probability density, for EI (blue), WECC (orange), and envelope of 12 subset cases (black), and normal fit (red).

transformer MVA base $\boldsymbol{S_B^{Txf}}$, which is related to the $\boldsymbol{X_{p.u.}}$ value used in the power flow case by the formula

$$\boldsymbol{X_{p.u.}^{Txf} = X_{p.u.} \cdot \frac{S_B^{Txf}}{S_B}}. \quad (12)$$

Analysis shows in the transformer reactance parameters a rather consistent distribution, when viewed in per-unit on the transformer base values. Figure. 20 shows the density functions for each case, along with a normal fit. It is typical for at least 80% of transformers to have a reactance value in the range [0.05, 0.2], and the distribution is roughly normal, centered around 0.12, with some variation as shown in the figure.

**Transformer MVA limit and X/R ratio.** Transformer MVA limit and X/R ratio statistics include outliers for large cases, because R and MVA limits for transformers are not absolutely essential to power flow studies. Sometimes a default

| | High Voltage Level (kV) | MVA Limit | | | X/R Ratio | | |
|---|---|---|---|---|---|---|---|
| | | 10% | Median | 90% | 10% | Median | 90% |
| EI | 69 | 10 | 42 | 115 | 10 | 20 | 50 |
| | 115 | 22 | 53 | 140 | 16 | 25 | 48 |
| | 138 | 39 | 83 | 239 | 19 | 30 | 54 |
| | 161 | 48 | 100 | 276 | 18 | 32 | 68 |
| | 230 | 63 | 203 | 470 | 25 | 44 | 84 |
| | 345 | 200 | 444 | 702 | 35 | 60 | 157 |
| | 500 | 215 | 812 | 1383 | 44 | 70 | 119 |
| WECC | 69 | 7 | 26 | 83 | 10 | 20 | 37 |
| | 115 | 17 | 37 | 118 | 15 | 25 | 50 |
| | 138 | 15 | 35 | 90 | 18 | 25 | 38 |
| | 161 | 30 | 63 | 125 | 19 | 27 | 46 |
| | 230 | 50 | 162 | 304 | 21 | 37 | 79 |
| | 345 | 160 | 336 | 672 | 33 | 59 | 139 |
| | 500 | 150 | 600 | 1233 | 32 | 70 | 140 |

Table 6. Transformer MVA limit and X/R statistics.

| Voltage Level (kV) | 90% | Median | 10% |
|---|---|---|---|
| 500 | 0.000210 | 0.000155 | 0.000121 |
| 345 | 0.000518 | 0.000360 | 0.000198 |
| 230 | 0.001550 | 0.000945 | 0.000343 |
| 161 | 0.003780 | 0.001828 | 0.000517 |
| 138 | 0.006295 | 0.002471 | 0.000596 |
| 115 | 0.006387 | 0.003398 | 0.000796 |

Table 7. Transmission line per-km, per-unit X, for EI.

small R value is used, so that the X/R ratio appears to be 10000 or more, which is unlikely to be accurate. However, for many transformers the data is reliable.

It is found that the transformer high voltage level is well correlated with both of these characteristics. Thus the analysis is printed in Table 6 organized by voltage

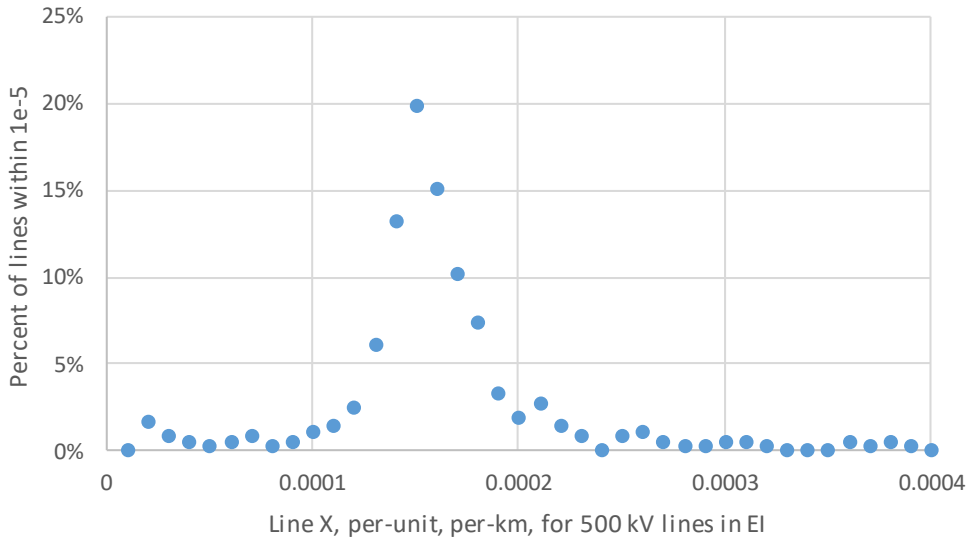


Figure 21. Discrete probability transmission line impedance characteristics. 500 kV lines in the EI are shown.

level for both the EI and WECC. The main objective is to see the common range of values for each level of transformer.

The validation criteria for MVA limit and X/R ratio are based on the median value, as well as the 10th and 90th percentile values. Cases should have at least 80% of transformer values within the 10th and 90th percentiles, and at least 40% above and 40% below the median. The less constrained of the EI and WECC values can be used.

**Transmission line reactance.** Transmission line parameters are organized by voltage level, since many aspects of transmission line design depend on the voltage level. The per-unit reactance depends heavily on the length of the transmission line, which, while not available exactly, can be approximated from the geographic distance between the two substations it connects. This distance will always be shorter than the actual right-of-way length, but serves as an approximation, especially for longer lines.

| Voltage | X/R Ratio | | | MVA Limit | | |
|---|---|---|---|---|---|---|
| Level (kV) | 90% | Median | 10% | 90% | Median | 10% |
| 500 | 26.0 | 17.0 | 11.0 | 3464 | 2598 | 1732 |
| 345 | 16.0 | 12.0 | 9.0 | 1494 | 1195 | 897 |
| 230 | 12.5 | 9.0 | 6.4 | 797 | 541 | 327 |
| 161 | 10.0 | 6.0 | 4.1 | 410 | 265 | 176 |
| 138 | 9.1 | 5.7 | 3.0 | 344 | 223 | 141 |
| 115 | 8.3 | 4.6 | 2.5 | 255 | 160 | 92 |

Table 8. Transmission line X/R ratio and MVA limit, for EI.

Transmission line per-km impedances at a certain nominal voltage level typically have a unimodal distribution with heavy tails corresponding to outliers, as shown in Figure 21. Some of the outliers may be due to smaller transmission lines for which the per-distance metric is less accurate. Similar to the transformer parameters, the transmission line statistics used are the 10th percentile, the 50th percentile (median) and the 90th percentile. This encompasses most transmission lines. Table 7 shows these percentages. Data on the distribution of transmission line parameters is also significantly impacted by the number of conductors bundled together in a phase, with 2- and 3- conductor bundling reducing the 345 and 500 kV lines.

**Transmission line X/R ratio and MVA limit.** In the same way, the 10/50/90 percentiles were calculated for transmission line X/R ratio and MVA limit, for major voltage levels, as shown in Table 8. These statistics do not consider transmission lines whose R values or MVA limits are not given. It is noticeable how narrow the 10-90 window is in each statistic, indicating the relatively consistent range in which realistic line parameters fall. The rule-of-thumb for validation, allowing for some variability, is for at least 70% of lines to fall inside the 10-90 window. Synthetic

transmission lines are also validated during construction if they are synthesized from actual conductors and tower configurations, as described in the last section and done for synthesized cases.

### 4.2 Network topology structure

The next set of metrics relate to the most-studied aspect of power grid synthesis: the transmission line topology. While the complex network literature has approximated the topology analysis with random models such as small-world [4], [7], [9]-[10], others have discussed the limitations of such a model because of its deviations from node distribution and its highly-designed, static topological nature [6], [14], [12].

It is important to define how the power system is viewed as a graph. Because bus modeling, aggregating circuit nodes, can vary within a substation and be more dependent upon breaker configuration, the focus is on substation topology, where substations are the graph vertices and actual transmission lines connecting different substations are the edges. Since there is a special distinction and connectivity limitation between branches of different nominal voltage levels, most of the transmission line topology statistics are also based on individual networks at a single nominal voltage level. Statistics were created by dividing the studied cases into their line topologies, using substations as the graph vertices at 115 kV, 230 kV, 345 kV, 500 kV, etc.

Geography plays a major role in the approach to validating the network topology, since power systems are strongly geographically constrained. Based on the substation geographic coordinates, the geometric minimum spanning tree and

| Case | Largest network 201+ kV | | Largest network 115-200 kV | |
|---|---|---|---|---|
| | Lines / Substations | Line Length / MST | Lines / Substations | Line Length / MST |
| Area 1 | 1.26 | 2.07 | 1.41 | 2.57 |
| Area 2 | 1.29 | 2.49 | 1.25 | 1.84 |
| Area 3 | 1.18 | 1.64 | 1.24 | 2.03 |
| Area 4 | -- | -- | 1.21 | 1.95 |
| Area 5 | 1.21 | 1.99 | 1.20 | 1.70 |
| Area 6 | -- | -- | 1.15 | 1.43 |
| Area 7 | 1.32 | 2.37 | 1.27 | 2.16 |
| Area 8 | 1.16 | 1.69 | -- | -- |
| Area 9 | 1.41 | 2.98 | 1.26 | 2.07 |
| Area 10 | 1.36 | 2.12 | 1.3 | 1.84 |
| Area 11 | 1.2 | 1.85 | 1.21 | 1.83 |
| Area 12 | 1.2 | 1.81 | 1.28 | 2.17 |

Table 9. Ratio of lines to substations and length to minimum spanning tree.

Delaunay triangulation are used to capture the near-neighbor effects of this constraint on the system's topological structure.

**Ratio of transmission lines to substations, at a single nominal voltage level.** The first fundamental statistic, ratio of lines to substations, is measured for grids at a certain nominal voltage level, and expresses the expected number of transmission lines present, given the number of substations containing the voltage level. This topological metric encompasses the density and redundancy of the graph, as well as average nodal degree. For actual cases, this was evaluated by looking at subset networks with at least 50 substations at voltage levels of 115 kV and higher, as shown in Table 9. The result was that all networks fall roughly in the range of 1.1-1.4 for the ratio of transmission lines to substations.

**Percent of lines on the minimum spanning tree.** The Euclidian minimum spanning tree (MST) is the minimum distance graph which connects all substations at a voltage level. This statistic, along with the following Delaunay triangulation statistics, helps to capture the geographic constraints of transmission line networks. Reference [30] shows the fraction of actual lines which come from MST, Delaunay, and Delaunay neighbors in EI and WECC, with the MST percentage around 50%.

**Distance of transmission lines along the Delaunay triangulation.** The Delaunay triangulation is calculated from a set of coordinates, dividing the plane into triangles, in which no triangle's circumcircle contains another point. As shown in [29], which appears to be the first application of this technique to power grid synthesis, most transmission lines have a very short distance along it, and this is an excellent metric of the geographic constraints of transmission line topologies. This reference shows about 75% of lines are on their Delaunay triangulation, about 20% are second neighbors, and about 5% are third neighbors. The number of lines that are fourth neighbors and higher is consistently below 1%.

There are a variety of topology-related graph theory statistics, including the distribution of nodal degrees, clustering coefficient, and average shortest path length, for which transmission networks have distinguished characteristics that have been explored in previous work [5]-[15]. References [29]-[30] have shown that matching the Delaunay triangulation statistics often encompasses the key graph characteristics observed on actual cases, in addition to respecting the geographic constraints of power grids, since transmission lines in general connect nearby substations.

**Ratio of total length of all lines to the length of the minimum spanning tree.** This metric compares line length at a nominal voltage level to the minimum length needed to connect all the substations, i.e. the length of the minimum spanning tree. These values are shown in Table 9. For networks above 100 kV and larger than 50 substations, most have this ratio between 1.4 and 2.6. In addition to the relative consistency in this ratio, the driving intuition is that it measures the relationship between the actual size of a power grid and the theoretical geographic minimum required.

### 4.3 Power flow solvability and complexities

This section discusses validation metrics that are of additional importance when large system ac power flow solutions are involved, especially with regard to voltage and reactive power. The distribution of voltage magnitudes is an important consideration as synthetic grids move from a dc power flow, where all voltages are assumed unity, to an ac power flow. Load voltage is important because that is the delivery point of power, with the main objective being to keep the voltage close to nominal, higher, and flat across the system.

Voltage objectives in base and contingency conditions are met in grids by adding and adjusting shunt reactive power devices, such as capacitors, reactors, static var compensators (SVCs) and synchronous condensers. This work focuses on capacitors and reactors, which are most common, however the method is general and could be applied to other devices as well. Additional validation metrics are given for generator reactive power limits, tap-changing transformers, and phase-shifting

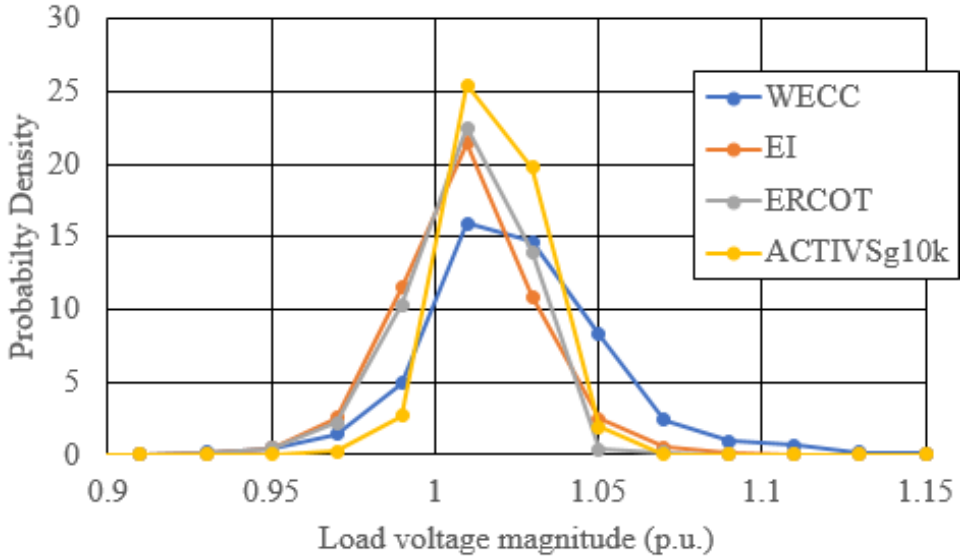


Figure 22. Load bus voltage magnitude distribution. Data shown for the synthetic 10k case (yellow) compared to the distribution found in actual cases. The EI, WECC, and ERCOT cases studied are all summer peak base planning cases, from 2012, 2013, and 2016, respectively. Although the voltage profiles can change throughout the year, the 10k synthetic case is designed to be a summer peak planning case, so it is compared to these.

transformers. In addition to these, aspects of power system solvability such as regions of convergence, impact of generator limits, and complexities such as impedance correction tables will be investigated.

**Voltage magnitudes.** The distribution of voltage magnitudes is an important consideration as synthetic grids move from a dc power flow, where all voltages are assumed unity, to an ac power flow. Load voltage is important because that is the delivery point of power, with the main objective being to keep the voltage close to nominal, higher, and flat across the system. These are exactly the characteristics observed in Figure 22 for the distributions of nominal voltage magnitude for load buses in three major North American interconnects, matched by the profile of the

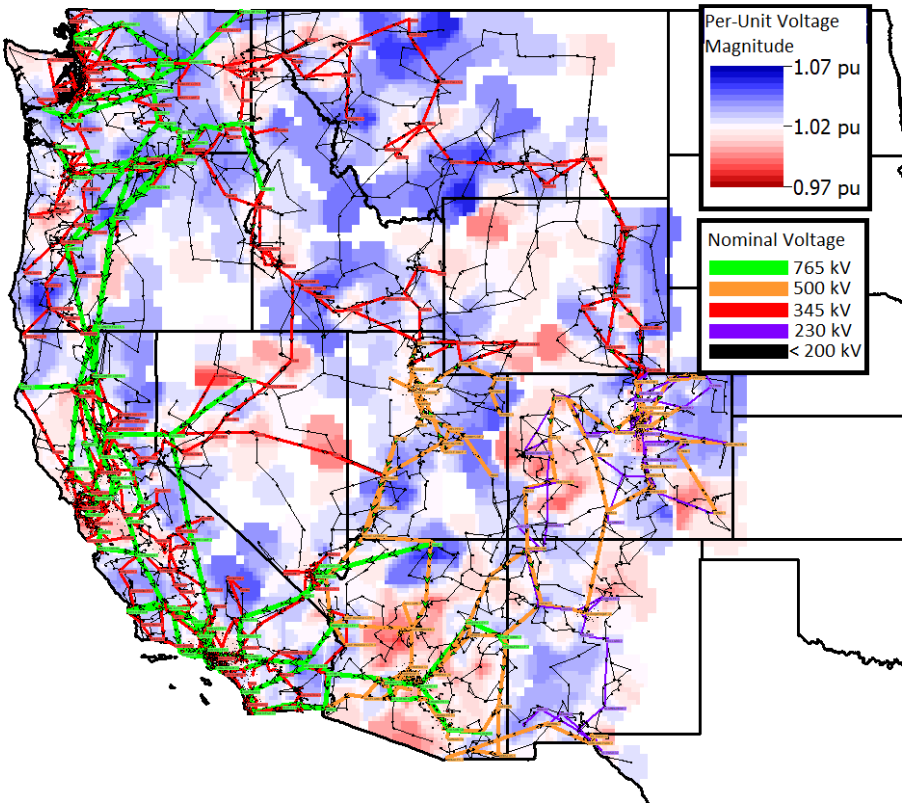


Figure 23. Voltage magnitude contour for the 10K case. The contour shows geographically the distribution of Figure 22, with load bus voltages varying in the range [0.97-1.07].

synthetic 10k grid. This distribution is also seen geographically on the one-line diagram of Figure 23, using a contour. Table 10 gives statistics of the distribution, showing the mean and standard deviation, which vary significantly among real interconnects.

**Shunt compensating devices.** Voltage objectives in base and contingency conditions are met in grids by adding and adjusting shunt reactive power devices, such as capacitors, reactors, static var compensators (SVCs) and synchronous condensers. This section focuses on capacitors and reactors, which are most common, however the method is general and could be applied to other devices as well, simply by converting the remaining temporary generators to the appropriate actual reactive

| Case | No. Buses | Voltage magnitude mean (p.u.) | Voltage magnitude standard deviation |
|---|---|---|---|
| EI | 62650 | 1.009 | 0.038 |
| WECC | 20131 | 1.014 | 0.105 |
| ERCOT | 7003 | 1.010 | 0.034 |
| Synthetic 10k | 10000 | 1.020 | 0.012 |

Table 10. Statistics of load bus voltage magnitude distribution.

| Case | No. Buses | % All Subs with Shunts | % Subs > 200 kV with Shunts | % Subs > 300 kV with Shunts | % Subs > 400 kV with Shunts |
|---|---|---|---|---|---|
| EI | 62650 | 22 | 36 | 46 | 59 |
| WECC | 20131 | 17 | 41 | 66 | 65 |
| Synthetic 10k | 10000 | 8 | 32 | 32 | 63 |

Table 11. Shunt reactive power devices by substation.

power source or sink. With a substation-oriented approach, the statistics given in Table 11 for actual cases list what percentage of a case's substations have a capacitor or reactor modeled. Table 11 also shows the dependence upon voltage level of the expected percentage of substations with reactive power compensation.

As can be seen, EI and WECC have 22% and 17% of all substations containing reactive power devices. There will be localized differences in specific areas' needs, modeling detail, and engineering design philosophies. When it comes to nominal voltage level, the trend is that higher-voltage substations are considerably more likely to have reactive devices. The synthetic 10k case matches these properties. The total percentage is slightly lower than observed in EI and WECC, which allows for extra shunt devices to be added as needed for special-case situations or future developments.

**Generator voltage regulation and reactive power limits.** The reactive power capability of synchronous generators is defined by a capability curve, where the boundary of possible reactive power supply is defined by the limiting factor of stator and rotor thermal limits and stability limits. However, in power flow analysis it is common to specify only a reactive power maximum and minimum value. These values vary substantially; for the purposes of these synthetic networks they are assigned by fuel type as a fraction of active power capacity.

The voltage set point of a generator assumes a voltage regulating scheme where the generator is attempting to maintain a certain bus voltage to some value. According to [63] and [64], power flow typically models generators which have a step-up transformer (GSU) as regulating the high-voltage bus to which the GSU is connected. In the EI and WECC cases, at least half of generators which are connected to a bus with a nominal voltage of 10-60 kV regulate a bus other than their own terminal.

**Tap-changing and phase-shifting transformers.** Another important control device for system voltage is the under-load tap-changing transformer (LTC), as well as transformers which maintain a fixed off-nominal tap ratio that cannot be changed under load. For synthetic power grids, the transformer placement is done in the Substation planning; what is done in Stage 2 of the algorithm is to choose the devices that have off-nominal tap ratios and the subset of these which control the low-side voltage, plus the parameters on this control. Statistics on the prevalence of these two types of devices in actual cases is given in Table 12, including the matching values for the 10k system.

| Property | EI | WECC | Synthetic 10k |
|---|---|---|---|
| Percent of network transformers with off-nominal tap ratio | 68% | 66% | 63% |
| Percent of network transformers which are regulating low side voltage magnitude | 44% | 35% | 33% |
| Percent of network transformers which control active power with phase angle regulation | 1.6% | 0.8% | 0.4% |
| Percent of network transformers with impedance correction table | 7.1% | 0.7% | 0.4% |

Table 12. Network transformers off-nominal tap control.

**Contingency analysis.** The 10k system was tested in both base conditions and under 12,000 single element outage contingencies. In contingency conditions, the switched shunts were treated as discrete stepped devices, where first the solution solves with shunts in their base state, then shunts are able to switch discretely and iteratively to control the voltage at the device terminal.

Initially 300 contingency violations were found, either with transmission branch overloads or voltage out-of-range. Manual adjustments were made to address them, leading to a secure base case. Although contingency analysis was not included in the reactive power planning process, the system performed well in contingency conditions largely because of the heuristic point system that encourages more than one reactive power resource path from a bus.

| Metric | Actual Systems | | | Synthetic Systems | | |
|---|---|---|---|---|---|---|
| | EI | WECC | ERCOT | 70K | 20K | 5000 |
| $n$ | 36,187 | 9398 | 3827 | 34,999 | 9524 | 2941 |
| $\bar{d}$ | 2.61 | 2.58 | 2.61 | 2.74 | 2.67 | 2.71 |
| $\bar{c}$ | 0.044 | 0.058 | 0.032 | 0.048 | 0.034 | 0.031 |
| $\bar{\ell}$ | 29.2 | 18.9 | 14.2 | 36.7 | 20.3 | 13.8 |
| $\bar{b}$ | 0.083 | 0.21 | 0.40 | 0.11 | 0.22 | 0.50 |

Table 13. Summary of complex network properties. Each power flow dataset was pre-processed to identify the substations as a single vertices, connected by transmission lines. $n$ is the number of substation vertices. The average vertex degree $\bar{d}$ can be calculated as $\bar{d} = 2m/n$, where $m$ is the number of transmission line edges. The Watts-Strogatz clustering coefficient $\bar{c}$ is calculated by averaging, for each vertex, the fraction of possible connections between neighbors that actually exist [5]. Vertices with degree 1 are ignored for the purpose of calculating $\bar{c}$, since there are no possible interconnections between pairs of neighbors. The average shortest path length $\bar{\ell}$ is the average number of hops between any two pairs of vertices. The average betweenness centrality quantifies what percentage of these shortest paths pass through the average vertex.

### 4.4 Complex network analysis

To begin the complex network analysis of the real and synthetic datasets, a critical first decision is how the systems are viewed as a graph. This subtle choice may contribute significantly to the disagreements in literature over complex network properties. While the electrical circuit is a graph-based model, with circuit nodes (buses) connected by branch elements (edges), there are multiple ways to model the same system which may be electrically equivalent but are not topologically identical and thus may have different metric properties. It is common in planning cases, for example, to have a single substation bus for each voltage level, grouping all the substation connections together. But for operational cases, all elements might be modelled, with 20 times as many nodes and lots of low-impedance edges. If a branch

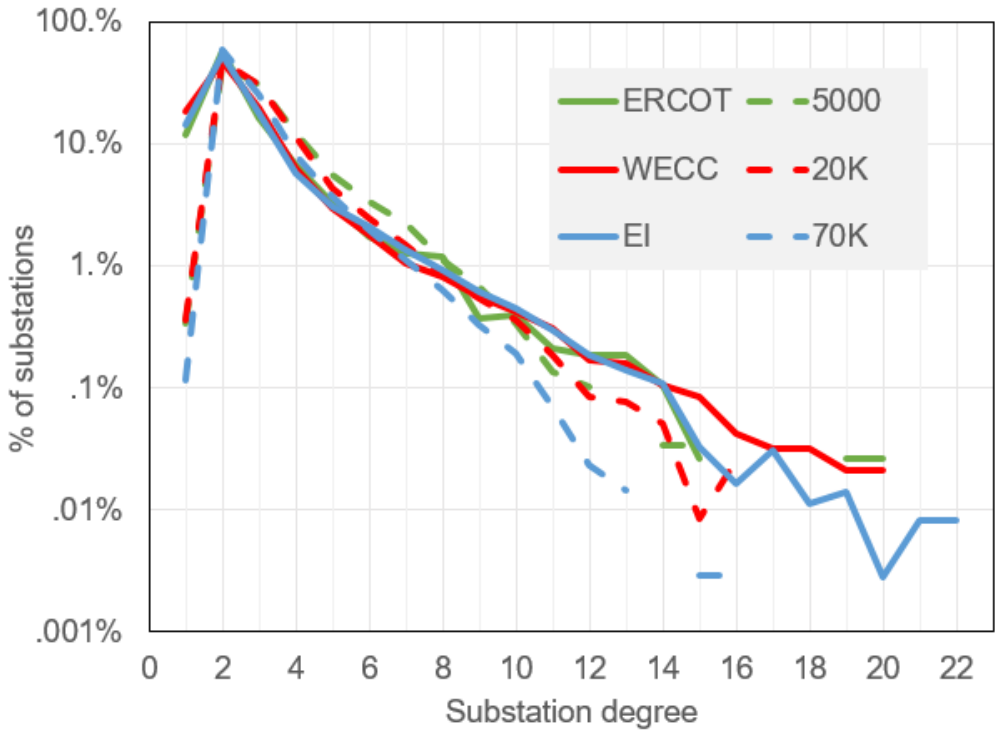


Figure 24. Degree distribution for real and synthetic grids. The graph shows the probability distribution function for the number of transmission line edges connecting to each substation vertex. Color indicates the size of the case; the solid lines are for actual grids and the dashed ones are for synthetic.

is divided into four segments, explicitly including elements such as breakers and switches, there will be many more edges with degree two, for example. Another important consideration is whether generators are modeled at the transmission bus or behind their own step-up transformer, which would add many radial vertices.

To minimize these concerns and get to the core of the power system structure, we consider each substation as an combined vertex, with edges being actual transmission lines that connect two substations in a single section. For the synthetic cases, the substation identities are known; for the actual data cases they must be inferred. While electrical and geographic metrics are certainly important and have been

considered in building the synthetic grids, for the purpose of this dissertation's analysis we focus only on the graph topology.

Summary metrics are given in Table 13, for the three actual system datasets and the corresponding synthetic datasets. The systems' substation vertices correspond to about two buses each on average. The average degree $\bar{d}$ is in effect a design parameter, since it relates the number of lines to the number of substations. But the actual systems show remarkable consistency, within 1% of an average degree of 2.6 for all three North American grids. This value fits comfortably in the range reported by literature, and validates a design choice for the synthetic grids: how many transmission lines can be placed to meet the other objectives.

The degree distribution has been frequently discussed, and we concur with some that an exponential distribution fits the data well, as shown in Figure 24, with the exception that there are fewer vertices with degree one (radial substations) than an exact exponential distribution would predict. While some have claimed that there is a similar scarcity of degree two vertices, these results show degree two vertices to be the most common kind. The results here confirm prior analysis that rejects the scale-free model for actual power grids, and show the similarity of synthetic grid data to the data for the actual grid.

The Watt-Strogatz small-world model focuses on the combination of the clustering coefficient and the shortest path length metrics. As Table 13 shows, the average clustering coefficients for all six systems are in the same neighborhood of 0.03 to 0.06. It is unclear the exact usefulness of this metric for power systems: for most

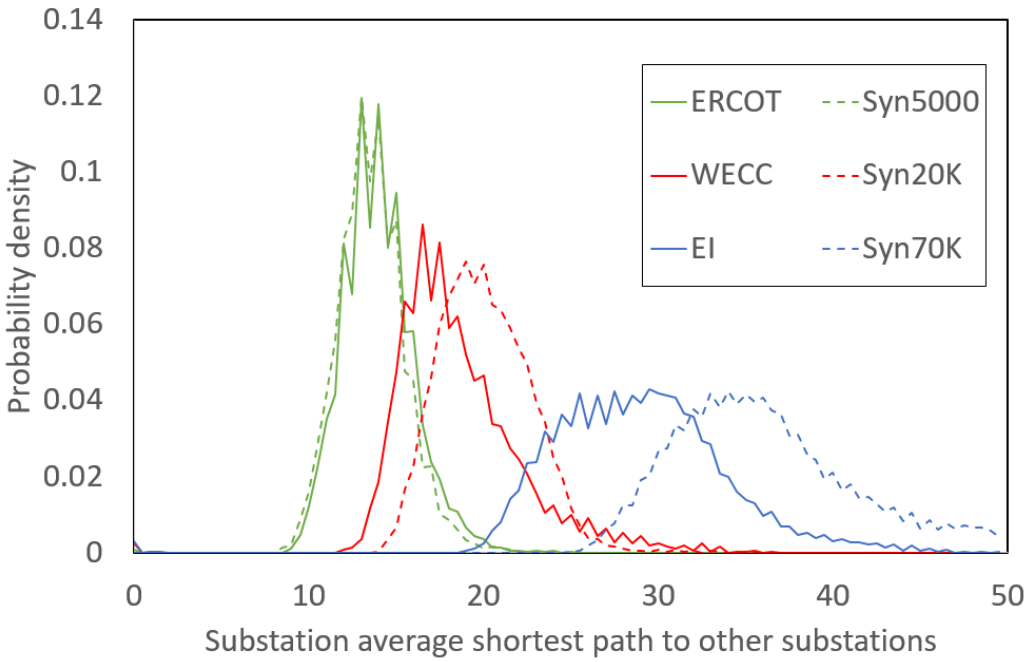


Figure 25. Distribution of substation average shortest path to other substations. The plots show the probability distribution function of the average shortest path between substation vertices, traveling along the combined substation graph. Color indicates the size of the case; the solid lines are for actual grids and the dashed ones are synthetic.

nodes the clustering coefficient is zero, with the exact magnitude depending largely on the few nodes whose neighbors are interconnected.

Shortest path length does scale with system size, but sub-linearly. The average shortest path length as shown in Table 13 remains very low even for the large EI case, qualitatively fitting the small-world idea. Figure 25 shows the distribution of shortest path length for nodes in the real and synthetic systems. The 5000 bus Texas model seems to match the distribution for ERCOT almost exactly, and both the other synthetic cases have slightly larger means than the corresponding real case. But the overall shape of all three distributions is similar. These differences in mean can be accounted for slight differences in the level of detail and modeling done: as opposed

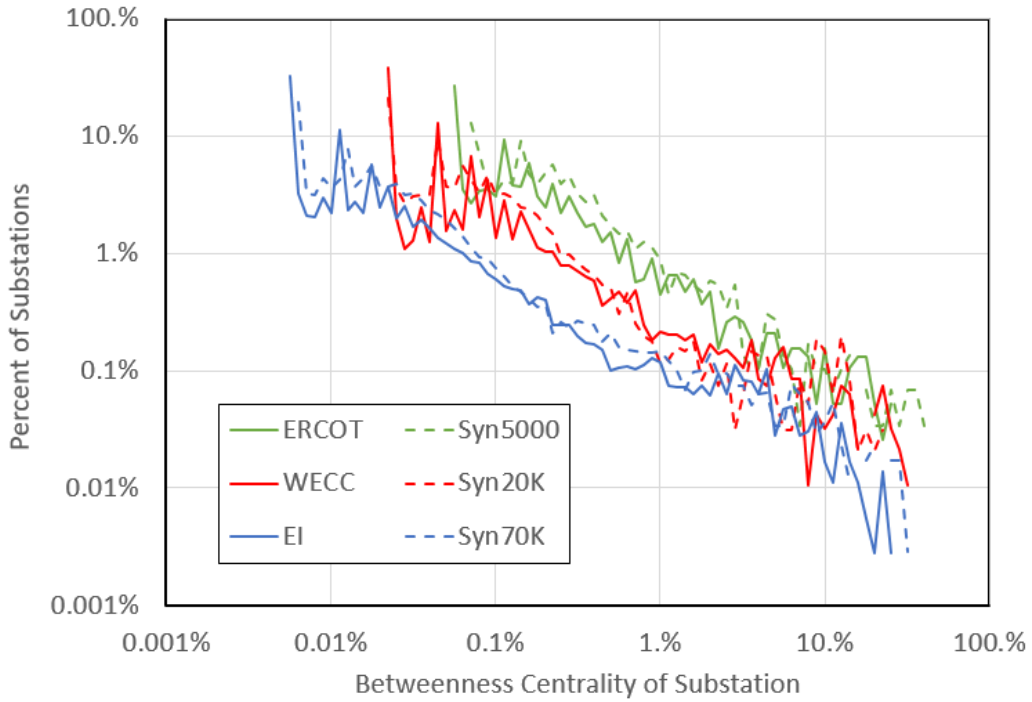


Figure 26. Distribution of betweenness centrality. The plots show the probability distribution function of the betweenness centrality metric. This metric is defined by the percentage of the shortest paths between each pair of nodes which passes through a given node, traveling along the combined substation graph. Color indicates the size of the case; the solid lines are for actual grids and the dashed ones are synthetic.

to the 70K synthetic case, the EI case has part of Canada modeled and has several areas with only the highest voltage grids modeled.

Betweenness centrality measures how central each node is by the shortest paths passing through it. There are $n^2$ possible combinations of nodes, and a node's betweenness centrality measures the percentage of these paths it includes. The distribution of this parameter shows the fraction of nodes that have various levels of centrality, as shown in Figure 26. These log-log plots show significant similarity between the synthetic and actual cases of each size. Power systems tend to have a small

number of substations which are very central topologically, while many nodes have much lower betweenness.

The analysis of this subsection shows that the complex network properties of degree distribution, cluster coefficient, average shortest path length, and betweenness centrality indicate strong similarities topologically between the synthetic 5000, 20K, and 70K cases and the actual cases with which they share a size and geographic footprint.

# 5. APPLICATIONS OF SYNTHETIC GRIDS*

Synthetic grids open up many opportunities in power systems research, since realistic power system data is available. This section combines a few supplemental research objectives which both support and are supported by research in building and validating synthetic grids.

## 5.1 Studies of geomagnetically induced current (GIC)

Geomagnetically induced current (GIC) is a result of low-frequency disturbances to the earth's magnetic field, caused either by solar storms or by a high-altitude nuclear blast. These geomagnetic disturbances (GMDs) and electromagnetic pulses (EMPs) induce quasi-dc GICs on power transmission systems which in extreme scenarios may damage equipment and cause voltage collapse on the grid.

Preparing for and mitigating the effects of geomagnetic disturbance (GMD) scenarios and associated geomagnetically induced currents (GICs) on electric power grids is in important aspect of grid resiliency, emphasized by recent regulatory agencies in the United States [68]-[70]. Developing synthetic, realistic test cases capable for GMD planning studies are beneficial to the research community for purposes of

* Parts of this section are reprinted, with permission, from A. B. Birchfield and T. J. Overbye, "Techniques for drawing geographic one-line diagrams: Substation spacing and line routing," *IEEE Transactions on Power Systems*, to appear, 2018. Copyright © 2018, IEEE.

Parts of this section are reprinted, with permission, from A. B. Birchfield, T. J. Overbye, and K. R. Davis, "Educational applications of large synthetic power grids," *IEEE Transactions on Power Systems*, to appear, 2018. Copyright © 2018, IEEE.

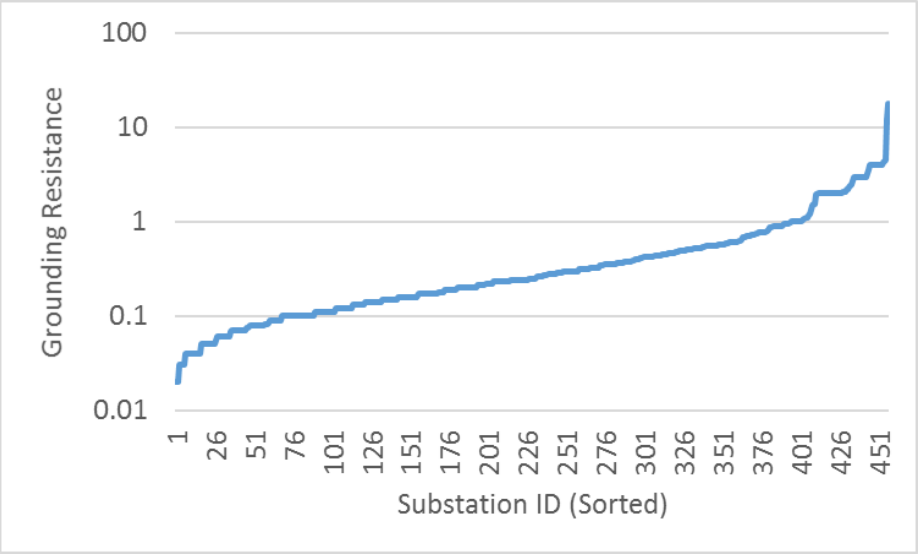


Figure 27. Sorted substation grounding resistance values, actual.

reproducing results and testing algorithms that can then be applied to actual grid models [71].

For simple studies, a spatially-uniform electric field can be appropriate, and this is allowed by the NERC planning guide [70]. The NERC reference event [72] is a commonly-used scenario, a time series based on the 1989 Quebec storm.

To aid in modeling these events, synthetic power grids can be augmented with the data necessary to perform the associated studies. Synthetic grids are particularly useful in studying EMPs due to the sensitive nature of that topic. GIC analysis requires substations be defined with substation coordinates, which are integral to the created synthetic grids anyway, making them ideally suited. The additional data which must be generated includes substation grounding resistance and transformer grounding configuration. With these additions, based on actual data statistics, GIC methodologies can be applied and tested on the synthetic grids.

| Substation grounding resistance, Ω | Percent of substations, actual case | Percent of substations, validation range |
|---|---|---|
| 0.01-0.05 | 5.7% | 0-10% |
| 0.06-0.10 | 13.8% | 5-20% |
| 0.11-0.20 | 22.8% | 10-30% |
| 0.21-0.50 | 29.5% | 15-40% |
| 0.51-1.00 | 15.5% | 10-25% |
| 1.01-5.00 | 12.3% | 5-20% |
| 5.01-20.00 | 0.4% | 0-5% |

Table 14. Substation grounding resistance, statistics and validation range.

The network synthesis method in this dissertation begins with a geographic footprint and substation grouping. Thus these values are intrinsic to the base case specifications. Transmission line lengths, as is typical, are approximated by the shortest path between the two substations it connects, although in actuality the true transmission line right of way is always longer than the straight-line distance.

**Substation grounding resistance.** Previous work has shown that substation grounding resistance typically falls in the range 0.01 to 2 Ω, and that GIC solutions can be highly sensitive to these values [73]-[75]. This data is rarely included with power flow data, and so it must be added to supplement the base case.

Statistics are derived from an actual power system test case which includes the substation grounding resistance for 457 substations in WECC. These are actual measured values. Substations have a wide range of values, as can be seen in Figure 27. While there is very minor correlation between the voltage levels, number of buses, and other characteristics and the grounding resistance, these factors are not sufficiently tied to use for approximating these values with any degree of realism. Allowing for a

| | Auto | Non-Auto |
|---|---|---|
| Secondary voltage >= 100 kV | 245 | 9 |
| Secondary voltage < 100 kV | 62 | 305 |

Table 15. Statistics on autotransformers.

reasonable range of tolerance, Table 14 represents some categories that should be represented by any synthetic set of substation grounding resistances.

.

**Autotransformers.** Autotransformers are not usually identified in standard power flow cases, but their modeling is important to GIC calculations. The statistics in this document are based on an actual GIC model in the Western Interconnect, where actual data is given for 621 transformers, of which 307 are known to be autotransformers, and 314 are known to not be autotransformers. It is found that the secondary voltage level is highly correlated with autotransformer status, as shown in Table 15. The validation criteria is that at least 75% of the first type of transformer should be autotransformers, whereas at least 75% of the second type should be non-auto.

**Transformer Configuration.** For non-auto transformers, the grounding configuration of the windings is important to GIC modeling, as a wye winding allows a path to ground for GICs while a delta winding does not. This data is also available for 305 of the known non-auto transformers. Of these, 75% have a wye-configured primary side and a delta-configured secondary side, which is typical for generator step-up transformers (GSUs) and some step-down transformers. 16% of the transformers have a wye-connected secondary and delta-connected primary, while 9% have wye-

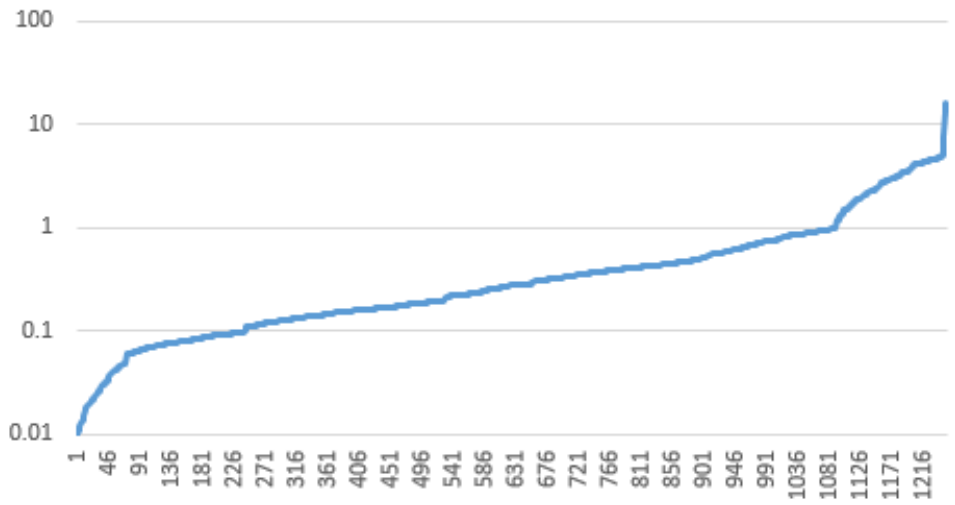


Figure 28. Ordered substation grounding in 2000 bus Texas case. Horizontal axis gives the ordered substation number, and the vertical axis gives the substation grounding resistance in Ohms on a logarithmic scale.

connected configurations on both windings. So the validation criteria is that at least 80% of the non-auto transformer should be wye-delta or delta-wye.

Transformer winding dc resistance values can be calculated from the ac resistance values using the formulas given in [73]. The ac resistance values are specified as part of the base case.

Figure 28 shows the ordered substation grounding resistance for the 2000 bus Texas case, formed by assigning from the distribution given in Table 14. The addition of GIC parameters to this case allows for example studies. Figure 29 shows the voltage magnitude contour under a 5 V/km electric field for the Texas case, and Figure 30 shows the 10,000 bus case under an EMP.

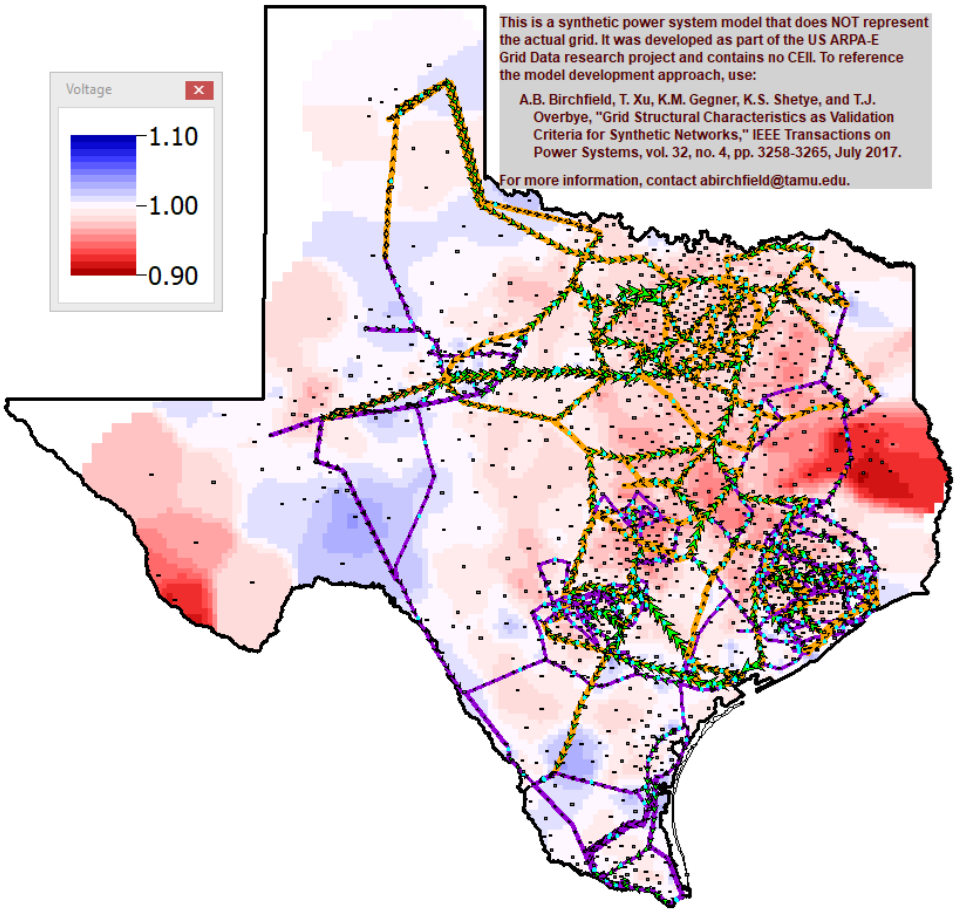


Figure 29. GMD voltage profile on 2000-bus Texas case. The voltage magnitude profile begins to sag under a 5 V/km electric field on the synthetic 2000-bus Texas case.

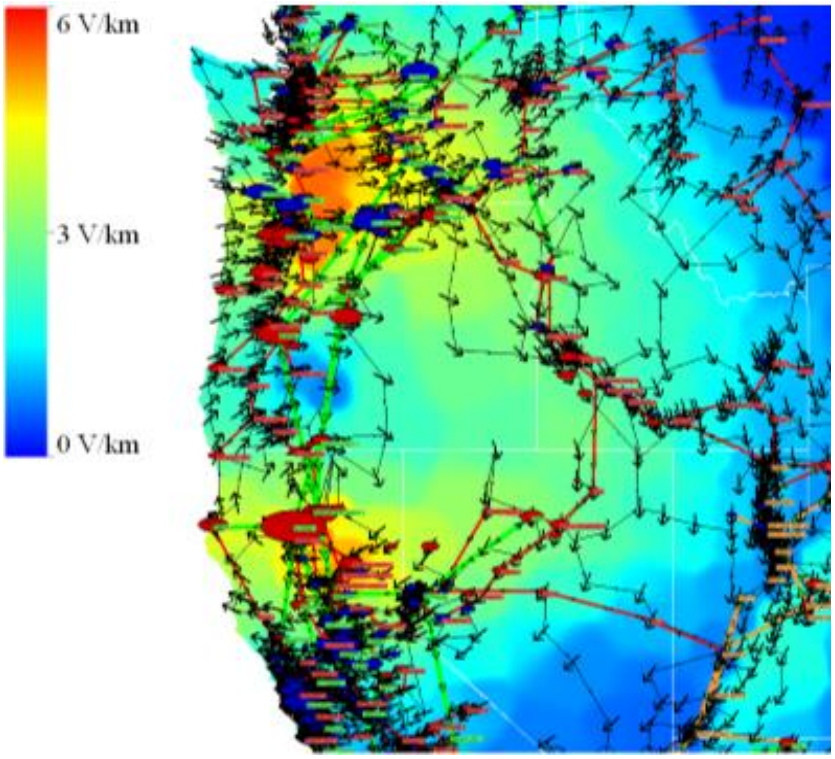


Figure 30. Example results from EMP study on the 10,000 bus synthetic grid. The contour shows the electromagnetic field magnitude varying spatially. Black arrows at each substation show the direction of the electromagnetic field. Red and blue ovals show the magnitude and direction of GIC induced in substation neutrals.

### 5.2 One-line diagrams and visualization platforms

Diagrams for electric power networks aid system analysis; they complement numerical data with a visual context. One-line diagrams are often carefully maintained through extensive labor, detailing system substations, generators, loads, and the transmission branches which connect them. Engineers studying a transmission system which has a high-quality system diagram have an additional platform to diagnose problems, skim large datasets for abnormalities, and communicate their results clearly. The target platform here is a computer-navigated diagram that can be zoomed and panned in planning and analysis software.

This section addresses the problem of quickly creating automated single line diagrams for large bus-branch power system models, with substations defined and geographically tagged. The present work employs graph drawing approaches in a geographic context to arrange substations and route transmission branches. The methodology is useful for a quick look at power flow cases which do not otherwise have a diagram, such as those from an energy management system or synthetic grid-making algorithm. It can also provide an excellent starting point for higher-grade diagrams of networks real or synthetic.

In drawing a single line diagram, physical correctness should balance logical function. The approach of this section reflects physical layout mainly in the substation positioning, keeping the substations drawn in a similar spatial arrangement as they are on a map. Latitude and longitude data is now becoming routinely available, driven in part by geomagnetic disturbance studies. The challenge addressed here is to implement

this balance in placing and sizing the substations, with enough room to show the substation configuration and interconnections and minimal distortion to the geographic positioning.

The second part of the section's methodology, the line routing algorithm, puts the system bus and branches in and between the drawn substations. This step does not consider the actual right-of-way positioning of the transmission lines, since the system functionality can be displayed well without it, and often this data is not available. The objective in this step is to route the branches to show the network topology in a way which is visually understandable and logically accurate. The approach builds on previous work, with a geometric-based method for transmission line branches which relies on the Delaunay triangulation.

**Background on power system single-line diagrams.** Ways to draw a graph are as diverse as the types of graphs there are to be drawn. From computer networking and social networking to life sciences and microelectronics, researchers of all sorts have graphs complex and informative, and they need them to be visualized [76]. The method depends on the properties of the graph, and involves trade-offs between the desired characteristics of the drawing and computational complexity [77].

The force-directed approach to graph drawing [78]-[79] shows how an analogy to the physics of particle interactions can be used to position nodes in a way that shows graphs aesthetically. These drawings show edges as straight lines and vertices as freely-mobile points that come to a visually pleasing equilibrium where edge distances are short but node spacing is comfortable. Extensions to this approach abound, for better computation [80], clustered graphs [81], graphs with many edges that should be

bundled [82] or spaced [83]. Methods for non-straight lines include using general segmented paths [84], curves [85], and orthogonal segmented paths [86]. Computational geometry concepts such as the Delaunay triangulation are sometimes used for geometrically-constrained methods [87]. With large graphs, interactive viewing that scales between levels [88] can be an appropriate strategy. Keeping aesthetics in mind, competing objectives for a graph drawing include symmetry, minimal crossing, and minimal bends [89]-[90].

Power system models are graphs of various kinds, depending on whether transmission or distribution is considered, whether the representation is bus-branch or node-breaker, and any equivalencing done in a particular case. Studies of the graph theory properties of power networks, evaluated some in earlier parts of this dissertation, show that power systems are sparse connected graphs, with an exponential node degree distribution averaging 2-3, a high clustering coefficient and a small topological diameter [91]-[94]. Recent work has documented that transmission lines have a short and consistent distribution of topological distance along the geographic Delaunay triangulation graph, since power systems are geographically constrained [95]-[96].

In power systems visualization, early work on one-line diagrams includes [97], which discusses the usefulness of diagrams of individual substations for evaluating the security of these substations. One approach is to automatically visualize the neighborhood of a bus of interest [98], to help a user get the context of the issue being addressed. One-line diagrams can be augmented with additional visualization techniques [99] to show line loading, power transfer distribution factors, and voltage

magnitude contours superimposed on the network. For node-breaker models, [100] and [101] propose methods to diagram substations automatically. For distribution systems, which are more radial in nature, and numerous, [102] proposes a layout technique that scales up well. Reference [103] points out the benefits of hierarchical diagramming with relative coordinates. Some work applies methods similar to the force-directed graph drawing approach to draw networks [104], including using geographic information to initialize the positioning [105]. Reference [106] discusses power-flow related graphical objectives that improve the utility of diagrams, and [107] uses a multidimensional scaling of electric distance metrics to visualize a network's power flow characteristics.

**Substation drawing and spacing: challenges in spacing and size.** Drawing substations as rectangles in their geographic location works for much of the system, but challenges come when substations are large enough and close enough together that they overlap. Substations are drawn much larger than they actually are, and to show rural substations without extensive zooming and whitespace requires a size that in urban areas results in a indiscernible mess.

The automated approaches discussed by this section determine the substation size and location to best display the functionality of the grid while keeping geographic context. That is, they maximize visual spacing between neighboring substations while minimizing distortion from their actual coordinates. Two approaches are discussed: one which is based on the force-directed graph drawing approach and one which uses a greedy approach.

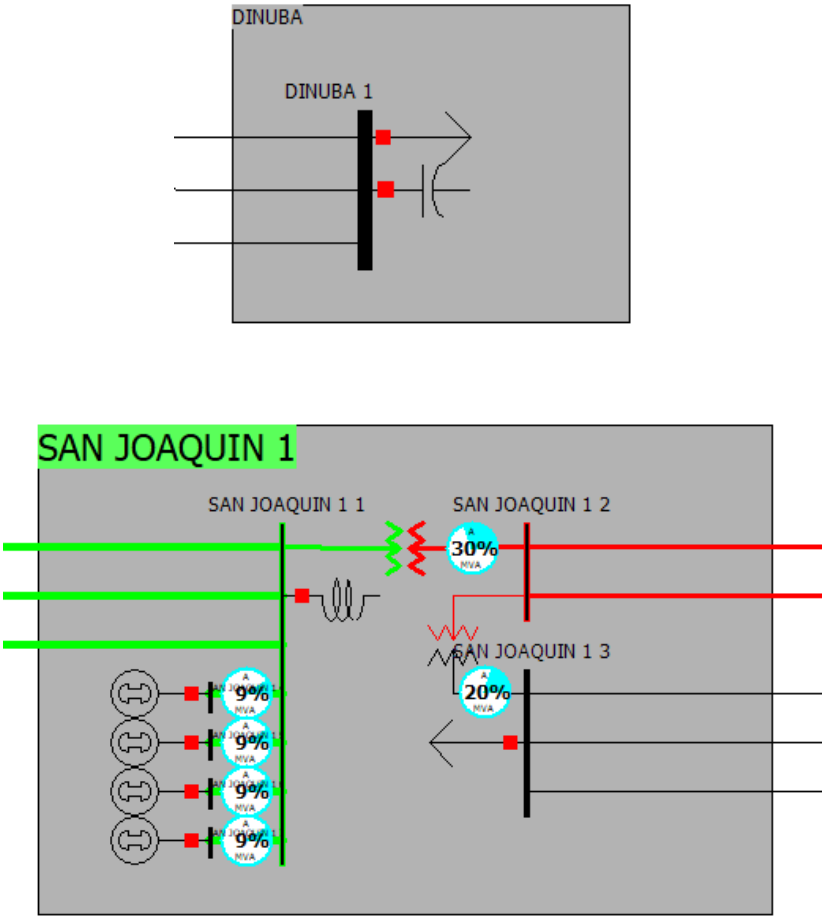


Figure 31. Example substation internal layouts from 10K case. The layout shows buses, lines, transformer, shunts, generators, and loads.

Note that substations are drawn on a coordinate system corresponding to a standard unit Mercator projection, where (0,0) is the intersection of the prime meridian with the equator, the longitude scales linearly to $\pm\mathbf{1}$ at the international date line, and the latitude scales non-linearly to $\pm\mathbf{1}$ at about the eighty-fifth parallels.

**Internal layout of substations.** The spacing algorithm must know as an input the dimensions of the substation, which depends on how the substation is laid out

internally. The substation will be scaled from there, but it is better to scale all or most of the substations uniformly so that the features appear the same size on a diagram. In addition, for routing the transmission lines in the next discussion, this internal layout will determine the starting points for entering and exiting the substation.

The bus-branch data for substation configuration is often aggregated from the underlying node-breaker model, so no attempt is made to show the spatial layout of the substation (this is not always available anyway). Rather, the logical layout is used. Thin rectangles are the typical symbol for buses, and lines of varying thickness and color show the branches, aggregating three phase conductors and any neutral wires. Branches which are transformers are shown symbolically, and loads, generators, and shunt devices are indicated with symbols attaching to the corresponding bus. These conventions are all retained in the present approach, with a box identifying the substation. This box contains all elements in the substation, with inter-substation branches crossing the boundary, as shown in Figure 31.

The internal drawing of each substation is decoupled; that is, each one can be done independently. Although a co-optimization with the line routing might improve the routing by determining, for example, which side of the substation a line ought to exit, such considerations can also make the internal configuration less clear. The present approach draws the buses vertically in one or two columns, with height proportional to the number of connections to that bus. They are placed in order of nominal voltage.

In adding elements to the substation diagram, first any external branches are put on the outward-facing connecting points. Then any internal branches are placed,

straight across if possible (such as the green-red branch in Figure 31), otherwise with two right angles (as the red-black branch in Figure 31). Single-port elements are placed last in remaining spots. Radial buses with nominal voltage below transmission levels, such as the low side of generator step-up transformers, can be placed in-line as an element of the bus to which they are attached, such as the generators in Figure 31.

**Force-directed approach to substation spacing.** The force-directed approach is a general method common to graph drawing problems that models each vertex as a particle subject to physical forces, which, when simulated to equilibrium, balance the various constraints desired for a graph drawing. The common basic formulation is to use a Coulomb-like inverse square repulsion force between vertex-particles to enforce spacing and a Hooke-like spring force attracting connected nodes. What is different about the present problem is that a reasonably good starting spot is known for each substation (its geographic point) and a key constraint is to minimize its displacement from that point. So the formulation proposed here is to keep the Coulomb repulsion and add a Hooke attraction between each point and its actual location. As is typical, static equilibrium is desired so the forces are modeled as inducing velocity rather than Newtonian acceleration. The algorithm steps are as follows:

Initialize each substation for $\boldsymbol{i = 1{:}N}$

$\boldsymbol{X_i = X_{i0}}$

$\boldsymbol{Y_i = Y_{i0}}$

Loop of $\boldsymbol{M}$ iterations

For $\boldsymbol{i = 1{:}N}$

$\boldsymbol{F_{Xi} = C_1 \cdot (X_{i0} - X_i)}$

$\boldsymbol{F_{Yi} = C_1 \cdot (Y_{i0} - Y_i)}$

For $\boldsymbol{i, j \in [1, N], i \neq j}$

$$\boldsymbol{F_{mag} = \frac{C_2}{\sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}}}$$

$\boldsymbol{F_{angle}} = \mathrm{ATan2}(\boldsymbol{Y_i - Y_j, X_i - X_j})$

$\boldsymbol{F_{xi} \mathrel{+}= F_{mag} \cdot \cos(F_{angle})}$

$\boldsymbol{F_{yi} \mathrel{+}= F_{mag} \cdot \sin(F_{angle})}$

For $\boldsymbol{i = 1{:}N}$

$\boldsymbol{X_i \mathrel{+}= C_3 \cdot F_{xi}}$

$\boldsymbol{Y_i \mathrel{+}= C_3 \cdot F_{yi}}$

In the equations above, there are $\boldsymbol{N}$ substations with properties for substation $\boldsymbol{i}$ of position ($\boldsymbol{X_i}$, $\boldsymbol{Y_i}$), original position ($\boldsymbol{X_{i0}}$, $\boldsymbol{Y_{i0}}$), and force ($\boldsymbol{F_{xi}}$, $\boldsymbol{F_{yi}}$). $\boldsymbol{F_{mag}}$ and $\boldsymbol{F_{angle}}$ are the magnitude and angle of the force. The parameters of the algorithm are $\boldsymbol{M}$, the number of iterations, $\boldsymbol{C_1}$, the Hooke constant, $\boldsymbol{C_2}$, the Coulomb constant, and $\boldsymbol{C_3}$, the mass constant. Experimentally, the algorithm was found to converge for about $\boldsymbol{M = 100}$, and good values for the constants were found to be $\boldsymbol{C_1 = R}$, $\boldsymbol{C_2 = R}$, and $\boldsymbol{C_3 = 0.1 \cdot R}$, where $\boldsymbol{R}$ is the desired spacing between substations, in the same units as $\boldsymbol{X}$ and $\boldsymbol{Y}$. To reduce the $\boldsymbol{N^2}$ computational order of the repulsion force checking, each particle need only check neighbors within a radius of $\boldsymbol{6R}$. Figures 32 and 33 illustrate the original spacing and results after the force-directed approach comes to an equilibrium.

**Multiple scales.** The size of substations shown in Figure 32 is good for most of the system, where substations are sparse, but for crowded areas (such as Figure 32

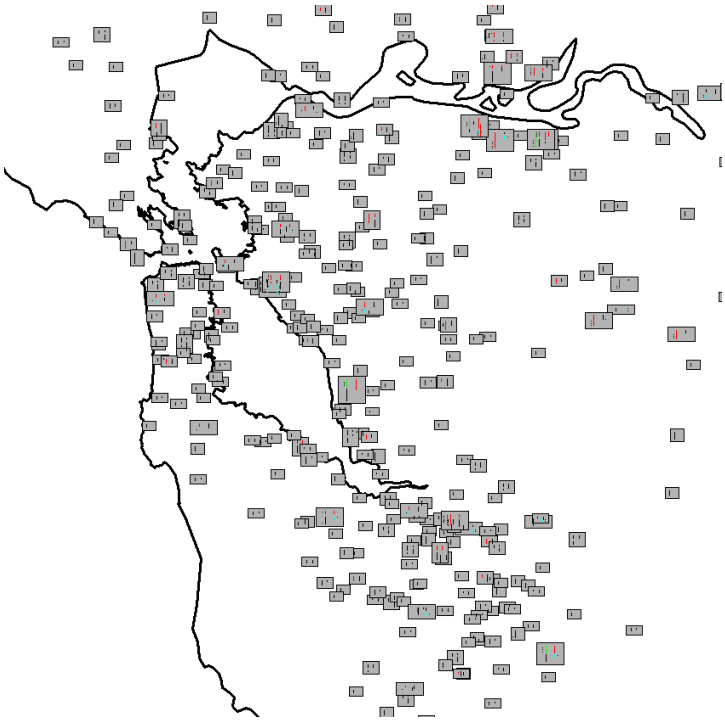

Figure 32. Original substation placement. Shown is a portion of the synthetic 10,000 bus case, with rectangles overlapping and crowded at many points. Black lines show outline of San Francisco Bay for context.

shows) the substation locations must be either significantly distorted in their location or shrunk to allow viewing them and routing lines between them. This section discusses a shrinking process in preparation for the greedy approach discussed below.

A substation's size is defined in blocks, where, for example, the substations of Figure 32 are sized 12-by-15 and 20-by-30 blocks. The scale of a substation defines how large a block is on the screen. The default scale is 1e-5, measured on the unit Mercator projection. This scale is such that at the latitude of San Francisco a 12-by-15 substation would cover about 1.2-by-1.5 km of map space. Two other scales are defined at 4e-6 and 1.5e-6, about two-thirds reduction in size each, for crowded parts of the grid.

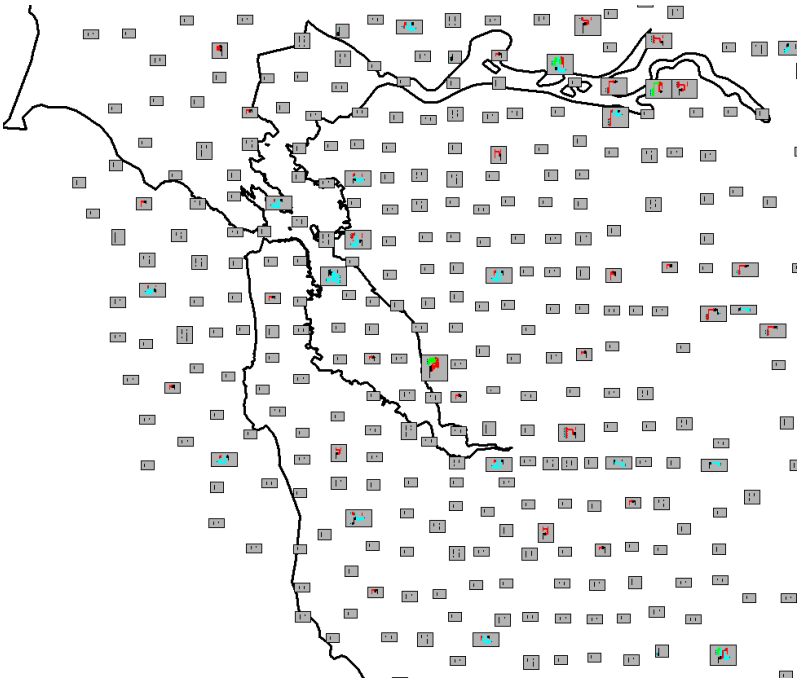

Figure 33. Force-directed approach to substation spacing.

The scale for each substation is determined by a simple check of the number of neighboring substations in a square radius of 25e-4 (on unit Mercator measure, approximately equal to 25-by-25 km), reducing to the second level if there are 30 substations in this area and to the third level if there are at least 200. Then the differences in size only appear at the intersection of the crowded regions and sparse regions. At an appropriate zoom level, similar elements will appear the same size within these regions. This method is general enough to be appropriate for various densities in transmission and sub-transmission modeling.

**Greedy approach to substation spacing.** Once an appropriate scale for each substation has been found, the following algorithm is used to place each substation. It takes the greedy approach of assigning a location for each substation closest to its

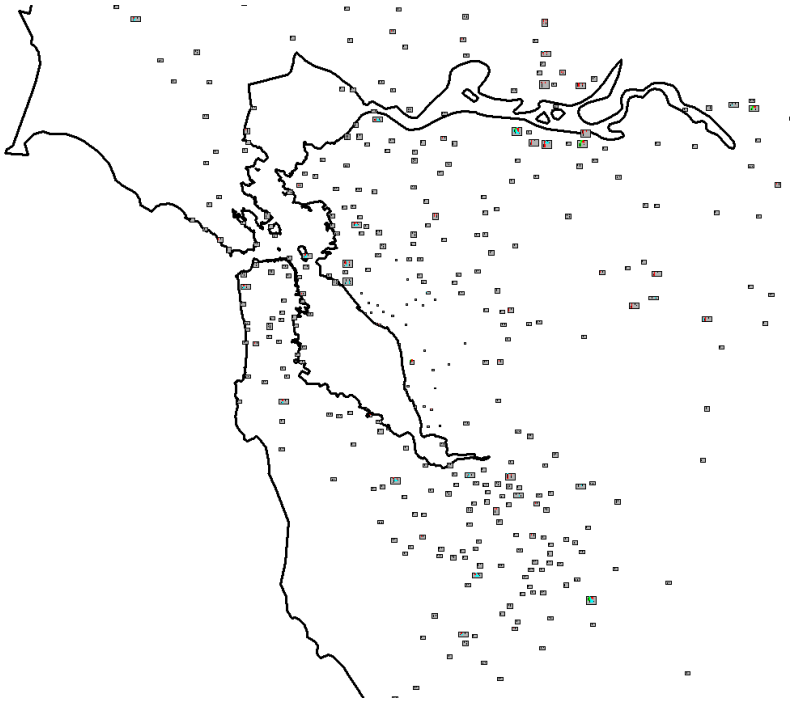

Figure 34. Greedy approach to substation spacing. While more sparse substations (outside the figure) were kept the original size, these were shrunk and slightly moved to reduce the crowding and maximize geographic context.

actual location without violating a buffer region around any of its neighbors that have already been placed.

For each substation for $\boldsymbol{i = 1{:}N}$
For $\boldsymbol{j, k \in [-10, 10]}$
$\boldsymbol{X_{iT} = X_{i0} + j \cdot R}$
$\boldsymbol{Y_{iT} = Y_{i0} + k \cdot R}$
If $(\boldsymbol{X_{iT}, Y_{iT}})$ intersects with spacing buffer of any already-placed substation
$\boldsymbol{(X_i, Y_i) = (X_{iT}, Y_{iT})}$ only if closer than existing and no feasible point found yet
Else continue
Else if $(\boldsymbol{X_{iT}, Y_{iT}})$ is closer to $(\boldsymbol{X_{i0}, Y_{i0}})$ than $(\boldsymbol{X_i, Y_i})$ is
$\boldsymbol{(X_i, Y_i) = (X_{iT}, Y_{iT})}$

In this algorithm, 121 points $(X_{iT}, Y_{iT})$ are tested for overlap with already-placed substations and the closest acceptable one to the original location is chosen. An example of the results is shown in Figure 34. This method has several advantages over the force-based approach, and is the one selected when continuing to the transmission line routing algorithm.

**Transmission Line Routing.** Goals and preliminary considerations. Given substations properly spaced with the greedy algorithm above, the next step is to draw the transmission lines which connect them. If straight lines are drawn to connect them, these lines can cross over substations and produce a web that is challenging to understand. The goals for line routing are to add segmented waypoints to avoid overlapping with substations and each other, while minimizing the increased length of the line and avoiding sharp bends. Figures 35-37 show examples of the straight-line approach and the Delaunay approaches explained below.

One heuristic that it used as a first step is applied when a line is headed the opposite direction as its exit from the origin substation. Since the line must route around its origin anyway, a first orthogonal corner is made to route the line north or south around the substation.

**Delaunay-based approach.** From computational geometry, the Delaunay triangulation is a planar graph which connects geometric points into nicely-shaped triangles, where nicely-shaped is defined as having the triangle's circumcircle empty of other points. This algorithm uses segments of the substation Delaunay triangulation as routing channels through which the spacing of the lines routes is managed and the lines avoid intersecting with the substations. The steps are given below:

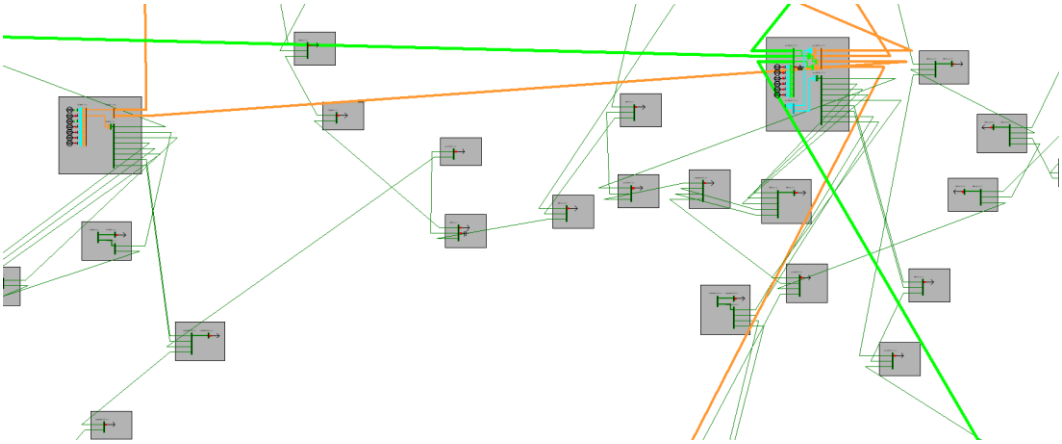

Figure 35. A straight-line approach to drawing transmission lines. Crowding and overlap can occur.

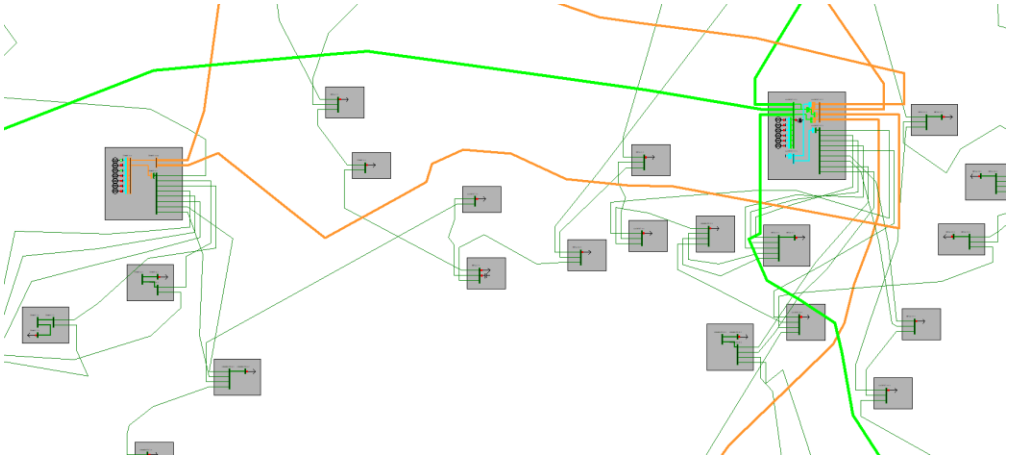

Figure 36. A Delaunay approach to drawing transmission lines. This approach avoids overlapping substations but causes too many bends in the high-voltage lines.

*1. Set up the Delaunay triangulation*

The Delaunay triangulation can be computed for a set of points (in this case the substation centers) in $\boldsymbol{O(n \log n)}$ time. Each edge of this graph is established as

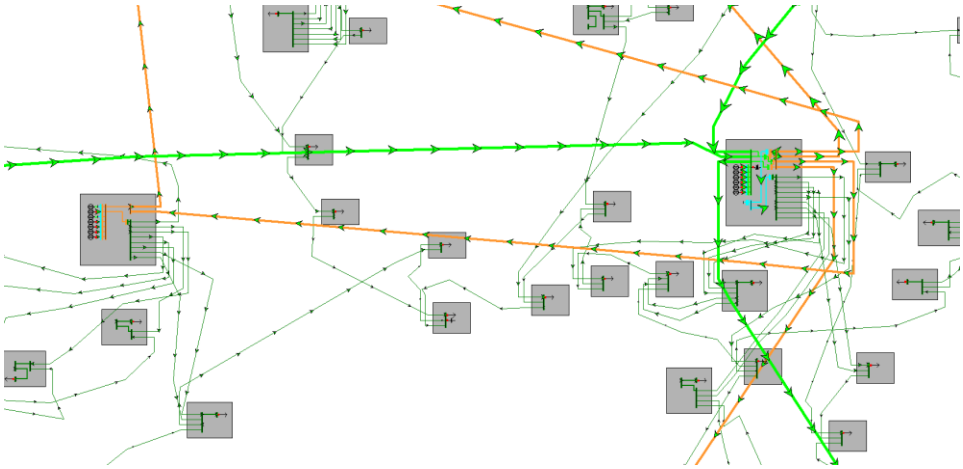

Figure 37. A two-layer Delaunay approach to drawing transmission lines. This approach avoids both overlapping substations and excessive bends in the high-voltage lines.

a routing channel that will record lines which pass through it, so in the data structure it is connected with its neighboring edges on each end.

*2. Route lines straight through the Delaunay triangulation*

Lines have an initial point which is located near to a substation within a triangle in the Delaunay triangulation. The next step is to trace each line's path through the triangulation, inserting a waypoint at each intersection with a Delaunay edge. This path can be traced quickly since from each waypoint there are only two possible edges which could form the next one. Figure 38 illustrates this step.

*3. Adjust the waypoints to ensure good spacing.*

Next, the routing channels, which correspond to Delaunay edges, are each analyzed in turn. There may be one or more waypoints registered on it, and the goal is to minimize the changing of these waypoints while keeping a buffer around the

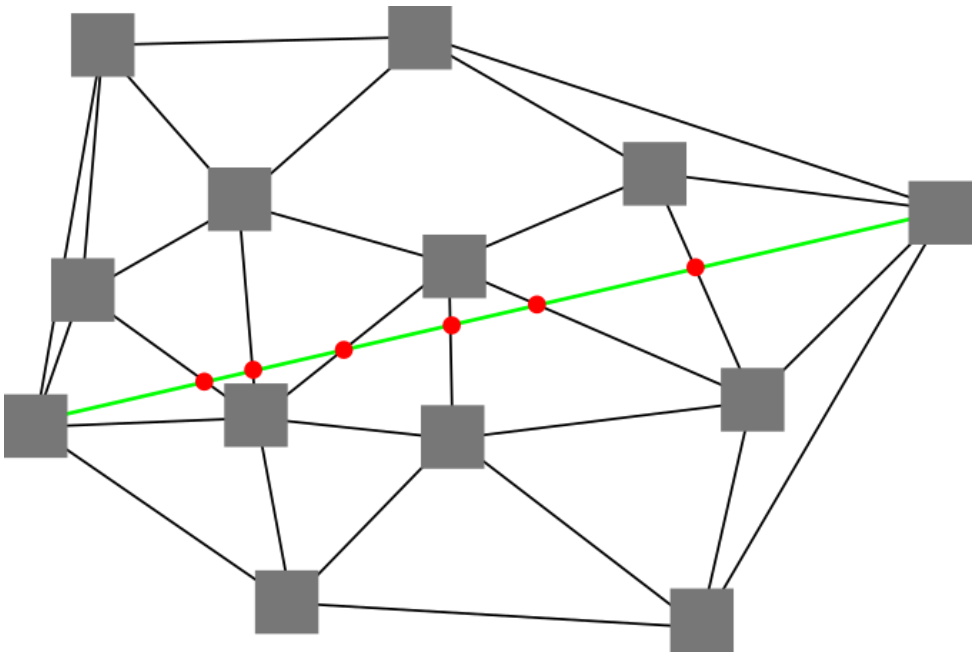

Figure 38. A Delaunay triangulation of substations and routing through channel. The red dots are waypoints which the channels will adjust to improve the spacing.

substations on either end and to keep multiple lines adequately spaced along the corridor.

Ordered from one end of the channel, each waypoint is given the spot closest to its desired location (given by the point along the straight line between its two adjacent waypoints on the same transmission line) which is acceptably far from either substation and any already placed line. Thus this step takes a greedy approach to spacing the channels.

*4. Iterate the process*

Step 3 is repeated at least 10 times, allowing neighboring channels to iteratively coordinate.

**Two-level Delaunay approach.** The results of the Delaunay approach can be seen in Figure 37, where the lines are routed to avoid collision with each other and the substations. But the high-voltage lines (light green and orange) snake back and forth to avoid lower-voltage substations, which causes too much distraction for the more important visual components. The approach presented by this section does the algorithm above twice: once using only extra-high-voltage (200+ kV) substations and lines, then again using all substations and remaining lines. The result (Figure 37) has a much more understandable high-voltage network without much cost in overlap.

## 5.3 Engineering education

Graduates of an engineering program who go to work in the electric power industry should be ready to study large power systems. Though very small examples are useful for teaching the analysis principles and modeling basics, there are natural benefits to additionally exposing students to power systems that are realistic in size and complexity. A key reason undergraduate power systems courses often do not utilize large grids is that such data sets can be hard to obtain with the appropriate permission for classroom usage. Though there are some existing test cases, much actual power grid information is not fully public due to legitimate security concerns.

This section discusses how synthetic power grids can fill this gap by providing publicly available test cases that match the size and complexity of actual grids. The large system assignments described here have been implemented in a undergraduate class, titled "Power System Operation and Control," at Texas A&M University. To date they have been used in Fall 2017 and Spring 2018, both times with an class of 70

primarily undergraduates divided into six 12 student lab sections. The lab portion of this class now includes, in addition to exercises with physical equipment and small system models, assignments that analyze a 2000-bus synthetic power grid in the commercial software package PowerWorld Simulator. A variety of planning and operations studies using the 2000-bus synthetic case are assigned throughout the term. Topics include power flow sensitivity, contingency analysis, optimal power flow, and transient stability.

**Background on simulation assignments in power engineering education.** Computer simulations have formed a part of power systems education for over four decades [108]-[109]. Since analyzing practical systems quickly becomes difficult for hand calculations, over the years software tools have been created and used for teaching power engineering. An early example is [110], which dedicated a course to computer applications in power system control centers. By the late 1980s many universities were balancing hardware lab assignments with computer simulations in power systems courses [111]. New and more feature-rich software tools and their educational applications were developed in the following decade, as engineering education could integrate analysis for security, transients, and control [112]-[115]. Advances in computer graphics made tools more user-friendly and interactive [116]-[117], and towards the turn of the century educational software for power systems was becoming more general and widely used [118]-[120]. Reference [121] is an example of the use of simulation in training industry professionals. In the context of power system simulation, [122]-[123] highlight the importance of targeting industry needs in university education.

In the last decade, simulation in power systems education has advanced to envelop new grid technologies [124]-[127], and to integrate with common programming interfaces and open source platforms [128]-[130]. With increased computing capability, power systems labs now greatly benefit from exercises using state-of-the-art software [131]. Recent publications have shown that new lab assignments are including hardware, software, and real-time simulations such as hardware-in-the-loop [132]-[133]. Software simulations are broadening in topic also to cyber infrastructure [134] and the applications extend even to secondary education [135].

Throughout this extensive literature documenting power system simulation in education, the test cases consistently tend to be small, usually on the order of ten buses or fewer. Building on a long tradition of power systems education, this section shows how teaching modeling, analysis, and control concepts can be supplemented with demonstration and assignments that involve large power systems. The main motivation for using large systems in undergraduate education is that when they enter the power industry students will be dealing with large systems. In addition, there are unique aspects to studying large systems, which include a system diagram that cannot be displayed on the whole screen at once, multiple areas that involve aggregating hundreds of generators, and inter-area oscillations in frequency response characteristics.

| Wk. | Lecture Topics | Lab Assignment |
|---|---|---|
| 1 | Introduction, complex power, 3-phase, per-unit | No lab |
| 2 | Power system structure, history, operations | No lab |
| 3 | Modeling of transmission lines and transformers | Power calculations (Matlab) |
| 4 | Loads and generators | Three-phase circuits (PowerWorld) |
| 5 | Y-bus matrix | Power system operations |
| 6 | Power flow problem | No lab (Exam #1) |
| 7 | Numerical solutions | Synchronous generator parameters |
| 8 | Sensitivity, large systems | Synchronous generator operation |
| 9 | Economic dispatch | **Power flow analysis, sensitivity** |
| 10 | OPF, SCOPF | **Economic dispatch, contingencies** |
| 11 | Power system stability | **OPF, SCOPF** |
| 12 | Power system stability | No lab (Exam #2) |
| 13 | Power system controls | Power markets |
| 14 | Distribution systems | **Transient stability, dynamics** |
| 15 | Emerging topics | No lab |

Table 16. Course outline and lab assignments.

**Learning objectives and methodology for power systems course with large grids.** The objectives of this fifteen-week semester course, which included mostly senior-level undergraduate students and a few first-year graduate students, is to teach how power systems are modeled, analyzed, and managed. Prerequisites for the course include circuit theory and linear signals and systems. Table 16 gives the lecture topics and associated lab assignments for each of the fifteen weeks of the course. The assignments shown in boldface type use a large synthetic grid model.

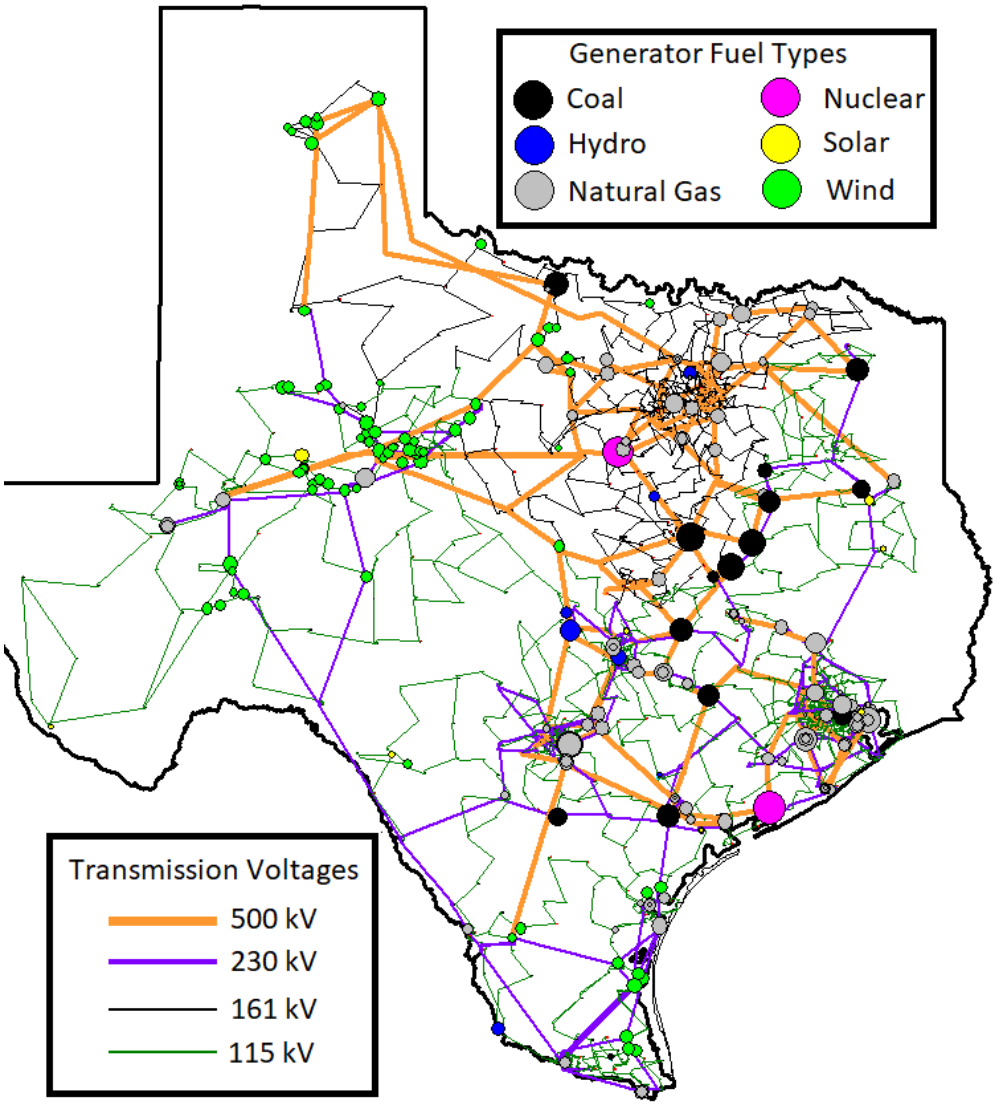


Figure 39. Synthetic 2000-bus test case oneline diagram. Transmission voltages are shown, as well as circles representing the generators, with color indicating fuel type and size indicating the relative MW capacity. This grid is fictitious and does not represent the actual Texas grid.

The weekly lab assignment component demonstrates the planning and analysis concepts taught with exercises using a combination of software and hardware tools. The first three labs use very small systems, first in a Matlab toolbox to show active and reactive power, phasors, and imbalance; then in PowerWorld Simulator, with an

introduction to single line diagrams, power flow solutions, maximum loading, and area control error (ACE). The fourth and fifth labs demonstrate calculation of synchronous generator parameters and the equivalent circuit for simulation modeling using hardware exercises. Students do open-circuit and short-circuit tests, and observe the effect of excitation control on voltage and power when a generator is operated under variable load.

Starting in the second half of the semester, labs begin to integrate both small and large power system simulation, as will be discussed below. The first of these covers fixing overloaded branches with sensitivity analysis, volt-var control, capacitors, tap coordination, and phase-shifting transformers. Then the following labs introduce the economic operation of power systems and system dynamics.

The synthetic 2000-bus case (Figure 39) is a natural fit for engaging student interest in large power systems at Texas A&M, since its geographic footprint follows the Electric Reliability Council of Texas (ERCOT), which serves the majority of the U.S. state of Texas. While at first glance the grid looks real, it contains no actual lines since it was built with a synthetic methodology. As Figure 39 shows, four levels of high-voltage networks connect eight areas and 1250 substations. The case's 544 generators roughly correspond to public information about actual plants, with fuel types including coal, hydro, natural gas, nuclear, solar, and wind.

In addition to power flow data and the single line diagram, the synthetic data associated with this case includes quadratic cost curves for each of the generators and transient stability models. Synchronous machine models, excitation systems, governors, and system stabilizers are all specified. While this course only introduces

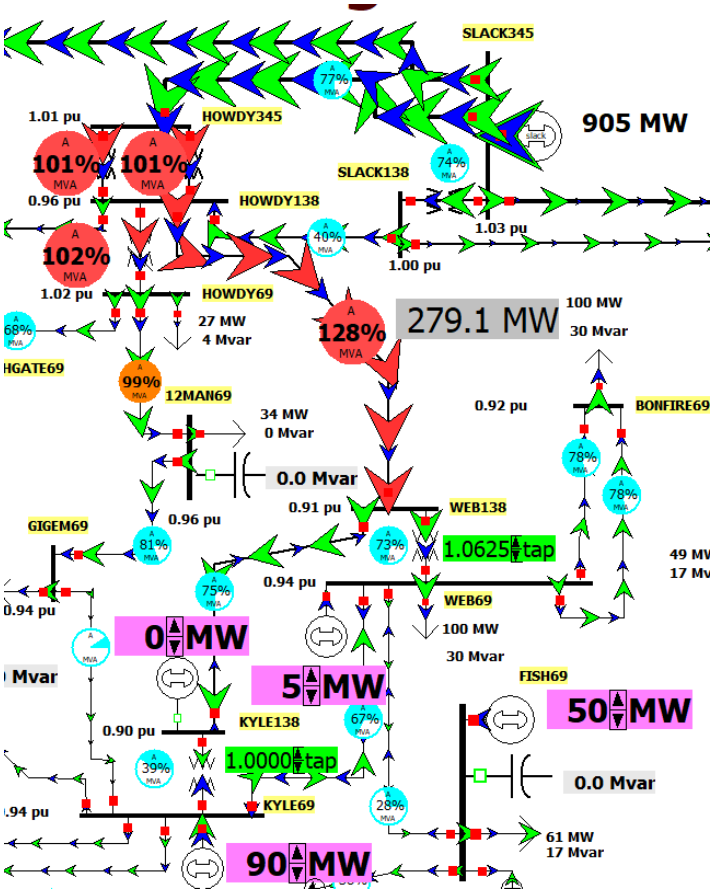


Figure 40. 37-bus case used for first part of lab assignments. For the power flow lab, several lines are overloaded (red pie charts [136]) following a generator failure (the magenta label showing 0 MW). The other magenta rectangles show controllable generators, which can be re-dispatched to fix the violations. The arrows show real power flow (green), real power flow on violating lines (red), and reactive power flow (blue).

some of these models, the principles of transient stability can be shown in the effect on 2000-bus system frequency response.

Analyzing a power system of this scale in a two-hour lab session is challenging, especially as students are still learning the software and the underlying modeling concepts. One key to making the cases accessible is that each lab instruction guide is detailed and shows with specific steps how to use the software to perform the studies

under consideration. In addition, specially tailored single line diagrams with prominent labels and controls, which are highlighted later in the section, aid students in visually comprehending what is happening across the system and focuses their attention on the concept at hand.

**Power Flow analysis lab.** The first introduction to operating a larger power system comes in the sixth lab, which focuses on power flow analysis studies. The objectives of this lab are for students to gain insight and experience with the power flow solution, sensitivities, the effects of various controls, and how to mitigate line overloads and excessive losses. There are two parts: one with a 37-bus case and the other with the 2000-bus synthetic Texas grid. A key objective in adding the 2000-bus grid to this exercise is to present students with a large system and allow them to learn by doing that they could apply techniques presented in lecture to solve problems with the larger grids they will encounter in industry.

For the first portion of this lab, the instruction guide acts as a tutorial that describes the starting situation in the 37-bus case for the fictional utility, part of which is shown in Figure 40. The starting situation is that a 175 MW generator in the center of the system has failed. This causes several overloaded lines as the lost power is supplied by the slack bus, which represents an external intertie. The oneline diagram is configured to show these overloads prominently, with MW fields for two particular lines given in large gray boxes. The other controllable generators are shown with magenta fields that can be edited to change their dispatch value. There are also control options for load tap changing transformers (LTCs), switched shunts, and a phase angle

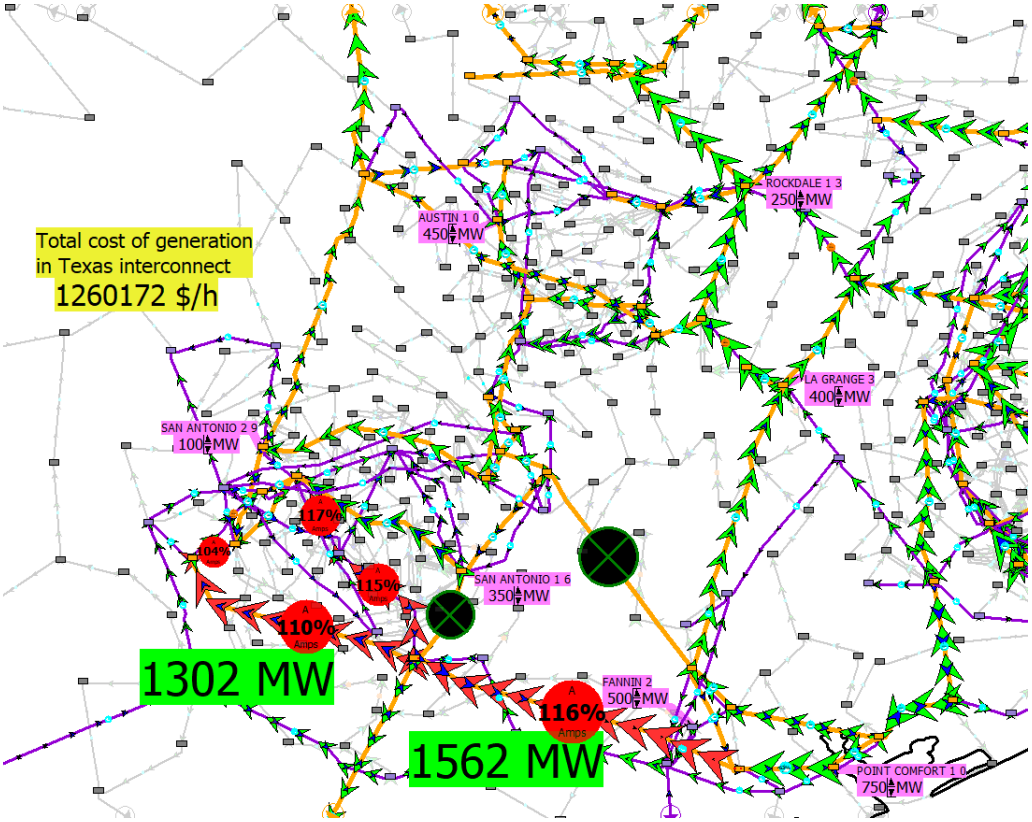


Figure 41. Zoomed-in display of 2000-bus case for power flow lab. The green fields show the power flow through two 500 kV lines which are overloaded (red pie charts [136]) following a double outage contingency (black circles). The magenta rectangles show controllable generators, which can be re-dispatched to fix the violations.

regulating (PAR) transformer. There are additional fields giving bus voltages, substation names, and the total system load and losses.

The students' assignment for this first portion is to develop, justify, and implement a strategy to fix the line overload violations and minimize the system losses. They start by collecting data on the sensitivity of two line overloads to a change in selected generators' dispatch setting and the change in the PAR setting. (For example, if the 5 MW generator is increased to 10 MW and the line flow reduces from 279.1 MW to 277.1 MW, the sensitivity will be -0.40.) This analysis will show which controls

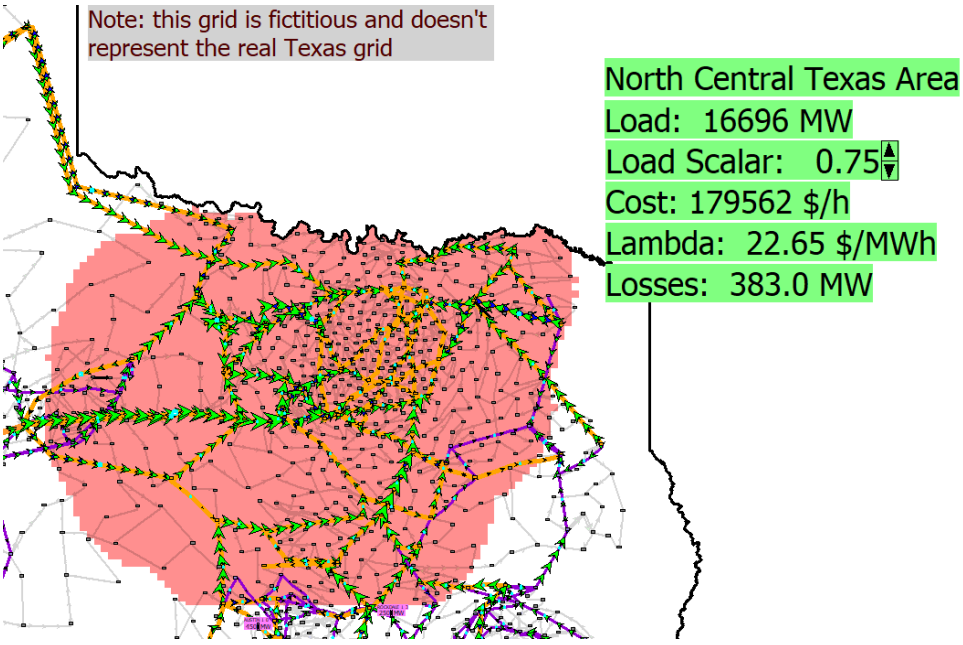


Figure 42. Economic dispatch area of North Texas. The green fields give data relevant to the economic dispatch solution.

are most useful for correcting the violations, and how much they should be changed. Once the changes are implemented, the directions point out circulating reactive power flows that are contributing to a higher than necessary level of active power losses. By coordinating the LTCs and switching in capacitors, the losses can be reduced to an empirical minimum.

Next, the students are directed to the 2000-bus system representing the fictitious Synthetic Texas Grid Company, which has an analogous situation in which a double line outage has caused overloads in the transmission system. It is a much larger case, but the customized diagram presented in Figure 41 shows the overloads similarly to the 37-bus case just completed, and highlights some controllable generators with magenta fields. Just as before, the students calculate the sensitivities and develop an

action plan to resolve the emergency situation. Though cost is not a crucial concern in this exercise, students observe how the system cost of operation has increased with these remedial actions.

**Economic dispatch and contingency analysis lab.** The next lab exercise introduces the economic operation of power systems, and contingency analysis. Again the assignment starts with the 37-bus case and moves to the 2000-bus case with assignments that are similar but have important distinctions unique to the large system.

In this lab, students focus on the North Central area of the system, highlighted in red in Figure 42. They explore the effects of changing the load by a constant scalar from 70% of peak in 5% increments to full peak. The software adjusts the generation with an economic dispatch solution, leading to a system marginal cost (lambda) value and a new power flow solution. In learning about economic dispatch, the students quantify and explain how the total cost, losses, and lambda have changed with the load.

For the contingency analysis portion of the lab, students run ac power flow solutions on 77 single-element outage conditions using the automatic contingency analysis software tool. In the 70% peak condition, there is one violation, and in the 100% peak condition there are two violations. Students sort the list of contingency results and report the outaged lines and the corresponding overloaded lines.

Next students are assigned to investigate both of these contingency conditions and restore the system to secure operation. They find each of the violating areas and reproduce the contingent situation. Then they observe the overload and determine what control actions can be taken to mitigate the violation. Figure 43 shows one of

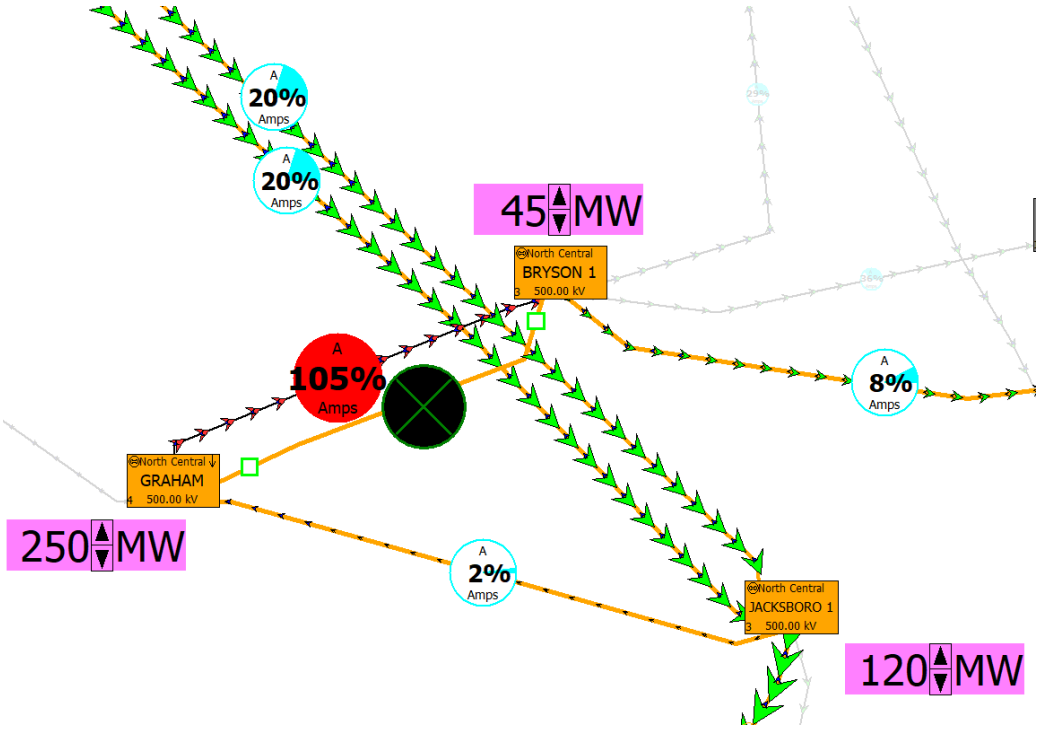


Figure 43. Zoomed-in view of a contingency violation in the 2000-bus lab.

these as an example, where the outaged 500-kV line, indicated with the black circle, causes the neighboring 161-kV line to overload at 105%. It turns out that in reducing the generation at one substation and increasing it at the neighbor (against the economic dispatch), this overload can be eliminated. The connection visible here between economic dispatch and secure operation of the power system sets the stage for the security constrained optimal power flow analysis of the next lab.

**Optimal power flow lab.** The eighth lab of the semester aims to familiarize students with optimal power flow (OPF), security constrained OPF (SCOPF), and locational marginal prices (LMPs) on both power systems. For the 37-bus case, the manageable size allows investigation of each individual generator to verify its profit and LMP. A few binding constraints appear as the load is increased, so that the

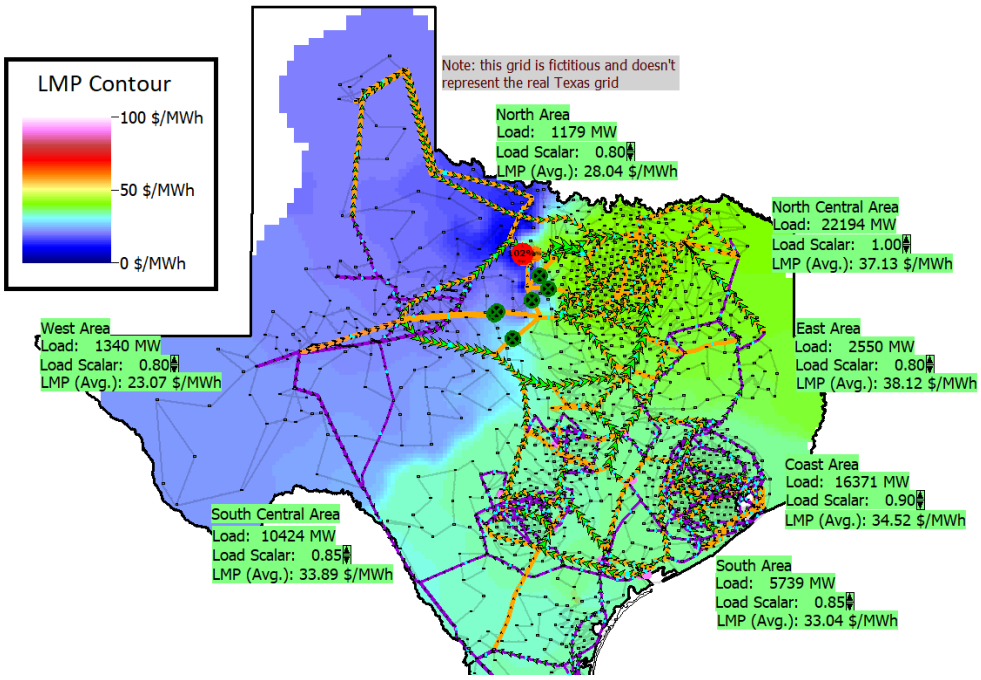


Figure 44. Diagram display for optimal power flow lab. Green fields provide controls for the load scalar in seven of the system areas, and report the average LMP for these areas. The background contour [137] shows the locational marginal prices.

solution moves from what is essentially an economic dispatch to a LMP-based constrained solution. Students are instructed to isolate a single generator and plot its profit as a function of its bid offer. With a low enough bid, the generator is the marginal unit and profit increases with the bid, until another generator becomes more profitable. Finding this maximum, and possibly multiple local maximums, introduces the operational decisions generator owners make when bidding into a LMP-based market.

The 2000-bus system gives the added benefit that locational marginal prices are associated with system areas and a geographic span. In this portion of the lab, the situation is that several lines are out along an east-west corridor, leading to the potential for congestion as load varies throughout a summer day. The diagram shown in Figure

44 gives students control of the simulated load scalar for each of seven areas in the system, which are initialized to 80% of peak. The OPF solution for this case yields near-uniform LMPs, as in the 37-bus case, and again as the load changes congestion is introduced. Students investigate which areas have the most significant impact on LMPs and which combinations of load scalars are particularly troublesome. The typical results are that the prices are higher in the east since the cheaper wind generation in the west cannot be fully transmitted to the load centers.

Some conditions of load near peak lead to extremely high LMPs and even unenforceable constraints in the OPF solution. Recognizing that load shedding is a last resort, the next part of this lab assigns students the task of determining which loads, under emergency conditions, would have the most favorable effect on LMPs if shed. This data indicates the loads' appropriateness for relieving the excessive congestion.

The final step in this analysis is to run an SCOPF solution, considering 349 single element outage contingencies. This number was selected to capture the high voltage network contingencies, since including all N-1 events, as would be done in an actual analysis, would take too long for the lab period. But with these contingencies considered the results show the change in LMPs across the system due to binding security constraints.

**Transient stability lab.** Having small and large systems provides the opportunity to teach different aspects of power system transient stability analysis in

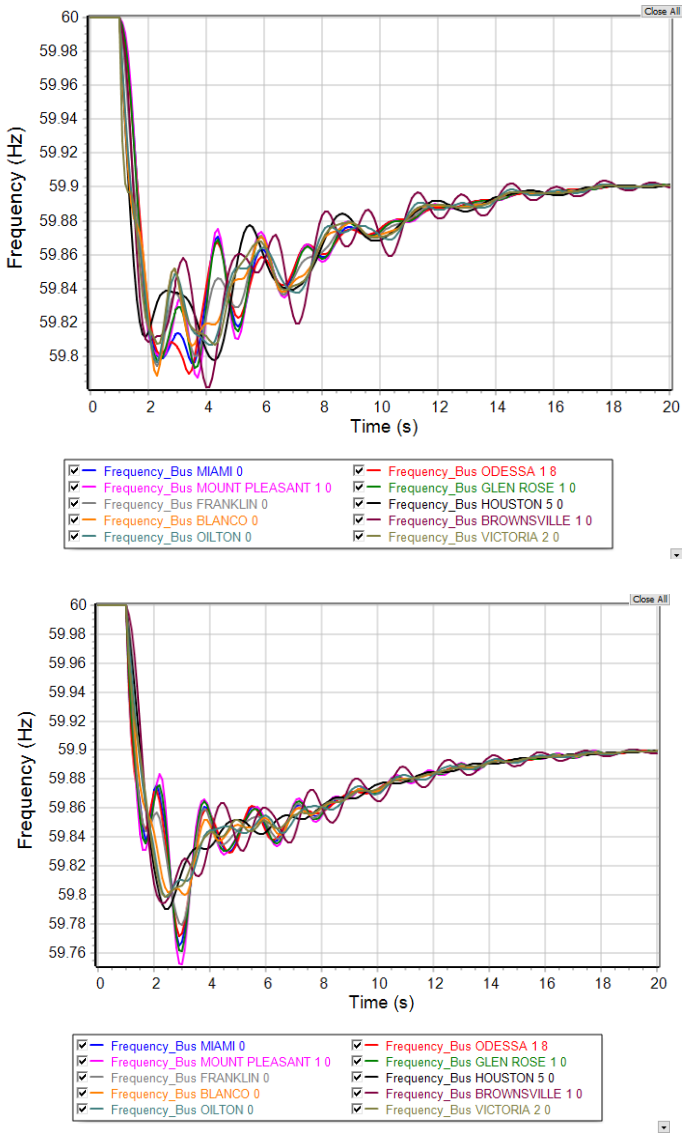


Figure 45. Frequency response plot for transient stability lab. Ten selected bus frequencies are shown. The upper plot is the contingency of the loss of 2700 MW of generation at the same location near eastern side of the system, and the lower plot shows the loss of 2700 MW of generation, with half in the east and the other half in the far west.

the lab setting. This lab has two parts for each case: one in which students run planning studies and the other which simulates a real-time operations scenario.

The 37-bus case transient stability runs very fast and has a nearly-uniform frequency response due to its small size. Thus it is well suited to calculating critical clearing time. Students are assigned the task of simulating the transient stability response to a line fault and opening, adjusting the clearing time and observing the range in which the response is stable. This process is repeated for two different contingencies, and two different loading conditions. In addition, the impact of the generator inertia constant $\boldsymbol{H}$ is investigated. When $\boldsymbol{H}$ is increased, the critical clearing time is longer, since the increased inertia improves stability in the system.

These insights carry well into the analysis of the 2000-bus system, for which the transient stability simulations take longer to run and frequency varies noticeably among parts of the systems. In this lab, students investigate the frequency and voltage response of the system to a contingency corresponding to the loss of a large amount of generation. The plots in Figure 45 show the overall behavior of the system, as the frequency drops quickly and settles below 60 Hz due to the governor droop. Students first investigate the difference between the loss of 1350 MW and 2700 MW, showing that a larger disturbance has a more severe frequency response.

The next task, illustrated by Figure 45, is to compare two contingencies that both correspond to the loss of 2700 MW of generation. The first involves the loss of all 2700 MW at the same location on the eastern side of the case, whereas the second contingency involves the simultaneous loss of 1350 MW in that spot and another 1350 MW far to the west. While both contingencies are stable and settle to a frequency of about 59.9 Hz, the response is noticeably different. The second condition has a lower nadir frequency and shows different and more intense inter-area oscillation.

Then the transient simulation is rerun, with 2000 MW of coal generation removed from the system. This study connects the class material to a current event facing the actual ERCOT system [138]. This reduction in system inertia can be observed in the impact on frequency response, analogously to the changing of the ***H*** constant in the 37-bus case.

Finally, this lab includes a portion which involves simulated real-time operation, where transient stability simulations are run in an interactive environment and students attempt to maintain system stability throughout the scenario. In both the 37-bus case and the 2000-bus case, the scenario is that a tornado is assumed to move through the system, taking out three lines in rapid succession. These changes induce oscillations in system voltage and frequency, and cause other lines to be overloaded. If overload, under-frequency, or over-excitation violations are not fixed quickly enough, the modeled protection schemes will trip those lines or generators, exacerbating the problem. The task is to stabilize the system with minimum load shedding, and avoid a blackout, where the system has deteriorated to the point that the simulation can no longer solve.

The controls provided in these scenarios are load shedding, generator set points, and line switching. The interface diagram for the 2000-bus case is shown in Figure 46. Line flows arrows show how power is being transferred in the system, and pie charts indicate line loading levels [136]. Black circles show lines which are already opened, and red circles indicate overloaded lines that will open if further intervention is not done. The background contour shows where voltage issues are occurring [137]. System frequency is indicated by a strip chart in the upper left corner. The gray boxes,

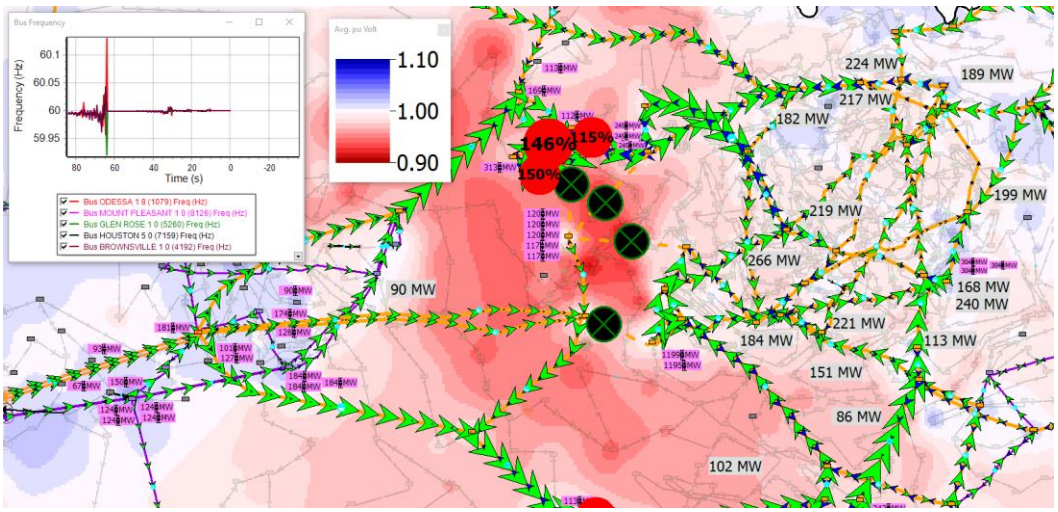


Figure 46. Interface diagram for dynamic simulation of the 2000-bus case. This snapshot is halfway through without any intervention. This is a zoomed-in view that shows four opened lines (black circles), and three severely overloaded lines (red circles). The gray boxes show loads that could be shed, and the magenta boxes show generators which can be opened or adjusted. The contour shows voltage magnitude [136]. The strip chart in the upper left plots the frequency at five buses as the simulation plays out in real time.

mainly to the right, represent loads that can be shed, and the magenta fields correspond to controllable generators.

With a large 2000-bus case, there are many generating units and loads, so this exercise helps to develop an intuitive insight into how power is flowing across the system and what changes can be made to improve the system state. This exercise integrates what has been learned throughout the course, in power flow sensitivity, contingency analysis, and transient stability. Most students were able eventually to avoid the Figure 46 voltage collapse by changing generation and load, albeit it took most several iterations.

**Student Perception and Extension**. The course was highly rated in student end-of-semester assessments, and students found the additional challenge and realism

of large-system lab exercises to be valuable. One student commented that the lab "has allowed me to work with big systems that are similar to what industry power engineers use. The case is laid out geographically, so it is beneficial to see how an outage at a substation in one area can affect neighboring substations' reliability." Another student said the labs helped her "appreciate how complex a full-sized grid can be."

After becoming familiar with the synthetic 2000-bus system through half a semester in class, four students also used the case for a senior thesis project. The project looked at planning for the future of the synthetic Texas grid under additional load and renewable generation, and making the system more resilient to severe weather and geomagnetic disturbances. These students presented their work in a poster at the *2018 IEEE Texas Power and Energy Conference (TPEC).*

# 6. SYNTHETIC GRIDS PRODUCED

This section gives an overview for each of the nine created cases associated with this dissertation. The sizes are 200, 500, 2000, 5000, 10,000, 20,000, 25,000, 70,000, and 100,000 buses. All are geo-located on a subset of the United States footprint.

## 6.1 Medium-sized systems: up to 1000 buses

The 200-bus base test case is built for power flow studies and general analysis and research purposes. The geographic footprint is fourteen counties in Central

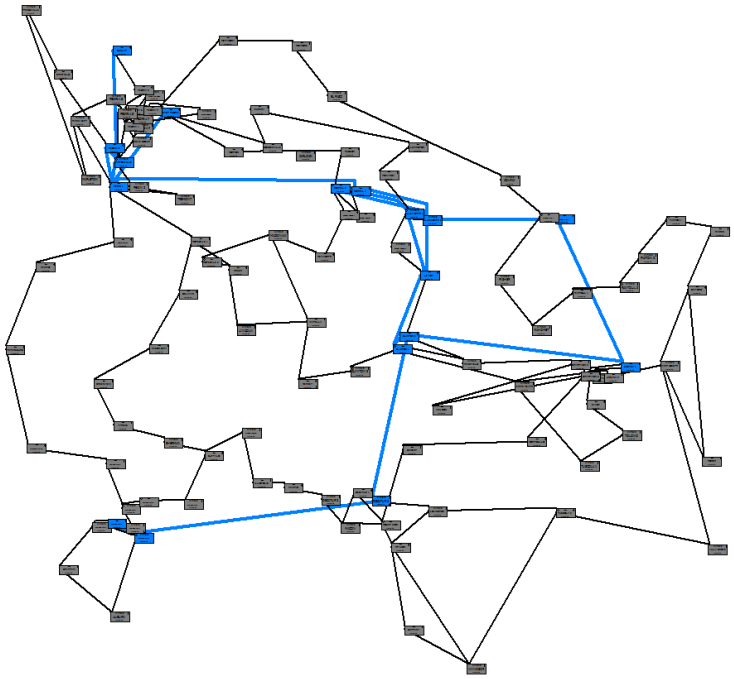

Figure 47. One-line diagram of the 200-bus case. The 230 kV lines are in blue and the 115 kV lines in black.

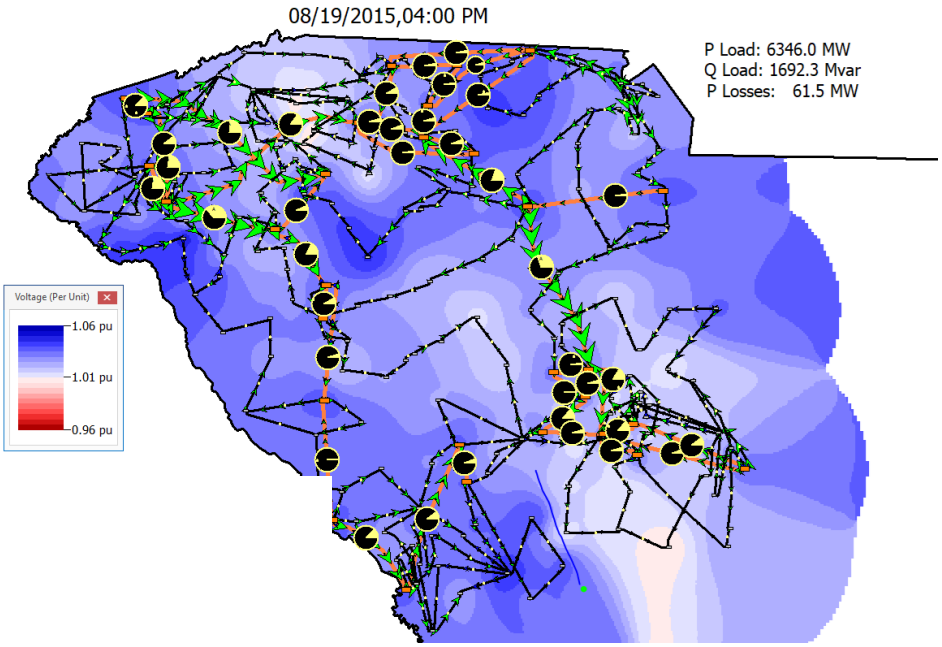


Figure 48. One-line diagram of the 500-bus case. The 345 kV lines are in red and the 138 kV lines are in black. This figure shows a voltage contour during the time series simulation.

Illinois, an area with a population of about 1.1 million and generation that exceeds the local need (net export region). The geography is simple, as there are no major features, and only one area was created for such a small region. The transmission network was built with two voltage levels: 230 kV and 115 kV. This section provides details and validation statistics for this case. Figure 47 shows the one-line diagram of the case.

The 500-bus base test case is built for power flow studies and general analysis and research purposes. The geographic footprint is 21 counties in western South Carolina, an area with a population of about 2.6 million and generation that exceeds the local need (net export region). The geography is simple, as there are no major features. The transmission network was built with two voltage levels: 345 kV and 138 kV. Its diagram is shown in Figure 48.

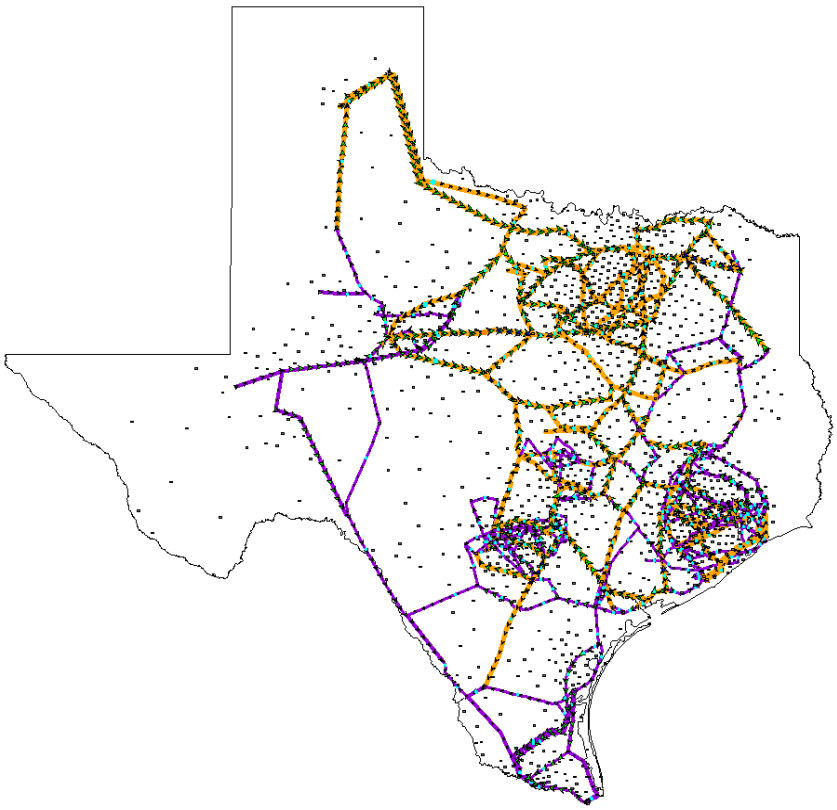

Figure 49. One-line diagram of the 2000-bus case. The 500 kV lines are in orange and the 230 kV lines are in violet. The lower voltage lines are not shown, and the small gray boxes are the substations.

### 6.2 Large systems: up to 10,000 buses

The geographic footprint for the 2000-bus test case is the portion of the state of Texas served by the Texas interconnection and the Electric Reliability Council of Texas (ERCOT), an area with a population of about 22.3 million that corresponds to an actual self-contained electric interconnection. Eight areas are used: North, East, North Central, South Central, Coast, South, West, and Far West. However, for optimal power flow (OPF) purposes the entire case is considered and dispatched together. The transmission network was built with four voltage levels: 500 kV, 230 kV, 161 kV, and 115 kV, mixed and matched among the eight regions. There was also considerable

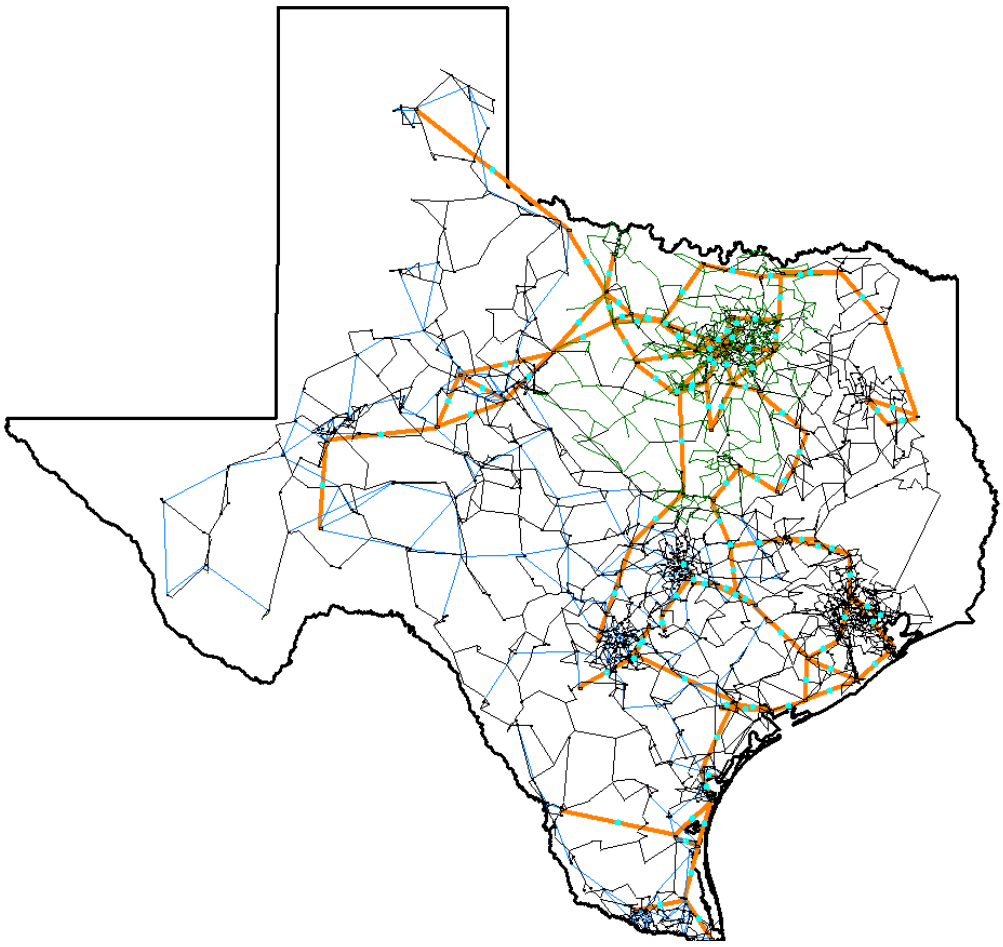

Figure 50. One-line diagram of the 5000-bus case.

manual adjustment for generating a realistic power flow scenario, OPF solution, and valid statistics. A one-line diagram of the case is visible in Figure 49.

The 5000 bus case is shown in Figure 50. It is on the same footprint as the 2000-bus case with the same areas, except that it is modeled with a number of buses more commiserate with the actual ERCOT case, though this case is still totally synthetic. Voltage levels used are 69 kV, 100 kV, 115 kV, 161 kV, 230 kV, 500 kV. This is one of the cases used in the complex network analysis discussion of section 4.4.

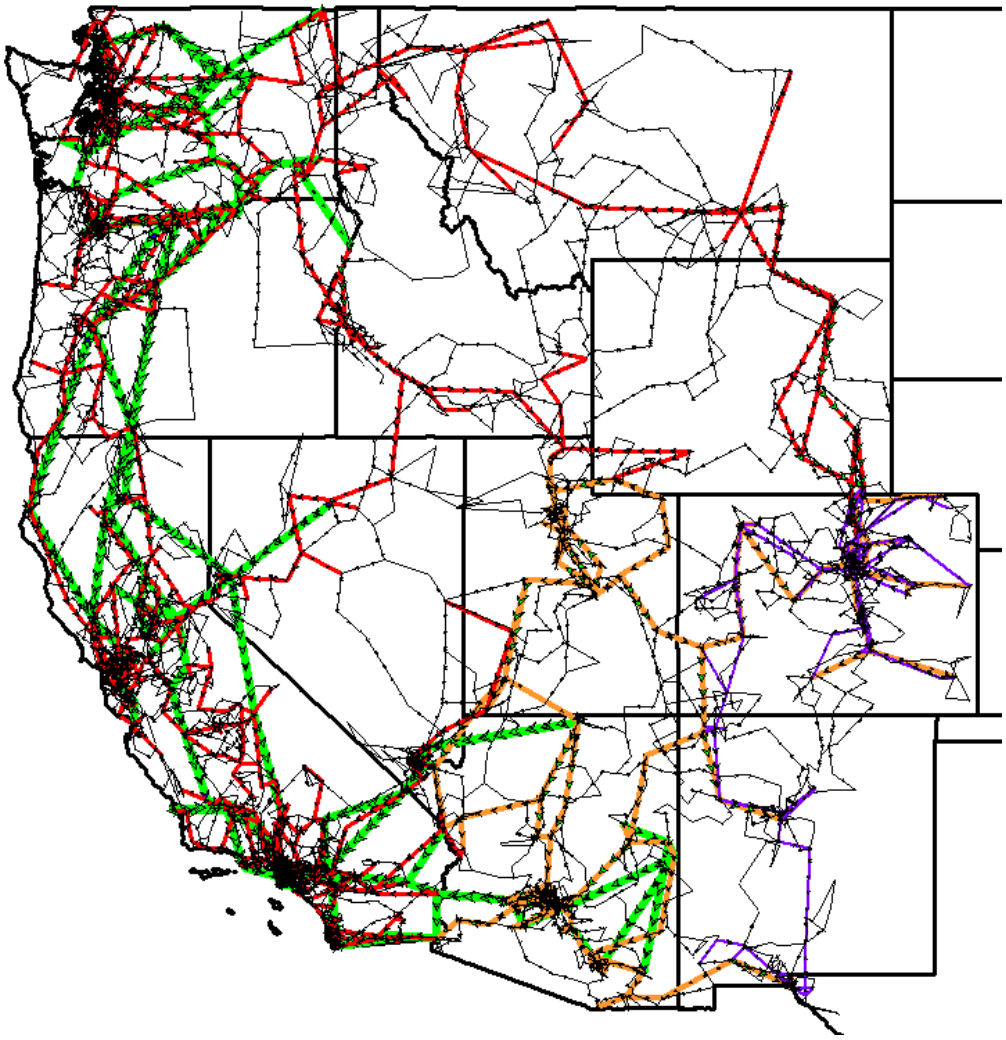

Figure 51. One-line diagram of the 10,000 bus case. The 765 kV lines are shown in green, 500 kV lines in orange, 345 kV lines in red, 230 kV lines in purple, and lower voltage lines in black.

The 10,000-bus base test case geographic footprint is the USA portion of the North American Western Interconnection (WECC). The geography is complex and diverse, and there are 16 areas: Washington, Oregon, Idaho, Montana, Wyoming, Utah, Colorado, Nevada, Northern California, Bay Area, Central California, Southwest California, Southeast California, Arizona, New Mexico, and El Paso. The transmission

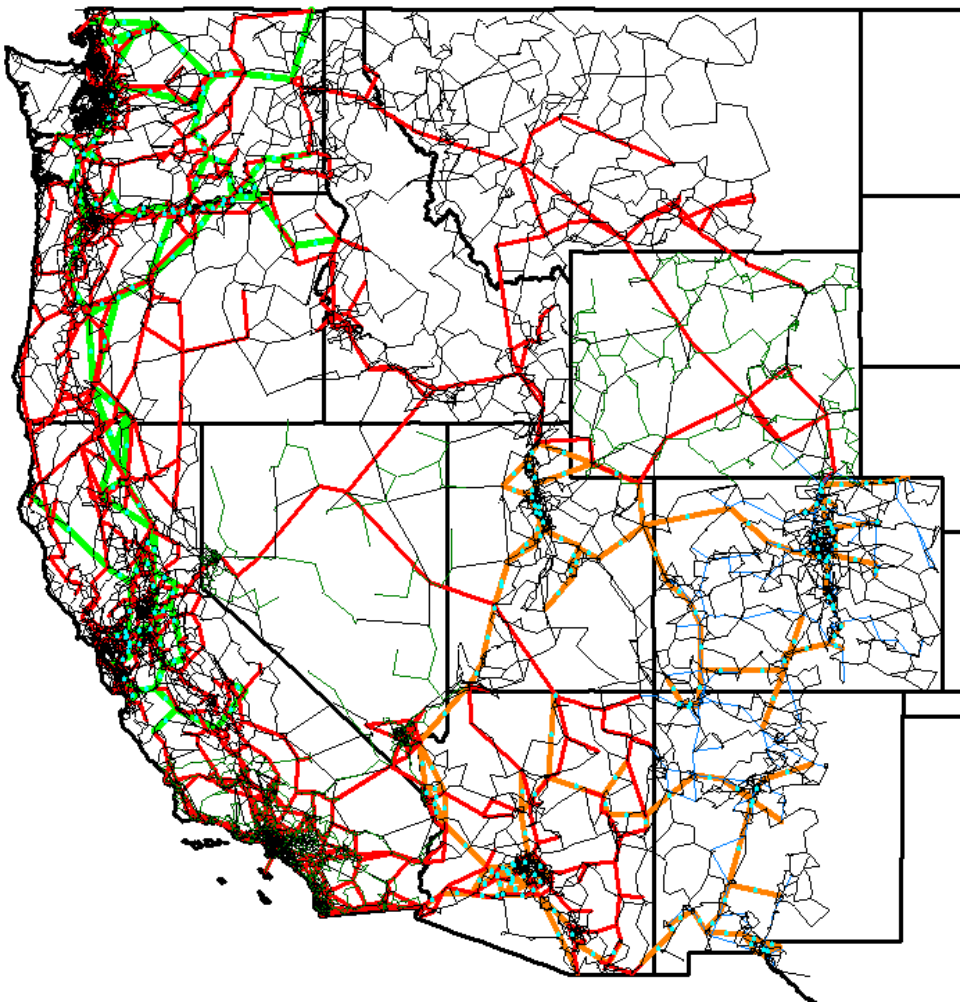

Figure 52. One-line diagram of the 20,000 bus case. The 765 kV lines are shown in green, 500 kV lines in orange, 345 kV lines in red, 230 kV lines in purple, and lower voltage lines in black.

network was built with seven voltage levels: 765 kV, 500 kV, 345 kV, 230 kV, 161 kV, 138 kV, and 115 kV, and it is shown in Figure 51.

### 6.3 Very large systems: up to 100,000 buses

The 20,000 bus test case uses the same footprint and areas as the 10,000 bus case, as shown in Figure 52, except the number of buses is closer to the actual WECC

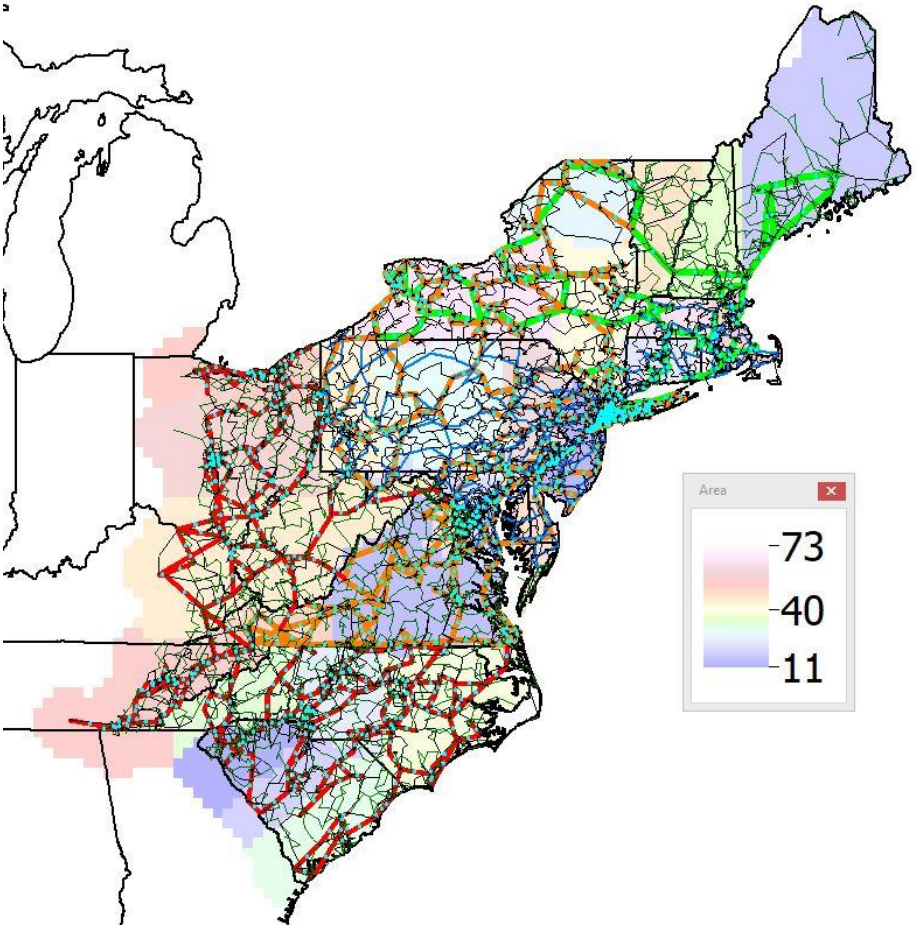


Figure 53. One-line diagram of the 25,000 bus case. The 765 kV lines are shown in green, 500 kV lines in orange, 345 kV lines in red, 230 kV lines in purple, and lower voltage lines in black or dark green. The light contour shows the 31 areas.

case, though this case is still completely synthetic. The voltage levels are the same, plus additional 100 kV and 69 kV in some areas.

The 25,000-bus base test case geographic footprint is the USA Atlantic seaboard. The geography is complex and diverse, and there are 31 areas from all or part of 18 states. The system has a higher bus-load density than some previous cases, and was the first to include lower voltage levels. The transmission network was built

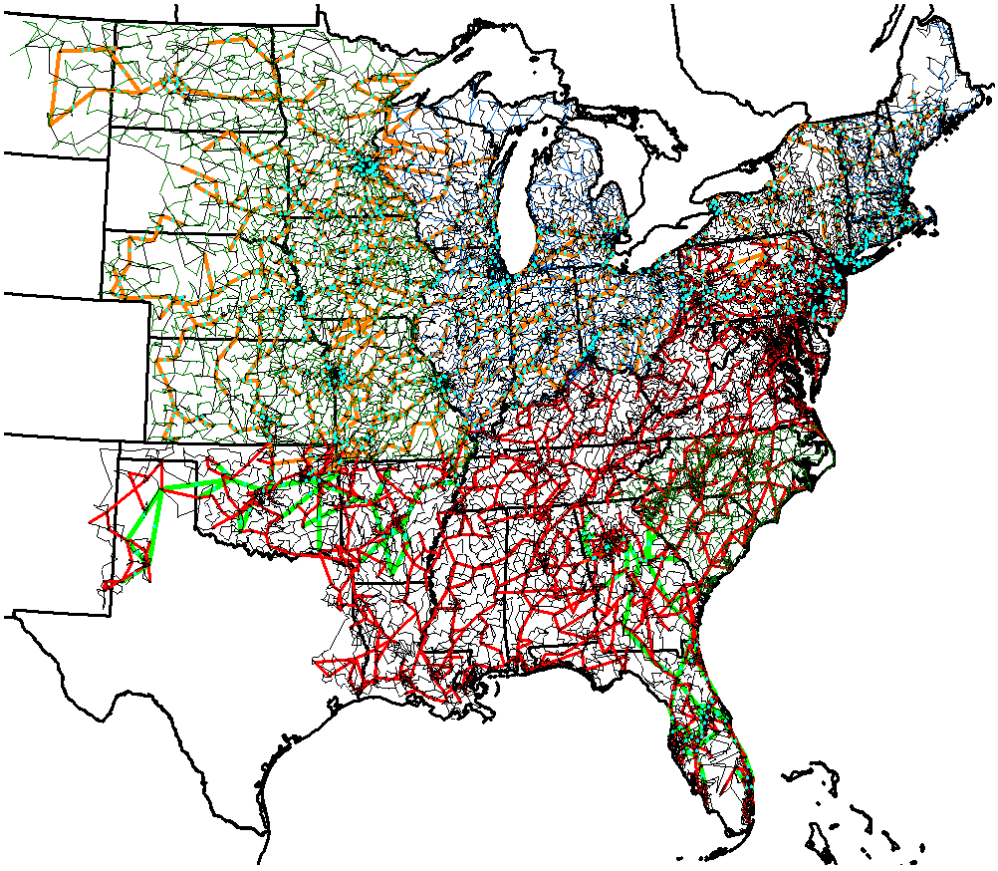

Figure 54. One-line diagram of the 70,000 bus case. The 765 kV lines are shown in green, 500 kV lines in orange, 345 kV lines in red, 230 kV lines in purple, and lower voltage lines in black or dark green.

with nine voltage levels: 765 kV, 500 kV, 345 kV, 230 kV, 161 kV, 138 kV, 115 kV, 100 kV, and 69 kV. The one-line diagram of this case is shown in Figure 53.

The 70,000-bus test case, shown in Figure 54, covers the footprint of the U.S. portion of the eastern interconnect in North America, a region covering over 30 states. It uses the same nine voltage levels.

The 100,000-bus test case, built to demonstrate the scalability of the grid synthesis algorithms, is shown in Figure 55. It covers the entire continental United States as one synchronous interconnect.

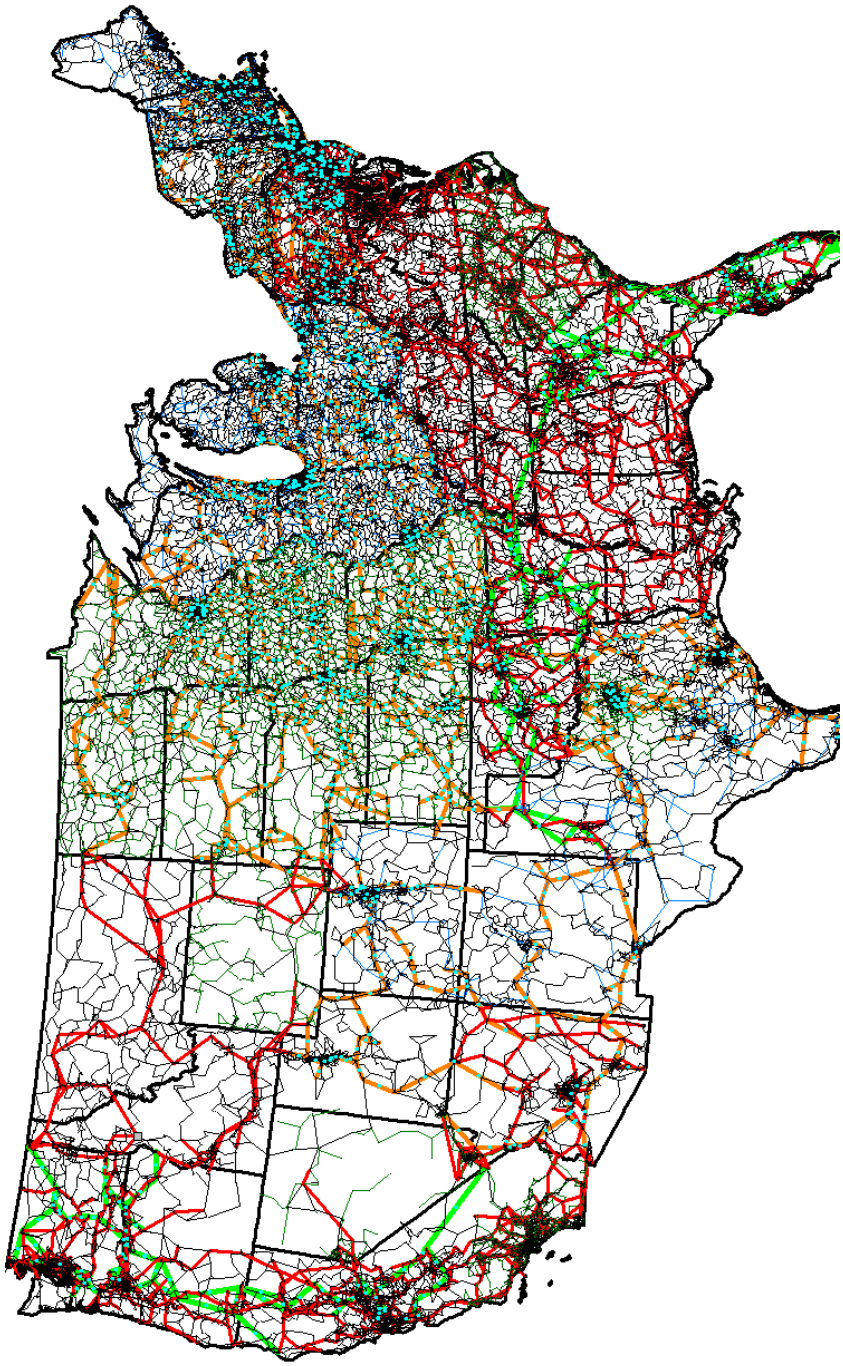

Figure 55. One-line diagram of the 100,000 bus case.

| Case | Date Built | Computation Time (hrs) | Manual adjustment time (hrs) | Comment |
|---|---|---|---|---|
| 200 | Nov. 2016 | < 1 | 4 | No automatic contingency analysis |
| 500 | Dec. 2016 | < 1 | 6 | |
| 2000 | Apr. 2017 | < 1 | 20 | |
| 10K | Jun. 2017 | 2 | 10 | Second-stage ctg. analysis |
| 25K | Nov. 2017 | 3 | 4 | |
| 70K | Apr. 2018 | 40 | 2 | Faster solver, contingency analysis at each iteration |
| 5000 | Jun. 2018 | 6 | 0 | |
| 20K | Jun. 2018 | 20 | 0 | |
| 100K | Jun. 2018 | 100 | 0 | |

Table 17. Estimated approximate computation times.

### 6.4 Discussion of computation times

While generating synthetic power grids tends to be an offline application, and thus computation speed is not a crucial concern, it is useful to include here a brief discussion of the computation times for building the synthetic grids, particularly when considering the manual intervention added and the possibility to automate building many cases for stochastic testing. The approximate computation and manual adjustment times are given in Table 17. Computation time is related to the size of the system, but also depends on the algorithms used and the number of iterations run. For the latest built cases, a full N-1 contingency analysis is performed at each iteration, greatly increasing the run time but also improving the quality of the case in terms of contingency violations. With the present run times, the possibility of generating hundreds or thousands of large cases is not tractable. However, there are several opportunities for future improvement in run times, including increased parallelization of the contingencies and the potential to use Cholesky update-downdate techniques for contingency matrix decomposition.

# 7. CONCLUSION

The work has as its main contribution new methodologies for the creation of synthetic grids. These methodologies are significantly refined, improved, and extended compared to previous work, scalable to 100,000 buses with additional complexities, an accomplishment which is well beyond what has previously been done. The work also contributes a validation framework for these cases, pointing out the metrics important to check to quantify how realistic a synthetic grid is. Secondary contributions are the applications to GIC studies, visualization, and engineering education. The final, tangible contribution is the set of nine test cases validated and published.

The main impact of these contributions is new test cases for power systems research, which have higher quality, more complexity, and more realism that existing test cases. These datasets will improve the ability to cross-validate published research result, supporting the scientific principle of reproducibility of results in this research field. These cases will also add to security in working with and publishing on sensitive research topics. The other major impact of this work is that these cases serve as demonstration platforms that have no restrictions related to data confidentiality. These are useful for a variety of applications from engineering education to showing new innovations in power system analysis and visualization. Secondarily, new insights into power system structure and properties related to the creation and validation of the grids make an additional impact in understanding these electric systems we study.

# REFERENCES


[1] IEEE and NSF. "Report on the first IEEE workshop on the future of research curation and research reproducibility," Nov. 2016. [Online]. Available: https://www.ieee.org/content/dam/ieee-org/ieee/web/org/ieee_reproducibility_workshop_report_final.pdf

[2] National Institutes of Health (NIH). "NIH strategic plan for data science," 2018. [Online]. Available: https://commonfund.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

[3] United States Department of Agriculture (USDA). "Reproducibility and rigor in research, education, and economics (REE)'s portfolio of research," Sept. 20, 2016. [Online]. Available: https://www.usda.gov/sites/default/files/documents/nareeeab-reproducibility-rigor-report.pdf

[4] The National Academies of Sciences, Engineering, and Medicine. *Enhancing the resilience of the nation's electric system.* The National Academies Press, Washington, 2017.

[5] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[6] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 28–39, 2010.

[7] S. Soltan, A. Loh, and G. Zussman, "A learning-based method for generating synthetic power grids. *IEEE Systems Journal* (in press, 2018).

[8] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the Topological and Electrical Structure of the North American Electric Power Infrastructure," *IEEE Systems Journal*, vol. 6, no. 4, pp. 616–626, Dec. 2012.

[9] M. Newman, "The Structure and Function of Complex Networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, Jan. 2003.

[10] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," *Phys. Rev. E*, vol. 69, no. 2, Feb. 2004.

[11] P. Hines, S. Blumsack, E. Cotilla Sanchez, and C. Barrows, "The Topological and Electrical Structure of Power Grids," in *Proc. 2010 43rd Hawaii Int. Conf. System Sciences*, Koloa, HI, USA, Jan. 2010, pp. 1–10.

[12] E. Cotilla-Sanchez, P. D. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure. *IEEE Systems Journal*, vol. 6, 2012, pp. 616-626.

[13] Y. Yang, T. Nishikawa, and A. E. Motter, "Small vulnerable sets determine large network cascades in power grids. *Science*, vol. 268, 2017, p. 886.

[14] B. Shafer, D. Witthaut, M. Timme, and V. Latora, "Dynamically induced cascading failures in power grids. *Nature Communications*, vol. 9, 2018, p. 1975.

[15] G. A. Pagani and M. Aiello, "The Power Grid as a complex network: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 11, pp. 2688–2700, Jun. 2013.

[16] Power Systems Test Case Archive. [Online]. Available: https://www2.ee.washington.edu/research/pstca/

[17] N. Hutcheon and J. W. Bialek, "Updated and validated power flow model of the main continental European transmission network," *2013 IEEE PowerTech Conference, Grenoble, France*, 2013, pp. 1-5.

[18] W. Medjroubi, U. P. Muller, M. Sharf, C. Matke, and D. Kleinhans, "Open data in power grid modelling: new approaches towards transparent grid models," *Energy Reports*, vol. 3, pp. 14-21, Nov. 2017.

[19] K. M. Gegner, A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "A Methodology for the Creation of Geographically Realistic Synthetic Power Flow Models," in *Proc. 2016 IEEE Power and Energy Conf. at Illinois*, Champaign, IL, Feb. 2016.

[20] S. Soltan and G. Zussman, "Generation of synthetic spatially embedded power grid networks," *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, 2016, pp. 1-5.

[21] J. Hu, L. Sankar and D. J. Mir, "Cluster-and-Connect: An algorithmic approach to generating synthetic electric power network graphs," *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, 2015, pp. 223-230.

[22] B. Cloteaux, "Limits in modeling power grid topology," *2013 IEEE 2nd Network Science Workshop (NSW)*, West Point, NY, 2013, pp. 16-22.

[23] S. H. Elyas and Z. Wang, "Improved synthetic power grid modeling with correlated bus type assignments," in *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3391-3402, Sept. 2017.

[24] M. H. Athari and Z. Wang, "Introducing Voltage-level Dependent Parameters to Synthetic Grid Electrical Topology," in *IEEE Transactions on Smart Grid*, in press, 2018.

[25] R. Espejo, S. Lumbreras and A. Ramos, "A Complex-Network Approach to the Generation of Synthetic Power Transmission Networks," in *IEEE Systems Journal*, in press, 2018.

[26] D. Deka, S. Vishwanath and R. Baldick, "Analytical Models for Power Networks: The Case of the Western U.S. and ERCOT Grids," in *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2794-2802, Nov. 2017.

[27] A. B. Birchfield, "The creation of synthetic grids: preliminary considerations," M.S. thesis, Dept. of ECE, Univ. of Illinois at Urbana-Champaign, Urbana, IL, 2016.

[28] K. M. Gegner, A. B. Birchfield, Ti Xu, K. S. Shetye and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," *2016 IEEE Power and Energy Conference at Illinois (PECI)*, Urbana, IL, 2016, pp. 1-6.

[29] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye and T. J. Overbye, "Statistical considerations in the creation of realistic synthetic power grids for geomagnetic disturbance studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502-1510, March 2017.

[30] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258-3265, Jul. 2017.

[31] U.S. Census Bureau. 2010 Census Gazetteer Files: ZIP Code Tabulation Areas. [Online]. Available: https://www.census.gov/geo/maps-data/data/gazetteer2010.html

[32] [U.S. Energy Information Association. Form EIA-860, 2014. [Online]. Available: http://www.eia.gov/electricity/data/eia860/index.html

[33] D. B. West, Introduction to Graph Theory. Upper Saddle River, NJ: Prentice Hall, 1996.

[34] F. M. Preparata and M. I. Shamos. Computational Geometry: An Introduction. New York: Springer-Verlag, 1985.

[35] M. S. Smit, "Epidemic Delaunay," 2009. [Online]. Available: http://graphics.tudelft.nl/matthijss/epidemicdelaunay/paper.pdf.

[36] T. Xu, A. B. Birchfield, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Application of large-scale synthetic power system models for energy economic studies," *2017 50th Hawaii International Conference on System Sciences*, January 2017.

[37] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of Synthetic Electric Grid Models for Transient Stability Studies," *Bulk Power Systems Dynamics and Control Symposium (IREP 2017)*, Espinho, Portugal, September 2017.

[38] T. Xu, A. B. Birchfield and T. J. Overbye, "Modeling, tuning and validating system dynamics in synthetic electric grids," in *IEEE Transactions on Power Systems*, to appear, 2018.

[39] H. Li, A. L. Bornsheuer, T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Load modeling in synthetic electric grids," *2018 Texas Power and Energy Conference,* Feb. 8-9 2018.

[40] A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "Building synthetic power transmission networks of many voltage levels, spanning multiple areas," *2018 51st Hawaii International Conference on System Sciences*, January 2018.

[41] A. B. Birchfield, T. Xu, and T. J. Overbye, "Power flow convergence and reactive power planning in the creation of large synthetic grids," *IEEE Transactions on Power Systems*, to appear, 2018.

[42] A. B. Birchfield, E. Schweitzer, H. Athari, T. Xu, T. J. Overbye, A. Scaglione, and Z. Wang, "Validation metrics to assess the realism of synthetic power grids," *Energies*, vol. 10, no. 8, p. 1233, Aug. 2017.

[43] A. B. Birchfield and T. J. Overbye, "Techniques for drawing geographic one-line diagrams: Substation spacing and line routing," *IEEE Transactions on Power Systems*, to appear, 2018.

[44] A. B. Birchfield, T. J. Overbye, and K. R. Davis, "Educational applications of large synthetic power grids," *IEEE Transactions on Power Systems*, to appear, 2018.

[45] B. Stott, J. Jardim, and O. Alsac, "Dc power flow revisited," *IEEE Transactions on Power Systems*, vol. 24, pp. 1290-1300, 2009.

[46] R. J. Bennon, J. A. Juves and A. P. Meliopoulos, "Use of Sensitivity Analysis in Automated Transmission Planning," in *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 1, pp. 53-59, Jan. 1982.

[47] W. F. Tinney and C. E. Hart, "Power flow solution by Newton's method," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-86, no. 11, pp. 1449-1460, Nov. 1967.

[48] S. Abe, N. Hamada, A. Isono, and K. Okuda, "Load flow convergence in the vicinity of a voltage stability limit" *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, No. 6, pp. 1983-1993, Nov. 1978.

[49] S. S. Tripathy, G. D. Prasad, O. P. Malik, and G. S. Hope, "Load-flow solutions for ill-conditioned power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 10, pp. 3648-3657, Oct. 1982.

[50] S. Iwamoto and Y. Tamura, "A load-flow calculation method for ill-conditioned power systems," *IEEE Transactions on Power Apparatus and Systems*, vol PAS-100, no. 4, pp. 1736-1743, Apr. 1981.

[51] C. L. DeMarco and T. J. Overbye, "Low voltage power flow solutions and their role in exit time based security measures for voltage collapse," *Proc. of the 27th IEEE Conference on Decision and Control*, 1988.

[52] J. S. Thorp and S. A. Naqavi, "Load flow fractals," *Proc. of the 28th IEEE Conference on Decision and Control*, 1989.

[53] W. Murray, T. T. De Rubira, and A. Wigington, "Improving the robustness of Newton-based power flow methods to cope with poor initial points," *Proc. 2013 North American Power Symposium (NAPS)*, 2013.

[54] T. J. Overbye, "A power flow measure for unsolvable cases," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1359-1365, Aug. 1994.

[55] S. Granville, J. C. O. Mello, an A. C. G. Melo, "Application of interior point methods to power flow unsolvability," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 1096-1102, May 1996.

[56] D. K. Molzahn, V. Dawar, B. C. Lesieutre, and C. L. DeMarco, "Sufficient conditions for power flow insolvability considering reactive power limited generators with applications to voltage stability margins," *2013 IREP Symposium-Bulk Power System Dynamics and Control (IREP)*, Aug. 2013.

[57] J. Deng, T. Zhao, H. Chiang, Y. Tang, and Y. Wang, "Convergence regions of Newton method in power flow studies: numerical studies," *2013 IEEE International Symposium on Circuits and Systems (ISCAS 2013)*, 2013.

[58] K. Iba, "Reactive Power Optimization by Genetic Algorithm," in *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 685-692, May 1994.

[59] K. Y. Lee and F. F. Yang, "Optimal Reactive Power Planning Using Evolutionary Algorithms: A Comparative Study for Evolutionary Programming, Evolutionary Strategy, Genetic Algorithm, and Linear Programming," in *IEEE Transactions on Power Systems*, vol. 13, no. 1, pp. 101-108, Feb. 1998.

[60] W. Zhang, F. Li and L. M. Tolbert, "Review of Reactive Power Planning: Objectives, Constraints, and Algorithms," in *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2177-2186, Nov. 2007.

[61] J. C. Lopez, J. Contreras, J. I. Munoz, and J.R.S. Mantovani, "A Multi-Stage Stochastic Non-Linear Model for Reactive Power Planning Under Contingencies," in *IEEE Transactions on Power Systems*, vol 28, no. 2, pp. 1503-1514, May 2013.

[62] National Academies of Sciences, Engineering, and Medicine, *Analytic research foundations for the next-generation electric grid*. The National Academies Press, Washington, DC, 2016.

[63] C. W. Taylor, "Reactive power today: best practices to prevent blackouts," in *IEEE Power and Energy Magazine*, vol. 4, no. 5, pp. 101-104, Sept. 2006.

[64] S. Noguchi, M. Shimomura, J. Paserba, and C. Taylor, "Field verification of an advanced high side voltage control at a hydro power station," in *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 693-701, May 2006.

[65] "Regional Power Control Device Coordination Study: Final Report" IESO, MISO, NYISO, and PJM. June 1, 2011.

[66] I. Kormaz, "Gladstone Phase Shifter Qualification," Tri-State Generation and Transmission Association, Inc. Sept. 30, 2014.

[67] "Transmission and Security Desk Operating Procedure Manual" ERCOT. Feb. 25, 2014.

[68] "Special Reliability Assessment Interim Report: Effects of Geomagnetic Disturbances on the Bulk Power System," NERC, Feb. 2012.

[69] *Reliability Standard for Geomagnetic Disturbance Operation*, FERC, Docket No. RM14-1-000, Order No. 797, Jun. 2014.

[70] *Transmission System Planned Performance for Geomagnetic Disturbance Events*, NERC Std. TPL-007-1, Jun. 2014.

[71] R. Horton, D. H. Boteler, T. J. Overbye, R. Pirjola, and R. C. Dugan, "A Test case for the calculation of geomagnetically induced currents," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2368–2373, Oct. 2012.

[72] "Benchmark Geomagnetic Disturbance Event Description," NERC, 2014.

[73] T. J. Overbye, T. R. Hutchins, K. Shetye, J. Weber, S. Dahman, "Integration of geomagnetic disturbance modeling into the power flow: A methodology for

large-scale system studies," in *Proc. 2012 North American Power Symp.,* Champaign, IL, USA, Sept. 2012, pp.1-7.

[74] U. Buy, T. J. Overbye, K. S. Shetye, H. Zhu, and J. Weber, "Geomagnetically induced current sensitivity to assumed substation grounding resistance," in *Proc. 2013 North American Power Symp.*, Manhattan, KS, 2013, pp. 1-6.

[75] "IEEE Guide for Safety in AC Substation Grounding," in IEEE Std. 80-2013, pp. 1-226, May 2015.

[76] R. Tamassia, *Handbook of Graph Drawing and Visualization*, Boca Raton, Florida: CRC Press, 2014.

[77] I. G. Tollis, G. Di Battista, P. Eades, and R. Tamassia, *Graph Drawing: Algorithms for the Visualization of Graphs*, 1st ed., 1998.

[78] P. Eades, "A Heuristic for Graph Drawing," *Congressus Numerantium*, vol. 42, pp. 149-160, 1984.

[79] T. M. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and experience*, vol. 21, no. 11, pp. 1129–1164, 1991.

[80] Y. Hu, "Efficient, high-quality force-directed graph drawing," in *The Mathematica Journal*, vol. 10, no. 1, pp. 37-71, 2006.

[81] C. Walshaw, "A multilevel algorithm for force-directed graph drawing" in *Journal of Graph Algorithms and Applications*, vol. 7, no. 3, pp. 253-285, 2003.

[82] D. Holten, "Hierarchical edge bundles: visualization of adjacency relations in hierarchical data," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 741-748, 2006.

[83] C. Lin, H. Yen, "A new force-directed graph drawing method based on edge-edge repulsion" in *Journal of Visual Languages and Computing*, vol. 23, no. 1, pp. 29-42, 2012.

[84] P. Eades, Q Feng, and X. Lin, "Straight-line drawing algorithms for hierarchical graphs and clustered graphs," in *Proc. Symposium on Graph Drawing*, pp. 113-128, 1996.

[85] M. Fink, et al., "Drawing metro maps using Bezier curves," in *International Symposium on Graph Drawing*, pp. 463-474, 2012.

[86] S. Kieffer, T. Dwyer, K. Marriot, and M. Wybrow, "HOLA: Human-like orthogonal network layout," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 349-358, 2016.

[87] W. Cui, et al., "Geometry-based edge clustering for graph visualization," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1277-1284, 2008.

[88] M. Zinsmaier, U. Brandes, O. Deussen, and H. Strobelt, "Interactive level-of-detail rendering of large graphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, pp. 2486-2495, 2012.

[89] H. C. Purchase, R. F. Cohen, and M. James, "Validating graph drawing aesthetics," in *Proc. Symposium on Graph Drawing*, pp. 435-446, Passau, Germany, 1995.

[90] W. Huang, M. L. Huang, C. Lin, "Evaluating overall quality of graph visualizations based on aesthetics aggregation," in *Information Sciences*, vol. 3300, no. 10, pp. 444-454, 2016.

[91] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[92] M. Newman, "The Structure and Function of Complex Networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, Jan. 2003.

[93] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the Topological and Electrical Structure of the North American Electric Power Infrastructure," *IEEE Systems Journal*, vol. 6, no. 4, pp. 616–626, Dec. 2012.

[94] G. A. Pagani and M. Aiello, "The Power Grid as a complex network: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 11, pp. 2688–2700, Jun. 2013.

[95] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye and T. J. Overbye, "Statistical considerations in the creation of realistic synthetic power grids for geomagnetic disturbance studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502-1510, March 2017.

[96] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258-3265, July 2017.

[97] B. J. Parker, R. F. Chao, J. K. M. Sabiston, and P. Locke, "An analytical technique to evaluate station one-line diagrams in a network context," in *IEEE Transactions on Power Delivery*, vol. 6, no. 4, pp. 1454-1461, 1991.

[98] P. M. Mahadev, R. D. Christie, "Minimizing user interaction in energy management systems: Task adaptive visualization," in *IEEE Transactions on Power Systems*, vol. 11, no. 3, 1996.

[99] T. Overbye and J. Weber, "New methods for the visualization of electric power system information," in *IEEE Symposium on Information Visualization*, pp. 131-160, 2000.

[100] Y. S. Ong, H. B. Gooi, and C. K. Chan, "Algorithms for automatic generation of one-line diagrams," in *IEE Proceedings – Generation, Transmission and Distribution*, vol. 147, no. 5, pp. 292-298, Sept. 2000.

[101] Z. Yongli and O. P. Malik, "Intelligent automatic generation of graphical one-line substation arrangement diagrams," in *IEEE Transactions on Power Delivery*, vol. 18, no. 3, pp. 729-735, 2003.

[102] J. C. Moreira, E. Miguez, C. Vilacha, and A. Otero, "Large-scale network layout optimization for radial distribution networks by parallel computing: Implementation and numerical results," *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1468-1467, 2012.

[103] G. Ravikumar, Y. Pradeep, and S. A. Khaparde, "Graphics model for power systems using layouts and relative coordinates in CIM framework," in *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3906-3915, Nov. 2013.

[104] S. C. Teja and P. K. Yemula, "Power network layout generation using force-directed graph technique," *2014 Eighteenth National Power Systems Conference (NPSC)*, pp. 1-6. Guwahati, India, 2014.

[105] A. de Assis Mota and L. T. M. Mota, "Drawing meshed one-line diagrams of electric power systems using a modified controlled spring embedder algorithm enhanced with geospatial data." *Journal of Computer Science*, vol. 7, no. 2, pp. 234-241, 2011.

[106] P. Cuffe and A. Keane, "Novel quality metrics for power system diagrams," *2016 IEEE International Energy Conference (ENERGYCON)*, Leuven, 2016, pp. 1-5.

[107] P. Cuffe and A. Keane, "Visualizing the electrical structure of power systems," in *IEEE Systems Journal*, vol. 11, no. 3, pp. 1810-1821, 2017.

[108] H. H. Woodson, F. C. Schweppe and G. L. Wilson, "Electric power systems engineering education in a modern curriculum," in *Proceedings of the IEEE*, vol. 59, no. 6, pp. 860-868, June 1971.

[109] F. E. Terman, "A brief history of electrical engineering education," in *Proceedings of the IEEE*, vol. 64, no. 9, pp. 1399-1407, Sept. 1976.

[110] A. P. Meliopoulos, A. Feliachi, A. G. Bakirtzis and G. Cokkinides, "Development of courses on power system energy control centers," in *IEEE Transactions on Education*, vol. 27, no. 2, pp. 66-72, May 1984.

[111] A. Domijan and R. R. Shoults, "Electric power engineering laboratory resources of the United States of America and Canada," in *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 1354-1360, Aug. 1988.

[112] P. W. Rush, "Developments in power system simulation education and training," *IEE Colloquium on Power System Simulation*, London, 1989, pp. 4/1-4/4.

[113] C. N. Lu and M. Unum, "Interactive simulation of branch outages with remedial action on a personal computer for the study of security analysis [of power systems]," in *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1266-1271, Aug 1991.

[114] C. I. Hatziadoniu, M. Daneshdoost, R. Y. Shaat and X. Cheng, "Application independent interactive environment for power systems education," in *IEEE Transactions on Power Systems*, vol. 6, no. 4, pp. 1418-1424, Nov 1991.

[115] J. H. Chow and K. W. Cheung, "A toolbox for power system dynamics and control engineering education and research," in *IEEE Transactions on Power Systems*, vol. 7, no. 4, pp. 1559-1564, Nov 1992.

[116] T. J. Overbye, P. W. Sauer, C. M. Marzinzik and G. Gross, "A user-friendly simulation program for teaching power system operations," in *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 1725-1733, Nov. 1995.

[117] T. Ohashi and T. Ono, "Visualization and interactive simulation for power system operation and education," *Fourth International Conference on Advances in Power System Control, Operation and Management, APSCOM-97*. (Conf. Publ. No. 450), 1997, pp. 802-807 vol.2.

[118] J. Yang and M. D. Anderson, "PowerGraf: an educational software package for power systems analysis and design," in *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1205-1210, Nov 1998.

[119] A. Abur, F. Magnago and Y. Lu, "Educational toolbox for power system analysis," in *IEEE Computer Applications in Power*, vol. 13, no. 4, pp. 31-35, Oct 2000.

[120] V. A. Levi and D. P. Nedic, "Application of the optimal power flow model in power system education," in *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 572-580, Nov 2001.

[121] M. Crow, G. Gross and P. W. Sauer, "Power system basics for business professionals in our industry," in *IEEE Power and Energy Magazine*, pp. 16-18,20, Jan/Feb 2003.

[122] P. W. Sauer, G. T. Heydt and V. Vittal, "The state of electric power engineering education," in *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 5-8, Feb. 2004.

[123] P. W. Sauer, M. Crow, and M. Venkata, "Manpower development: Industry and educators need to work together," in *IEEE Power and Energy Magazine*, pp. 30-33, Jan/Feb 2005.

[124] G. F. Reed and W. E. Stanchina, "Smart grid education models for modern electric power system engineering curriculum," *IEEE PES General Meeting*, Minneapolis, MN, 2010, pp. 1-5.

[125] A. S. Deese, "Development of smart electric power system (SEPS) laboratory for advanced research and undergraduate education," in *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1279-1287, May 2015.

[126] Q. Hu, F. Li and C. f. Chen, "A smart home test bed for undergraduate education to bridge the curriculum gap from traditional power systems to modernized smart grids," in *IEEE Transactions on Education*, vol. 58, no. 1, pp. 32-38, Feb. 2015.

[127] Q. Wang, Z. Tang, I. Knezevic, J. Yu and G. Karady, "Power system protection education and digital relay training based on a physical platform," *2016 North American Power Symposium (NAPS)*, Denver, CO, 2016, pp. 1-5.

[128] R. D. Zimmerman, C. E. Murillo-Sanchez and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," in *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12-19, Feb. 2011.

[129] S. Cole and R. Belmans, "MatDyn, A new Matlab-based toolbox for power system dynamic simulation," in *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1129-1136, Aug. 2011.

[130] L. Vanfretti and F. Milano, "Facilitating constructive alignment in power systems engineering education using free and open-source software," in *IEEE Transactions on Education*, vol. 55, no. 3, pp. 309-318, Aug. 2012.

[131] N. Mohan, W. P. Robbins and B. F. Wollenberg, "Power systems education based on CUSP™-curriculum," in *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1896-1902, July 2014.

[132] D. Strickland, L. Jenkins, S. Luke, J. Andrews and B. Wood, "Electrical power systems education for employment," *2017 52nd International Universities Power Engineering Conference (UPEC)*, Heraklion, 2017, pp. 1-6.

[133] P. C. Kotsampopoulos, V. A. Kleftakis and N. D. Hatziargyriou, "Laboratory education of modern power systems using PHIL simulation," in *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3992-4001, Sept. 2017.

[134] A. K. Srivastava, A. L. Hahn, O. O. Adesope, C. H. Hauser and D. E. Bakken, "Experience with a multidisciplinary, team-taught smart grid cyber infrastructure course," in *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2267-2275, May 2017.

[135] M. A. Cohen, G. O. Niemeyer and D. S. Callaway, "Griddle: Video gaming for power system education," in *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3069-3077, July 2017.

[136] T. J. Overbye and J. D. Weber, "Visualization of power system data," Proceedings of the *33rd Annual Hawaii International Conference on System Sciences*, 2000, pp. 7.

[137] J. D. Weber and T. J. Overbye, "Voltage contours for power system visualization," in *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 404-409, Feb 2000.

[138] Electric reliability council of Texas (ERCOT) report, "Generation resource retirements in the ERCOT region." November 2017. [Online]. Available: http://www.ercot.com/content/wcm/lists/114740/Retirements_in_ERCOT_FINAL_Nov2017.pdf