Physics-Aware Optimization and Data-Driven Methods for Low-Carbon Power Systems

by

Rabab Haider

B.Sc., University of Toronto (2018)
S.M., Massachusetts Institute of Technology (2020)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Author .....
Department of Mechanical Engineering
May 15, 2023

Certified by .....
Anuradha M. Annaswamy
Senior Research Scientist
Thesis Supervisor

Accepted by .....
Nicolas Hadjiconstantinou
Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Theses

Physics-Aware Optimization and Data-Driven Methods for Low-Carbon Power Systems

by

Rabab Haider

Submitted to the Department of Mechanical Engineering
on May 15, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

The US electricity sector is undergoing a transformation with aggressive targets to achieve 100% carbon pollution-free electricity by 2035. To achieve this objective while maintaining a safe and reliable power grid in the presence of intermittent renewable generation, new operating paradigms of computationally fast and accurate decision making in dynamic and safety-critical environments are needed. To this end, this thesis focuses on answering three questions: How can we enable dynamic (fast + frequent) decision making for safety-critical applications in the presence of integer constraints? How do we coordinate distributed grid-edge devices across multiple timescales and ownership boundaries? How do we develop and evaluate algorithms without access to real-world data? To address these questions this thesis proposes two physics-aware optimization frameworks that coordinate grid-edge resources towards meeting three goals: improving grid efficiency, ensuring grid operability, and supporting clean energy directives.

First, we propose Grid-SiPhyR (Sigmoidal Physics-Informed Rounding; *pronounced as: 'cipher'*), a physics-informed machine learning framework for end-to-end learning to optimize for combinatorial problems, and apply it to the dynamic grid reconfiguration problem. Grid-SiPhyR employs a novel physics-informed rounding approach to tackle the mixed integer nature of dynamic reconfiguration while satisfying salient safety-critical operating constraints. Offline training of the unsupervised framework on representative load and generation data makes dynamic decision making via the online application of Grid-SiPhyR computationally feasible. Second, we propose a physics-aware distributed coordination architecture for grid-edge devices, upon which two grid services are developed. We first develop a hierarchical coordination approach for voltage regulation to coordinate slow-timescale utility-owned devices with fast-timescale solar generation towards managing grid voltages. We then develop a load ramp mitigation service to coordinate the actions of distributed storage resources to provide aggregated support at the bulk level. Lastly, we address the third question through the development of synthetic datasets with representative load and generation characteristics.

Thesis Supervisor: Anuradha M. Annaswamy
Title: Senior Research Scientist

Acknowledgments

There are many people to thank in my academic journey and time at MIT. First, I'd like to thank my advisor Dr. Anuradha Annaswamy, without whom this research would not be possible. I've grown immensely during my PhD and am thankful for your support and mentorship. I appreciate the flexibility and support you've provided in allowing me to explore different research projects and develop a unique perspective in a very multidisciplinary research space. I would also like to thank the members of my committee, Dr. Marija Ilic and Dr. Faez Ahmed. You both have been supportive of me and my research from the first day, and have provided helpful perspective, direction, and suggestions.

I would like to thank my collaborators for the many interesting discussions, feedback, and impactful projects. Among them: Dr. Giulio Ferro and Dr. Michela Robba at the University of Genoa for our longstanding collaboration, enthusiastic discussions, and consistent encouragement. Dr. Amit Chakraborty and Dr. Biswadip Dey with whom I worked during my internship at Siemens. The direction you provided during my internship were critical to my thesis contributions. Weiwei Yang of Microsoft Research and Dr. Yixing Xu of Breakthrough Energy for your mentorship and support during my internship with Microsoft Research. I had a wonderful time collaborating with a larger team and am energized by each of our conversations. The Grid Innovation Lab at Avista Utilities for many useful discussions. To all: I look forward to continuing our collaborations for many years to come.

There are many in the MechE and MIT community who I would like to acknowledge. I'd like to thank the DEI Task Force which was led with openness and direction. I feel incredibly fortunate to have been on the Members at Large of the DEI Task Force, and to have had the opportunity to give back to the community. I'd additionally like to thank Muslim Chaplain Nada El-Alami and the MSA community, Rachel and the MIT Energy Initiative, the community at Sidney Pacific, and the team at MIT MindHandHeart. In particular, to Chaplain Nada – through your support we launched the Connect Sustainability program, and the Muslim Mental Health Dia-

logues series, projects which are very important to me.

I'd like to acknowledge the members of the Active Adaptive Control Laboratory at MIT with whom I have collaborated and enjoyed the PhD journey. A short list includes: David, Yohan, Patricia, Giulio, Yue, Vineet, Venkatesh, and Priyank. One of the best experiences of my graduate school journey has been working with others, especially the undergraduate and graduate researchers who I've mentored. To Bella, Adam, Swochchhanda, and Jules – it's been a privilege to be a part of your journeys, and to help you navigate the space of clean energy research. To my friends and colleagues who have joined me on my graduate school journey, thank you for providing much needed levity, grace, and support.

Finally, to my family for your infinite love and support. There are no words to describe everything you have done for me, and all the love and care you have showered me with. To my sister, thank you for your constant presence in every aspect of my life. To my husband who has seen first hand every part of this journey – you have always pushed me to be the best version of myself and to be confident and bold in everything I do. To my parents who have always supported me and pushed me forward, this accomplishment is as much yours as it is mine.

This thesis was supported by the MathWorks Mechanical Engineering Fellowship, MIT Energy Fellowship, and the U.S. Department of Energy under Award Number DE-IA0000025. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Contents

List of Figures

4-4 Consider an example with a single DER in the network and the LTC at the PCC. The linear equations of the CI model describes a polytope (indicated by the purple shaded area) showing the feasible space where voltages are within operating bounds, $V_t \in [0.95, 1.05]$. Note that in the illustration the integer points for the LTC are not shown for sake of simplicity in the diagram. Suppose the objective function is linear (for sake of illustration), where $f(V_t, P_t, Q_t)$ is a linear function, and that the integer points for the LTC coincide with the corner points of the polytope. This means the optimal solutions lie at one of the corners of the polytope (of the entire purple shaded area). When the optimization problem of Eq. 4.5c is solved the optimal solution reached is indicated by the purple dot, at $K_t^*, \{P_t^*, Q_t^*\}$ with objective value c_1. This result is that of a single centralized optimization over all possible device actuation, which gives the optimal solution. The direction of the red arrow indicates increasing cost, thus $c_1 < c_2 < c_3$ in this diagram (more generally, the costs may be less than or equal to each other). When the bi-level architecture without any additional coordination between upper and lower levels is used, the resulting solution is indicated by the blue dot, at $K_t^{UL}, \{P_t^{LL}, Q_t^{LL}\}$ with objective value c_3. This solution arises by first optimizing the tap operation and setting K_t^{UL}, then optimizing over the feasible DER operations which lie along the blue dashed line. Finally we consider the proposed approach of adding coordination between the upper and lower level optimization problems. The coordination restricts the operating range of the DER injections, as indicated by the new polytope boundaries in dashed purple. The new optimal solution with coordination is then indicated by the green dot, at $K_t^{Coor}, \{P_t^{Coor}, Q_t^{Coor}\}$ with objective value c_2. Notably the upper level optimization now considers the reduced DER operating range, resulting in a new optimal tap setting K_t^{Coor}. The lower level considers the reduced operating range (a subset of the original operating range) and the voltage at the PCC set by K_t^{Coor}, to arrive at the optimal DER injections $\{P_t^{Coor}, Q_t^{Coor}\}$. 152

List of Tables

Chapter 1

Introduction

The IPCC's Sixth Assessment Report asserts the unequivocal impact of human activities on climate change, pointing to the increase in atmospheric greenhouse gas concentrations due to emissions from human activities [70]. The need for action to mitigate and reverse the effects of climate change have been recognized by the UN Sustainable Development Goal 7 which advocates for affordable and clean energy access at a global scale [107]. To facilitate the growth and access to clean energy requires an unparalleled decarbonization and electrification effort. To achieve a low-carbon future, a transformation of the electricity sector is inevitable. Such a transformation of the U.S. electricity sector is already underway. The Biden administration has set aggressive targets to achieve 100% carbon pollution-free electricity by 2035, and 45% solar power generation by 2050 [121]. The resulting low-carbon grid would introduce new challenges in grid operation as controllable fossil-based generators are replaced with intermittent renewable resources. To maintain grid performance and stability (i.e. keep the lights on), additional flexibility is needed so clean energy is available 24/7. Within the distribution grid we look towards exploiting the flexibility offered by small-scale customer-owned distributed energy resources (DERs) which include rooftop solar, distributed batteries, and smart appliances. Although DERs are small in individual capacity, they can be used in aggregate to provide operational flexibility. New operating paradigms and coordination mechanisms are required to fully integrate DERs into the grid and utilize their operational flexibility.

The opportunity to provide new, green sources of power necessitates new decision-making paradigms. Current practices of decision making in power systems consist of static optimization tools at slower time-scales for applications in planning and markets, and dynamic tools at faster time-scales for applications in prediction, estimation, and control. The optimization tools currently used at the transmission level naturally presents themselves as a coordination mechanism for DER dispatch as well. However, there remain technical limitations to this operating paradigm with standard out-of-the-box optimization routines. First, the proliferation of DERs introduces dynamic signatures across the power system landscape, pushing for dynamic decision making (dynamic = fast + frequent decisions) which may be computationally prohibitive using current methods. Second, operational decisions must be made over continuous and discrete variables, presenting as a combinatorial mixed integer problem (consider switch statuses and discrete device setpoints). Third, the set of devices over which control decisions are made span multiple timescales (slow-actuating vs. fast-actuating) and multiple ownership boundaries (operator or utility-owned vs. privately- or customer-owned). Finally, and most importantly, the power grid can be viewed as a safety-critical system as electricity is an essential service and must be available, particularly for critical loads like hospitals and police/fire stations. All dispatch decisions – optimal or not – must satisfy power physics constraints and device operating constraints at all times, and must be treated as hard constraints.

Dynamic decisions in the presence of hard constraints and the other elements noted above then require new tools. These tools must ensure feasibility at all times while seeking optimality. The foundation of these tools must be the available knowledge of the physical system such as steady-state equations representing power physics. These algorithms must also extend to decisions across multiple independent agents without requiring complete data or information sharing. In this thesis we propose physics-aware optimization frameworks which embed within the decision architecture the salient physical equations and operating constraints, towards decision making in dynamic and distributed paradigms.

To this end, this thesis focuses on answering three questions: How can we enable

dynamic decision making (fast + frequent decisions) for safety-critical applications in the presence of integer constraints? How do we coordinate distributed grid-edge devices across multiple timescales and ownership boundaries? How do we develop and evaluate algorithms without access to real-world data? The frameworks proposed in this thesis coordinate grid-edge resources towards meeting three goals: improving grid efficiency, ensuring grid operability, and supporting clean energy directives.

1.1 Contributions of this Thesis

This thesis is organized as follows. Chapter 2 presents a primer on electric grid fundamentals required to understand the technical challenges of low-carbon power systems tackled in this thesis. This section also provides a brief review of the voltage regulation problem, and a discussion of machine learning for the power flow problem using physics-informed techniques.

Chapter 3 tackles the first question. The problem of distribution grid reconfiguration is introduced, an extension to the power flow problem as follows. Power flow searches for a feasible solution (of nodal voltages and angles) to known generator dispatch and load conditions. Grid reconfiguration searches for an optimal solution of both grid topology and generator dispatch (with feasible nodal voltages) to satisfy known load conditions. The shift from feasibility towards optimality is further discussed in the presentation of results. The grid reconfiguration problem is mathematically stated as a mixed integer linear program (MILP), with linear representation of power physics constraints and big-M relaxation of conditional constraints. The combinatorial nature of the grid reconfiguration task is highlighted, and a discussion of utility company perspectives on dynamic reconfiguration is presented. The first main contribution of the thesis is then presented: a physics-informed machine learning framework with suitable modifications to enable end-to-end learning for combinatorial problems. The proposed framework, denoted as Grid-SiPhyR (Sigmoidal Physics-Informed Rounding; *pronounced as: cipher*), is applied to the grid reconfiguration problem. Extensive testing is carried out to train multiple neural architectures

leveraging the physics-informed approach. Simulation results on two canonical distribution grids show that Grid-SiPhyR provides order of magnitude improvement in both optimality and feasibility error metrics as compared to other approaches for integer accommodation which do not leverage (or suitably accommodate) grid physics. Results also show that Grid-SiPhyR contributes to all three design goals: improves grid efficiency by reducing electrical losses, ensures grid operability by improving voltage profile, and supports clean energy directives by increasing solar energy utilization.

Chapter 4 tackles the second question. A physics-aware distributed coordination architecture is proposed as the substrate upon which multiple grid services can be realized. The framework formally integrates a linear power flow model for the unbalanced distribution grid, different DERs including flexible loads, distributed solar, and storage, and a distributed optimization approach to optimal power flow. Two grid services are developed upon this architecture: voltage regulation and load ramp mitigation. These two services highlight the flexibility of the proposed physics-aware distributed coordination architecture by addressing pressing concerns for grid operators at the local level (within the distribution grid) and bulk level (in the transmission grid). The local service of voltage regulation is designed as a bi-level optimization problem, with the lower level optimizing distributed DER dispatch and the upper level optimizing load tap changer (LTC) switch operations. A main contribution of this thesis is the introduction of a hierarchical coordination mechanism between the lower and upper levels: the proposed approach takes advantage of the timescale separation between fast-actuating DERs and slow-actuating LTC (as typically done in literature) and introduces minimal communication between these decision layers to coordinate optimal decisions (new to this approach). Offline simulation studies on a real utility feeder validate that the proposed hierarchical coordination for voltage regulation contributes to all three design goals: improves grid efficiency by reducing network losses and extending the lifespan of LTC devices, ensures grid operability by improving voltage profile, and supports clean energy directives by supporting DER integration and increasing the utilization of DER capabilities. Having shown the potential for DERs to provide services to the local distribution grid, the next ser-

vice considers aggregating DERs to provide services to the transmission grid. The bulk level service of load ramp mitigation coordinates distributed storage devices to flatten the net load profile seen at the transmission-distribution substation. The inter-temporal storage constraints are modelled within the distributed optimization framework which determines the optimal hourly dispatch signals for storage devices. The distributed optimization based service is evaluated against a local approach (i.e. without the physics-aware architecture), on a case study of San Francisco, California, using a modified IEEE-34 node grid under a high penetration of solar PV, flexible loads, and battery units. Simulation results show considerable reduction in load ramp can be achieved with the proposed service, which translates to reduced carbon emissions and reduced costs from operating gas-fired combustion "peaker" plants at the transmission level.

The third question is addressed in both Chapters 3 and 4, and is another main contribution of this thesis. Multiple datasets are generated for canonical distribution grids using standard grids found in literature (including IEEE test feeders). The developed datasets can support algorithm development, testing of common operational scenarios, and evaluation of proposed future operating paradigms. Each of the grids are modified to include moderate to high penetration of DERs, including distributed generation like inverter-based solar, electrical storage, and flexible loads. In Chapter 3 datasets for reconfiguration are developed for two canonical distribution grids with switches by introducing a moderate penetration of grid-edge devices and developing representative load and generation profiles. These datasets include both residential and commercial loads with weekend-weekday and seasonal variation, and introduce distributed solar resources throughout the grid at penetration levels consistent with current rates of adoption across the U.S.. In Chapter 4 datasets are developed for the unbalanced distribution grid in the presence of single-phase loads and generation. A combination of residential and commercial loads are included and various ownership boundaries are elucidated.

Chapter 5 presents contributions in three additional research topics, which extend the three main questions posed in this thesis. The first is on the design of

retail electricity markets which provide the transactive energy platform with which DER services – such as those proposed in Chapters 4 – can be realized. The second is the design of accelerated distributed optimization methods to improve the computational capabilities of distributed methods and increase the real-world applicability of distributed computations. The last is the design of energy management platforms for interconnected energy hubs (ex. electricity + heating + transportation) which may become more prevalent with increased electrification and smart city efforts. These projects are the outcome of extensive collaborations and present the depth and breadth of research completed in this thesis.

Chapter 6 presents a discussion on state-of-art and emerging technologies for optimization and control of future distribution grids. Comments are made as to how the algorithms developed in this thesis can be implemented in field, noting the need for situational awareness, widespread communication infrastructure, and concerted effort to develop operator trust in machine learning methods. Existing grid modernization projects are also presented. Finally, Chapter 7 presents conclusions and opportunities for future research directions that build upon the results of this thesis.

Chapter 2

Background

2.1 Electric Power Grid

The power grid has traditionally been composed of four parts: Generation, Transmission, Distribution, and Customers. In this traditional model, electric power flows in a single direction from large centralized generators to spatially distributed customers through the electricity delivery infrastructure (transmission + distribution). Generators are typically coal or gas fired combustion plants, nuclear power plants, or large-scale hydroelectric dams. These are all large infrastructure elements that are capital intensive and rely on economies of scale (constantly producing a lot of power by running at full capacity) to recover capital investment costs and operating costs (including fuel costs). Generators are located at a significant distance from load centers (i.e. where customers reside) and rely on the transmission and distribution grids to transport power to customers. The transmission system operates at a high voltage to efficiently transport power across long distances (think on the order of 500 km, or 300 miles which is the distance of approximately 11.5 marathons) from the generators to load centers. Once at the load center, the transmission system connects to the distribution system via a substation transformer where the electricity is 'stepped down' to a lower voltage level for safe distribution. The distribution system delivers the electricity to customers, where upon flipping a light switch or plugging in a laptop, the electricity is 'consumed'.

Electricity travels a long distance from where it is generated to where it is consumed. This may seem similar to other goods, like agricultural products which are grown on farms at a significant distance to where they are eventually sold, purchased, and consumed. However, unique to electricity is that generation and consumption are *instantaneous*: the electrons which flow through a light bulb upon flipping a switch are being generated at the same moment by a turbine spinning hundreds of miles away. **This presents the central challenge to electric power systems: generation and load must be balanced at all times**. Traditionally, this balancing has been done on the generation side, since combustion turbines and hydroelectric dams are controllable (also called 'dispatchable') and provided the requisite flexibility for load-following behaviours. This presents a very tight coupling between load and generation, and a singular mode of flexibility in the load-following paradigm. While this approach has been successful in operating the grid thus far, it is no longer a viable option in future power systems with high penetration of renewable resources like wind and solar.

2.1.1 Challenges for Low-Carbon Power Systems

While necessary to achieve a carbon pollution-free electric grid, renewable resources like wind and solar present a significant paradigm shift in how the electricity system operates.

Centralized / Distributed and Decentralized: The fleet of traditional generators consists of a small number of large scale power plants. Decisions of how much power each plant should produce at a given time is made using a centralized optimization which minimizes overall system generating costs. In contrast, renewable generation plants are typically smaller in capacity, more spatially distributed, and in larger numbers. In the U.S. there were 644 natural gas-fired combined-cycle units (a common traditional generating plant) in 2017, with most having a capacity of 600 to 700 MW[1]. In comparison, by 2018 the U.S. had over 2500 utility-scale solar photovoltaic (PV, a particular technology for solar electricity generation) generating

[1]Based on data collected by the Energy Information Agency (EIA)

facilities, with most having a capacity of 5MW or less[2]. Continuing to make decisions in a centralized paradigm for a significantly larger number of individual resources may be computationally prohibitive. It must also be noted that the centralized decision paradigm discussed above is for the transmission system. Within the distribution system there was historically little to no generation, and utilities needed very little visibility (ex. sensors, precise knowledge of load characteristics) into the distribution grid to operate it reliably. The emerging widespread distributed generation in the distribution grid (ex. rooftop solar, utility-owned devices) presents a significant shift in operating and control paradigms, with situational awareness becoming necessary towards maintaining grid reliability. To this end, new optimization tools must be developed to enable distributed and/or decentralized decision making while providing situational awareness and visibility into the grid.

Controllable and Dispatchable / Uncontrollable and Non-Dispatchable: Traditional generators rely on different fuel sources to generate power, such as coal, natural gas, and uranium. These fuels can be stored, either in a reserve or on-site, and provide a reliable and consistent supply of fuel[3]. Even hydroelectric generation facilities can store water in a reservoir to be used at a later time. With these technologies, generation facility operators have near full control over how much fuel is put through the generating unit at any given time to generate any amount of power at any given time (with limits on minimum operating levels and minimum operating times). Since traditional generators are fully controllable, they can be "dispatched" to produce a given amount of electricity at any given time. In contrast, renewable generation plants like wind and solar farms are uncontrollable, and can only generate when the wind is blowing or the sun is shining[4]. This phenomenon is known as

[2]Based on 2018 data from the EIA. Note that the EIA classifies utility-scale generating facilities to be those where total generation capacity is 1 MW or greater

[3]It should be noted that, as with any engineered system, there is no 100% reliability-of-supply guarantee, even with these fuels. Event such as the Coal Strike of 1902 in Pennsylvania, and geopolitical events such as the Oil Embargo of 1973 and the Russia-Ukraine war of 2022 impacts the availability or import/export of various fuels including coal, oil, and natural gas, and may cause shortages or inflated prices. Rather the reliability and consistency of supply come from the fact that these fuel sources can be stored for long periods of time.

[4]Similarly for other renewable resources like run-of-river hydro and tidal power. These have a significantly smaller penetration as compared to wind and solar.

intermittency and presents an immediate challenge to both the load-following and centralized (transmission supplies/supports all) paradigms.

Load-following / Load flexibility: As previously mention, the challenge of balancing generation and load was traditionally left to controllable generators. In this paradigm the loads were treated as uncontrollable and generation was used as the single mode of flexibility. However, the influx of intermittent and non-dispatchable renewable resources prohibits such a paradigm. The opposite to load-following could then be generation-following, wherein loads are regulated to follow generation patterns: i.e. when the sun sets, all loads are curtailed. Obviously this is an unacceptable option. Instead of relying on a single mode of flexibility with which to exert control, the paradigm of load flexibility must be used, wherein both generation and loads can be controlled, at least in part. Controllable resources on the generation side (ex. like fast-ramping combustion turbines, or emerging electric storage devices) and on the load side (ex. delaying turning on a dishwasher, setting a higher temperature on the thermostat during the summer, or managed charging of electric vehicles) must both be used. This paradigm of load flexibility presents a significant shift from traditional operations and requires customer engagement to ensure load side resources are available. This further presents a challenge to centralized decision making, as the number of actuators (i.e. generators and devices) and participating agents (ex. generator owner, residential customer) dramatically increases. Further, the actions of flexible loads in addition to intermittent distributed solar generation introduces dynamic signatures across the power system landscape. The optimal decision making tools used to dispatch flexibility across the grid must be capable of making dynamic (dynamic = fast + frequent) decisions.

2.2 Fundamentals of Electric Power Systems

This sections presents a brief review of basic concepts in electrical systems. An interested reader is referred to comprehensive texts on electric power systems and distribution grid modelling [57, 79, 78] for additional details.

Voltage (V): Difference in electric potential energy from a reference, between two nodes, or across an electric element (like a resistor)

Electric current (I): Measures the flow of electric charge through an electrical element (such as a wire, or a resistor such as a load).

Electric elements: These include resistors (with resistance R), capacitors (with capacitance C), and inductors (with inductance L). Resistors are passive elements with resistance values measured in Ohms (Ω). The voltage drop across a resistive element is described by Ohm's Law: $V(t) = RI(t)$. A capacitor is an electric energy storage device with capacitance measured in Farads (F), and describes the energy stored in the electric field between two parallel plates. An inductor is an electric energy storage device with inductance measured in Henrys (H), and describes the energy stored in the magnetic field of a coil. The voltage-current relationship for a capacitor and inductor is $I(t) = C\frac{\mathrm{d}}{\mathrm{d}t}V(t)$ and $V(t) = L\frac{\mathrm{d}}{\mathrm{d}t}I(t)$ respectively. The Fourier Transform is used to represent these electric elements in frequency domain with frequency ω, wherein we define the reactance: $X = \omega L - \frac{1}{\omega C}$; the impedance: $Z = R + jX$; and the admittance: $Y = 1/Z$.

Modeling electric lines: The simplest model of an electric wire is as a lossless element. The next model would include a resistive element which represents the power loss over the wire. The next model would further add an inductance due to the conductor's electrical and electromagnetic properties. Finally, for larger circuits such as a power grid, the electromagnetic interactions with the physical ground (i.e. Earth's surface) can be modeled as a capacitor, as the charge imbalance between the wire and ground can be roughly imagined as two capacitor plates with air as the dielectric medium. The lumped parameter model is frequently used to model electric lines, as it presents the simplest model wherein the electrical elements which exist through the entire length of an electric wire are approximated as a single resistor, inductor, and/or capacitor connected with a lossless wire.

Energy vs. Power: Power is the rate of change of energy.

Real Power (P): In a purely resistive circuit all the energy supplied by a source (say a battery) is dissipated across resistive elements, like electrical losses in the wires

and useful work like lighting a light bulb. $P = IV = I^2R = \frac{V^2}{R}$.

Reactive Power (Q): In a circuit with capacitance and inductances, some of the energy supplied by a source will be stored in the electric and magnetic fields of these elements. Thus, reactive power represents energy exchange between reactive loads and parasitic elements (such as inductors and capacitors), and does not contribute to work done. This "sloshing" back and forth of power increases the overall current flowing in a circuit without contributing to any work, resulting in increased thermal losses due to power line resistance.

Apparent Power (S): Describes the total power in an electric circuit, with $S^2 = P^2 + Q^2$. More generally, $S = V(t)I(t)^* = |I(t)|^2 Z = \frac{|V(t)|^2}{Z^*}$, where $V(t)$ and $I(t)$ may be time-varying, such as in AC systems where sinusoidal voltage and currents are used.

Kirchhoff's Current Law (KCL): fundamental law describing the conservation of charge in an electric circuit. KCL dictates that the sum of currents into a node equals the sum of currents out of a node: $\sum I_{in} = \sum I_{out}$.

Kirchhoff's Voltage Law (KVL): fundamental law describing the conservation of energy in an electric circuit. KVL dictates that the algebraic sum of voltage drops around a loop equals zero: $\sum V = 0$.

2.2.1 Voltage Regulation

Secure and reliable operation of the power system requires the system frequency and voltages to be within certain acceptable limits. The increasing penetration of renewable energy resources and their intermittent nature results in voltage fluctuations throughout the grid and is making this task increasingly challenging. Within the distribution grid, voltage regulation (i.e. keeping voltages within acceptable bounds) is a key operating task and is becoming more challenging with the growth of distributed solar generation. This section presents a brief introduction to voltage regulation. Interested readers are invited to refer to a recent survey paper [125] on voltage regulation in distribution grids for more details.

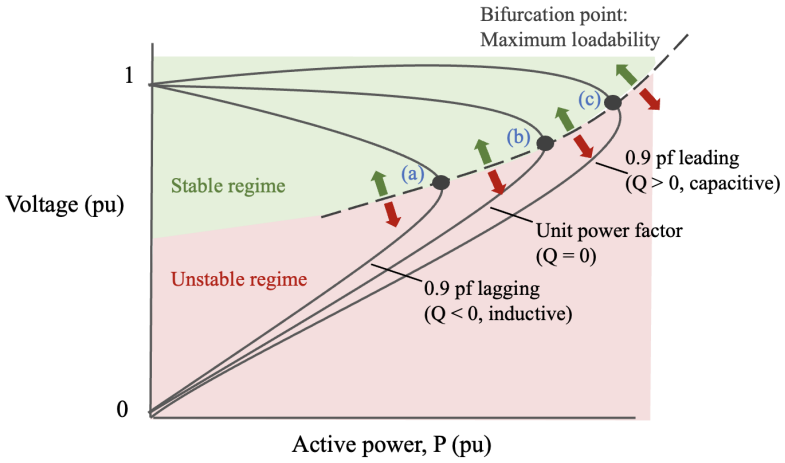


Figure 2-1: Nose curves demonstrating the critical active power values for different values of the reactive power.

Reactive Power and Voltage Regulation

While active power does 'useful work' (i.e. lighting a bulb or running a motor), reactive power is necessary to support system voltages and enable active power transfer. *If the network voltages are too low, real power cannot be supplied; reactive power is needed to provide the voltage levels necessary for real power transfer.* Figure 2-1 shows the P-V curve (often called the 'nose' curve) for different power factors. For a typical network, the power factor will be lagging and reactive power will be drawn from the network by the loads. This restricts the amount of active power that can be delivered to loads without violating the voltage stability, denoted on the figure as (a) on the maximum loadabilty curve connecting the voltage bifurcation points. To increase the active power transfer limit from (a) to (b) or (c), reactive power must be injected (capacitive or leading).

Reactive power compensation then becomes the basic strategy in maintaining grid voltages and voltage stability. As a result, voltage regulation problems are sometimes also labelled as volt-var control (VVC) or volt-var optimization (VVO). While voltage stability margins are monitored and calculated at the transmission level, the basic strategy for voltage stability is to have reactive power compensation located in distribution and load area transmission substations to minimize reactive power trans-

mission. This is particularly effective as reactive power travels poorly, and requiring reactive power to be carried across the transmission network to a load center will increase line currents, incur higher losses, and reduce the real power carrying capacity of lines.

State of Art for Voltage Regulation Architectures

Voltage stability has two timescales – short-term which is on the order of seconds, and long-term which is on the order of tens of seconds to minutes. Both short- and long-term voltage stability are imperative to the safe and reliable operation of the electric power system. Likewise, both fast and slow timescales for voltage regulation are necessary to maintaining stability. To address these varying timescales, voltage regulation can be broken into primary (component control) and secondary (area control and optimization) control.

Transmission system: For a generator in the transmission system, these can correspond to (1) primary: the control of the terminal voltage of the generator through automatic voltage regulation; (2) secondary: the reactive power setpoint to track, as obtained by an area control, through an additional signal (ex. can be computed via solving an optimization problem) to the primary voltage control set-point.

Distribution system: Voltages in the distribution grid have traditionally been controlled using devices such as load tap changing transformers (LTCs), shunt capacitors, and voltage regulators, which have discrete settings and electro-mechanical actuation. Voltage regulators and LTCs adjust the voltage at a particular node, such as the distribution substation. Shunt capacitors (generally, capacitor banks) inject reactive power. While these devices have been successful in managing voltages thus far, they were not designed to handle the new level of variability that comes with high penetration of intermittent distributed generation such as solar. Dynamic Volt-Ampere Reactive (VAR) devices such as smart inverters are thus an attractive addition for achieving voltage control. Many distributed devices, such as distributed solar and electric vehicles and storage are often accompanied by smart inverters, making them capable of providing dynamic VAR services. The proliferation of these devices

throughout the distribution grid enables many more opportunities for achieving voltage regulation. Chapter 4 presents a hierarchical voltage regulation framework which leverages both slow-timescale traditional devices (LTC) and fast-timescale solar to enable effective voltage regulation in the distribution grid.

Voltage Regulation as an Optimization Problem

The voltage regulation problem can be cast as an optimization of the form:

$$\begin{aligned} &\min_{x} \;\; f(x) \\ &\text{s.t.} \begin{cases} h(x) = 0 \\ g(x) \leq 0 \end{cases} \end{aligned} \tag{2.1}$$

where the decision variable x is the reactive power setpoints of the distributed devices (can also include the active power setpoints), and the tap settings of LTC, the latter giving rise to the mixed integer problem tackled in Chapter 4. The objective function f describes system-wide objectives such as real power losses, congestion of real flows, and maintaining adequate voltages throughout the system. The first of these, of loss minimization, is a typical objective in distribution grids since roughly 40% of total power system losses occur at the distribution level. The equality constraints h come from the power flow equations. The inequality constraints g represent the voltage limits, constraints on actuators, and flow limits of the lines. As distributed devices are becoming increasingly common in distribution grids, these actuator constraints can include active and reactive power injection limits of smart inverter-based resources (IBRs), charge or discharge limits of battery devices, or flexibility range of demand response loads for example. The optimization problem in (2.1) will be revisited in detail in Chapter 4.

Solution Methodologies for Voltage Regulation

The survey [125] presents a detailed classification and discussion of voltage regulation strategies. One particular classification is of the solution methodology, which deter-

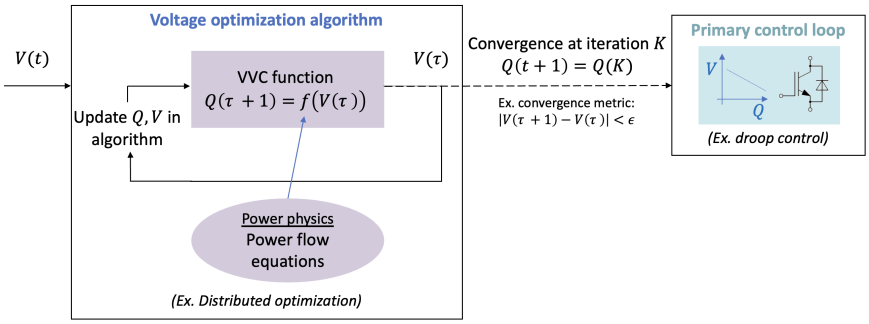
mines how the network measurements are utilized in the algorithmic updates of the voltage optimization problem. A brief discussion is presented here.

Open loop: In the *open-loop* or *offline* methodology shown in Figure 2-2(a), the algorithmic updates are computed using a model of the distribution system and the setpoints are implemented only after the algorithm converges or a certain stopping criterion (number of iterations/desired accuracy) is met. This methodology solves the power flow equations within the optimization loop, and may use the full AC power flow or a convexified or linearized version. In Chapter 4 an open-loop methodology is proposed wherein a distributed optimization algorithm is proposed to solve the voltage regulation problem over the set of distributed devices. The power physics are described using a linear model for the unbalanced distribution grid, in the presence of single-phase loads and distributed generation.

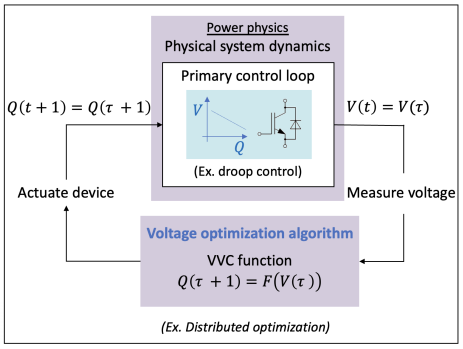
Feedback-based/Closed-loop: In the *feedback-based* or *closed-loop* methodology shown in Figure 2-2(b), the algorithmic updates are computed using voltage measurements of the real system and the setpoints are implemented at every iteration. This data-driven approach avoids dealing with the nonlinear power flow equations coupling V and Q, helping to overcome many of the computational burdens inherent in the AC power flow equations, and communication requirements of open-loop approaches. In this approach, the measurements are used as a *surrogate* or *proxy* for solving the AC power flow equations. It must be noted that the power system dynamics are assumed to be significantly faster than that of the computation loop, such that the physical system settles (i.e. dynamics die out) before the next voltage measurement is taken. With this, this approach is still a steady-state method (as is the open loop method) and does not deal with the underlying dynamics of the power system.

2.2.2 Power Flow Problem

The power flow (PF) problem is a fundamental task in power systems theory and applications, and asks the following question: "Given a set of known loads and generator dispatch points, what are the resulting system voltage magnitudes and angles?". The PF can also involve the calculation of nodal or line currents, however the voltages



(a) **Open-loop:** the optimal dispatch solution (i.e., Q setpoint) is implemented only after the algorithm converges. The power flow equations are used in designing the algorithm update rules (or the function f).



(b) **Feedback-based:** the dispatch solution (i.e. Q setpoint) is implemented at every iteration of the computation scheme. In doing so, feedback from the physical system by way of voltage measurement is used in updating Q at the next iteration/time instant. The implemented setpoint is not optimal at every iteration; rather the algorithm will track an optimal solution through successive iterations.

Figure 2-2: Open-loop and feedback-based solution methodologies for voltage control.

are the primary variables of interest. The PF involves the calculation of power flows (active and reactive) in an electrical network with known topology and line parameters, given some measured or forecasted data. Typically it is solved to determine the feasibility of generation dispatch to service network loads, for contingency analysis to ensure reliable grid operations under equipment failure (ex. loss of a line or generator), and state estimation. The PF problem can be represented mathematically as the relationship between network power (S), voltages (V), and currents (I):

$$f_{PF}(S, V, I) = 0 \tag{2.2}$$

The corresponding power flow equations describe the physical phenomenon of currents flowing through an electrical network (Kirchhoff's Current Law), the corresponding drop in electrical potential across an element (Ohm's Law), and the sum of potential differences about a closed electrical loop (Kirchhoff's Voltage Law). The physical phenomenon of power flow is well understood. However, the *AC power flow* equations are highly nonlinear and nonconvex, and significant efforts have been made to simplify the power flow equations through convexification strategies (such as second-order cone programming, semi-definite programming, and McCormick relaxations) and linearization techniques (such as Taylor expansions). These efforts aim to address the computational difficulty of nonconvexity and nonlinearity by finding exact convex relaxations for the power flow problem. Another area of research is the development of iterative methods to solve the power flow problem quickly, with low memory burden, or with sufficient condition for the non-singularity of the power flow Jacobian.

2.2.3 Optimal Power Flow problem

The optimal power flow (OPF) problem is an extension of the PF problem from feasibility to optimality. It asks the following question: "Given a set of loads what dispatch setpoint is required for each generator to satisfy loads subject to voltage limits and line flow limits, that minimizes a cost function". The cost function may include generating costs, electrical line losses, deviation from a desired voltage profile, disutility of reducing loads, or carbon minimization goals, among others. The design of the cost function and the OPF constraints must reflect the type of decision being made (ex. electricity market dispatch, voltage regulation, etc.). The problem statement can also be extended to include flexible loads where a set of upper and lower limits on loads are presented rather than fixed values, and to include temporal coupling constraints on device actuation for storage or load shifting devices.

2.3 Machine Learning for Power Flow Problem

Recent efforts have also proposed the use of machine learning methods to learn the nonlinear mapping between power flow variables, and offloading the computational burden to offline training using abundant system data. There is a growing body of literature employing ML techniques for various tasks in power systems, including optimal power flow (OPF) [112, 155, 35, 157, 34, 51, 151, 29], probabilistic power flow [146], security constrained unit commitment [62, 113, 142], fault isolation [86], and grid reconfiguration [76, 150, 147, 127].

We introduce a physics-informed neural framework in Fig. 2-3, where the power flow variable space is decomposed into x: training data, z: independent variables, and φ: dependent variables. The neural framework is then tasked with learning the (possibly nonlinear) relationship between x and z, as $z = f_{NN}(x)$, and the remaining variables are calculated by way of the power flow equations, $\varphi = \mathcal{P}(x, z)$. The exact decomposition of power flow variables is non-unique, and different power flow methods and formulations can be represented by such a framework. In each case, a component of the modeling and computational challenge is offloaded to the neural block; however, others may remain. The neural framework in Fig. 2-3 enforces the power flow equations at every iteration, simplifying the learning task for the neural network. This neural framework provides the conceptual substrate for the development of Grid-SiPhyR (Sigmoidal Physics-Informed Rounding; *pronounced as: 'cipher'*) in Chapter 3 to tackle constrained optimization in the presence of integer variables.

Two examples are provided below. In the first, the computational challenge of nonlinear power flow relations is offloaded to the neural block (i.e. the relationship between x and z is a nonlinear one), and system information (topology and line parameters) is used in $\mathcal{P}$ to calculate the remaining variables. In the second example, linearized power flow equations are used. The mapping of nodal loads to voltages is learned by the neural engine, and the properties of the Linearized DistFlow model (assumptions of radiality, balanced lines, and lossless power balance) make the system Jacobian accessible to $\mathcal{P}$. In the second example, the proposed neural framework does

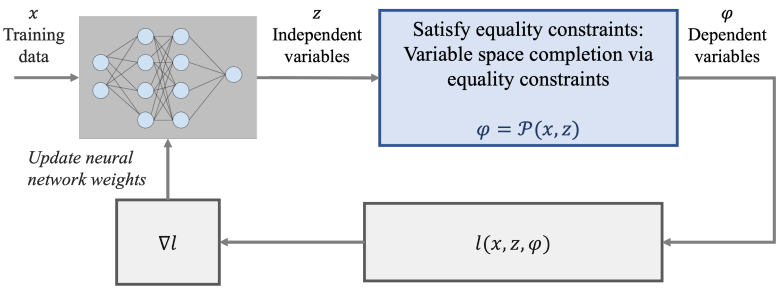


Figure 2-3: A general neural framework for power flow calculations, which decomposes the power flow variables into x : training data, z : independent variables, and φ : dependent variables.

not tackle the Jacobian learning problem [50]. Rather, the framework is well suited for optimal power flow (OPF) problems, where generator dispatch must be determined to satisfy all loads and optimize an objective function. In the OPF problem, there are multiple dispatch points which can satisfy the power flow equations. The resulting challenge is to both satisfy the power flow equations and to search the feasible space for an optimal solution. A brief discussion of the Jacobian learning problem is provided as well.

Backward-Forward Sweep Algorithm

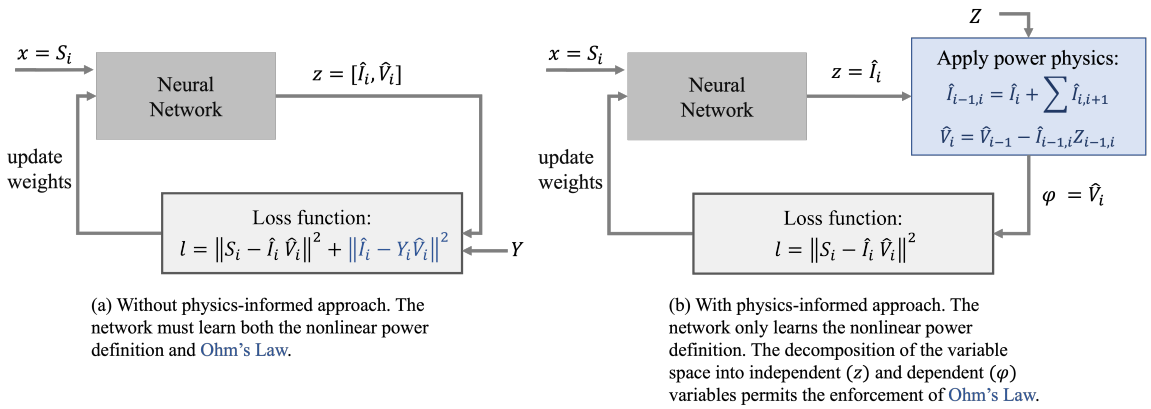
A common implementation of the backward forward sweep (BFS) algorithm for power flow is presented below.

$$I_i^{(k)} = \text{conj}\left(\frac{S_i}{V_i^{(k-1)}}\right) \tag{2.3}$$

$$I_{i-1,i}^{(k)} = I_i^{(k)} + \sum I_{i,i+1}^{(k)} \tag{2.4}$$

$$V_i^{(k)} = V_{i-1}^{(k)} - I_{i-1,i}^{(k)} Z_{i-1,i} \tag{2.5}$$

Figure 2-4 shows two neural frameworks for power flow based on the BFS algorithm. In the first framework, the neural network is estimating both the nodal currents I and voltages V, by learning two physical relationships: Generalized Ohm's law (Eq. 2.5 and the definition of power (Eq. 2.3). The physical relationship between I and V is implicit via a penalty term added to the loss function. In the second



(a) Without physics-informed approach. The network must learn both the nonlinear power definition and Ohm's Law.

(b) With physics-informed approach. The network only learns the nonlinear power definition. The decomposition of the variable space into independent (z) and dependent (φ) variables permits the enforcement of Ohm's Law.

Figure 2-4: Two neural frameworks for learning the power flow problem, via the backward forward sweep algorithm.

framework, the neural network is estimating only the nodal currents I, by learning the nonlinear power definition (Eq. 2.3). Note that the physical relationship between I and V is explicit in the framework: Generalized Ohm's law is enforced through the variable space completion step. In detail, the BFS algorithm can be represented by the proposed neural framework of Fig. 2-3, with $x := [S_i]$, $z := [I_i]$, $\varphi := [I_{i,i+1}, V_i]$. In this case the mapping between x and z is nonlinear, with $I_i = f_{BFS}(S_i)$ where f_{BFS} is an approximation of Eq. 2.3. The mapping between z and φ is linear, rendering $\mathcal{P}_{BFS}$ as a matrix.

Linear DistFlow Model

A common linearization of the power flow equations for radial and balanced distribution grids is given by the Linear DistFlow model [20], as below:

$$P_j^G - P_j^L = \sum_{k:(j,k)} P_{jk} - \sum_{i:(i,j)} P_{ij} \tag{2.6}$$

$$Q_j^G - Q_j^L = \sum_{k:(j,k)} Q_{jk} - \sum_{i:(i,j)} Q_{ij} \tag{2.7}$$

$$v_j - v_i = -2(R_{ij}(P_{ij} - P_{ji}) + X_{ij}(Q_{ij} - Q_{ji})) \tag{2.8}$$

The model describes lossless power balance at every node (2.6)-(2.7), and a linear approximation to Ohms' Law (2.8). The Linear DistFlow model assumes radiality of the

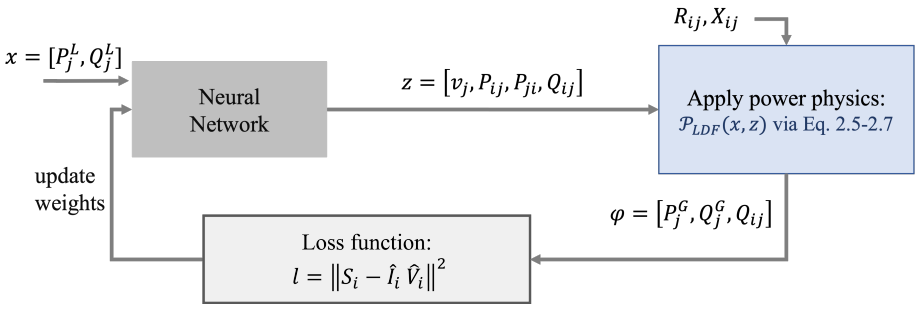


Figure 2-5: A neural frameworks for learning the Linear DistFlow equations, to be used towards optimal power flow predictions.

underlying power grid topology, balanced line parameters, and lossless power balance. The nodal power injection is separated into generation P_j^G and load P_j^L to represent the different decision variables for prosumers (co-located load and generation), and similarly for reactive power. The neural framework representation of the Linear DistFlow model (shown in Fig. 2-5) is as follows: $x := [P_j^L, Q_j^L]$, $z := [v_j, P_{ij}, P_{ji}, Q_{ij}]$, $\varphi := [P_j^G, Q_j^G, Q_{ji}]$. The neural engine will learn the mapping $z = f_{LDF}(P^L, Q^L)$, which is nontrivial, and the linear equations (2.6)-(2.8) can be used to calculate $\varphi = \mathcal{P}_{LDF}(x, z)$. In this case the mapping between z and φ is linear, and $\mathcal{P}_{LDF}$ can be represented as a matrix.

Relationship to Jacobian Learning Problem

Consider a linear system, parameterized by matrix A

$$z = Ax \tag{2.9}$$

In a quasi-static system, this equation defines a linear relationship between two sets of variables. In power systems this can be Ohm's law, $I = YV$. The Jacobian learning problem for Eq.2.9 is a regression problem, where the learning system approximates the Jacobian matrix A by $\hat{A}$, such that

$$\hat{z} = \hat{A}x \tag{2.10}$$

Next consider a nonlinear system, as below,

$$z = g(x) \tag{2.11}$$

The same Jacobian learning algorithm can be applied to the nonlinear system when linearized about an equilibrium point x_0, such that $\Delta z = A_0 \Delta x$, where A_0 is the system Jacobian about the equilibrium point. Then $\hat{A}$ in Eq.(2.10) is an approximation to A_0, and describes the linearized behaviour of the system about the equilibrium point x_0. Next consider representing the nonlinear relationship by a piece-wise linear function – this family of Jacobians represents the linearized behaviour at multiple equilibrium points. Finally, consider parameterizing the nonlinear relationship by a set of functions (such as the set of polynomial functions, trigonometric functions, etc.), whose weights must be learned. The resulting approximation $\hat{g}(x)$ represents the nonlinear relationship throughout the domain x. This is the neural network approach, where the set of functions are not restricted a priori, but are determined by the combination of neural network weights and activation functions. The approximation quality of the system Jacobian is determined by the amount of data samples, size and depth of the neural network, and the selection of neural network architecture (including activation functions, the use of recurrent or convolutional layers, etc). Figure 2-6 summarizes this discussion.

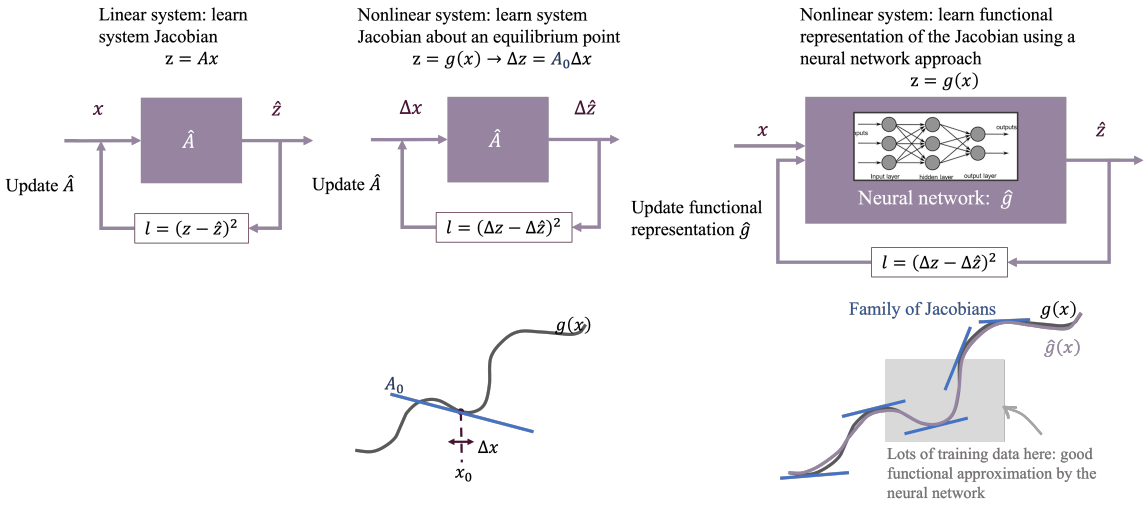


Figure 2-6: Jacobian learning as a regression or neural network problem. **Left:** a linear system can use regression. **Center:** nonlinear system linearized about an equilibrium point can similarly use regression. **Right:** the nonlinear system $g(x)$ can be approximated by a family of Jacobians (blue) which represent a piece-wise linear functional approximation. Finally, a neural network approach permits the functional approximation of $g(x)$ directly, with $\hat{(g)}(x)$ shown in grey.

Chapter 3

Grid-SiPhyR: End-to-End Learning to Optimize for Combinatorial Problems in Distribution Grids

3.1 Introduction

An important decision in power systems involves grid reconfiguration, a highly dynamic phenomenon. This problem asks the following question: "Given a distribution network with a set of switches and distributed generation, what is the most efficient grid topology and resource dispatch (power setpoints) that satisfies all loads, subject to power physics, generator, and topology constraints?" A naive method would conduct an exhaustive search of the space of switch status permutations and optimize the resource dispatch upon each topology to find the optimal switch on/off decisions (topology) and dispatch solution with minimal electrical line losses. Instead, this thesis proposes the use of a physics-informed framework for end-to-end learning to optimize for constrainted combinatorial problems, of which the grid reconfiguration problem is a member.

3.1.1 Distribution Grid Reconfiguration as a Dynamic Phenomenon

Grid reconfiguration involves the selection of switch-states (open or closed) in the grid to ensure that loads are met with available generation resources, while satisfying voltage and line flow operating constraints. At the transmission (bulk) level, reconfiguration is a typical response to contingencies, failure of a generator or line. At the distribution level, reconfiguration can be used to reduce electrical losses involved in carrying power from bulk resources and DERs to individual loads (commercial businesses, residential units). The latter, increasing efficiency in the distribution grid with high DER penetrations, is the focus of this work. Static reconfiguration (StatR) determines a fixed set of switch-states that optimize losses over a long term. But the introduction of DERs necessitates dynamic reconfiguration where local DERs supply loads in closer proximity to them, thus reducing losses and improving voltage profiles across the grid. Dynamic reconfiguration (DyR) identifies the switch-states (i.e. grid topology) which minimizes losses for the given load and generation conditions. This optimization involves solving a mixed integer problem (MIP), a combination of both continuous and binary variables. Exactly solving MIPs, even those with only linear constraints, is an NP-hard problem, requiring exponential time to solve. Thus the reconfiguration problem with DER dispatch is not possible with either traditional power systems approaches which rely on heuristics and operator experience, or traditional optimization approaches which can be computationally intractable. We will employ a neural architecture, such as the one introduced in Fig. 2-3 with suitable modifications to treat the binary nature of the switch status variables, towards solving the DyR problem.

3.1.2 Statement of the Reconfiguration MIP

The grid reconfiguration problem using the Linear DistFlow model of the distribution grid [20] is formulated as below.

$$\min_{\psi} \quad f(\psi) \tag{3.1a}$$

$$\text{s.t.} \quad v_j - v_i = -2(R_{ij}(P_{ij} - P_{ji}) + X_{ij}(Q_{ij} - Q_{ji})) \qquad \forall (i,j) \in \mathcal{T}_{\mathrm{D}} \setminus \mathcal{T}_{\mathrm{D}}^{sw} \tag{3.1b}$$

$$v_j - v_i \leq -2(R_{ij}(P_{ij} - P_{ji}) + X_{ij}(Q_{ij} - Q_{ji})) + M(1 - y_{ij}) \qquad \forall (i,j) \in \mathcal{T}_{\mathrm{D}}^{sw} \tag{3.1c}$$

$$v_j - v_i \geq -2(R_{ij}(P_{ij} - P_{ji}) + X_{ij}(Q_{ij} - Q_{ji})) - M(1 - y_{ij}) \qquad \forall (i,j) \in \mathcal{T}_{\mathrm{D}}^{sw} \tag{3.1d}$$

$$P_j^G - P_j^L = \sum_{k:(j,k)} P_{jk} - \sum_{i:(i,j)} P_{ij} \qquad \forall j \in \mathcal{B} \setminus j^{\#} \tag{3.1e}$$

$$Q_j^G - Q_j^L = \sum_{k:(j,k)} Q_{jk} - \sum_{i:(i,j)} Q_{ij} \qquad \forall j \in \mathcal{B} \setminus j^{\#} \tag{3.1f}$$

$$z_{i,j} \geq 0 \qquad \forall (i,j), (j,i) \in \mathcal{T}_{\mathrm{D}} \tag{3.1g}$$

$$z_{i,j} + z_{j,i} = 1 \qquad \forall (i,j) \in \mathcal{T}_D \setminus \mathcal{T}_D^{sw} \tag{3.1h}$$

$$z_{i,j} + z_{j,i} = y_{i,j} \qquad \forall (i,j) \in \mathcal{T}_D^{sw} \tag{3.1i}$$

$$\sum y_{i,j} = (N-1) - (M - \#sw) \tag{3.1j}$$

$$\sum_{j:(i,j)} z_{i,j} + z_{j,i} \geq 1 \qquad \forall j \in \mathcal{B} \tag{3.1k}$$

$$y_{i,j} \in \{0,1\} \qquad \forall (i,j) \in \mathcal{T}_D^{sw} \tag{3.1l}$$

$$z_{i,j}, z_{j,i} \in \{0,1\} \qquad \forall (i,j) \in \mathcal{T}_D \tag{3.1m}$$

$$0 \leq P_{i,j} \leq M z_{i,j} \qquad \forall (i,j),(j,i) \in \mathcal{T}_D \tag{3.1n}$$

$$0 \leq Q_{i,j} \leq M z_{i,j} \qquad \forall (i,j),(j,i) \in \mathcal{T}_D \tag{3.1o}$$

$$P_j^L = P_j^{L,0} \qquad \forall j \in \mathcal{B} \tag{3.1p}$$

$$Q_j^L = Q_j^{L,0} \qquad \forall j \in \mathcal{B} \tag{3.1q}$$

$$\underline{P_j^G} \leq P_j^G \leq \overline{P_j^G} \qquad \forall j \in \mathcal{B} \tag{3.1r}$$

$$\underline{Q_j^G} \leq Q_j^G \leq \overline{Q_j^G} \qquad \forall j \in \mathcal{B} \tag{3.1s}$$

$$\underline{v_j} \leq v_j \leq \overline{v_j} \qquad \forall j \in \mathcal{B} \setminus j^{\#} \tag{3.1t}$$

$$v_{j^{\#}} = 1 \tag{3.1u}$$

$$z_{i,j^{\#}} = 0 \qquad \text{(no export allowed)} \tag{3.1v}$$

$$\sum_{j:(j^{\#},j)} P_{j^{\#},j} = P_{j^{\#}}^G \qquad \text{(no export allowed)} \tag{3.1w}$$

$$\sum_{j:(j^{\#},j)} Q_{j^{\#},j} = Q_{j^{\#}}^G \qquad \text{(no export allowed)} \tag{3.1x}$$

In (3.1), we have denoted a general distribution grid as a graph $\Gamma(\mathcal{B}, \mathcal{T}_D)$, where $\mathcal{B} \coloneqq \{1, ..., N\}$ is the set of N nodes, $\mathcal{T}_D \coloneqq \{(i,j)\}$ is the set of M edges where $\mathcal{T}_D^{sw} \subset \mathcal{T}_D$ are the set of lines with switches, and M_{sw} is the number of switches; node $j^{\#}$ is the point of common coupling (PCC) of the distribution grid to the bulk transmission grid; $\psi = \left[P^G\ P^L\ Q^G\ Q^L\ v\ P_{ij}\ Q_{ij}\ y_{ij}\ z_{ij}\ z_{ji}\right]$ is the decision vector for

the reconfiguration problem; P^L, Q^L are real and reactive power loads at every node j; P^G, Q^G are real and reactive power generation at every node j, and generation at $j^{\#}$ indicates import from the bulk transmission grid; $P_{ij}, P_{ji}, Q_{ij}, Q_{ji}$ are the directed real and reactive power flows through a distribution line $(i,j) \in \mathcal{T}_D$; v denotes the squared magnitude of voltage at every node j; z_{ij}, z_{ji} are binary variables indicating the direction of power flow through a line $(i,j) \in \mathcal{T}_D$; y_{ij} is a binary variable indicating the switch status $\forall (i,j) \in \mathcal{T}_D^{sw} \subset \mathcal{T}_D$, where $y_{ij} = 1$ is closed and $y_{ij} = 0$ is open.

Constraints (3.1b)-(3.1f) describe power flow using the Linearized DistFlow model. Constraints (3.1b)-(3.1d) describe Ohm's law across all lines. The big-M relaxation is used to describe the conditional constraints for line $(i,j) \in \mathcal{T}_D^{sw}$. Constraints (3.1e)-(3.1f) describe lossless power balance at every node.

Constraints (3.1g)-(3.1o) describe the topology selection through switch-status, grid radiality and connectivity constraints, enforce binary constraints on the y_{ij}, z_{ij}, z_{ji} variables, and restrict power flow through a line based on switch-status. In particular, constraint (3.1j) restricts the number of closed switches in the grid so it is radial with $N-1$ total branches, where L switches must be closed and the remaining $M_{sw} - L$ must be open. Constraint (3.1k) enforces connectivity by requiring power to flow into or out of a node along at least one line. Note that the connectivity constraint must be relaxed when a fault has occurred and islanding of sections of the grid is permitted, and for system restoration. It should be noted that typical reconfiguration problem statements also include an arborescence constraint [131], either explicitly or implicitly in the formulation of the radiality constraint. However, the increasing penetration of DERs voids this assumption, and multiple generating sources (roots of the tree) must be permitted. We have relaxed this arborescence constraint in (3.1k). Various other mathematical formulations of radiality and connectivity constraints include constraints on the determinant of the branch-to-node incidence matrix or spanning tree constraints and other graph theoretic approaches [85, 137, 17, 84]. However, many of these suffer from high computational requirements and additional complexity, and do not leverage the fact that grid connectivity can be ensured by power flow constraints under normal operation. Our formulation accounts for this.

Constraints (3.1p)-(3.1q) set the loads at node j, which are assumed to be inflexible. Constraints (3.1r)-(3.1s) describe generator operating limits, and (3.1t)-(3.1u) describes grid voltage limits where the voltage at the PCC is assumed to be fixed at 1pu, as is common practice in power systems. Finally, constraints (3.1v)-(3.1x) describe "no export" limits on the PCC, where net generation excess of net load in the distribution grid cannot be injected into the transmission grid. For regions where distribution grids are permitted to export power to the bulk grid, these constraints can be removed.

The mixed integer linear program (MILP, a combinatorial problem with linear constraints) detailed above is solved to minimize an objective function $f(y)$. For a modern distribution grid with high DER penetration, various objectives are sought after by grid operators. Some such objectives include minimizing electrical line losses (maximizing grid efficiency), minimizing costs for power generation, minimizing congestion, improving voltage profiles across the distribution feeder, reducing peak power demand, ensuring reliability of service (ex. higher capacity margins for feeders and supply transformers), and balancing load. Depending on the types of switches in the grid, operators may also minimize the cost incurred by actuating switches. In general, these objectives can be formulated using a convex function, thus retaining the uniqueness and global optimality of the MILP solution. Further extensions to the reconfiguration problem include distinctions between hard and soft constraints, considering the optimal switch change order to go from topology A to topology B, and considering grid outage conditions and subsequent generator restart and load recovery. Note that soft constraints can include lines which can exceed thermal limits for short periods of time during a reconfiguration activity.

3.1.3 Utility Perspective on Dynamic Reconfiguration

Discussions with a utility company provide the following perspectives on dynamic reconfiguration and the state of their distribution grid.

Network Topology and Switch Operations: A distribution substation will typically have 2-9 feeders, with 4 switches per feeder. The normally closed switches

(NCS) connect the feeders to the substation and across feeders. The normally open switches (NOS) are used for fault location, isolation, and service restoration (FLISR). In the operating room, the dispatchers can monitor the state of the feeders, and the system will display where and when feeders are in an abnormal state. If any switch operations are needed, the operator will follow a pre-determined switching order to reconfigure the network. All switches are remotely operated, and are typically only operated during faults or overload on one of the feeders. Switches are also operated during regular scheduled maintenance wherein operators will offload the entire feeder. Maintenance is scheduled during light load conditions, to reduce risk of capacity overloading. Dynamic reconfiguration is currently used for FLISR applications. Note that this is different from the application to line loss reduction that is investigated in this thesis. Grid-SiPhyR presents a dynamic decision making tool to efficiently operate distribution grids under normal operating conditions. Extensions can be made to the framework for fault conditions, but are outside the scope of the presented thesis.

Load and generation: Customers typically include residential and commercial loads. Feeders have moderate PV penetration at around 20% penetration of nameplate capacity to maximum real power load. These include utility owned PV installations and some residential units. The utility owned PV units are equipped with a smart inverter with power electronic control. The inverters typically operate at a fixed power factor (as per manufacturer setting; power factor is the ratio of real to reactive power), but can provide reactive power support through the inverter capacity. Future applications of power factor control are possible and can be investigated.

Communication and Control: The utility company has remote control of the switches throughout their distribution system. These Distribution Automation Control (DAC) systems are often equipped by automated switches and remote monitoring facilities. These DAC systems can enable optimal reconfiguration applications with switch actuation decisions being made by the distribution utility and grid state being monitored as well. While all utilities do not have these remote communication and control capabilities, many are exploring grid modernization initiatives to improve

visibility and control in both centralized and decentralized/distributed frameworks.

3.1.4 Lifespan of Switching Devices

The lifespan of switching devices (such as reclosers) is measured in mechanical and electrical endurance. The international standard for automatic circuit reclosers for AC systems up to 38kV is the IEEE/IEC C37.60 [6]. The standard for high voltage circuit breakers is IEEE C37.06 [5]. The mechanical endurance required for both reclosers and circuit breakers is 2,000 operations (no load). However vendors typically publish endurance capabilities of 10,000 operations. While the standards for reclosers do not specify electrical endurance at rated continuous current (compared to 100 for circuit breakers as per IEEE C37.06), most vendor reclosers have 10,000 operations (similarly for circuit breakers) with some vendors supporting 30,000 close-open operations. The loss of life per switch operation depends on the circuit through the switch, with regular load conditions differing considerably from fault conditions (which may see very high currents). Switching devices are expected to have 30 to 40 year lifespans. At a mechanical endurance of 10,000 operations, this corresponds to roughly 300 operations per year, or once daily. Note that switches in the distribution grid today typically see less than 50 operations per year.

3.2 SiPhyR: end-to-end learning to optimize for combinatorial problems

This section demonstrates a novel use of physics-informed machine learning suitably modified to incorporate physical constraints present in various optimization problems for physics systems, both equality and inequality, and optimize a relevant cost function. Where traditional solvers cannot take advantage of the structures present in repeatedly solving an optimization problem, and warm-start techniques may struggle when parameters vary rapidly - such as with solar generation forecasts, ML presents an opportunity to *learn* the underlying problem structure. In particular, when ML

is modified in a physics-informed approach, it can provide good and fast solutions which satisfy operating constraints.

This section proposes SiPhyR, a physics-informed machine learning framework for end-to-end learning to optimize for a class of combinatorial problems. Consider a mixed integer optimization problem which given input data $x \in \mathbb{R}^d$ finds ψ which minimizes $f(\psi)$ subject to a set of equality and inequality constraints,

$$\min_{\substack{\psi=[z_\tau, z_{\setminus\tau}, \varphi], \\ z_\tau \in \mathbb{Z}_2^m, [z_{\setminus\tau}, \varphi] \in \mathbb{R}^n}} \quad f_x(\psi) \tag{3.2}$$

$$s.t. \quad g_x(\psi) = 0 \tag{3.3}$$

$$h_x(\psi) \leq 0 \tag{3.4}$$

$$b_x(z_{\setminus\tau}) = 0 \tag{3.5}$$

where f, g, h are potentially nonlinear and nonconvex, and b has a particular structure to be discussed in coming sections. Note the presence of binary variables z_τ. Leveraging techniques from variable space reduction from optimization literature, the decision variable ψ has been separated into *independent* and *dependent* variables, $z = [z_\tau, z_{\setminus\tau}]$ and φ respectively. Knowledge of z and the function g permits the calculation of φ, i.e. $\varphi = \tilde{g}_x(z)$. In general, this decomposition is non-unique. It critically depends on the structure of the given problem, which determines the relationship between the sets of variables, and requires domain knowledge to exploit the underlying problem structure to produce good solutions.

This optimization problem can be cast as a learning problem with a neural network parameterized by θ as $\psi = N_\theta(x)$, with suitable accommodations made for the binary variables and constraints. To solve such a problem, we propose an unsupervised neural network shown in Fig. 3-1, composed of five key components described next.

Lightweight neural network: a simple neural network with a sigmoidal (Si) output layer. The neural network predicts the *independent variables*, denoted as $\hat{z}$, from the input data x. The output of the neural network is divided into two sets: a vector

of probabilities describing the likelihood that a binary variable takes on the value of 1, $\hat{z}_\tau = \mathbb{P}(z_\tau = 1)$; and the prediction for the continuous variables $\hat{z}_{\setminus\tau}$. The sigmoidal function is a cumulative distribution function of the logistic distribution, and establishes the interpretation of $\hat{z}$ as a probability.

Physics-informed rounding (PhyR): algorithm to recover integer solutions from the vector of probabilities. This layer embeds salient constraints describing the relationship between the binary variables which are used to convert probabilities $P(z_\tau)$ (input) to binary decisions z_τ (output).

Inequality constraint layers: the prediction from the neural network $\hat{z}_{\setminus\tau}$ are scaled onto box constraints, using a sigmoidal function mapping, with bounds selected as per the box constraints. This ensures that inequality constraints pertaining to the independent variables are have certified satisfiability. This layer acts in parallel to the physics-informed rounding. The output of this layer are the scaled variables $z_{\setminus\tau}$.

Equality constraint layers: leveraging techniques for variable space reduction from optimization, the equality constraints are used to calculate the *dependent variables* φ from the independent variables $z = \begin{bmatrix} z_\tau & z_{\setminus\tau} \end{bmatrix}$, as $\varphi = \tilde{g}_x(z)$. This ensures that equality constraints have certified satisfiability.

Loss function: the neural network is trained to *learn to optimize* by selecting a loss function which is composed of the objective function to be minimized during optimization (ex. line losses in the grid) and regularization to bias the solution against violating physical constraints (typically inequality constraints of dependent variables). This is done by way of a soft-loss penalty with hyperparameter λ_h. The resulting unsupervised loss function is $l = f_x(\psi) + \lambda_h \| \max\{0, h_x(z, \varphi)\} \|_2^2$. Note that the presence of nonlinear equations in g may result in equality constraint violations during the variable space completion. A corresponding penalty for equality constraints can also be included, as well as any soft constraints on the system. It must be stressed

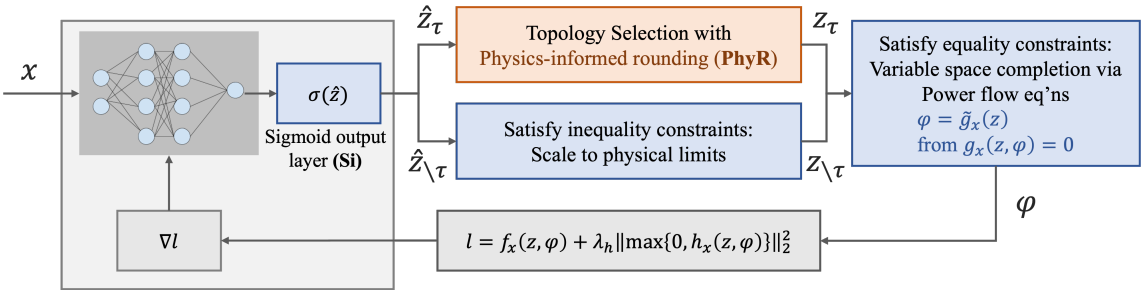


Figure 3-1: Si**PhyR**: A physics-informed machine learning framework for end-to-end learning to optimize for a class of combinatorial problems, applied to the grid reconfiguration problem to optimally determine grid topology and power dispatch for all generators

that the proposed neural framework does not have access to the optimal solutions. Rather, by using the unsupervised loss function, the neural network is only given access to the input training data (loads in the grid) and must determine a feasible and (ideally) optimal solution.

3.2.1 Physics-informed rounding (PhyR)

Traditional optimization literature deals with mixed integer programming using an array of heuristic methods involved in developing good upper and lower bounds, pruning solution branches, and selecting variables to round to nearest integer solutions. Taking inspiration from the class of rounding heuristics well-established in the MIP literature ([75], and others including [58, 95, 69]), we propose a physics-informed rounding algorithm. We consider the class of mixed integer problems with binary variables 3.2 where the function b, typically describing some physical relationships, uniquely defines a cutoff index L such that L variables have value 1, with the remaining $m - L$ variables having value 0.

Rounding Variables

In simple terms, rounding a number involves reducing the numerical accuracy for ease of calculations (ex. back of envelope calculations), controlling the degree of accuracy when using measured values (i.e. using significant digits), or embedding

logic statements (ex. item count in integer digits). Typical rules for rounding are straightforward: if a variable value is above half a decimal point, round up to the next integer number; else, round down. However, this can become more complex when rounding multiple variables at the same time, and when these variables are related to one another. This is particularly true in the latter category of logic statements, which may come from system constraints, with which we are concerned.

In the following, we restrict the discussion to numbers between 0 and 1 which must be rounded to integer values (i.e. binary 0 or 1), without loss of generality. The rounding function for such a variable can be thought of as a high pass filter: values above 0.5 "pass through" with value 1, and those below 0.5 are attenuated to value 0. Variables at the cutoff value of 0.5 can be rounded up or down, depending on the standard being used. For example, the IEEE 754 floating point standard will round values to the nearest even number, so 0.5 is also attenuated to 0. The parameter of this high pass rounding function is the cutoff parameter. It was chosen in the example to be 0.5, however there may be applications where asymmetry better represents the problem. Consider a screening procedure where only 20% of applicants with the highest scores can move to the next round; the cutoff parameter may then be closer to 0.75 or 0.8. Even in this screening procedure, selecting the value of the cutoff parameter without viewing all the data is nontrivial. A cutoff of 0.8 assumes that applicant scores are uniformly distributed between 0 and 1. A conservative approach may be to have a cutoff of 0.6, then view each data point to calculate the true cutoff, and proceed to screen. Perhaps data from previous application cycles show that most applicants have high scores, and the data can be used to estimate a cutoff of 0.85. In such applications, the design of the rounding function can itself be cast as a learning problem. Of note in the presented example of screening applicants, is that the actual value of the scores do not matter; rather, the relative value of the score is important. Consider a particular application cycle where all the scores are below 0.8. Then the cutoff parameter itself would be lower than 0.8. The rounding function can then be written as a *relative* filter, where the N scores are ranked then passed through a filter with a cutoff at $0.8N$. Observe that the relative filter is parameterized by

the number of applications, N, and the selected cutoff percentage, 0.8. When these two parameters are known, the relative rounding function is fully defined and easy to implement. This is the approach for the proposed physics-informed rounding for grid reconfiguration. In the next sections, the physics-based equations defining the two parameters (number of switches, and cutoff index) are described, and the proposed rounding heuristic is detailed.

Physics-informed rounding algorithm

Our physics-informed rounding approach is as follows: function b is used to define the cutoff index L of a *rounding function* $z_\tau = R_b(\hat{z}_\tau)$ that considers the relative value of probabilities across all variables z_τ, and sets the L variables with the largest probabilities to 1, and the remaining $m - L$ variables to 0. This is formally presented in Algorithm 2, and applied to the dynamic grid reconfiguration problem.

Algorithm 1: Physics-Informed Rounding **PhyR** for binary variables

Data: Probability $\hat{z}_\tau = \mathbb{P}(z_\tau = 1)$

Result: Binary variables z_τ

initialization: $L = f(b(\psi))$

Sort $\hat{z}_\tau$ in descending order; assign $\mathbb{I}_{\hat{z}_\tau}$ the sorted indices of $\hat{z}_\tau$

Assign L binary variables to have value 1: $z_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{1:L}\right] = \max\{\ \hat{z}_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{L}\right], 1\ \}$

Assign remaining $M - L$ binary variables to be 0:

$z_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{L+1:m}\right] = \min\{\ \hat{z}_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{L+1:m}\right], 0\ \}$

3.2.2 Extensions to the proposed SiPhyR method

Binary variables in the dependent set: The framework of Fig. 3-1 restricts binary variables to the independent set z. A simple extension is to binary variables in the dependent set (where $\varphi = [\varphi_\tau, \varphi_{\setminus\tau}]$ instead of the prior definition of z) which employ the PhyR-based rounding as a post-processing step to the prediction, similar to a corrective projection [112, 155] or gradient-based procedure [35] employed for continuous variables. However, the logical flow of decision making is then broken, where the prediction of $z_{\setminus\tau}$ and $\varphi_{\setminus\tau}$ are mathematically consistent with the *probabil-*

ities which are calculated using the variable space completion, not the final rounded variables. The decision to include binary variables in the independent set is not purely a mathematical one. Many mixed integer problems are formulated for sequential decision processes where the selection of binary/integer variables precedes the decision of continuous variables: ex. the grid topology is selected before enforcing power flow constraints; a set of workers are selected before allocating tasks. It must also be highlighted that the proposed SiPhyR framework optimizes decisions over both integer and continuous variables simultaneously, as these decisions are intricately linked through the constraints. That said, the SiPhyR framework is general and can be extended to use cases where binary variables are included in the dependent set – perhaps for applications where function b is conditionally dependent on other variables in the optimization problem. There may be a need to design modeling or neural layers before the completion step, which encourage integer variables in the dependent set or improve inequality constraint satisfaction. This is highly problem dependent, and the additional complexity from such neural layers should be well motivated by application.

Additional binary variables: If the model includes additional integer variables which impose *soft* constraints, these may be included in either the independent or dependent variable sets. These variables may be driven towards integer solutions using functional approximations of rounding functions (discussed more in the results, see the InSi framework) or penalized in the loss function.

Integer variables: The proposed Algorithm 2 can be extended to the case of integer variables $z_\tau \in \mathbb{Z}_\ell^m$ where the function b defines a set of cutoff values $\{L_1, L_2, ..., L_{\ell-1}\}$. The largest $L_{\ell-1}$ probabilities map to the largest integer variable in the set $\mathbb{Z}_\ell$, the smallest L_1 probabilities map to the smallest integer variable, and so forth. With R_b defined, the rounding operations can be suitably modified to round onto the set $\mathbb{Z}_\ell$. The discussions above can be extended similarly to integer variables.

Generalizing to set constraints: More generally, the proposed architecture can be extended to general set constraints which are treated within the PhyR layer. These set constraints may include operators including the union "$\cup$", intersection "$\cap$", or set exclusions "$\setminus$", along with various equality or inequality relations between set expressions (including "$=$", $\subseteq$, $\subsetneq$). This may include logical constraints (ex. $x \implies y$) and conditional constraints ($x \leftarrow y/2$ if $y\%2 = 0$, else $x \leftarrow 3x + 1$). Set constraints present a broad and general group of constraints which can represent various computing and engineering problems, ranging from language processing, type inference, and design problems. Observe that the grid reconfiguration problem includes conditional constraints (also called on/off constraints) on Ohm's law across the lines with switches. While we use the common big-M convex relaxation, the quality of the bound given by this formulation depends on the constant M. In practice for a general set of constraints, this may not be trivial to determine. SiPhyR can be extended to more a general framework for conditional constraints by also incorporating prior work from constrained conditional models (CCMs) in machine learning, frequently used for natural language processing and linear integer programming.

3.3 Grid-SiPhyR: Physics-Informed Machine Learning for Dynamic Reconfiguration

The reconfiguration problem is an MIP, thus using traditional optimization methods for DyR is intractable, especially when requiring fast decisions and high degree of accuracy to remain within operating constraints. We employ the SiPhyR framework described in the previous section towards solving the dynamic grid reconfiguration problem. The complete Grid-SiPhyR architecture is presented in Fig. 3-2.

3.3.1 Variable Decomposition Based on Equality Constraints

In general, this decomposition is non-unique, as discussed in Section 2.3. It critically depends on the structure of the given problem, which determines the relationship

between the sets of variables, and requires domain knowledge to exploit the underlying problem structure to produce good solutions. As will be seen below, the application to dynamic reconfiguration will require a further decomposition of the independent and dependent variables into continuous and binary variables, denoted by superscripts C and B, respectively.

Equality Constraints

Here we re-introduce the salient equality constraints from the MIP formulation of dynamic reconfiguration, presented in (3.1):

$$P_j^G - P_j^L = \sum_{k:(j,k)} P_{jk} - \sum_{i:(i,j)} P_{ij} \qquad \forall j \in \mathcal{B} \setminus j^{\#} \qquad (3.6)$$

$$Q_j^G - Q_j^L = \sum_{k:(j,k)} Q_{jk} - \sum_{i:(i,j)} Q_{ij} \qquad \forall j \in \mathcal{B} \setminus j^{\#} \qquad (3.7)$$

$$v_j - v_i = -2(R_{ij}(P_{ij} - P_{ji}) + X_{ij}(Q_{ij} - Q_{ji})) \qquad \forall (i,j) \in \mathcal{T}_{\mathrm{D}} \setminus \mathcal{T}_{\mathrm{D}}^{sw} \qquad (3.8)$$

$$z_{ij} + z_{ji} = 1 \qquad \forall (i,j) \in \mathcal{T}_D \setminus \mathcal{T}_D^{sw} \qquad (3.9)$$

$$z_{ij} + z_{ji} = y_{ij} \qquad \forall (i,j) \in \mathcal{T}_D^{sw} \qquad (3.10)$$

$$\sum_{(i,j)\in\mathcal{T}_D^{sw}} y_{ij} = L$$

The total number of equality constraints is: $2(|\mathcal{B}| - 1) + 2|\mathcal{T}_D \setminus \mathcal{T}_D^{sw}| + |\mathcal{T}_D^{sw}| + 1$, where notation $|\cdot|$ denotes the size of the set. Each equality constraint permits the calculation of a dependent variable, φ, from the relevant training data x and independent variables z.

The variable decomposition in the context of reconfiguration is summarized below:

$$\begin{aligned}
x &= \left[\left\{P_j^L, Q_j^L \mid j \in \mathcal{B} \setminus j^{\#}\right\}\right] \\
z^B &= [z_{ji}, \{y_{ij} \mid (i,j) \in \mathcal{T}_D^{sw} \setminus \{-1\}\}] \\
z^C &= \left[\left\{v_j \mid j \in \mathcal{B} \setminus j^{\#}\right\}, P_{ji}, P_{ij}, Q_{ji}, \{Q_{ij} \mid (i,j) \in \mathcal{T}_D^{sw}\}, P_{j^{\#}}^L, Q_{j^{\#}}^L\right] \\
\varphi^B &= \left[z_{ij}, \left\{y_{ij} \mid (i,j) \in \mathcal{T}_{D,-1}^{sw}\right\}\right] \\
\varphi^C &= \left[P_j^G, Q_j^G, \{Q_{ij} \mid (i,j) \in \mathcal{T}_D \setminus \mathcal{T}_D^{sw}\}\right]
\end{aligned} \qquad (3.11)$$

In (3.11) the set of independent and dependent variables are $z = \left[z^B, z^C\right]$ and $\varphi = \left[\varphi^B, \varphi^C\right]$ respectively.

Implication 1: PhyR enables higher prediction accuracy

Using the proposed SiPhyR framework, the grid topology is selected, then the power physics are enforced upon the topology via the variable space completion. This sequential decision process improves prediction performance by mimicking the simpler optimal power flow problem on a fixed topology, in every offline training iteration and every online prediction. This statement is corroborated by the experimental results presented in Section 3.7.

Implication 2: Certified satisfiability of inequality constraints

Here we highlight a key feature of the proposed variable decomposition (3.11): the decomposition results in certified satisfiability of inequality constraints describing voltage, line flow, and generator limits. The particular variable decomposition in (3.11) and the selection of which variables belong in z and φ are critical to satisfying inequality constraints in Grid-SiPhyR. Thus, a critical feature of the proposed variable decomposition is the certified satisfiability of inequality constraints describing voltage, line flow, and generator limits. By selecting voltages v as an independent variable, these are scaled onto the box constraints describing operating limits; for a grid operator, this means voltage limits across the grid will always be satisfied, a critically important aspect of power systems operation. This is inherent in our proposed structure, as compared to other methods which rely on projections, clipping, or penalties to enforce voltage constraints.

Implication 3: Accommodating different power physics models

The equality constraints of the Linearized DistFlow model are linear, and therefore the dependent variables can be determined trivially with zero error. Further extensions to multi-phase unbalanced grids can be carried out as well, in which case the appropriate ACOPF models and relaxations would be employed. There are existing work which

have proposed linear models for unbalanced grids, which result in less than 1.2% optimality gap compared to ACOPF [49, 64]. For problems which involve more complex (potentially nonlinear) equality constraints, such as nonlinear DistFlow (ex. Branch flow model) or the full AC power flow, the same variable space reduction techniques can be used, and programs like Newton's method can be leveraged to solve for the dependent variables. In the backpropagation step through the equality constraint layer, the Jacobians describing the derivatives can be explicitly written out and the implicit function theorem used to backpropagate through the dependent variables.

3.3.2 Physics-Informed Rounding

A key challenge in the DyR problem with simultaneous topology selection and DER dispatch is the mixed integer nature. We employ the PhyR approach as follows.

Equations Governing Grid Topology

Distribution grids in the US are operated with a radial structure, and during normal operation, all nodes must be connected. Various mathematical formulations of these radiality and connectivity constraints include constraints on the determinant of the branch-to-node incidence matrix or spanning tree constraints and other graph theoretic approaches [85, 137, 17, 84]. However, many of these suffer from high computational requirements and additional complexity, and do not leverage the fact that grid connectivity can be ensured by power flow constraints under normal operation. We integrate these radiality and connectivity constraints in the following manner. We first define $L = (N-1) - (M - M_{sw})$ where N is the number of nodes, M the number of lines, and M_{sw} the number of switches.

$$\sum_{(i,j)\in\mathcal{T}_D^{sw}} y_{ij} = L \tag{3.12}$$

$$\sum_{j:(i,j)} z_{ij} + z_{ji} \geq 1 \qquad \forall j \in \mathcal{B} \tag{3.13}$$

Constraint (3.12) restricts the number of closed switches in the grid so it is radial with $N-1$ total branches, where L switches must be closed and the remaining $M_{sw}-L$ must be open. Constraint (3.13) enforces connectivity by requiring power to flow into or out of a node along at least one line. It should be noted that typical reconfiguration problem statements also include an arborescence constraint [131], either explicitly or implicitly in the formulation of the radiality constraint. However, the increasing penetration of DERs voids this assumption, and multiple generating sources (roots of the tree) must be permitted. We have relaxed this arborescence constraint in (3.13).

Physics-Informed Rounding for Grid Reconfiguration

With the above constraints, we now use this explicit knowledge of L to inform our topology selection. In particular, constraint (3.12) restricts the number of closed switches in the grid so it is radial with $N-1$ total branches, where L switches must be closed and the remaining $M_{sw}-L$ must be open. As it pertains to the PhyR method, the radiality constraint (3.12) is the function b in Eq. 3.2. We define the cutoff index of the rounding function R_b as $L=(N-1)-(M-M_{sw})$.

The underlying concept of the SiPhyR framework is to explicitly embed discrete decisions into a neural framework. This presents a challenge to learning, since the implementations of PyTorch functions like round, max, and min all eliminate the gradient in the range where the function is applied. This prevents gradients from flowing through the backpropagation and makes learning challenging. The approach taken in Grid-SiPhyR is to only round $M_{sw}-2$ switch statuses (with $L-1$ being rounded up and the remaining $M_{sw}-L-1$ being rounded down), thereby retaining the gradient information for two switches.

By modifying the implementation of PhyR for topology selection in this way, we ensure gradient information flows through the backpropagation. The neural network predicts the probability of 'on' status of all but one switch, and is described by $\hat{z}_\tau$. Of these $M_{sw}-1$ switches, at least $L-1$ must be closed. The remaining L-th switch that is closed can belong either to $\hat{z}_\tau$ or to $\varphi_\tau = \left\{ y_{ij} \mid (i,j) \in \mathcal{T}^{sw}_{D,-1} \right\}$. We then close $L-1$ switches in $\hat{z}_\tau$ by setting the corresponding $z_\tau = 1$, and we open $M_{sw}-L-2$ switches

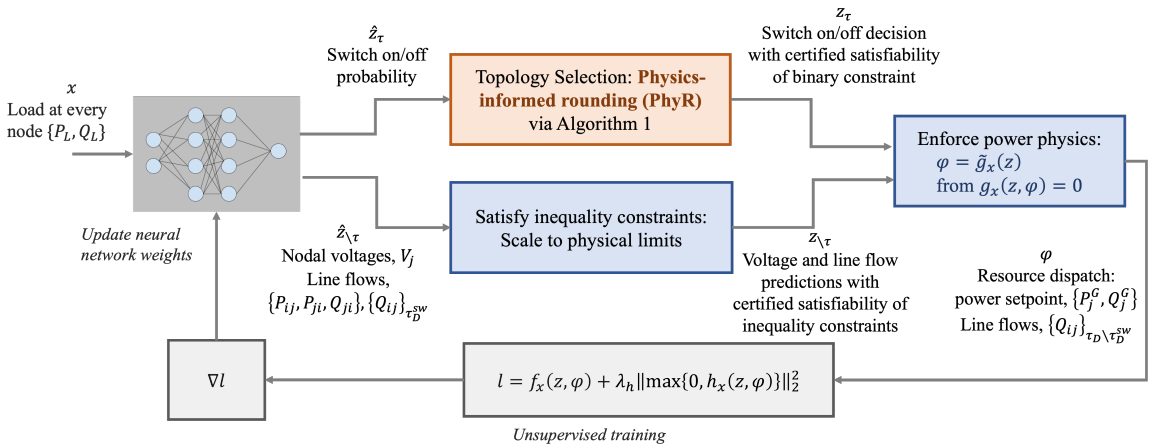


Figure 3-2: Grid-SiPhyR: A physics-informed ML-based framework for end-to-end learning to optimize, applied to the dynamic reconfiguration problem

in $\hat{z}_\tau$ by setting the corresponding $z_\tau = 0$. For the remaining two switches, the neural network training guides y_{ij} to integer solutions. The application of PhyR using the above constraints then permits the selection of a feasible (read: radial and connected) grid topology upon which the power flow describing the relationship between z^C and φ^C are satisfied (*Implication 1* in the prior section).

Algorithm 2: Physics-Informed Rounding for Topology Selection

Data: $\hat{z}_\tau$

Result: Binary variables z_τ for switch-state prediction and topology selection

initialization: $L = (N-1) - (M - M_{sw})$

Sort $\hat{z}_\tau$ in ascending order; assign $\mathbb{I}_{\hat{z}_\tau}$ the sorted indices of $\hat{z}_\tau$

Assign switches to be closed: $z_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{1:L-1}\right] = \max\left\{ \hat{z}_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{L-1}\right], 1 \right\}$

Assign switches to be open: $z_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{L+1:M_{sw}-1}\right] = \min\left\{ \hat{z}_\tau \left[\mathbb{I}_{\hat{z}_\tau}^{L+1:M_{sw}-1}\right], 0 \right\}$

3.4 Generating Datasets

In this section we present the datasets used in the evaluation of the Grid-SiPhyR framework. Of note, there are few good training datasets available for power system applications, and no known datasets for the grid reconfiguration problem. As a result, the creation of these datasets is a contribution of this research project. There are

three relevant parameters to each dataset: (i) network topology, line parameters, and location of switches; (ii) location, magnitude, and time-varying profile of loads; and (iii) location, capacity, and time-varying profile of distributed generation. Note that (i)-(iii) are the input to the reconfiguration problem, and determine the optimal configuration of the network. In the sections that follow, each network dataset and (i)-(iii) will be presented. The network data (i) are taken from literature, as are the location and magnitude of loads for a single time period. The time-varying load profiles (ii) and the solar resource locations and generation profile (iii) are developed as part of the dataset generation. Table 3.1 summarizes key parameters of the two distribution grids (BW-33 and TPC-94, presented in detail below).

	BW-33	TPC-94
$\lvert\mathcal{B}\rvert$	33	94
$\lvert\mathcal{T}_{\mathrm{D}}\rvert, \lvert\mathcal{T}_{\mathrm{D}}^{sw}\rvert$	30, 7	83, 14
# Discrete vars - topological, y_{ij}	7	14
# Discrete vars - power flow, z_{ij}, z_{ji}	74	194
# Continuous vars	248	671
# Equality constraints	134	369
# Inequality constraints	545	1465
# Training data, $\lvert x\rvert$	64	186
# Independent variables, $\lvert z\rvert$	195	510
# Dependent variables, $\lvert\varphi\rvert$	134	369

Table 3.1: Size of reconfiguration problem for two distribution grids

3.4.1 33-Node Baran-Wu Grid (BW-33)

The BW-33 grid presented in [20] is a canonical grid used in the reconfiguration literature. The grid is very lossy, with losses up to 8% of total load, and voltage profile violating voltage limits. These characteristics make the BW-33 grid an excellent test case for dynamic reconfiguration, with the objective function to reduce line losses. The data available in [20] is provided in Appendix A. The network is shown in Fig. 3-3a.

Network Topology and Parameters

The grid consists of 33 nodes ($N = 33$), 37 lines ($M = 37$) of which 5 are tie lines (normally open switches, NOS) and the remaining 32 are typically assumed to be sectionalizing switches (normally closed switches, NCS). Topology and line parameter data are presented in Table A.1. To restrict the problem to a simpler test case, only a subset of the lines are considered switchable – these include the 5 tie lines (numbered 33 to 37) and 3 NCS lines (line numbers 4, 10, and 26). These are highlighted in green in Fig. 3-3a, with dashed lines representing NOS and solid lines representing NCS.

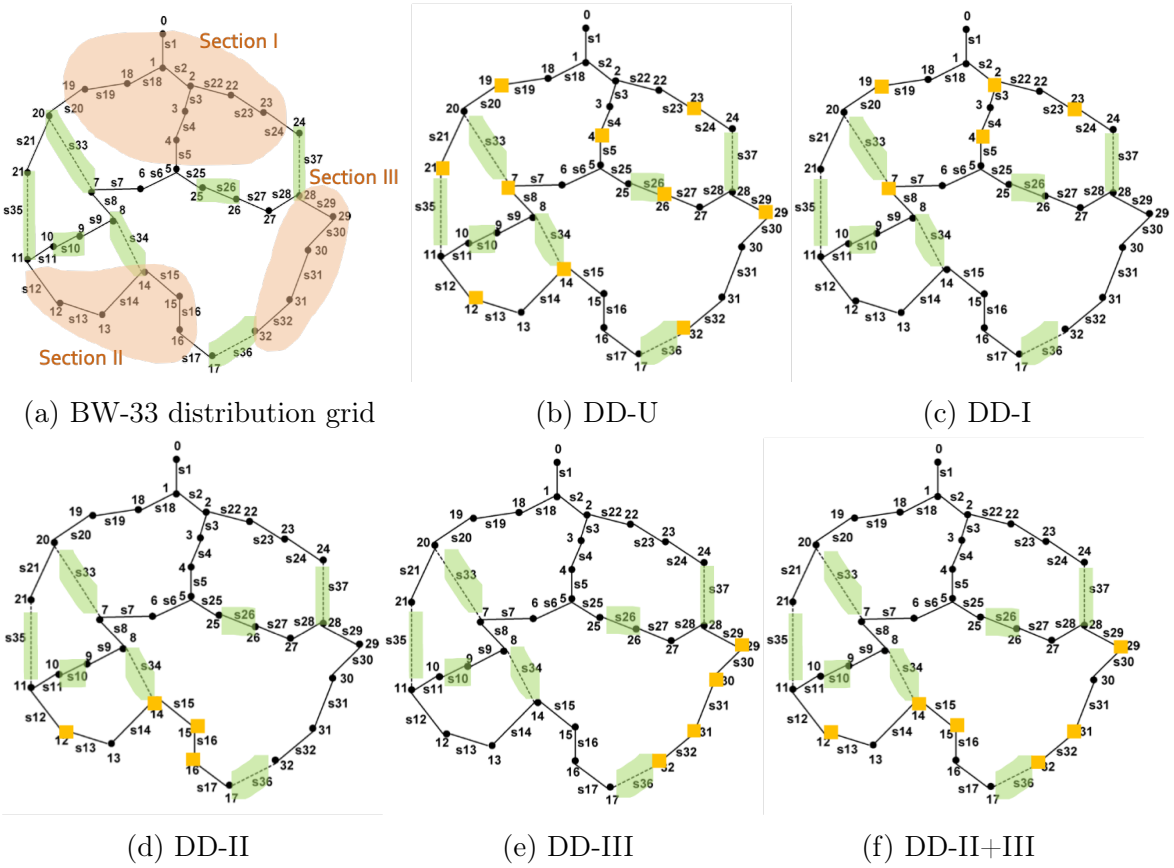
Load Location and Profiles

The location of loads and their nominal power demand (P and Q) for a single period is presented in Table A.2. To develop a diverse set of training data, the maximum load perturbation at each node is restricted to 70% deviation from nominal value (i.e. $P' = \delta P, \delta = [0.3, 1.7]$). The power factor of the loads, which describes the relationship between the real and reactive power as $pf = P/(\sqrt{P^2 + Q^2})$, is kept constant to the pf in the nominal data. This is a common approach used in literature [51, 151, 29].

Distributed Generation: Community Solar Dataset

We add a range of community solar facilities (each <5MW), up to a penetration of 25.3% of nameplate capacity to baseline load. This is a modest DER penetration compared to that which we would expect in the future grid, and reflects solar uptake now and over the next few years. We divide the grid into sections, based on the location of switches, and vary the location of community solar farms amongst these sections. We denote the distribution of these DERs (DD) as follows: (i) DD-U: uniform distribution of solar throughout the grid; (ii-iv) DD-I, DD-II, DD-III: all facilities are in Sections I, II, or III of the grid respectively; (v) DD-II+III: all facilities are in Sections II and III of the grid. The DD and location of each community solar

facility is shown in Fig. 3-3, as indicated by the yellow squares. Different DDs are used to consider effect on grid reconfiguration, line losses, and voltage profiles. The location and nameplate capacity of each solar facility is provided in Appendix A in Table A.3.



(a) BW-33 distribution grid (b) DD-U (c) DD-I

(d) DD-II (e) DD-III (f) DD-II+III

Figure 3-3: BW-33 distribution grid from [20]. The switches are highlighted in green, sections of the grid highlighted and labelled in orange, and location of community solar DERs noted in yellow squares.

The solar generation data is taken from NREL's System Advisory Model (SAM) tool [106]. The data is of a 185kW distributed commercial solar PV facility, located in Phoenix, AZ, using the SunPower SPR-E19-310-COM module, and SMA America (STP 60-US-10, 400V) inverter. The DC to AC ratio is set to the default of 1.2. The desired array size is set to 220kWdc, giving a total AC capacity of 179.580kWac. All other parameters are left unchanged in the SAM setup.

Dataset

The BW-33 grid dataset consists of 8760 data points, using hourly load and solar generation over a year.

3.4.2 94-Node Distribution Grid (TPC-94)

The TPC-94 grid is presented in [126], and is a practical distribution network of Taiwan Power Company. The data available in [126] is provided in Appendix A. The network consists of 11 feeders which are able to share load and generation by using tie line switches. The grid was modified by adding different loads and generation. The resulting grid is shown in Fig. 3-4, with the locations of switches, residential and commercial loads, and distributed PV installations marked.

Network Topology and Parameters

The grid consists of 94 nodes ($N = 94$), of which 11 are the T-D substations for feeder heads A thru K, and 83 are the remaining nodes in the network. The grid has 97 lines ($M = 97$) of which 14 are tie lines (NOS) and the remaining 83 are typically assumed to be sectionalizing switches (NCS). Topology and line parameter data are presented in Table A.4. To restrict the problem to that of dynamic reconfiguration during nominal operation (i.e. full network is connected), only the tie lines are considered as switches: these include the original 13 tie lines and a new tie line introduced between nodes 59 and 67. This tie line is introduced to allow the 11 feeders to be connected to one another; in the original network, there are two sub-networks consisting of feeders A, G and H, and the remaining feeders B, C, D, E, F, I, J, and K. The new tie line connects feeders H and I.

Load Location and Profiles

The location of loads and their nominal power demand (P and Q) for a single period is presented in Table A.5. In the TPC-94 grid dataset, three different load datasets are generated.

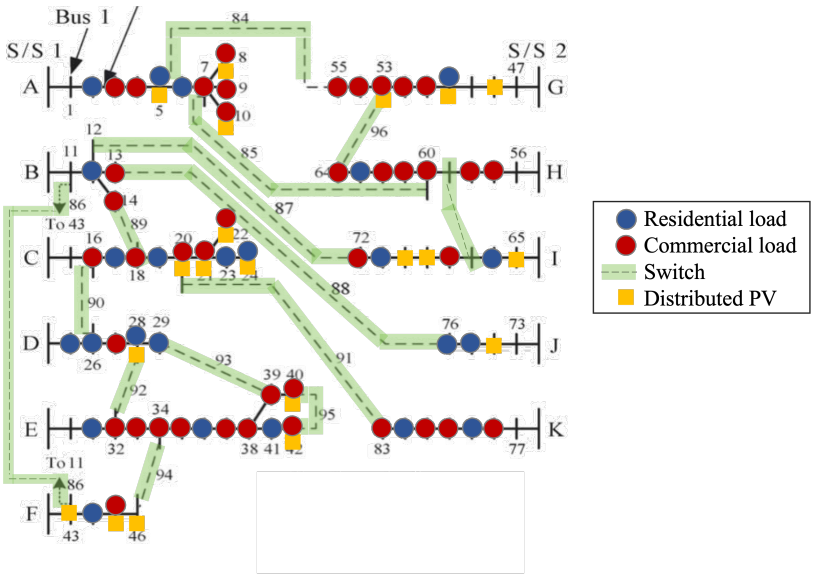


Figure 3-4: 94-node distribution grid from [126]. The switches are highlighted in green. The location of residential and commercial loads are noted by blue and red circles respectively. The location of community solar DERs are noted in yellow squares (for the S1 dataset). In the L1 and L2 dataset, all commercial loads (red) are assumed to follow residential profiles (blue). In the L3 dataset, the residential and commercial loads are located as per this figure. Note that the selection of locations of residential loads, commercial loads, and DERs, and their corresponding load/generation profiles, are not available in the original network data.

Perturbed dataset, L1: The first load dataset generates random load perturbations, similar to the BW-33 grid. The maximum load perturbation at each node is restricted to 70% deviation from nominal value (i.e. $P' = \delta P, \delta = [0.3, 1.7]$). The power factor of the loads, which describes the relationship between the real and reactive power as $pf = P/(\sqrt{P^2 + Q^2})$, is kept constant to the pf in the nominal data. Sample data over 6 days is shown in Fig.3-5(a).

Residential dataset, L2: The second load dataset assumes all loads to be residential loads, which follow one of six profiles. The power factor of the loads is kept constant to the pf in the nominal data. The six load profiles are shown in Fig.3-5(b). The six residential load profiles are described below:

- Nominal: Typical residential load profile from [39] exhibiting the characteristic bimodal distribution for residential load. The first mode occurs in the morning hours between 6-9am, when residential customers wake up and begin consuming electricity. The second mode occurs in the evening, after 6pm, when residents are returning home from work.

- Early riser: A variation of the nominal profile with both modes shifted earlier in the day. This profile represents a residential customer which wakes earlier in the morning, and retires earlier in the evening.

- Weekend/Late night: A variation of the nominal profile for weekend residential consumption. This profile has higher electricity usage throughout the day, with a unimodal distribution of afternoon and evening consumption.

- Early-Covid19 (March and April 2020): The residential load profile shifted visibly during the Covid-19 pandemic, in particular during the early months of March and April. Analysis on residential electricity demand for the province on Ontario (Canada) show significant increase in daily electricity consumption, with a delayed morning peak and a higher maximum peak in the evening hours. The overnight consumption remains typical to the nominal residential profile. Residential load profile reported by the Ontario Independent Electricity System Operator (IESO) is presented in [16], and used to inform the load profile.

- Massachusetts peak summer and winter load: A report on residential customers in Massachusetts prepared by Guidehouse Inc. presents the summer and winter peak day load patterns [12]. The summer peak day is a unimodal distribution with a steady increase in electricity usage from 5am to 8pm, and a high afternoon load. In contrast, the winter peak load profile has lower daily electricity consumption and exhibits two modes, with a peak around 8am, and an evening peak around 10pm. The primary difference in summer and winter load is the HVAC (heating, ventilation, and air conditioning) load which increases substatially in the summer months. Notably, the state of Massachusetts does not have widespread electrified heating, which would otherwise result in higher winter loads. These two profiles are adopted here.

Note that while data for the Early-Covid19 profile and the Massachusetts peak data are from different geographical regions, both of these regions experience similar climate, with 4 seasons, similar temperatures and precipitation, and similar solar irradiation. As a result, these load profiles are representative of a general distribution grid which may be located in a similar climate.

Mixed load dataset, L2 + L3: The third load dataset considers a mix of residential loads (as in L2) and commercial loads, which are selected from one of five types. The five commercial loads profiles were selected to be different from one another and different from the residential load profiles of L2. In this way, they introduce new load patterns at different nodes in the grid, and change the net load characteristic as well. The commercial load data is taken from the NREL ComStock^{TM} dataset [15] for the city of Chicago, which occupies Weather Zone 4A, similar to most of Massachusetts. The location of commercial loads was selected by matching the nominal load data from [126] to the peak hourly load from the commercial load data. Some nodes have multiple commercial loads of the same type (ex. multiple retail stores, restaurants, or office buildings at a single node). The location and type of commercial loads are shown in Table A.6. Sample data over the first 6 days of January is shown in Fig.3-5(c). The five commercial load profiles are described below:

- Hospital: The load profile has high temporal characteristics, with higher electricity consumption from 6am to 6pm. The minimum hourly load remains higher than other commercial facilities, at half the peak load. The peak hourly load of the hospital is 1700 kW, corresponding to a hospital of approximately 530,000 sq ft.

- Medium office building: The load profile has a high temporal characteristic with highest load during the morning hours of 5-7am. This commercial building has minimum load during weekends and 10pm-5am. The minimum hourly load is around 20% of peak load. The peak hourly load is 200 kW.

- Quick service restaurant: The load profile has a unique profile in that the load through the day cyclically increases and decreases. There is significant difference in electric load throughout the year, likely depending on customer load through different times of the year. The peak hourly load is 40 kW.

- Stand-alone retail space: The load profile for this commercial building is complimentary to the residential load profiles. The load during the day is high for the retail space, corresponding to lower residential loads. The demand begins reducing earlier in the day, approaching the evening residential peak. Similar to the quick service restaurant, there is significant difference in electric load throughout the year. The peak hourly load is 66 kW.

- Warehouse: This commercial building has the fewest hours of load from the selected profiles, with sustained high load during the middle of the day. The peak hourly load is 60 kW.

Distributed Generation: Solar Dataset

We add a range of distributed solar facilities to the network, up to a penetration of 23% of nameplate capacity to baseline load. This is a modest DER penetration, which once again reflects the solar uptake now and over the next few years. The location of solar facilities was selected to encourage both local utilization of generation, as

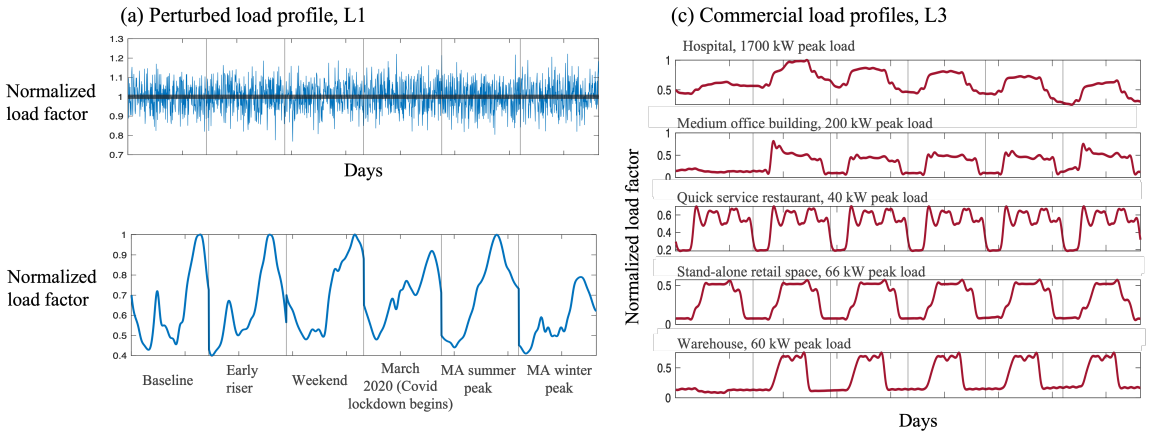


Figure 3-5: Load profiles for the perturbed dataset (L1), residential dataset (L2), and commercial load profiles for the mixed dataset (L3).

well as exports from one feeder to another. Two sets of solar location and capacity data are created: S1 corresponding to the training data (see Table A.7) and 'Solar Error' corresponding to a set of test data (see Table A.8). The location and nameplate capacity of each solar facility is provided in the corresponding tables, with each column representing a feeder of the network. Feeders B, H, and K do not have any solar facilities in the training data, while Feeder K has two solar facilities in the test data. The 'Solar Error' dataset emulates the reality that system operators have incomplete information of solar PV location and installed capacity, particularly in regions with lower rates of solar adoption, or new growth in installed solar. The lack of DER visibility is a known concern [11]. In the 'Solar Error' dataset the location and capacity of the PV units may be different than what the operator thought they were (i.e. the training data available). Of note, Feeder G is assumed to have very good information so there are no errors in solar installation data, while the operator has no visibility into the two installations at Feeder K. This dataset can also represent changes in solar adoption over time, and tests the ability of the machine learning algorithm to perform accurate predictions when system conditions change. Compared to the 23% penetration in the S1 dataset used for neural training, the 'Solar Error' has a penetration of 26% of nameplate capacity to baseline load.

To match the load and solar data, the solar generation data is taken from NREL's

Solar Power Data for Integration Studies, which consist of synthetic solar PV power plant data points for the United States representing the year 2006. The dataset for Massachusetts was used, for a 12MW distributed PV unit. The dataset provides solar generation at 5-min intervals. The particular file used is *Actual_42.55_-72.55_2006_DPV_12MW_5_Min.csv*. Figure 3-6 shows a sample generation profile of the solar facility, over 6 days.

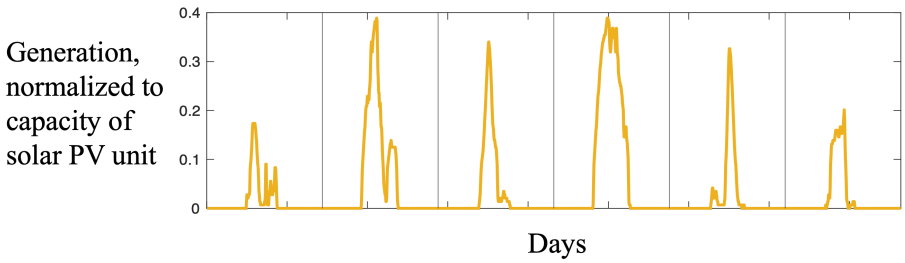


Figure 3-6: Sample solar PV generation profile for distributed PV located in Massachusetts, across 6 days.

TPC-94 Datasets

Using the load and generation data described above, the following five datasets were constructed for neural network training and testing. Each dataset has 17280 data points, generated by using load and solar generation data at 5-minute intervals over 60 days. It must be noted that the load data above includes weekend-weekday variations in demand (most notably for the commercial load profiles), and both load and generation data include seasonal variations (most notably for the solar generation).

- DS-1 (Perturbed): Uses perturbed load dataset L1 and solar generation S1, with solar resources introduced in the TPC-94 grid as per Table A.7

- DS-2 (Residential): Uses residential load dataset L2 and solar generation S1, with solar resources introduced in the TPC-94 grid as per Table A.7

- DS-3 (Mixed): Uses mixed load dataset L2+L3 and solar generation S1, with solar resources introduced in the TPC-94 grid as per Table A.7

- DS-4 (Solar error): Uses residential load dataset L2 and solar generation S1, with solar resources introduced in the TPC-94 grid as per Table A.8

- DS-5 (Flat solar): Uses residential load dataset L2, with solar resources introduced in the TPC-94 grid as per Table A.7. Solar resources are assumed to be generating at nameplate capacity at all times

3.5 Power Systems Implications of Dynamic Reconfiguration

In this section the value of dynamic reconfiguration is investigated by simulating optimal reconfigurations of the BW-33 and TPC-94 grids. The datasets developed and presented in the prior section are used as input data to the reconfiguration MILP (Eq. 3.1). The MILP is solved using Gurobi, a state-of-art commercial solver for mixed integer programming. The impact of dynamic reconfiguration is assessed in terms of grid efficiency, operability, and clean energy directives. The corresponding metrics are as follows:

- **Grid efficiency:** line losses incurred in delivering power to loads. The losses are calculated as $R_{ij}\frac{(P_{ij}-P_{ji})^2+(Q_{ij}-Q_{ji})^2}{v_i}$, where v_i is the squared voltage magnitude of the parent node for line $\{i,j\} \in \mathcal{T}_D$ in a radial network [20]. Line losses are measured in units of power or energy;

- **Grid operability:** voltages must remain within operating limits (ANSI limits of $\pm 5\%$ of the base system voltage in North America, or IEC limits of $\pm 10\%$ in Europe). Any under- or overvoltage violations must be mitigated. We consider the voltage distribution, number of undervoltage events ($\sum_{j\in\mathcal{B}} \mathbb{I}_{\sqrt{v_j}<0.95}$), and average grid voltage ($\frac{1}{N}\sum_{j\in\mathcal{B}}\sqrt{v_j}$);

- **Clean energy directives:** amount of available PV generation that is dispatched to meet loads. The higher the PV utilization, the more clean energy is used. PV utilization is measured as the amount of PV generation dispatched

as a ratio of the available PV resource: $PV_{util} = \frac{P^{G*}}{\overline{P}^{G}}$. Similarly, the PV curtailment is measured as the amount of PV generation curtailed as a ratio of the available PV resource: $PV_{curt} = \frac{\overline{P}^{G} - P^{G*}}{\overline{P}^{G}}$. Trivially, $PV_{util} + PV_{curt} = 1$.

The following three reconfiguration approaches will be considered:

No reconfiguration: the default grid topology as given in the datasheet is used. All switches are assumed to be in their default position (normally open NOS, or normally closed NCS).

Static reconfiguration (StatR): determines a fixed set of switch-states that optimize losses over a long term, such as a few months or a year. The load and generation forecasts over the period will be considered, and a robust optimization approach can be taken to determine the optimal topology across all grid conditions.

Dynamic reconfiguration (DyR): determines the switch-states which minimizes losses for a given load and generation condition for a particular period. This period is shorter in length than for StatR, such as a few minutes, hours, or days. The introduction of DERs necessitates DyR where local DERs supply loads in closer proximity to them, thus reducing losses, improving voltage profiles, and increasing PV (renewable solar energy) utilization.

The three outcomes (loss reduction, voltage improvement, and increased PV utilization) are investigated below, on simulations of the BW-33 and TPC-94 grids.

3.5.1 Dynamic Reconfiguration Reduces Electrical Line Losses

The following set of results are for the BW-33 grid, a very lossy network: recall that losses average about 8% of total load. Table 3.2 shows the significant loss reduction enabled by dynamic reconfiguration for the BW-33 as DER locations are varied, upwards of 23% PV penetration. The StatR closes tie line 35 and opens NCS 10, and keeps this topology fixed. The DyR selects primarily between two states: (a) closing tie line 35 and opening NCS 10, and (b) closing tie lines 35 and 36 and opening NCS 10 and 26. For a grid without any DERs versus a grid with DERs, the loss reduction from reconfiguration is higher without DERs; this can be attributed to greater losses

Table 3.2: Loss reduction using StatR and DyR

	StatR vs. no reconfig % Loss reduction, MW saved per year *	Grid-SiPhyR vs. StatR MW saved per year *
No DERs	23%, 370 MW	0 MW
DD-U	20%, 300 MW	23 MW
DD-I	20%, 320 MW	31 MW
DD-II	19%, 270 MW	20 MW
DD-III	21%, 310 MW	27 MW
DD-II+III	20%, 280 MW	28 MW

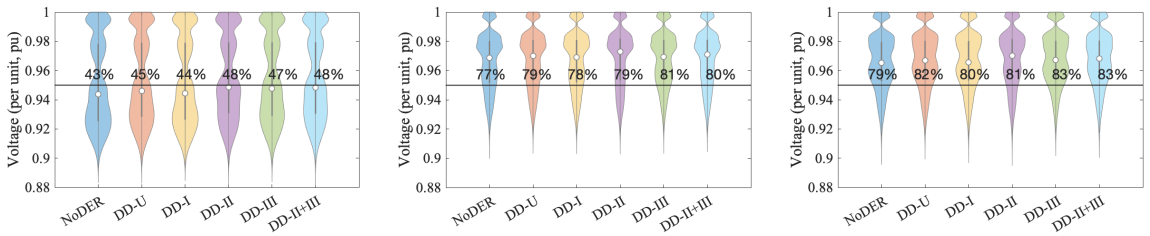
*The MW (power) saved is equal to the MWh (energy), as the simulation is run for every hour of the year and loss reduction summed for every test case

without leveraging local generators which are located closer to loads and thus incur lower losses when supplying those loads. The second column compares DyR to StatR, showing savings up to 30 MW for a single distribution feeder.

While this is a modest 2.5% improvement of DyR upon StatR, this is a nontrivial reduction for distribution grid operators and utilities. Current industry standard methods which aim to reduce the load on a distribution grid, such as Conservation Voltage Reduction (CVR), typically reduce peak demand by a modest 2 to 2.5%; for a utility with a peak load of 100 MW, this translates to savings of US$200,000 per year [18]. Further, it should be noted that this was obtained with a small test case (33 nodes). As the dimension increases, with increasing penetration of DERs and switches, with disparate patterns and topologies, it is expected that this difference will be significantly larger.

3.5.2 Dynamic Reconfiguration Improves Voltage Profile Across the Grid

The following set of results are for the BW-33 grid, a grid with significant voltage violations. Figure 3-7 plots the voltage distribution across the grid for an entire year. The ideal distribution is the shape of a short ice cream cone - wide on top and narrow on the bottom, with the tip above the line indicating the ANSI minimum voltage limit of 0.95pu. The width of the plot indicates the total number of voltage



(a) No reconfiguration (b) StatR (c) Grid-SiPhyR

Figure 3-7: Voltage distribution over a year (8760 hours). Solid line is ANSI lower voltage limit.

observations at the y-axis value, i.e. wider indicates more voltage observations. On each plot is printed the percentage of time the voltages are within the ANSI limits, where higher numbers are better. We make the following key observations: (i) without reconfiguration, the grid performs very poorly, violating ANSI limits 50-60% of the time, and voltages drop to 0.88pu (outside of ANSI limits); (ii) reconfiguration (Stat or Dy) significantly improves the voltages across the grid, with minimum voltage improving to 0.9pu, and ANSI limits satisfied 77-83% of the time (IEC limits are always satisfied); (iii) DyR reduces the number of voltage violations throughout the year by 2%, as compared to StatR, which is a significant improvement as undervoltage can result in brownouts and even lead to blackouts. We note that since the grid chosen for this test case is very lossy, in our simulations we enforce a lower voltage limit of 0.87pu to ensure feasibility of loading conditions (instead of 0.95pu), which our physics-informed framework always satisfies. While it is interesting to note the simple case study of the BW-33 grid does not imply a preferred DD over others, higher dimensions will expect to naturally suggest an optimal DD.

3.5.3 Dynamic Reconfiguration Enables Better PV Utilization, by Connecting Generation with Loads

The following set of results are for the TPC-94 grid. Figures 3-8 and 3-9 show the dispatch results aggregated at the feeder level for the DS-2 and DS-3 datasets. The load in each feeder can be met by importing power from the bulk grid at the local T-D substation (i.e. the PCC of the same feeder), importing power from neighbouring

feeders, or using local PV generation. Most feeders serve the load using imports from the bulk grid, while Feeders A and E import significant amounts of power from neighbouring feeders. Feeder F has very high PV penetration and when solar generation is high, can meet most of its load locally.

Figure 3-10 shows a configuration change in the TPC-94 grid, where Feeder G connects to neighbouring Feeder H when solar generation becomes available. Power from Feeder G is exported to Feeder H and A. While the remaining network configuration does not change and the direction of power flow remains the same (i.e. the same networks export power), the amount of power transferred across the tie lines reduces substantially. Local PV generation supplies power to local loads, and the configuration change allows the feeders to meet their remaining load in an efficient way. Notably, the configuration change allows PV utilization to increase in the network.

Impact of DyR on PV Utilization

The PV utilization results are summarized in Table 3.9, for StatR and DyR applied to the DS-2 case, and DyR applied to the DS-3 case. The comparison of StatR and DyR show that the optimal reconfiguration of the TPC-94 grid can reduce PV curtailment from 23% to 19%. This corresponds to an overall reduction in PV curtailment by 17%, corresponding to an increase of 250MWh of solar energy used annually. To put into perspective, this is enough solar energy to power approximately 23 US households for a year[1], and results in a decrease of 107 metric tons of CO_2[2]. It must be stressed that this increase in solar energy does not require new solar installations. Instead, a change in operating paradigm from static to dynamic permitted by Grid-SiPhyR allows solar PV to be used more effectively, while simultaneously increasing operating efficiency of the distribution grid.

[1]This is approximated by 2021 EIA data, where the average annual electricity consumption for a U.S. residential utility customer was 10,632 kilowatthours (kWh), averaging around 886 kWh per month.

[2]The CO_2 emissions in Massachusetts from electricity generation in 2021 was 974 lbs/MWh. This value was used to approximate the CO_2 abatement from DyR.

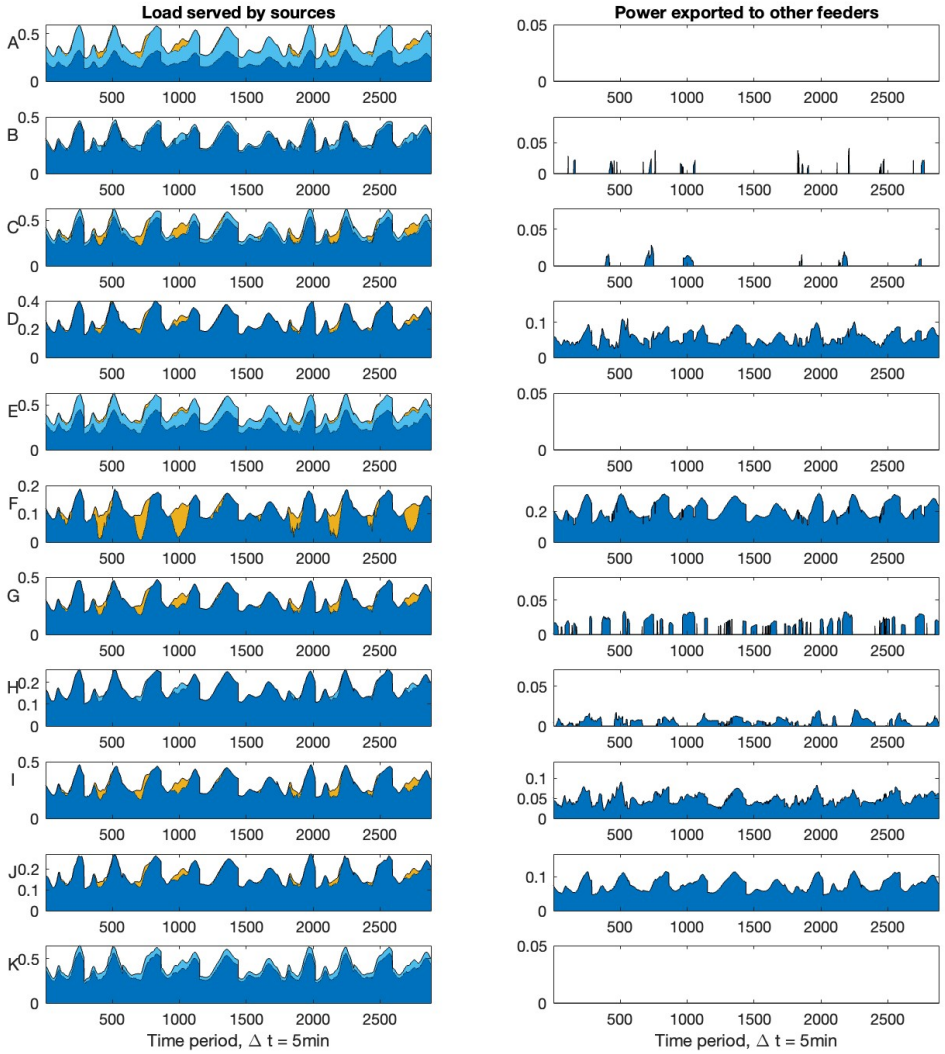


Figure 3-8: Optimal dispatch results are shown for the TPC-94 network with DS-2 (all residential loads), over a period of 10 days. The optimal reconfiguration is determined every 5-minutes, resulting in 2880 periods. All plots are in per unit measurements. **(Left)** A breakdown of the load served at each feeder, A thru K. Dark blue: load served by the bulk system at the T-D substation of the same feeder; Light blue: load served by power imported from neighbouring feeders; Yellow: load served by local distributed solar generation. **(Right)** Power exported to neighbouring feeders.

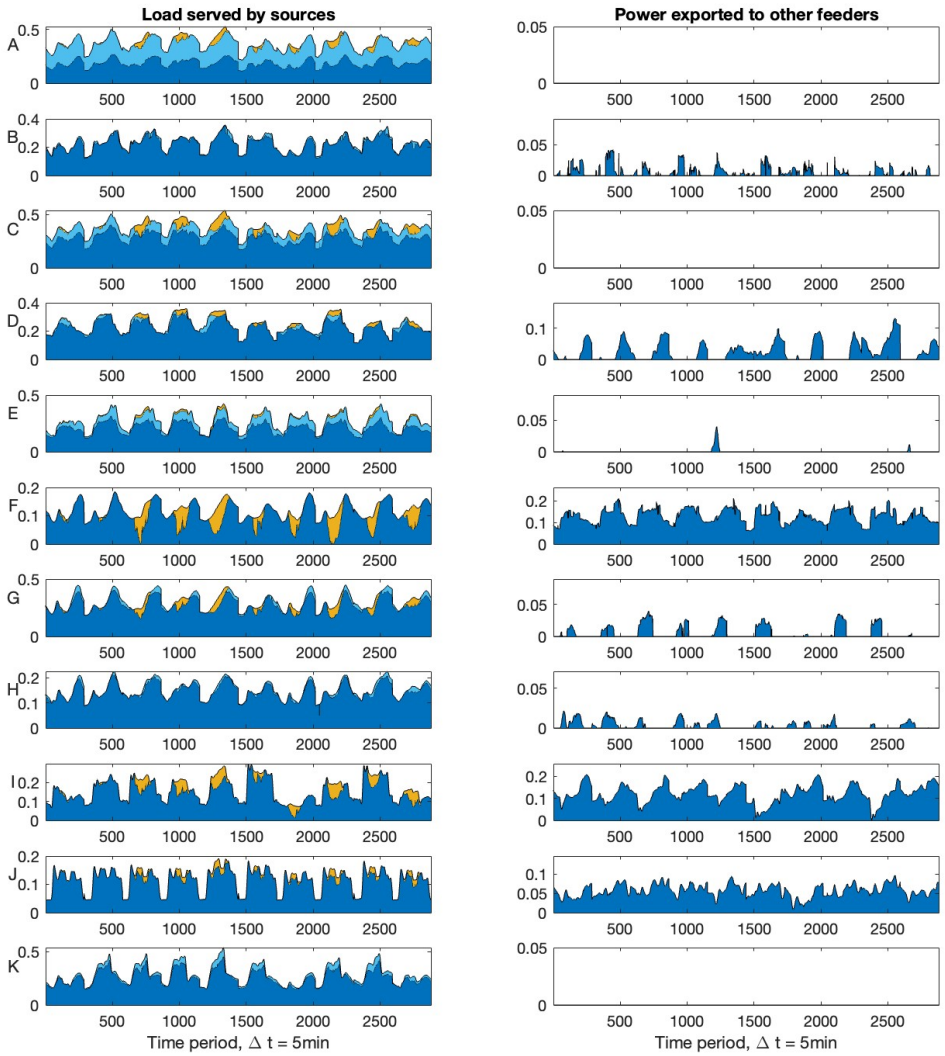


Figure 3-9: Optimal dispatch results are shown for the TPC-94 network with DS-3 (mixed residential and commercial loads), over a period of 10 days. The optimal reconfiguration is determined every 5-minutes, resulting in 2880 periods. All plots are in per unit measurements. **(Left)** A breakdown of the load served at each feeder, A thru K. Dark blue: load served by the bulk system at the T-D substation of the same feeder; Light blue: load served by power imported from neighbouring feeders; Yellow: load served by local distributed solar generation. **(Right)** Power exported to neighbouring feeders.

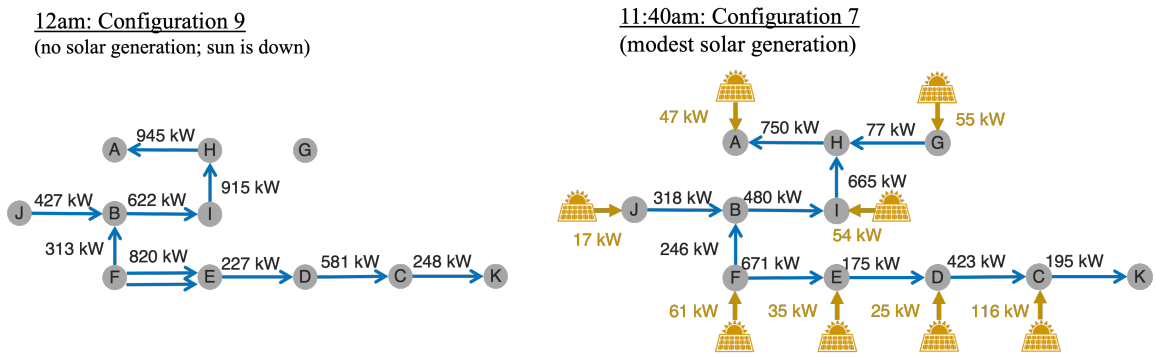


Figure 3-10: Two configurations for the TPC-94 network showing the switches connecting neighbouring feeders. These results correspond to periods 1 and 140 in the dispatch results of Fig. 3-8.

Impact of Mixed Load Profile

Overall the change in load pattern does not result in substantial differences in the resource dispatch decision. The amount of imported electricity increases in Feeder A when commercial loads are introduced in the network. Feeders A, C, E, and K remain importers of electricity. Meanwhile the power exported from Feeder D and F decreases, while power exported from Feeder I increases. Most notably, Feeders I and J have distinctly different profiles due to the composition of the commercial loads introduced (warehouse and an office building in Feeder I, and an office building and quick service restaurant in Feeder J). In particular, the quick service restaurant in Feeder J dominates the load profile, with the distinct load pattern through the day. As a result of the change in load type, the power exported from Feeder J also varies more than in the DS-2 case, with Feeder J connected to and exporting power to Feeders B and I. The impact of mixed loads (DS-3) on the PV utilization is negligible as compared to residential load (DS-2), with 20% and 19% curtailment respectively.

3.5.4 Frequency of Reconfiguration Events Depends on System Conditions

The following set of results are for the TPC-94 grid. Figures 3-11 to 3-13 show the reconfiguration results when using different load profiles, corresponding to datasets

Table 3.3: Summary of reconfiguration results for the TPC-94 grid, for different load profiles.

	DS-1	DS-2	DS-3
# optimal configurations over 6 days	27	10	27
# optimal configurations over 60 days	230	12	37
Avg. time between reconfiguration events	5 minutes	1.5 hours	1 hour

DS-1, DS-2, and DS-3. Each figure plots the total system load (top), the total solar generation (middle), and the optimal configuration number (bottom). The configuration numbers identify which optimal topology the grid should be in given the load and generation conditions. The reconfiguration results are summarized in Table 3.3. Note that the configuration number (bottom plot of each figure) is different for each figure; i.e. configuration 1 of Fig. 3-11 may not be the same as configuration 1 of Fig. 3-12. What is of note across the three figures is the number of configurations triggered and the different patterns visible in when configuration changes are triggered.

From these figures and the table, the impact of load models on the reconfiguration problem is immediately obvious. The perturbed load data in DS-1 results in an unrealistically large number of unique optimal configurations in the 60 day period, and reconfiguration events on average every 5-minutes. In comparison, the DS-2 and DS-3 datasets which use real load profiles only have 27 and 37 unique configurations respectively (compared to 230 for DS-1), and result in reconfiguration events every 1.5 and 1 hour, respectively. These results motivate the need for representative datasets which reflect real system load and generation characteristics. The importance of representative datasets will be revisited when evaluating the performance of the proposed physics-informed neural framework.

Further, the configuration number plots in Fig. 3-12 and Fig. 3-13 motivate the need for data-driven methods. Visual analysis of Fig. 3-12 can derive the dominant reconfiguration rules from the temporal coupling of load and solar: Rule (1) is to use configuration 9 when low solar generation, Rule (2) is to use configuration 4 when high solar generation, and Rule (3) is to use configuration 7 or 8 at all other times. Even more obvious is that a neural framework for dynamic reconfiguration

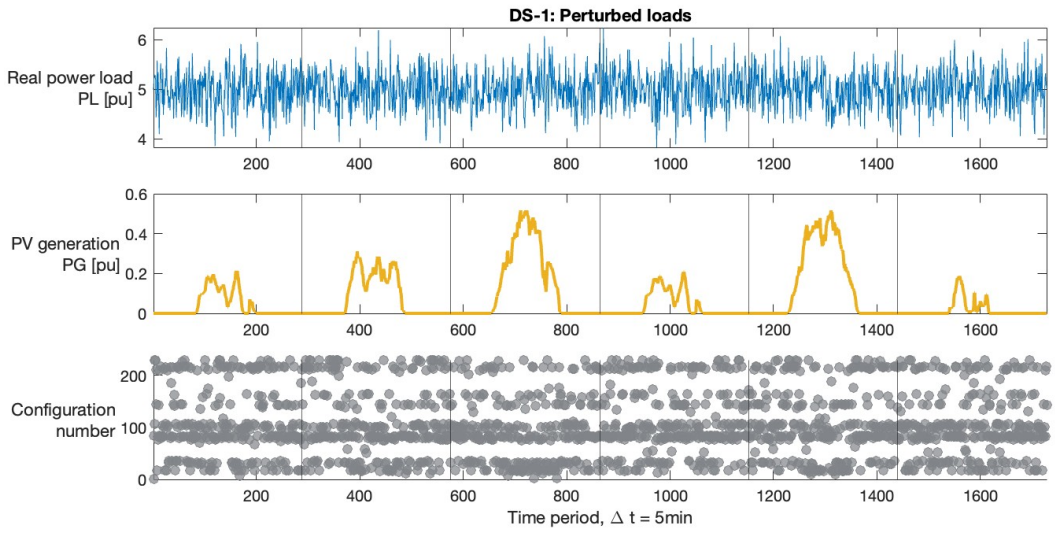


Figure 3-11: Reconfiguration results for the TPC-94 grid with DS-1 dataset. Results are presented over 6 days (1728 periods). Note that given the random nature of the load data, the configuration results over all 60 simulated days are difficult to view.

may not be needed for DS-2, where there are three dominant configurations: 7, 8, and 9. Rather, heuristic operational schemes can be deduced from these simulations. However, a similar visual analysis of Fig. 3-13 is not as trivial. The load profiles are heterogeneous and a larger number of configurations triggered (37 versus 12). The result is a configuration map which is more complex, requiring dynamic reconfiguration and a suitable optimization framework to be able to make these decisions in a timely manner.

3.5.5 Statistical Analysis of Results

In this section we investigate the statistical significance on the results presented above, pertaining to loss reduction, voltage improvements, and increased PV utilization. The analysis of variance (ANOVA) test is used to determine statistical significance, and the results are presented below.

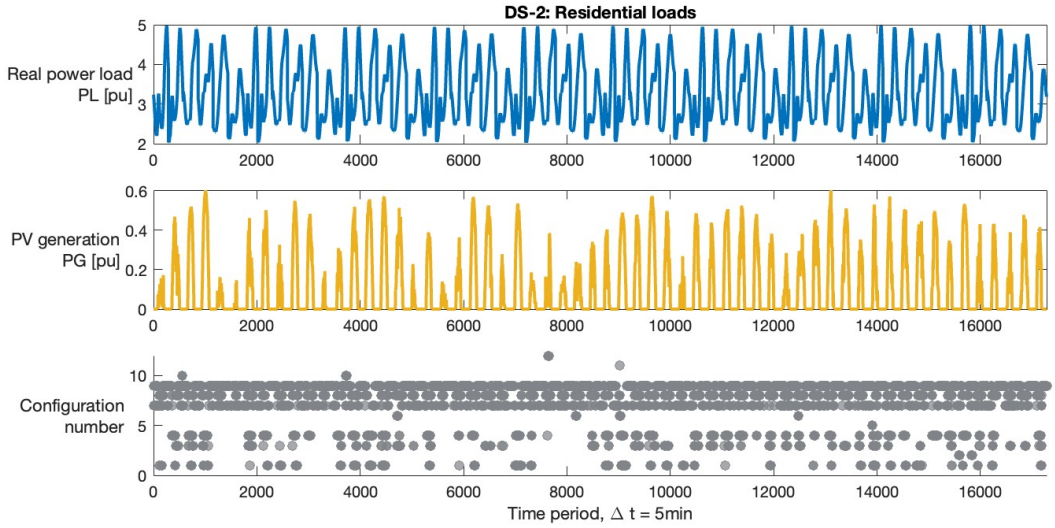


Figure 3-12: Reconfiguration results for the TPC-94 grid with DS-2 dataset. Results are presented over 60 days (17280 periods).

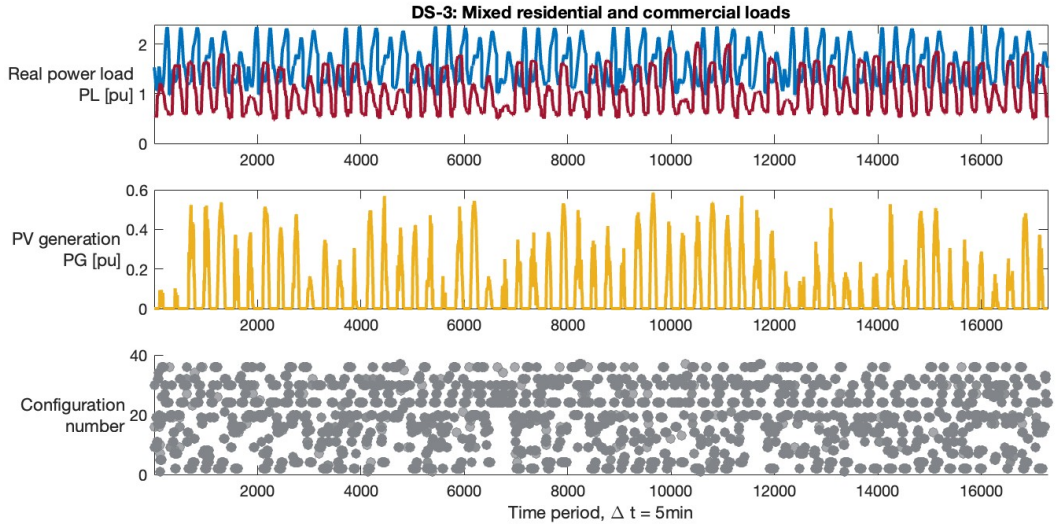


Figure 3-13: Reconfiguration results for the TPC-94 grid with DS-3 dataset. Results are presented over 60 days (17280 periods).

Table 3.4: Summary of line losses in BW-33 for DD-U

	No reconfiguration	StatR	DyR
Total loss (MWh)	1486.5	1178.2	1154.9
Mean loss per hour (MWh)	0.1697	0.1345	0.1318
Number of samples (1-hr intervals)	8760		

Table 3.5: ANOVA table for line loss reduction in BW-33 for DD-U

Source	Sum of Squares	Degrees of Freedom	Means Squares	F
Factor (Between)	7.8199	2	3.9099	2.2514
Error (Residual)	45634	26277	1.7366	
Total	45642			

Statistical Analysis of Line Loss Reduction

The null hypothesis is as follows: $\mathcal{H}_0$: on average the line loss reductions offered by reconfiguration on the BW-33 grid are not significant. The alternative hypothesis is as follows: $\mathcal{H}_1$: the mean line loss across the different methods is significant. The ANOVA test will determine whether to accept or reject $\mathcal{H}_1$. Table 3.4 presents a summary of the experiment data for the BW-33 grid, and Table 3.5 presents the resulting ANOVA table. For a significance level of $\alpha = 0.05$ the critical value is 3. As seen in the ANOVA table, the F statistic is 2.25. Thus, there is no statistically significant evidence at $\alpha = 0.05$ to show that there is a difference in mean line loss across the different methods. We accept the null hypothesis $\mathcal{H}_0$. Note that the same result holds at significance levels of $\alpha = 0.01$ and $\alpha = 0.10$.

Statistical Analysis of Voltage Improvements

Two sets of tests are carried out, for the number of voltage violations per period and the average network voltage. The null hypothesis are as follows:

- $\mathcal{H}_0^1$: on average the number of voltage violations is the same across all methods;
- $\mathcal{H}_0^2$: on average the mean grid voltage is the same across all methods.

The alternative hypothesis are as follows:

Table 3.6: Summary of voltage data in BW-33 for DD-U

	No reconfiguration	StatR	DyR
Average network voltage (pu)	0.9077	0.9354	0.9354
Total number of undervoltage violations	157873	59999	53160
Average undervoltage violations per hour	18.0220	6.8492	6.0685
Number of samples (1-hr intervals)	8760		

Table 3.7: ANOVA table for number of voltage violations in BW-33

Source	Sum of Squares	Degrees of Freedom	Means Squares	F
Factor (Between)	7.8352e5	2	3.9176e5	0.0033
Error (Residual)	3.1343e12	26277	1.1928e8	
Total	3.1343e12			

- $\mathcal{H}_1^1$: on average the number of voltage violations are different across the three methods. Reconfiguration is significant in reducing the number of voltage violations;

- $\mathcal{H}_1^2$: on average the mean grid voltage is different across the three methods. Reconfiguration is significant in improving the average grid voltage.

Table 3.6 presents a summary of the experiment data for the BW-33 grid pertaining to grid voltages, and Tables 3.7 and 3.8 present the resulting ANOVA tables. For a significance level of $\alpha = 0.05$ the critical value is 3. At this significance level, we accept $\mathcal{H}_0^1$ and reject $\mathcal{H}_0^2$: there is no statistically significant evidence that reconfiguration reduces the number of undervoltage violations, but there is evidence that reconfiguration improves the average grid voltage.

Table 3.8: ANOVA table for average network voltage in BW-33

Source	Sum of Squares	Degrees of Freedom	Means Squares	F
Factor (Between)	4.3392	2	2.1696	295.2281
Error (Residual)	193.1048	26277	0.0073	
Total	197.4439			

Table 3.9: Summary of PV utilization in TPC-94

	StatR | L2	DyR | L2	DyR | L2 + L3
Total available PV generation (MWh)	1042	1042	880.3
Total utilized PV generation (MWh)	802.7	843.7	701.0
Total curtailed PV generation (MWh)	239.5	198.5	179.4
Percentage of curtailed PV generation	23%	19%	20%
Average utilized PV generation per hour (kWh)	557.4	585.9	486.8
Average curtailed PV generation per hour (kWh)	166.3	137.9	124.5
Number of samples (5-min interval)		17280	

Table 3.10: ANOVA table for PV curtailment in TPC-94

Source	Sum of Squares	Degrees of Freedom	Means Squares	F
Factor (Between)	6.9904	1	6.9904	11.9325
Error (Residual)	20245	34558	0.5858	
Total	20252			

Statistical Analysis of Increased PV Utilization

The null hypothesis is as follows: $\mathcal{H}_0$: on average the PV curtailment is the same across the two methods (StatR and DyR). The alternative hypothesis is as follows: $\mathcal{H}_1$: on average the PV curtailment is different across the two methods (StatR and DyR), and DyR is significant in decreasing PV curtailment (i.e. increasing utilization). Table 3.9 presents a summary of the experiment data for the TPC-94 grid, and Table 3.10 presents the resulting ANOVA table. For a significance level of $\alpha = 0.05$ the critical value is 3.84. As seen in the ANOVA table, the F statistic is 11.9. Thus we reject the null hypothesis $\mathcal{H}_0$, and affirm that there is statistically significant evidence that DyR increases PV utilization on the TPC-94 grid.

3.6 Experimental Setup

The results in the prior section motivated the need for DyR, to reduce line losses, improve grid voltages, and increase PV utilization. To enable DyR in the distribution grid, new computational tools are needed, such as the proposed physics-informed

neural framework for end-to-end learning to optimize. In the following sections the performance of the neural framework will be investigated. This section presents the experimental setup and key performance metrics.

3.6.1 Simulation Parameters

The proposed physics-informed framework leverages grid physics and constraints to enable the end-to-end training of a machine learning tool for optimization, while retaining a light-weight neural network. As such, the neural network has only two hidden layers. The size of input and output layer is determined by x and $\hat{z}$. Each layer applies a linear transformation with bias, batch normalization, and ReLU activation. The weights are initialized with He initialization. Backpropagation uses an adaptive learning rate (Adam), with parameter $\gamma = 0.001$. The learning rate provided to the Adam optimizer is selected as the largest possible learning rate which permits neural training (typically $lr = 0.001$). The soft loss penalty on inequality constraint violation is set to $\lambda_h = 100$, chosen to enforce a high penalty on violating inequality constraints, while still allowing the underlying objective function of minimizing network losses to be improved upon. The datasets are split as 80% training, and 10% testing and validation each, using mini-batching with 200 batch size. All neural architectures were developed and tested using PyTorch on an Apple M2 Max with 12-core CPU and 96GB RAM.

Committee Machine

A committee machine approach is taken in evaluating the neural architectures, using *ensemble averaging*, also called the *ensemble committee of networks* approach. For each configuration and set of hyperparameter values, a set of 10 models are trained with independent weight initialization. The predictions from these 10 trained models (also called *predictors* or *experts*) are linearly combined to produce the final prediction. A simple linear combination is used, where each predictor is equally weighted. The ensemble approach can help reduce variance of neural networks by training mul-

tiple models and thereby reducing the impact of stochastic weight initialization and learning inherent in neural networks [23, 68].

Baselining Against Optimal Solution

The datasets for both BW-33 and TPC-94 are solved using Gurobi, the state-of-art commercial solver for mixed integer programs. The reconfiguration MILP is implemented via the Yalmip *optimizer* environment in MATLAB. The default parameters in Gurobi are used. The solution from Gurobi is treated as the baseline optimal solution, (z^*, φ^*) for every x input data. This optimal point is used to train a supervised network (see below) and evaluate the performance of the neural networks (see below).

3.6.2 Neural Architectures Tested

Figure 3-14 illustrates the set of architectures tested. These are described below.

Without Physics-Informed Rounding

This approach favours values of z^C which are zero or close to 1, but does not enforce integrality. The resulting switch status predictions may still be non-integer values, and radiality may not be enforced.

Sigmoidal activation for integer variables (InSi): The proposed physics-informed rounding is compared with a differentiable relaxation of the step function, using a steep sigmoid activation function at the output layer: $\sigma_{InSi}(z) = \left[2\frac{1+\mu}{\mu+e^{-\gamma z}} - 1\right]_+$, where γ, μ are free parameters ([27]). This integer sigmoid activation is used for the binary variables $\hat{z}^B$, while the continuous variables $\hat{z}^C$ still pass through the traditional sigmoid function. We set the parameters $\mu = 1$ and $\gamma = 5$, where γ governs the sharpness of the sigmoidal function - and hence how well the sigmoid approximates the step function. Larger values of γ better discriminate between binary values, but render the function less differentiable and thus learning more challenging. The InSi approach favours values of z^C which are zero or close to 1, but does not enforce integrality. The resulting switch status predictions may still be non-integer values.

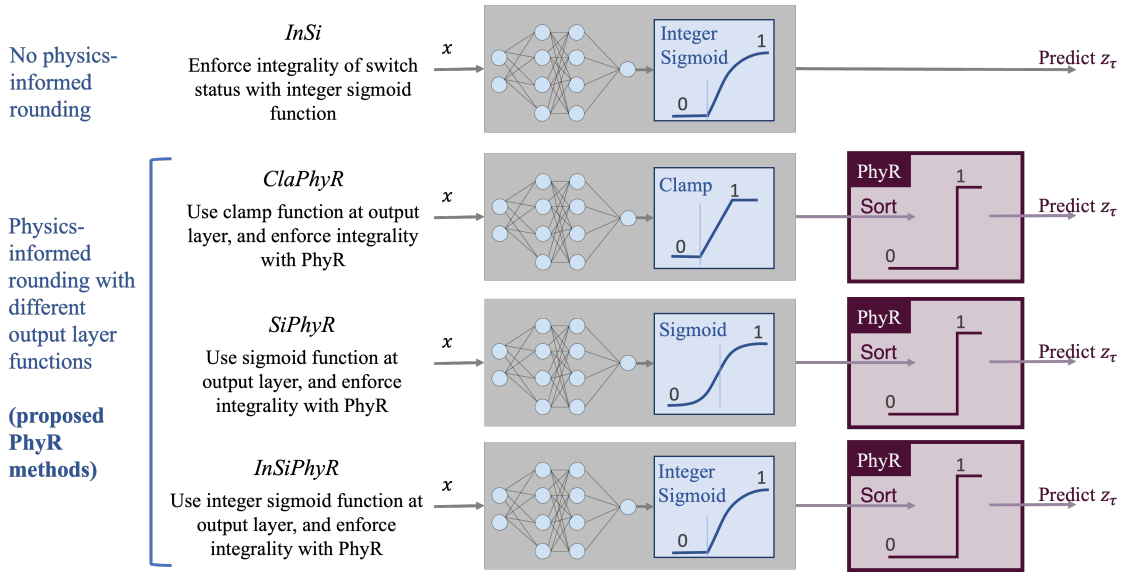


Figure 3-14: Different neural architectures tested, with and without PhyR

Sigmoidal activation for integer variables with rounding during testing (InSi2R): Use the InSi architecture during training. During testing the predictions z_τ are rounded to binary values before the variable space completion.

With Physics-Informed Rounding

Different output layer functions are tested with the proposed physics-informed framework. The output layer functions render different switch status probabilities (all between 0 and 1), while the use of PhyR enforces integrality. The output layer function is used for the binary variables $\hat{z}^B$, after which the topology variables $\hat{z}_\tau$ pass through the PhyR function. The resulting switch status predictions will always be integer values, and the necessary condition for radiality will be enforced.

Clamp at output layer (ClaPhyR): use a clamp function at the output layer: $\sigma_{Cla}(z) = \min\{\max\{0, z\}, 1\}$.

Clamp at output layer (SiPhyR): use a sigmoidal activation function at the output layer: $\sigma_{Si}(z) = \frac{1}{1+\exp(-z)}$.

Clamp at output layer (InSiPhyR): use the integer sigmoid function at the output layer (see InSi above).

Unsupervised learning

The InSi and PhyR-based methods are all unsupervised frameworks. The loss function for the dynamic reconfiguration problem is chosen to minimize the power loss over the lines, which is approximated as:

$$f_x(z, \varphi) = \sum_{(i,j) \in \mathcal{T}_D} (P_{ij}^2 + P_{ji}^2 + Q_{ij}^2 + Q_{ji}^2) R_{ij} \tag{3.14}$$

The soft loss penalizes inequality constraint violation on dependent variables with hyperparameter λ_h. Note that Linear DistFlow power flow model is composed of all linear constraints, so the completion step is trivial and results in no equality constraint violations. For nonlinear power flow models, a penalty for equality constraints can also be included. The resulting convex loss function is: $l = f_x(z, \varphi) + \lambda_h \| \max \{0, h_x(z, \varphi)\} \|_2^2$.

Supervised Learning

Performance is also compared to a supervised framework, where the neural network is given full information about the optimal solution. The input training data remains the same (load at each node), but the regression loss function is introduced with knowledge of the optimal solution (z^*, φ^*).

Supervised-x: method 'x' is trained in a supervised framework. The loss function used is a mean squared error as for a standard regression problem:

$$l_{sup}(z, \varphi) = \|(V_j - V_j^*)^2 + (P_j^G - P_j^{G*})^2 + (Q_j^G - Q_j^{G*})^2 \|_2^2 + \|(y_{ij} - y_{ij}^*)^2 \|_2^2 \tag{3.15}$$

Notably, the variables of interest to topology selection (y_{ij}), resource dispatch (P^G, Q^G), and pertinent grid constraint (V) are penalized in the supervised loss function. The remaining variables (line flows) can also be penalized, but the power flow equations already enforce the relationship between variables in the variable space completion step. If the predictions for topology, resource dispatch, and voltages are optimal, then there will be no errors in the remaining variables.

Supervised-x-pen: method 'x' is trained in a supervised framework. The loss function used is an augmented mean squared error, which includes a soft loss penalty on inequality constraint violation:

$$l_{sup-pen}(z, \varphi) = l_{sup}(z, \varphi) + \lambda_h \| \max \{0, h_x(z, \varphi)\} \|_2^2 \tag{3.16}$$

Parameter Tuning Studies

Significant testing of the proposed physics-informed framework with PhyR is conducted. The different neural architectures tested are discussed in the next section. Additional hyperparameters tested include:

- the width of the neural network (number of neurons in each layer)
- the number of data points in the dataset (each split as 80/10/10)
- the dataset used in training (for the TPC-94 grid, either DS-1 or DS-2)
- the value of the soft loss penalty λ_h

3.6.3 Performance Metrics

Three sets of performance metrics are defined below to evaluate the performance of the proposed PhyR-based neural method. Lower values are better for all metrics described below. Figure 3-15 illustrates the performance metrics.

Optimality Metrics

The optimality metrics evaluate the neural network's ability to predict the optimal solution, by measuring the distance of the prediction ψ from the optimal solution ψ^* which is provided by Gurobi.

Dispatch error: the mean squared error in optimal generator dispatch

$$\frac{1}{N} \sum_{j \in \mathcal{B}} (P_j^G - P_j^{G*})^2 + (Q_j^G - Q_j^{G*})^2 \tag{3.17}$$

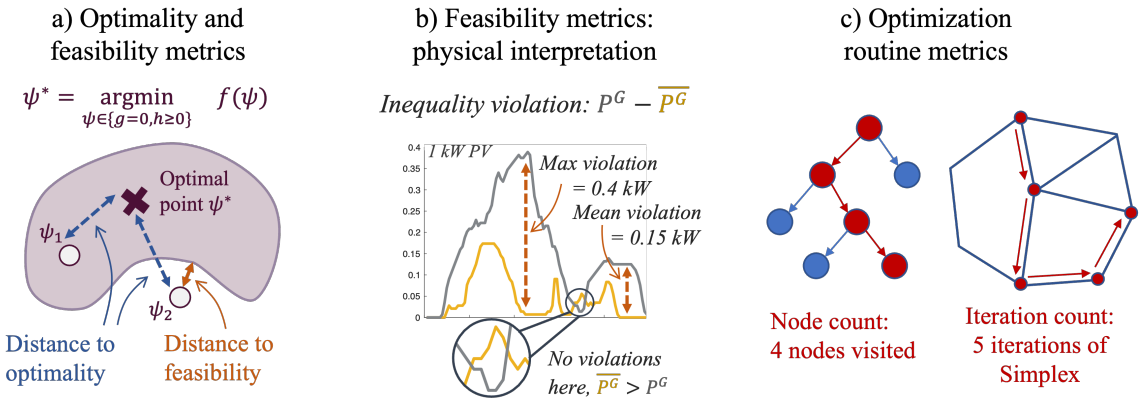


Figure 3-15: Illustration of the performance metrics. **(a) (Left)** A depiction of the feasible set is shown by the purple shape. The optimal point ψ^* as given by Gurobi is contained within the feasible set. Consider two predictions by the neural network, ψ_1 and ψ_2. The distance between these predictions and the optimal point is measured by the optimality metrics. Note that ψ_1 is within the feasible set, and so has a distance to feasibility of zero. However, prediction ψ_2 is outside of the feasible set, and has a nonzero feasibility error as measured by the feasibility metrics. **(b) (Center)** The feasibility metrics can be interpreted in units of power (kW for real power, kVA for apparent power). The illustration shows a dispatch schedule predicted by the neural network in grey (top line) and the actual available solar generation in yellow (bottom line). Wherever the prediction exceeds the available generation (grey above yellow), an inequality violation occurs. In the magnified circle there are no inequality violations as the available solar generation exceeds the dispatch command. **(c) (Right)** The node count in the branch-and-cut method and the iteration count of the resulting simplex method subroutine are shown.

Voltage error: the mean squared error in nodal voltage prediction

$$\frac{1}{N}\sum_{j\in\mathcal{B}}(V_j - V_j^*)^2 \tag{3.18}$$

Topology error: the Hamming distance of two topologies, calculated as the ratio of switch decisions not in the optimal position

$$\frac{1}{M_{sw}}\sum_{(i,j)\in\mathcal{T}_D^{sw}}(y_{ij} - y_{ij}^*)^2 \tag{3.19}$$

For the proposed PhyR-based methods where the switch on/off decision is a binary variable, the topology error is equivalent to using an indicator function which returns 1 if the switch status is not optimally selected, and 0 otherwise.

$$\frac{1}{M_{sw}}\sum_{(i,j)\in\mathcal{T}_D^{sw}}\mathbb{I}_{y_{ij}\neq y_{ij}^*} \tag{3.20}$$

Optimality gap: objective value distance between the best known solution f^* calculated at point (z^*, φ^*) from Gurobi, and the prediction f calculated at point (z, φ)

$$\frac{f - f^*}{f^*} = \frac{f_x(z,\varphi) - f_x(z^*,\varphi^*)}{f_x(z^*,\varphi^*)} \tag{3.21}$$

The objective function $f_x(z, \varphi)$ is the line losses, as in Eq. 3.14. Note that the optimality gap is always positive for a *feasible prediction*, as the best known solution f^* is a lower bound on the optimal objective value.

Feasibility Metrics

The feasibility metrics evaluate the neural network's ability to predict feasible solutions, by measuring the distance of the prediction from the feasible space. It must be stressed that the equality constraints describing the power physics and any inequality constraints on the independent variables z have certified satisfiability. These are key features of the proposed PhyR-based framework, which include the variable space

completion (to satisfy equality constraints), and the particular variable decomposition proposed. By selecting voltages v as an independent variable, these are scaled onto the box constraints describing operating limits; for a grid operator, this means voltage limits across the grid will always be satisfied, a critically important aspect of power systems operation. This is inherent in the proposed structure, as compared to other methods which rely on projections, clipping, or penalties to enforce voltage constraints. The feasibility metrics are then calculated for the inequality constraints pertaining to the dependent variables, φ.

Mean violation: the average magnitude of violations in constraint set

$$\frac{1}{|h_k|} \sum_k \max\{0, h_k(\psi)\} \tag{3.22}$$

Max violation: the maximum magnitude of violations in constraint set

$$\max_k \{\max\{0, h_k(\psi)\}\} \tag{3.23}$$

Number of violations exceeding a threshold: the number of inequality constraints which are violated by more than a threshold value ϵ

$$\sum_k \mathbb{I}_{\max\{0, h_k(\psi)\} > \epsilon} \tag{3.24}$$

Optimization Routine Metrics

The optimization routine metrics evaluate the performance of the Gurobi optimization solver when solving the MILP. It must be noted that the proposed PhyR-based framework is an end-to-end learning to solve the reconfiguration problem, and does not require an external solver. However, Gurobi is used to generate the data for the supervised framework. Additionally, a set of experiments are conducted to evaluate the PhyR-based framework as a warm-start technique for traditional optimization solvers. This experiment is evaluated based on the below optimization routine metrics.

Solve time: time in seconds to solve the optimization problem, through the YALMIP interface

Node count: number of branch-and-cut nodes explored in the most recent optimization

Iteration count: number of simplex iterations performed during the most recent optimization. Note that the reconfiguration problem is a linear program (LP) when the binary variables are fixed. Thus for every node in the branch-and-cut algorithm, the resulting optimization problem is an LP, efficiently solved using the simplex method.

3.7 Performance of Physics-Informed Machine Learning

Extensive evaluation of the proposed physics-informed framework were carried out. The key results are presented below:

Result 1: Physics-informed methods enable higher prediction accuracy in both optimality and feasibility metrics;

Result 2: There is a clear trade-off between optimality versus feasibility;

Result 3: Supervised learning outperforms unsupervised learning in optimality metrics, but underperforms in feasibility metrics;

Result 4: Datasets used for algorithm development must be representative of real system load and generation characteristics;

Result 5: Model variance is a challenge in reconfiguration prediction, with significant spread in prediction accuracy of topology error across multiple predictors

Result 6: Smoothness of the output layer function impacts training and prediction performance, with SiPhyR outperforming other architectures

Result 7: Warm-start of MILP is challenging: warm-start with the neural prediction can improve worst-case performance, but additional tuning is needed to

further reduce solve time. Even then, there are no guarantees on optimizer performance.

The key results from the ablation studies are summarized here:

- Neural network width for BW-33 is 5 neurons per hidden layer, and for TPC-94 is 300 neurons per hidden layer. This captures the tradeoff between training speed and prediction performance on optimality and feasibility metrics
- InSi-based methods require lower learning rates
- Datasets with real load data (DS-2 and DS-5) require lower learning rates

The results and detailed discussions are presented below. We begin with the main results assessing performance of physics-informed learning on optimality and feasibility metrics for the two grids and across datasets (for the TPC-94 grid) (Result 1 thru 4). We then present discussions on model variance and output layer functions (Result 5 and 6), followed by a discussion on warm-start of MILPs (Result 7). Finally we present the results of the ablation studies, which were used to inform the architectures trained and tested. **Notably, it is based on Results 6 and 7 that SiPhyR is selected as the preferred method, lending itself to the Grid-SiPhyR name.**

For all experiments below, a committee machine approach was taken where 10 models were trained with independent weight initialization (following He initialization) for each parameter. The results presented are the average over the 10 predictors.

3.7.1 Neural Architecture Performance on Optimality and Feasibility Metrics

Based on the ablation studies (presented in the following sections), the following parameters were used:

BW-33 grid: 2 hidden layers with 5 neurons each. The training data set had 7008 data points. The validation and test data sets had 876 data points.

TPC-94 grid: 2 hidden layers with 300 neurons each. The training data set had

7200 data points, with 900 validation points. The test datasets had 8640 data points for each dataset (DS-1 thru DS-5). Different committee of networks were trained on DS-1, DS-2, and DS-5. The prediction performance is evaluated within the same dataset (ex. trained and tested on DS-1), and generalization to other datasets with (ex. trained on DS-1 and tested on DS-2).

The tables below present the simulation results evaluating the prediction performance on test data for the BW-33 and TPC-94 networks. Results 1 thru 4 are derived from these tables, and a discussion is presented below.

Result 1: Physics-informed methods enable higher prediction accuracy in both optimality and feasibility metrics

The benefit of the physics-informed approach is clearly demonstrated in these simulation results. On all metrics - both optimality and feasibility - the PhyR-based methods consistently outperform the InSi methods, in both supervised and unsupervised frameworks (Table 3.11), and in generalizing to other datasets (Tables 3.12 thru 3.14). On the BW-33 grid, the PhyR-based methods outperform InSi methods by 12% on dispatch error, 51% on voltage error, 61% on mean inequality violation, and on the number of inequalities violated (0.8% versus 2.6%). To translate the feasibility metrics into units of power, the 61% lower inequality error corresponds to an error of 2.2kVA versus 5.7kVA. For a 10kW solar unit, the PhyR-based methods may request 12.2kW of power while the InSi based methods may request 15.7kW of power. While this may seem negligible, this violation would significantly impact dispatch decisions across multiple generating units, and may result in an imbalance between load and supply – a severe condition in the electricity grid. The results are similar for the TPC-94 grid where dispatch errors are a 1000 times lower, voltage errors are 10 times lower, and mean inequality violation is 10 times lower, for the PhyR-based methods. The inequality violation corresponds to 24kVA for PhyR-base methods, and 180kVA for InSi methods. Overall the topology error is comparable across all methods.

Result 2: There is a clear trade-off between optimality versus feasibility

The challenging nature of the reconfiguration problem is the simultaneous optimization of grid topology and DER dispatch. These two decisions are tightly coupled in that the physics of power flow must be satisfied on a grid topology. A change in topology then changes the constraints which must be satisfied. In the Linear Dist-Flow model, the voltage constraints are the key constraints which must be satisfied, and are relaxed for the reconfiguration problem using big-M formulation. The feasible search space for the reconfiguration problem is highly nonlinear due to the integer constraints. This challenge in simultaneous optimization manifests in the neural framework as a trade-off between optimality and feasibility.

In Table 3.11 the trade-off can be generalized as follows:

- The PhyR methods perform best on feasibility metrics
- The InSi methods outperform the supervised methods on feasibility metrics, but worst on optimality metrics
- The supervised methods perform best on the optimality metrics, in particular the topology error, but worst on feasibility metrics

In Table 3.13 the trade-off is also clearly visible in the test results on DS2 and DS3. The predictors have higher optimality errors on DS2 but lower feasibility errors. When tested on DS3, the optimality errors are improved upon (including the topology error from 48.6% to 44.0% for PhyR-based methods), at the expense of feasibility metrics which increase from 22 violations to 65 violations (from 1.5% to 4.4% of inequality constraints which are violated). Similar behaviour is observed when comparing the performance of predictors across the datasets DS1 thru DS5 (further discussed in Result 4).

At this point it is necessary to distinguish between optimality and feasibility. The power system is a critical infrastructure and meeting electricity demand with available generation is a critical action. Feasibility is vastly more important than optimality: only when we have feasibility, can we begin to think about optimality. This clear and

indisputable prioritization gives a clear criterion for evaluating the performance of the different neural architectures. The physics-informed approaches which prioritize feasibility explicitly in their design are therefore the preferred architecture, further corroborated by their lower feasibility errors and higher prediction accuracy (decidedly in dispatch and voltage, similar in topology).

Result 3: Supervised learning outperforms unsupervised learning in optimality metrics, but underperforms in feasibility metrics

Table 3.11 presents simulation results for the BW-33 grid in both unsupervised and supervised learning frameworks. The results show the order of magnitude improvement in optimality metrics obtained when using a supervised framework wherein the neural network has full access to the optimal solutions during training, with loss functions eq.(3.15) and eq.(3.16). In the Supervised-x methods, the topology error is considerably lower for the supervised methods, however this comes at the expense of higher feasibility errors (6.8% versus 2.8% for InSi method, and 3.7% versus 1.1% for PhyR-based methods). To reduce the inequality error, a penalty is introduced, and the results are presented as Supervised-x-pen. With the inequality penalty, the inequality errors are further reduced, reaching comparable performance with the unsupervised frameworks. However this increases the topology error, *even though the optimal topology is being penalized in the supervised loss function.* Result 2 holds for both supervised and unsupervised frameworks.

Another observation is that the InSi method outperforms PhyR-based methods in topology error when using supervised loss functions. This is expected: when using a supervised framework the neural network has full access to optimal solutions which indirectly enforce integer solutions for the switch statuses. Supervised-InSi is then able to reduce both topology error and dispatch errors while moving towards integer solutions. Supervised-PhyR still performs better than PhyR because full information is available, but tries to enforce integer solutions throughout training. Integer solutions are then enforced prematurely without fully leveraging the available optimal data, and falling into the trap of Result 2. It should be noted that both

Supervised-PhyR and Supervised-PhyR-pen outperform their InSi counterparts on all other metrics. Thus Result 1 holds for both supervised and unsupervised frameworks.

Result 4: Datasets used for algorithm development must be representative of real system load and generation characteristics

Results in Table 3.12 and Table 3.13 show the performance of neural predictions when trained on DS1 and DS2 respectively. Error metrics are considerably lower when trained on realistic datasets (i.e. DS2) with representative load and generation data. Further these models are able to generalize better to unseen data when the true load characteristics may be different than in the data or may change over time, for different load profiles (DS3) and generation profiles (DS4). Comparing the error metrics for DS1 on DS3 (Table 3.12) with DS2 on DS3 (Table 3.13), the errors are significantly reduced when trained on DS2: 40% lower dispatch error, 95% reduction in voltage error, 87% reduction in mean inequality violation, and the number of inequality violations reduces from 8% to 4%. It should be noted that the topology errors increase slightly from the range of $40-45\%$ to $45-50\%$ when trained on DS1 vs DS2. However, this is a small change and is likely insignificant.

Results in Table 3.14 further considers the case where realistic load profiles are available but solar generation profiles are not. Comparing the results here with the previous tables it's clear that access to load data dominates the performance improvement, while access to generation data provides marginal improvement. Taking for example the predictors trained on DS1, DS2, and DS5, and tested on DS2, select error metrics are as follows: voltage errors of 27V, 14V, and 15V; mean inequality violation of 24 kVA, 3kVA, and 5kVA; and number of inequality violations of 7.3%, 1.5%, and 2.7%. A considerable error reduction is obtained simply by using representative load profiles. It should be noted that the TPC-94 grid is an 11.4 kV feeder; the voltage errors above then correspond to 0.23%, 0.12%, 0.13%.

Table 3.11: Performance results for the BW-33 grid. All networks were trained and tested on perturbed data. Tested on 876 data points.

Method	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	3.24e-2	2.30e-3	49.7%	1.53e-3	0.148	16.3	21.7
InSi2R	3.24e-2	2.30e-3	48.6%	1.02e-3	0.114	14.1	21.7
ClaPhyR	2.86e-2	1.12e-3	52.4%	3.80e-4	2.40e-2	4.42	16.4
SiPhyR	2.89e-2	1.69e-3	41.5%	4.79e-4	4.23e-2	5.72	17.9
InSiPhyR	3.18e-2	1.13e-3	44.6%	5.00e-4	4.88e-2	6.02	17.5
Supervised-InSi	8.61e-4	4.65e-4	14.2%	5.52e-3	0.817	37.4	42.9
Supervised-PhyR	1.51e-4	2.36e-4	29.9%	4.29e-3	0.888	19.6	43.5
Supervised-InSi-pen	1.00e-3	2.80e-3	25.8%	1.55e-3	0.173	15.7	38.3
Supervised-PhyR-pen	5.78e-4	1.35e-3	33.6%	5.49e-4	4.84e-2	7.26	36.0

Table 3.12: Results for the TPC-94 grid. All networks were trained on DS-1 (perturbed loads) and are **tested** on datasets DS-1, DS-2, DS-3, and DS-4. Tested on 8,640 data points.

Method (tested on DS-1, perturbed)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.41	3.31e-2	44.3%	3.21e-2	3.01	162	8.72
InSi2R	1.41	3.31e-2	43.6%	3.20e-2	2.98	161	8.72
PhyR	1.78e-2	1.52e-2	45.4%	8.11e-4	5.25e-2	40.5	8.46
ClaPhyR	1.15e-2	1.31e-2	44.1%	7.67e-4	4.63e-2	41.1	8.72
SiPhyR	1.12e-2	1.51e-2	45.4%	7.27e-4	4.25e-2	37.5	8.85
InSiPhyR	1.10e-2	1.39e-2	44.4%	7.87e-4	4.62e-2	42.5	8.42
Method (tested on DS-2, residential)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.03	1.47e-2	41.6%	3.22e-2	2.61	166	20.7
InSi2R	1.03	1.47e-2	40.9%	3.20e-2	2.61	165	20.7
PhyR	1.71e-2	4.80e-3	43.7%	4.32e-3	0.960	107	2.94
ClaPhyR	1.77e-2	3.55e-3	43.3%	4.43e-3	1.00	107	3.53
SiPhyR	1.72e-2	2.91e-3	46.6%	4.33e-3	0.959	106	3.09
InSiPhyR	1.71e-2	5.18e-3	44.0%	4.25e-3	0.963	103	2.96
Method (tested on DS-3, mixed)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.10	1.02e-2	42.8%	3.71e-2	3.09	167	29.0
InSi2R	1.10	1.02e-2	41.6%	3.69e-2	3.09	166	29.0
PhyR	1.63e-2	3.61e-2	45.0%	4.35e-3	0.960	117	4.54
ClaPhyR	1.99e-2	3.41e-3	44.4%	5.01e-3	0.988	122	5.72
SiPhyR	1.67e-2	2.16e-3	44.7%	4.47e-3	0.958	118	4.77
InSiPhyR	1.64e-2	3.87e-3	42.4%	4.31e-3	0.962	114	4.58
Method (tested on DS-4, solar error)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.03	1.31e-2	41.2%	3.22e-2	2.62	166	21.0
InSi2R	1.03	1.31e-2	40.5%	3.21e-2	2.62	165	21.0
PhyR	1.71e-2	4.06e-3	43.7%	4.33e-3	0.960	107	2.99
ClaPhyR	1.78e-2	3.01e-3	43.3%	4.42e-3	0.959	107	3.59
SiPhyR	1.73e-2	2.38e-3	46.5%	4.33e-3	0.959	106	3.14
InSiPhyR	1.71e-2	4.45e-3	44.0%	4.25e-3	0.963	103	3.01

Table 3.13: Results for the TPC-94 grid. All networks were trained on DS-2 (residential loads) and are **tested** on datasets DS-1, DS-2, DS-3, and DS-4. Tested on 8,640 data points.

Method (tested on DS-2, residential)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.12e-2	3.98e-3	47.8%	2.29e-3	0.324	73.1	3.82
InSi2R	1.12e-2	3.98e-3	45.7%	1.94e-3	0.148	71.2	3.82
ClaPhyR	1.00e-2	2.38e-3	48.6%	5.07e-4	5.57e-3	21.6	2.91
Method (tested on DS-1, perturbed)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	0.449	1.80e-2	47.5%	3.80e-2	2.95	192	5.42
InSi2R	0.449	1.80e-2	45.5%	3.77e-2	2.95	190	5.42
ClaPhyR	0.148	2.35e-2	46.2%	2.35e-2	2.36	168	4.80
Method (tested on DS-3, mixed)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	2.98e-2	2.22e-3	46.6%	5.58e-3	0.488	110	5.91
InSi2R	2.98e-2	2.22e-3	44.3%	5.36e-3	0.358	108	5.91
ClaPhyR	9.987e-3	1.91e-3	44.0%	1.40e-3	0.125	64.7	4.28
Method (tested on DS-4, solar error)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.14e-2	3.27e-3	47.5%	2.37e-3	0.327	74.0	3.87
InSi2R	1.14e-2	3.27e-3	45.4%	2.02e-3	0.153	72.0	3.87
ClaPhyR	1.00e-2	1.96e-3	48.2%	5.09e-4	5.55e-2	21.9	2.95

Table 3.14: Results for the TPC-94 grid. All networks were trained on DS-5 (flat solar) and are **tested** on datasets DS-2, DS-3, and DS-4. Tested on 8,640 data points.

Method (tested on DS-2, residential)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.05e-2	3.86e-3	47.8%	1.80e-3	0.314	66.2	3.43
InSi2R	1.05e-2	3.86e-3	44.8%	1.45e-3	0.120	64.3	3.43
ClaPhyR	9.18e-3	2.71e-3	47.9%	8.92e-4	9.26e-2	39.1	2.83
Method (tested on DS-3, mixed)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.29e-2	2.22e-3	46.5%	3.33e-3	0.340	105	5.18
InSi2R	1.29e-2	2.22e-3	44.3%	3.04e-3	0.230	104	5.18
ClaPhyR	9.94e-3	2.08e-3	46.3%	1.78e-3	0.133	79.2	4.33
Method (tested on DS-4, solar error)	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	1.05e-2	3.10e-3	47.8%	1.82e-3	0.314	67.1	3.48
InSi2R	1.05e-2	3.10e-3	44.8%	1.47e-3	0.121	65.2	3.48
ClaPhyR	9.16e-3	2.18e-3	47.9%	9.07e-4	9.37e-2	39.7	2.87

3.7.2 Investigation of Committee Machine Ensemble Method and Output Layer Functions

This section investigates the model variance experienced in the committee of networks and the impact of different output layer functions. Results 5 and 6 are derived from this investigation, and a discussion is presented below.

Result 5: Model variance is a challenge in reconfiguration prediction, with significant spread in prediction accuracy of topology error across multiple predictors

The results in the prior section (and the following set of ablation studies) all consider a committee of networks using a simple averaging ensemble method, wherein the predictions are averaged to produce the final result. This approach was taken to reduce the impact of stochastic initialization and learning. In this section, the model variance across trained predictors is investigated. Figure 3-16 plots the distribution of topology error across the test dataset (of 876 data points) for each trained model, 1 thru 10, for each method trained on the BW-33 grid. The boxplot diagrams show the minimum, first quartile, median (central red line), third quartile, and maximum values of topology error. Any outliers are marked with a red '+' marker. Note that in some of the plots, the median may coincide with the minimum or the maximum. Each of these plots show the high variance in topology error across predictors. In particular, predictor 6 for PhyR, and predictor 1 for SiPhyR indicate significantly better performance and a lower spread than other predictors. The variance across the methods is highest for the InSi approach: since the method does not enforce integrality, the search space upon which the neural network is training appears continuous. Comparatively, the PhyR methods have very structured topology error values, which is expected by definition of the topology error metric.

The predictor with the best performance on topology error is selected for each method, indicated by the grey box around the model. The results are presented in Table 3.15, and show significant improvement in prediction accuracy across all

metrics, and in particular for the topology error. The topology error for PhyR reduces from 41.5% to 13.7% on average, with a maximum error of 25.0%. As in the previous discussion, the PhyR-based methods outperform the InSi methods, particularly for the inequality metrics. While the performance of InSi2R in approaches that of the PhyR-based methods with a topology error of 20%, the PhyR-based methods also have a lower maximum error.

Result 6: Smoothness of the output layer function impacts training and prediction performance, with SiPhyR outperforming other architectures

The performance for the different PhyR-based methods can be compared. The plots in Figure 3-16 clearly show that the ClaPhyR and InSiPhyR methods have more difficulty improving upon the topology error. Table 3.15 shows the best performing predictor for these two methods underperforms against SiPhyR, with ClaPhyR having the worst performance of the group. It should be noted that the dispatch and voltage errors are marginally higher for SiPhyR, which may be expected given Result 2 (tradeoff in optimality and feasibility).

Figure 3-17 plots the topology error during training for the TPC-94 network, when using different PhyR-based methods. Notably, the ClaPhyR and InSiPhyR methods show jumping behaviour that is more erratic than the SiPhyR architecture. This may be caused by the non-smooth nature of the output layer functions, which have elbow points at 0 and 1. In comparison, the SiPhyR method uses a smooth sigmoid which is differentiable at every point. Further, with the SiPhyR method the prediction from the neural network $\hat{z}_\tau$ can be more readily interpreted as a probability, since the sigmoid function is a cumulative distribution function of the logistic distribution. While the prediction performance averaged across all 10 predictors is comparable (see Table 3.11), the results presented in this ablation study show the SiPhyR method outperforms other approaches, where the continuous nature of the sigmoid function enables better prediction.

It is based on Results 6 and 7 that SiPhyR is selected as the preferred method, lending itself to the Grid-SiPhyR name.

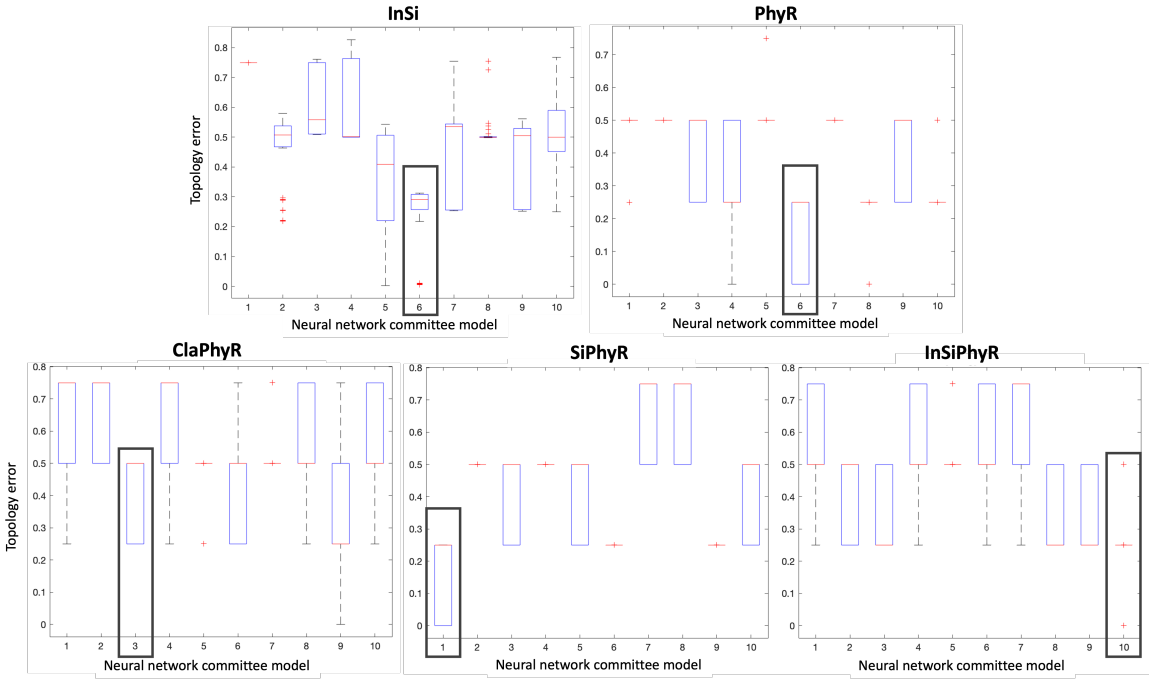


Figure 3-16: Boxplot showing high model variance phenomenon across multiple trained predictors on the topology error, on test data for BW-33. The mean topology error is indicated in red, with the quartiles indicated by the upper and lower box margins. Each predictor has the same neural architecture and is trained on the same dataset. The predictor with the best performance of the committee is outlined in a grey box, and is used to evaluate performance on the test set. Note that the PhyR-based approaches have discrete values for the topology error, leading to the fixed structure visible across the different models.

Table 3.15: Performance results for the BW-33 grid, for a single predictor (best of the committee). All networks were trained and tested on perturbed data. Tested on 876 data points.

Method	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
InSi	3.44e-2	4.01e-3	23.2% (max 103%)	2.22e-3	0.190	18.3	24.1
InSi2R	3.44e-2	4.01e-3	19.9% (max 79.5%)	1.34e-3	0.147	15.3	24.1
ClaPhyR	3.29e-2	1.21e-3	42.1% (max 75.0%)	3.06e-4	2.06e-2	2.94	14.2
SiPhyR	8.92e-3	3.06e-3	13.7% (max 25.0%)	3.56e-4	2.56e-2	3.94	18.7
InSiPhyR	3.72e-2	6.68e-4	21.6% (max 50.0%)	5.06e-4	5.67e-4	5.82	18.9

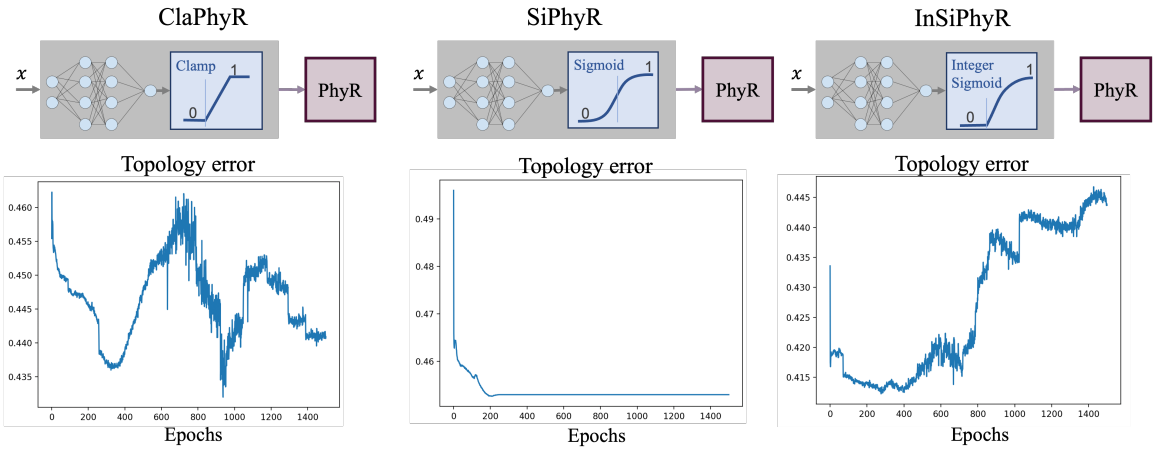


Figure 3-17: Plot of topology error during training on the TPC-94 network with $N_{nn} = 300$, on DS-1 with 9000 dataset size, and $lr = 0.001$.

3.7.3 Warm-start Performance on Optimization Routine Metrics

Prior work has looked at developing machine learning models to predict warm-start points for traditional optimization solvers. This section investigates the ability for the PhyR-based model to act as a warm-start predictor. Before presenting the warm-start experiment details and result, it is necessary to note that warm-start points are *generally* effective techniques to reduce optimization time. However, the benefits seen for general linear and nonlinear optimization are not replicated for mixed integer optimization problems. The techniques employed to solve the class of mixed integer problems, of which grid reconfiguration is a member, rely on multiple heuristics such as branch and bound, cutting planes, node presolve, and symmetry detection (among others). The selection of which techniques are used, the order in which they are used, and the techniques themselves are stochastic in nature. In addition, the optimization solver still needs to prove optimality of any solution (up to a tolerance level), which is itself a difficult task. For these reasons, there are no guarantees that providing an initial solution to an MIP solver will reduce computational time or effort.

Experiment Setup

All simulations of the reconfiguration MILP were carried out with Gurobi, using the Yalmip interface. Different optimization solvers have different requirements for warm-start points, specifying whether the point must be feasible or not, and whether a full set of variables must be provided or not. For Gurobi MIPs the initial point must be feasible, but does not need to be complete for all variables. Grid-SiPhyR guarantees feasibility of any prediction for the equality constraints and a subset of the inequality constraints. Translating the Grid-SiPhyR prediction to a warm-start point can be done quite easily by simply omitting any variables which violate the inequality constraints.

Three sets of simulations are carried. Case 1: the reconfiguration problem is solved without any warm-start. Case 2: the reconfiguration problem is solved by providing the optimal solution (from Gurobi) as the warm-start point to Gurobi. This provides a baseline comparison for the performance of warm-start on the reconfiguration problem. Case 3: the reconfiguration problem is solved by providing the Grid-SiPhyR prediction as the warm-start point to Gurobi. The optimization routine metrics are measured for each of these simulations. If the warm-start approach is effective for the reconfiguration MILP, it's expected that case 2 will significantly outperform both cases 1 and 3, and preferably case 3 will outperform case 1.

The simulations in case 2 were used to determine the set of variables to provide as a warm-start, denoted with the subscript '0'. When all variables $X_0 = \left[P^G\ Q^G\ V\ P_{ij}\ Q_{ij}\ y_{ij}\ z_{ij}\ z_{ji}\right]$ are provided as a warm-start point, the optimization solver is unable to identify it as a feasible solution. By testing different variables, the following were selected to provide as a warm-start: $X_0 = \left[P^G\ Q^G\ V\ y_{ij}\ z_{ij}\ z_{ji}\right]$. Any variables that violated inequality constraints were excluded.

Result 7: Warm-start of MILP is challenging: warm-start with the neural prediction can improve worst-case performance, but additional tuning is needed to further reduce solve time. Even then, there are no guarantees on optimizer performance.

Tables 3.16 and 3.17 present the results on the BW-33 and TPC-94 grids respectively. The key result from these results is that warm-start for the reconfiguration MIP is an ineffective technique, and may even result in poorer optimizer performance (along all three metrics). Interestingly, on average case 2 slightly increases solve time for the BW-33 grid while decreasing the node and iteration count, but is otherwise comparable. This difference across the metrics may be due to the size of the optimization problem and the relative time taken to reconstruct the full variable space and certify feasibility. These steps are not explicitly measured by the solver, but it has been noted by Gurobi developers that these actions can take considerable time. For the TPC-94 grid, case 2 improved optimizer performance on all metrics, by a larger margin than the BW-33 grid. Again, this may be due to the relative time taken in showing feasibility and optimality, versus searching the feasible space. Further, the search space for the TPC-94 grid is significantly larger than the BW-33 grid – in topology alone, there are 2^{14} possible unique topologies for TPC-94 as compared to 2^7 for BW-33. So although the optimality errors in neural prediction are comparable for the two grids (see Tables 3.11 to 3.14), the warm-start in case 2 for the TPC-94 grid is more meaningful by starting closer to the optimal solution.

The performance of warm-start with case 3 is quite inconclusive. For the BW-33 grid, case 3 outperforms both case 1 and 2 on both mean and median solve times, which is unexpected. For the TPC-94 grid, case 3 underperforms against case 2 (expected) and is comparable to case 1. Notably the worst-case solve time in case 3 is significantly reduced over both cases 1 (desired) and 2 (unexpected). Figure 3-18 plots the warm-start metrics for all three cases, evaluated on the TPC-94 grid. The desired shape in the plot is an asymmetrical 'V', where the left branch is longer than the right. This plot gives more insight into the performance of the different warm-

Table 3.16: Warm-start results for the BW-33 grid

Metric	Case 1 Without warm-start	Case 2 Warm-start at optimal	Case 3 Warm-start at Grid-SiPhyR prediction
Solve time	Mean: 0.2015 Median: 0.1586 Max: 0.7534	Mean: 0.2057 Median: 0.1661 Max: 0.8775	Mean: 0.1753 Median: 0.1371 Max: 1.0414
Node count	Mean: 2.2094 Median: 1	Mean: 1.8233 Median: 1	Mean: 2.2215 Median: 1
Iteration count	Mean: 2781 Median: 2629	Mean: 2770 Median: 2656	Mean: 2796 Median: 2663

Table 3.17: Warm-start results for the TPC-94 grid

Metric	Case 1 Without warm-start	Case 2 Warm-start at optimal	Case 3 Warm-start at Grid-SiPhyR prediction
Solve time	Mean: 1.4694 Median: 1.3322 Max: 6.6272	Mean: 0.4516 Median: 0.3993 Max: 8.0498	Mean: 1.4633 Median: 1.4157 Max: 3.4846
Node count	Mean: 182 Median: 137	Mean: 94.95 Median: 51	Mean: 182.4 Median: 140
Iteration count	Mean: 11405 Median: 10260	Mean: 5290 Median: 3605	Mean: 11453 Median: 10352

start methods. Looking at the solve time, the desired shape is emerging, with case 2 outperforming the others and forming the vertex of the 'V'. The shorter right side shows that while case 3 does not offer a reduction in solve time on average, it does permit a smaller in spread in solve times. This can be meaningful in applications where decisions must be made within a sensitive time window that is violated when using case 1, such as dynamic reconfiguration or electricity market clearing. Similar results are also seen for the other two metrics. It is possible that improving the neural prediction performance will further improve the warm-start performance of case 3, with a lower bound in solve time provided by case 2.

3.7.4 Parameter Studies: Neural Network Width

Parameter studies were carried out both the BW-33 and TPC-94 grids to determine neural network size and depth.

BW-33 grid, Unsupervised

Figure 3-19 shows the validation results during neural training of the PhyR-based methods with one hidden layer across different depths N_{nn}, and Fig. 3-20 shows the

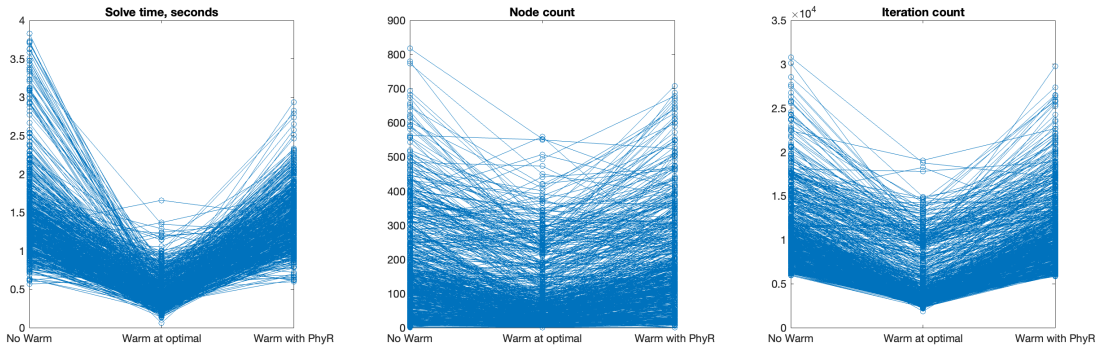


Figure 3-18: Warm-start results on the TPC-94 grid, plotting the optimization routine metrics. The warm-start experiment included 8640 data points, of which results for 700 randomly selected data points are plotted here.

same for a two hidden layer network. The deeper neural network (two layers vs. one) shows better performance on all metrics, reaching lower errors for dispatch and voltage, and inequality constraint errors at a faster rate. Based on the universal approximation theorem it is known that a neural network with a single hidden layer can represent any linear function, and one with two hidden layers can represent any arbitrary nonlinear function. While the underlying power flow constraints are linear (the reconfiguration problem is using the Linear DistFlow model), the presence of binary variables introduces a nonlinearity to the overall problem. The network with 2 hidden layers is expected to outperform those with a single hidden layer, as seen in the figures. Notably, by having 2 hidden layers the performance improvement is more substantial for the the narrow networks with $N_{nn} = \{5, 25\}$ neurons per layer.

Figure 3-21 shows similar results for the InSi method with 2 hidden layers. It must be noted that the InSi method requires a lower learning rate than the PhyR-based methods (using $lr = 0.0001$ for InSi vs. $lr = 0.001$ for PhyR). Overall the training plots for both PhyR and InSi methods are comparable. The faster convergence in validation loss of the PhyR method may be attributed to the higher learning rate. The jumping behaviour in the topology error plot of the PhyR method as compared to the smoother plot of the InSi method can be explained by the nature of the physics-informed rounding: the PhyR method enforces integer values for switch status predictions, so a change in the topology prediction may result in step-like behaviour

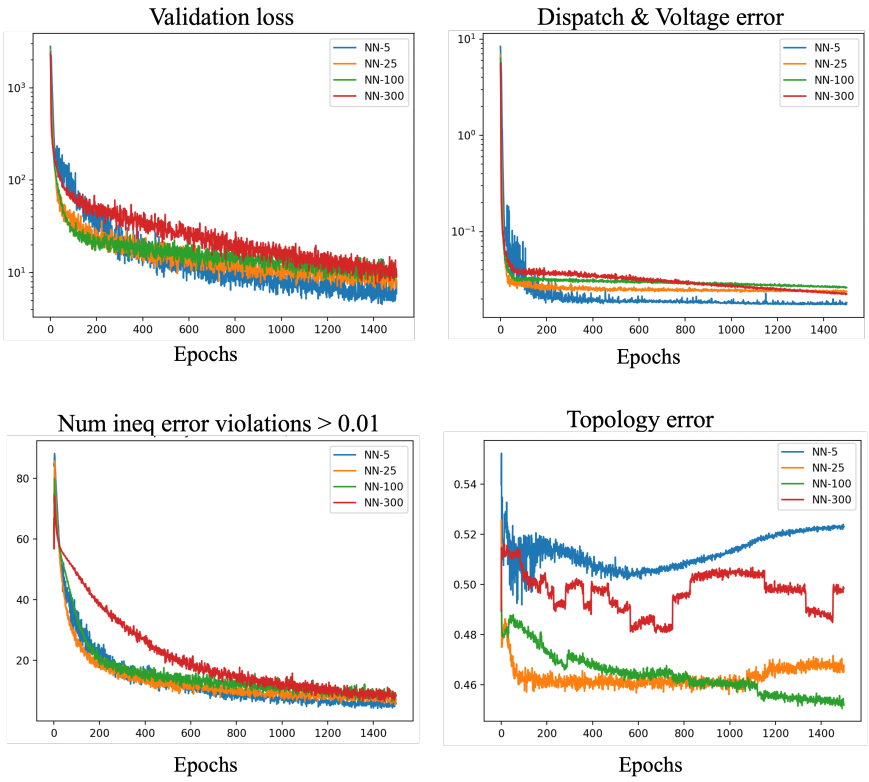


Figure 3-19: Ablation study on BW-33 grid for 1 hidden layers, across different neural network widths for the PhyR method. The learning rate is set to be $lr = 0.001$.

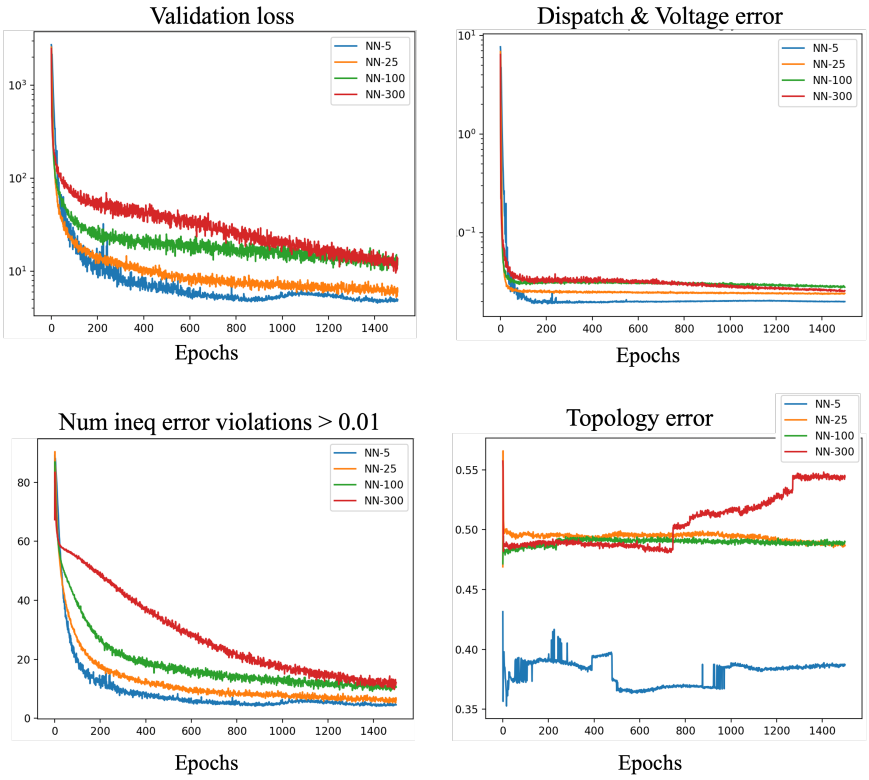


Figure 3-20: Ablation study on BW-33 grid for 2 hidden layers, across different neural network widths for the PhyR method. The learning rate is set to be $lr = 0.001$.

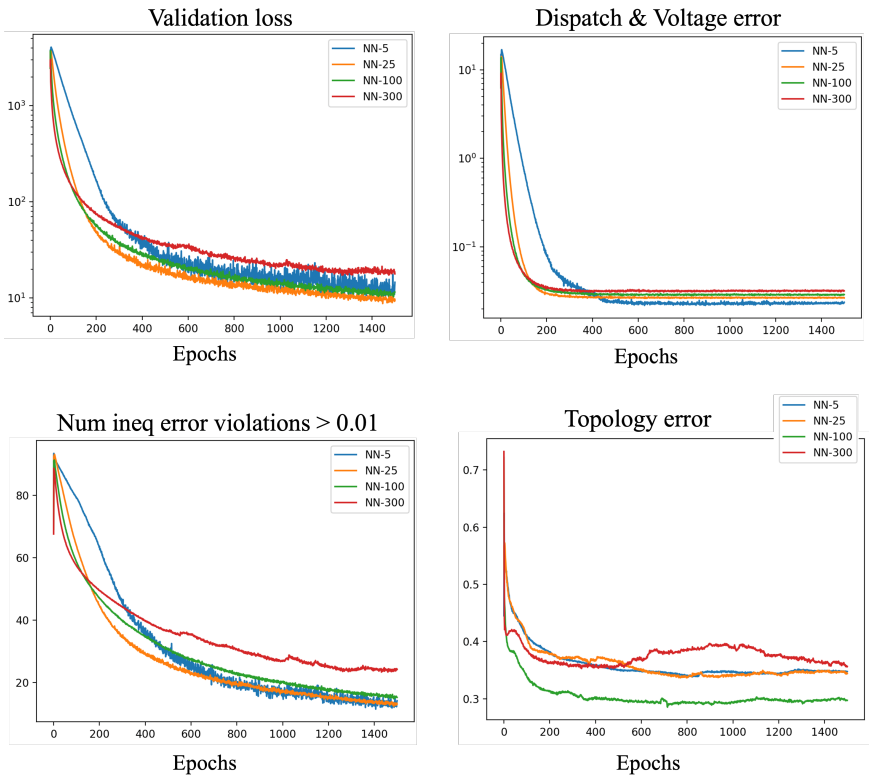


Figure 3-21: Ablation study on BW-33 grid for 2 hidden layers, across different neural network widths for the InSi method. The learning rate is set to be $lr = 0.0001$.

of the corresponding error metric. In comparison, the InSi method does not enforce integer solutions, so the corresponding metric is not expected to have a step-like behaviour. Overall the prediction performance with $N_{nn} = 5$ is reasonable across all three plots, providing a tradeoff between fast convergence, low error, and having a light-weight small network. This width is selected for the BW-33 network for all further testing.

BW-33 grid, Supervised

Figures 3-22 and 3-23 show the training plots for the Supervised-PhyR-pen and Supervised-InSi-pen architectures respectively. The networks consist of 2 hidden layers with 5 neurons each, as determined by the ablation studies in the previous section. The results show similar behaviour for both the PhyR and InSi approaches. The main difference is regarding the topology error: The Supervised-PhyR-pen set-

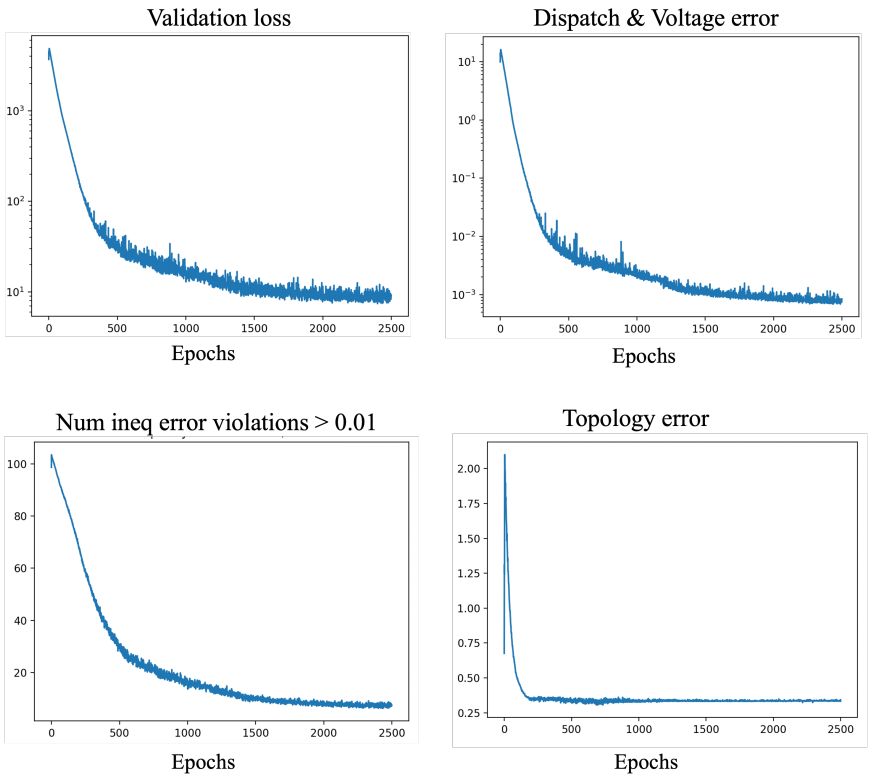


Figure 3-22: Results for Supervised-PhyR-pen on BW-33 grid for 2 hidden layers. The learning rate is set to be $lr = 0.0001$.

tles on a topology error within the first 250 epochs, making no further reductions in error as the training proceeds. All improvements are made with respect to the inequality metrics, and the other optimality metrics (dispatch and voltage error). The Supervised-InSi-pen has most topology error reductions within the first 500 epochs (reaching lower errors than the PhyR-based method), but has gradually increasing topology error after 500 epochs as the inequality and dispatch errors continue to improve. The tradeoff between optimal topology and feasibility (Result 2) is apparent here, and will be further discussed in the coming sections.

Figure 3-24 shows the training plots for the Supervised-PhyR architecture, where the loss function does not include a penalty for inequality constraint violation. The plots show the that dispatch and voltage errors can be reduced by an order of magnitude, as compared to the Supervised-PhyR-pen architecture, however the network performs more poorly on selecting the optimal topology. The dispatch errors from

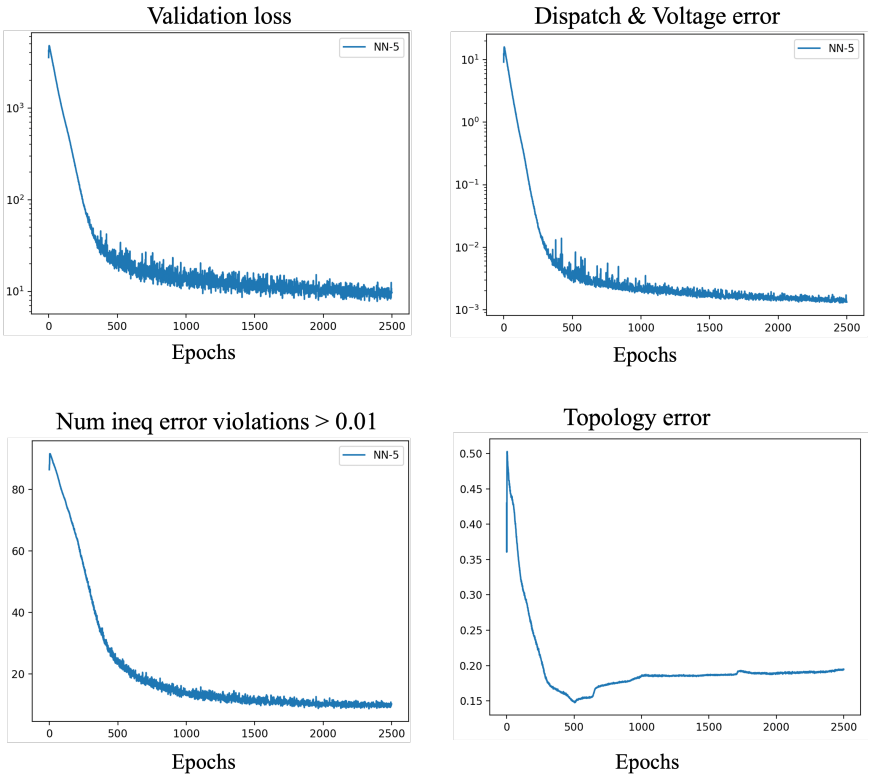


Figure 3-23: Results for Supervised-InSi-pen on BW-33 grid for 2 hidden layers. The learning rate is set to be $lr = 0.0001$.

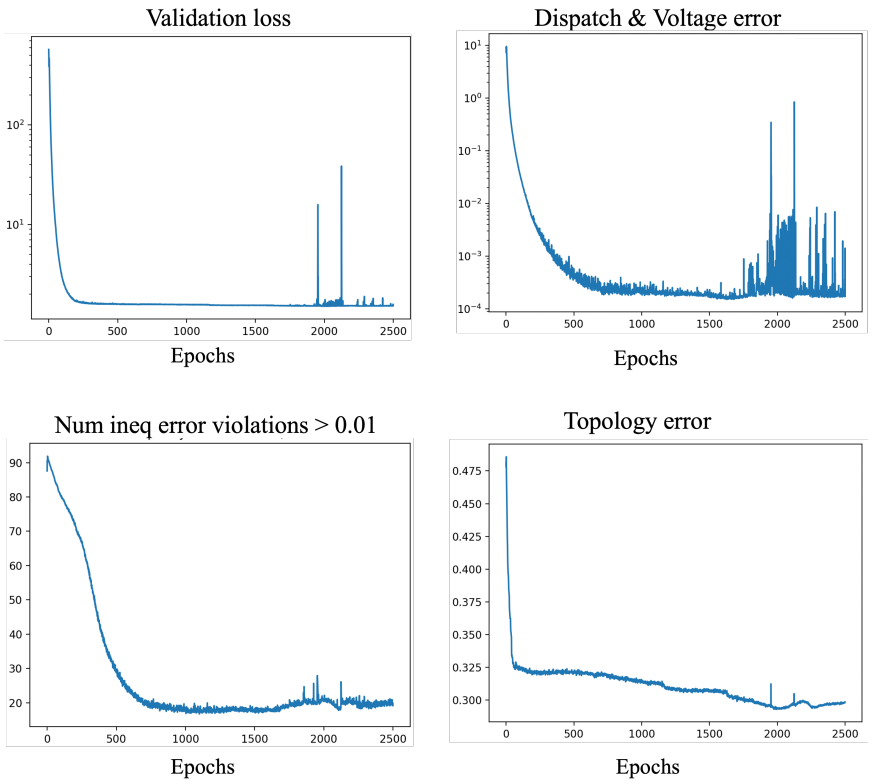


Figure 3-24: Results for Supervised-PhyR on BW-33 grid for 2 hidden layers. The learning rate is set to be $lr = 0.0005$.

epochs 1800 onward also show very noisy signal, despite the decreasing topology error. Notably the decrease in topology error around epoch 2000 is accompanied by an increase in the number of inequality violations, and corresponds to the highest noise in the dispatch and voltage errors. The binary nature of the topology decision variables makes training difficult, even when the neural network is given full information of the optimal solution. A lower learning rate may allow additional improvements, similar to those enjoyed by the Supervised-InSi architecture in Fig. 3-25 with learning rate $lr = 0.0001$. It is interesting to note that in Fig. 3-25 the topology error levels out around 10%, despite additional improvements to the other optimality metrics and inequality violations. The challenge of end-to-end learning to optimize for the combinatorial problem, even with only 7 switches, is clearly motivated by these results.

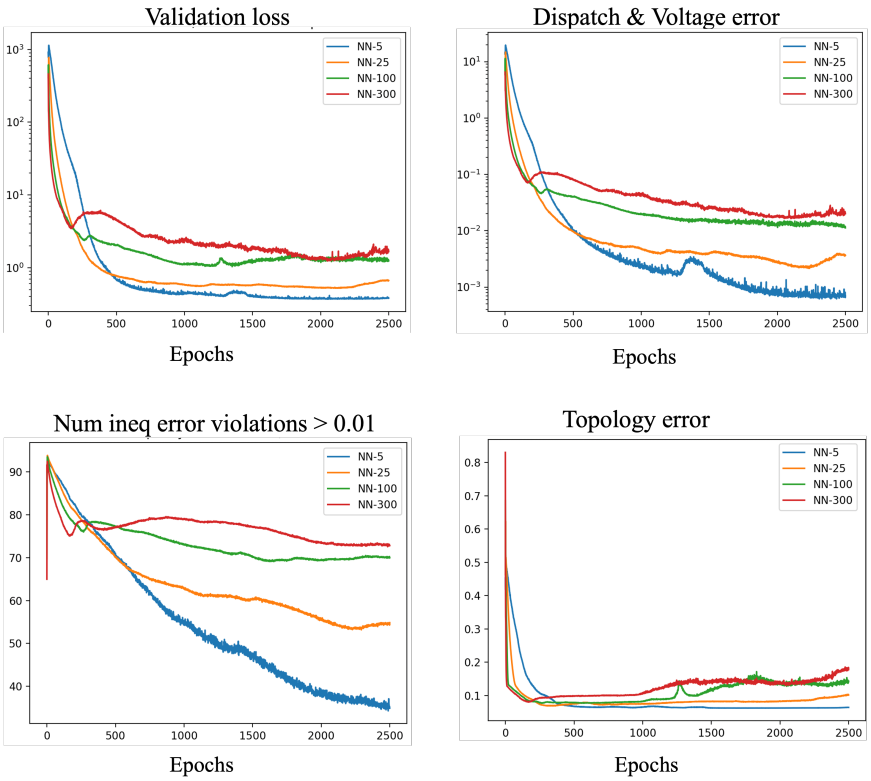


Figure 3-25: Results for Supervised-InSi on BW-33 grid for 2 hidden layers. The learning rate is set to be $lr = 0.0001$.

TPC-94 grid

Figure 3-26 shows the ablation study to determine the best N_{nn} for the PhyR method on TPC-94. The prediction performance of $N_{nn} = \{300, 700\}$ are best for the inequality violations and dispatch and voltage errors, while the largest network $N_{nn} = 1500$ performs the best on topology error. This wider network however requires a lower learning rate, and so converges more slowly on the other performance metrics. The tradeoff between convergence time and overall prediction performance is shown here. The same experiment was conducted for the InSi method, showing similar performance across different N_{nn}. Of note, the InSi method required a lower learning rate ($lr = 0.0005$ or $lr = 0.0001$). Overall the prediction performance with $N_{nn} = 300$ is reasonable, and this width is selected for all further testing. It should be noted that the InSi method requires a lower learning rate than the PhyR method. The profile datasets (DS-2 thru DS-5) also require lower learning rates than the perturbed dataset DS-1.

Figure 3-27 shows the performance of the InSi method during training, supporting Result 2 (tradeoff between optimality and feasibility). Around epoch 1750 the neural network prediction has moved towards a different topology that is closer to the optimal topology identified by Gurobi. However this drop in topology error is accompanied by an increase in the number of inequality violations and correspondingly, an increase in the loss function value. This phenomenon is easily explainable: the new neural prediction has updated the topology variables, but the corresponding update to the power variables is smaller. As a result, the power dispatch that had low feasibility error for the prior topology is less feasible for the new topology and has a resulting higher error. This behaviour is expected with the penalized loss function being used: it is difficult to improve upon the topology error when the loss function is dominated the by feasibility penalty which increases when the topology prediction is changed. The results can also be compared with the performance on the supervised methods – even with penalties on the topology error the neural network struggles to identify the optimal topology while learning optimal dispatch results. These two decisions are

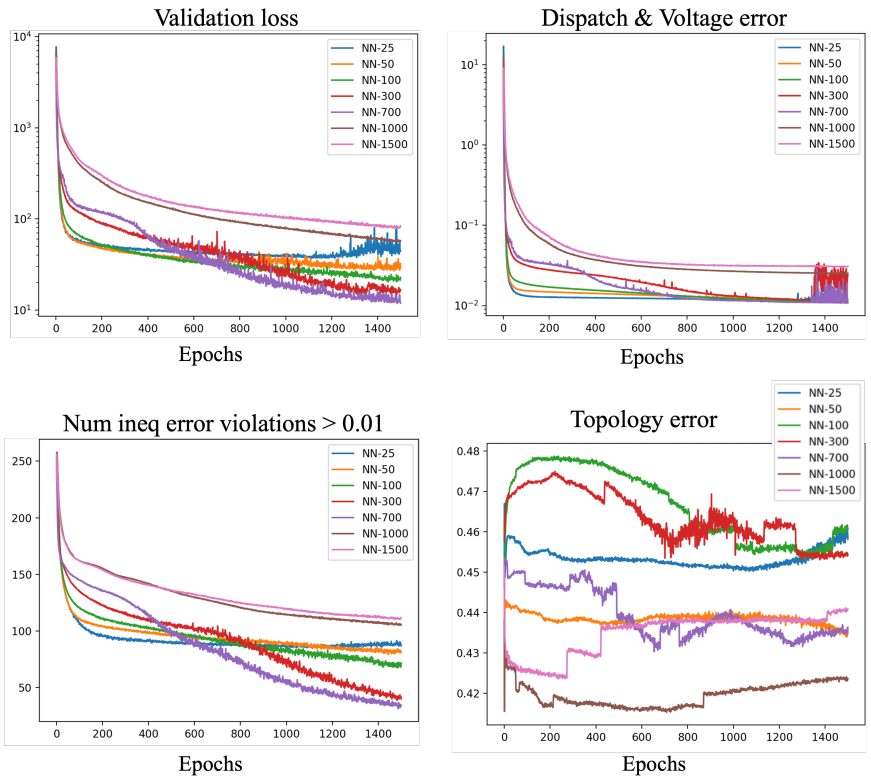


Figure 3-26: Ablation study using PhyR on TPC-94 grid across different neural network widths, on DS-1 with 9000 dataset size. The learning rate is set to be $lr = 0.001$ for all N_{nn} except $N_{nn} = \{1000, 1500\}$ for which $lr = 0.0001$.

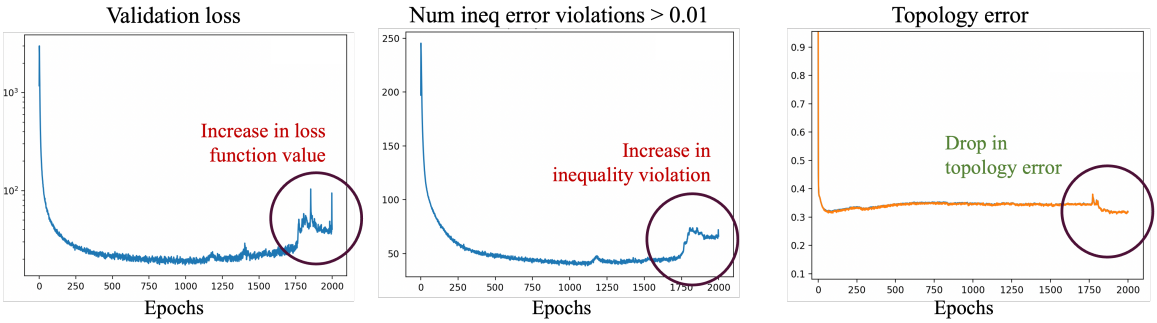


Figure 3-27: Plot of validation results of InSi method during training on the TPC-94 grid with $N_{nn} = 300$, on DS-2 with 9000 dataset size, and $lr = 0.0001$. This training plot shows the Result 2 (tradeoff between optimality and feasibility).

tightly coupled: a change in topology influences the power flow constraints which are immediately enforced in the proposed framework, and affects the inequality constraint bounds. The integer nature of the topology selection renders a nonlinear search space which the neural network is exposed to directly.

3.7.5 Parameter Studies: Number of Data Points

The following experiments were conducted to select the width of the neural network and determine the impact of dataset sizes, using the PhyR method. The results in Fig. 3-28 show validation prediction performance against optimality and feasibility metrics for the TPC-94 grid. Taking the committee machine approach a total of 120 models were trained (12 paramater sets, 10 models each). The results presented are the average over the 10 predictors. The learning rates were selected as follows: $lr = 0.001$ for $N_{nn} = 300$, and $lr = 0.0001$ for $N_{nn} = \{700, 1000\}$. Each model was trained for 2000 epochs, with batch size 200.

In each metric, the wider networks need larger datasets to reach comparable performance. This is easily visible from the purple highlighted area showing the upward and rightward trend of smaller errors. Immediately above this line the error is much higher, and further to the left the errors are smaller. This suggests a transition point in dataset size: When the dataset is too small the network overfits to the training data and results in good performance in the small validation dataset. When the dataset

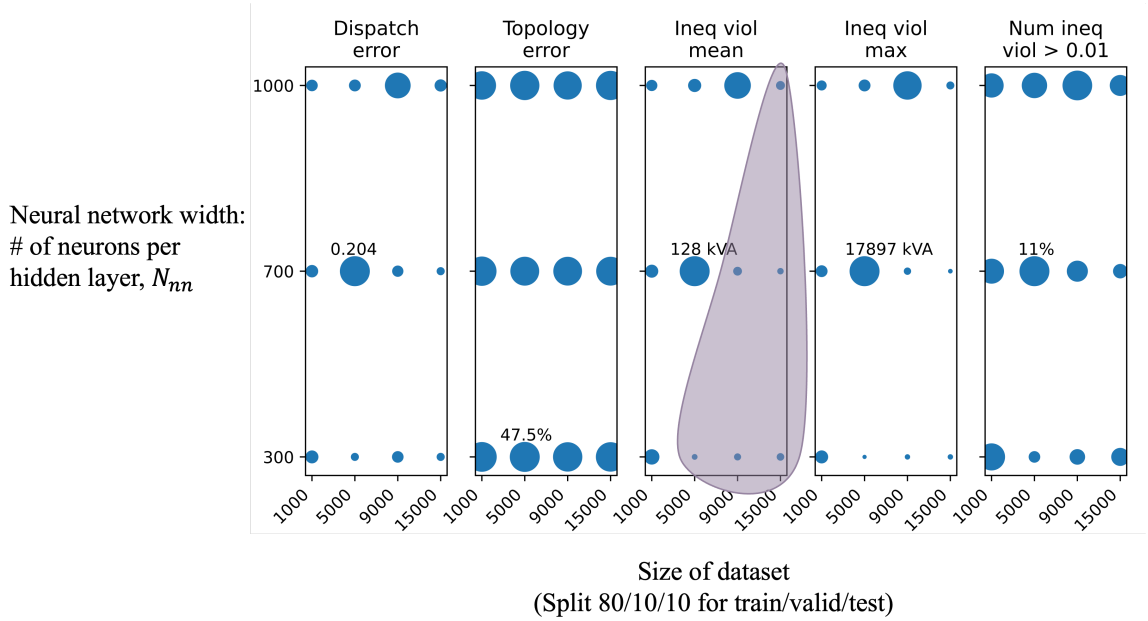


Figure 3-28: Ablation study on TPC-94 grid across different neural network widths $N_{nn} = \{300, 700, 1000\}$, and different dataset sizes $\{1000, 5000, 9000, 15000\}$. Smaller circles are better for each metric. The largest value is marked on the plot.

is large enough, the network learns across the dataset and is able to perform well on the validation data. However, at the transition point, the network does not have enough data to perform well on the validation set, and the validation set is too large to show good performance (as in the small size case). To put the training dataset size into perspective, 1000 data points corresponds to 3.5 days of load and generation data at 5-min intervals; 5000 to 17.4 days, 9000 to 31.2 days, and 15000 to 52 days. Notably, the topology error is about the same across all parameters tested, corroborating Result 2 (trade-off between optimality and feasibility). Overall the prediction performance with $N_{nn} = 300$ and dataset size 9000 is reasonable. These parameters are selected for all further testing.

3.7.6 Ablation Studies: Additional Architecture Designs

Additional architectures were tested to see if improvements in prediction accuracy could be easily achieved. The neural networks have two or five hidden layers, each with 5 neurons. Each layer applies a linear transformation with bias, batch normalization,

Table 3.18: Performance results for the BW-33 grid, with different network architectures. All networks were trained and tested on perturbed data. Tested on 876 data points.

Method	Metric						
	Dispatch error (MSE)	Voltage error (MSE)	Topology error	Ineq viol (mean)	Ineq viol (max)	Num ineq viol >0.01	Opt. gap (%)
SiPhyR-2D	2.22e-2	1.01e-3	56.6%	3.18e-4	2.29e-2	3.28	7.99
SiPhyR-5	3.61e-2	8.36e-3	52.7%	4.02e-4	2.57e-2	4.95	10.1
SiPhyR-5D	2.57e-2	3.65e-3	44.1%	3.24e-4	2.44e-2	3.46	7.90

ReLU activation (as before) and is tested with and without 30% dropout on each layer. The notation used describes these additional architectures is as follows: 'x-5D' is a 5 layer network with dropout using method x. Table 3.18 presents the prediction accuracy results. Overall these architectures provide minimal improvement in performance, and further tuning is not done.

3.8 Limitations and Extensions of the SiPhyR Framework

The proposed Grid-SiPhyR framework requires full knowledge of the underlying distribution grid including the topology and line parameters. This may be a restrictive requirement as many distribution utilities do not have full knowledge of the grid topology. Research effort has been made to develop topology and phase identification tools using sparse system measurements [138, 124, 21, 123] and can be used towards identifying system topology prior to training Grid-SiPhyR. Also, in the current implementation, Grid-SiPhyR cannot be used to generalize predictions to different grid topologies. The size of the input and output layers of the neural network and the variable space completion are limited to a single grid topology, which is used during both testing and training. While less common, the addition of new distribution lines may occur in some systems as DER penetration continues to increase. Further the generalizability would permit trained models to be transferred to other distribution systems, potentially reducing the amount of data required and training time. The current implementation also assumes the number and location of switches remains

the same. This is a reasonable assumption, but does potentially preclude using Grid-SiPhyR when a switch is not available (ex. switch is undergoing maintenance or must be on/off for other operational needs) or new switches are added to the network (ex. as part of a grid modernization project). Different approaches can be taken to address this limitation, such as training on augmented datasets with different switch conditions, or re-training the network on a new topology or set of switches (ex. using meta-learning or transfer learning approach). Employing a graph neural network (GNN) architecture in place of the lightweight neural network can also be investigated to enable generalizability across new grid topologies and switch locations [74].

More generally, the SiPhyR framework is restricted to problem structures of Eq. 3.2, where a function b can be used to explicitly and uniquely define a rounding function. However there are applications where a unique rounding/selection function may not exist. One such example is the unit commitment (UC) problem in electricity market dispatch which asks the following question: "Given a list of generators and their operating constraints and costs, what is the least cost dispatch solution which satisfies loads in the network?" The UC problem does not require a fixed number of generators to be on/off, and there are no comprehensive heuristics which can define an explicit function to select generators. However, it may be possible to define a family of rounding functions defined by the lower and upper limits on the number of generators which can be selected (based on generating costs, generating limits, ramp constraints, and load requested). Further design of such families of rounding or selection functions is an open question and may permit the extension of SiPhyR to a broader class of optimization problems.

The underlying concept of the SiPhyR framework is to explicitly embed discrete decisions into a neural framework. This presents a challenge to learning, since the implementations of PyTorch functions like round, max, and min all eliminate the gradient in the range where the function is applied. This prevents gradients from flowing through the backpropagation and makes learning challenging. The approach taken in Grid-SiPhyR is to only round $M_{sw} - 2$ switch statuses (with $L - 1$ being rounded up and the remaining rounded down), thereby retaining the gradient information for two

switches. A "leaky" approach can be taken wherein gradient information is retained for the backward pass of the neural training (similar to a leaky ReLU which solves the gradient saturation problem inherent in the ReLU operator, at the expense of sparsity) and may improve learning capabilities. More generally, the rounding function can be thought of as a projection operator. The further development and integration of differentiable projection operators into the SiPhyR framework can further improve the training performance and permit application of SiPhyR to other problems.

3.9 Concluding Remarks

This chapter presented Grid-SiPhyR, a physics-informed machine learning framework for end-to-end learning to optimize for combinatorial problems in the distribution grid. Grid-SiPhyR is applied to the problem of dynamic reconfiguration, an emerging paradigm for distribution grid control under high DER penetration. The proposed framework introduces a novel physics-informed rounding approach to tackle the mixed integer nature of the reconfiguration problem and satisfy salient operating constraints of the grid. The Grid-SiPhyR framework is tested on two canonical distribution grids across a range of operating conditions and data availability, including synthetic residential and commercial data and solar generation profiles. In each of these cases, Grid-SiPhyR provides order of magnitude improvement in both optimality and feasibility error metrics as compared to other approaches for integer accommodation which do not leverage grid physics, and is better able to identify the optimal grid topology and DER dispatch setpoints. Grid-SiPhyR uses an unsupervised machine learning approach, thus eliminating the need for expensive optimization computations to generate training data, which can be prohibitive for real-world systems.

The results in this chapter show that representative datasets (of load and generation profiles) have a considerable impact on prediction performance. Since real-world data of network topologies, utility models, and customer load data are not widely available due to data privacy concerns and system security, the creation of synthetic datasets with representative load profiles and DER penetration requires considerable

effort, but is a valuable and necessary task. The datasets generated as part of this thesis are a key contribution to the power systems community. The development and open-source sharing of such datasets is necessary to support the development, testing, and comparison of algorithms for power systems control and optimization.

The results in this chapter have shown that Grid-SiPhyR can be the enabling algorithmic technology for dynamic reconfiguration in distribution grids. Simulations on the two networks show improvements in three key metrics of grid efficiency, grid operability, and clean energy directives. Specifically, dynamic reconfiguration results in up to 23% reduction in line losses, significant improvement in voltage profile, and up to 17% reduction in PV curtailment, as compared to static reconfiguration. This latter metric corresponds to a reduction of 107 metric tons of CO_2 per year. Thus Grid-SiPhyR is an enabling technology for essential grid operations in future decarbonized energy systems.

Chapter 4

Physics-Aware Distributed Coordination Architecture for Grid Services

4.1 Introduction

The current state-of-art for managing, operational, and planning activities in the transmission grid consist of an all-knowing central agent; this central agent collects data from all resources, forecasts loads and renewable generation, and makes dispatch decisions which are then communicated to the resources. This centralized approach has been successful because the participating generators are a smaller number of large dispatchable generating units. However, this approach becomes challenging when integrating a large number of small-capacity DERs, as in distribution grids. Further, the lack of visibility of DERs in the distribution grid means it is harder to forecast net loads and distributed generation, with the increased presence of rooftop solar. These challenges can be addressed by using fast-acting generators or utility-scale storage which provide the requisite flexibility in load following; however, these options are not economically feasible. Instead we look towards flexibility from DERs to provide grid services at fast timescales and low cost.

DERs in the distribution grid are typically equipped with smart inverters, local intelligence, and computational abilities. Many are also integrated into home energy management systems that can optimize local storage devices, dynamic charging of electric vehicles, and control of smart appliances (like smart thermostats) subject to consumer constraints and preferences. The increased presence of actuators with smart sensing, light-weight computing capabilities distributed throughout the grid, and communication protocols enabled by Internet of Things (IoT), all of which are occurring rapidly in the distribution grid, provides great opportunities for enabling grid services. Traditional methods, which were purely reactive and responsive, can now be replaced with increasingly 'smart' strategies which are predictive, prescriptive, and increasingly autonomous. In general, this enables 'smart' grids of the future to adopt widespread automation for data acquisition, grid operations, and decision making.

The coordination of such a large number of these spatially distribution agents can render centralized optimization intractable, especially for online and real-time applications. An attractive alternative to centralized perspectives is a distributed one. Herein, each agent has access to local information and a limited amount of information shared by its neighbours, which it uses to optimize its actions towards meeting local and system-level goals. In the distributed paradigm, DERs can be coordinated to meet system-level goals without the aid of a central authority. Recent research efforts look towards decentralized and distributed approaches for optimization and decision making, which are enabled by the increased presence of grid-edge intelligence, computing resources, and peer-to-peer (P2P) communication networks (see [101, 61, 140, 148] for reviews). Where necessary, centralized control and decision making can also be augmented with distributed schemes to provide more visibility into the grid, reduce communication requirements to the central agent, and parallelize computational efforts. These paradigms of distributed decision making and augmenting centralized methods (commonly used in today's distribution grid management) are explored in this thesis.

4.2 Physics-Aware Distributed Coordination Architecture

To enable services from DERs situated throughout the distribution grid requires a physics-aware distributed coordination architecture, as shown in Fig. 4-1. This architecture is composed of three layers:

- **Electric grid layer**: modeling the underlying distribution grid enables physics-aware decision making. In doing so, the optimal actuation decisions will account for grid constraints (such as voltage limits or line thermal limits) and can co-optimize for system-level objectives.

- **Device layer**: distributed devices can provide much needed operational flexibility. However, the vast majority of these devices are privately owned. This layer encodes the locational information of devices (with respect to the underlying electric grid), the device-level constraints, and user preferences. In doing so, the optimal actuation decisions will also account for device and user constraints (such as charging times for electric vehicles or temperature preferences for smart thermostats) and can co-optimize for individual objectives.

- **Communication layer**: peer-to-peer communication enables coordination decision making and the co-optimization of local and system-level objectives subject to both local and system-level constraints. By carefully designing the computational and communication architectures (ex. which agents must communicate with others, the specific data to be shared, frequency of data sharing) the individual decision making agents can work together towards achieving a common goal, while retaining individual decision making autonomy and control. When appropriately designed, these coordinated decision mechanisms can also have privacy preserving characteristics to protect sensitive customer or device data.

This physics-aware distributed coordination architecture provides the substrate upon which distribution grid automation projects can be developed to enable DER

grid services. In the proposed grid services that follow, the individual decision agents located at each node of the electric grid will arrive at the optimal real and reactive power injections corresponding to (i) the optimized voltage profile over the grid; or (ii) the optimized net load at the T-D substation. Through iterative computation and communication, system-level objectives can be met while satisfying both device and grid constraints. Such a distributed approach leverages the computational and communication capabilities of intelligent grid edge devices. Further, if well designed, they can be resilient to communication link and single-point failures, and can preserve the private information of the DERs while still realizing network-level objectives.

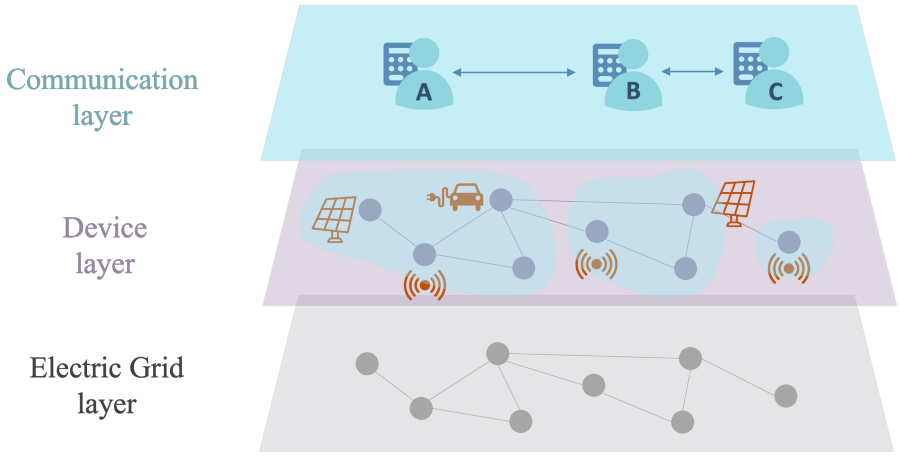


Figure 4-1: Proposed physics-aware distributed coordination architecture enables coordinated decision making across a heterogeneous group of devices owned by individual agents. The architecture enables optimization over system-wide and local objectives subject to system-wide and local constraints.

The following sections present details of the physics-aware distributed coordination architecture. First, a discussion on the electric grid layer where a power system model of the underlying unbalanced distribution grid is presented, denoted the Current Injection (CI) Model. The CI model is a linear model which leverages convex relaxations to make computationally tractable the unbalanced AC power flow. Second, the DER models within the CI-OPF problem are presented. Third, a PAC-based distributed CI-OPF algorithm is presented to coordinate DER actions through iterative communication and computation steps.

4.2.1 Electric Grid Layer: Current Injection Model

Two distinct types of models have been used to represent the physics of the distribution grid: modeling branch variables which leads to the Branch Flow model [83, 41], or modeling nodal variables which leads to the Bus Injection model [89, 31]. The Branch Flow model based on second order cone programming (SOCP) has proven to be advantageous in providing tight convex relaxations to the original AC-OPF problem with exactness under some conditions, and is shown to be more computationally stable than the Bus Injection model. However, both Branch Flow and Bus Injection models are typically limited to networks with radial topologies and balanced networks, and extensions to these models for unbalanced distribution grids are only valid for a small range of angle imbalances. This is a limiting assumption for many grids, especially with increasing penetration of DERs located on single-phase lines.

A recent model denoted as the Current Injection (CI) model [46, 48], avoids this assumption and so is an ideal candidate for representing unbalanced grids with various single-phase loads and generation. The CI model uses nodal variables, similar to the Bus Injection model, and represents all loads and generators as nodal current injections, with all power, current, and voltage phasors represented in Cartesian coordinates. The 3-phase impedance matrix is used to describe the self and mutual inductance between phases to model the coupling of phases that are common to a distribution grid. More importantly, the key obstacle of non-convexity of the AC-OPF and the subsequent nonlinearity of SOCP and SDP convexification strategies are dealt with in the CI approach by leveraging McCormick Envelope (MCE) based convex relaxation [97] for the bilinear power relations (VI terms). The MCE uses the convex hull representation of bilinear terms to render a linear OPF model. This representation requires adequate bounds on the nodal voltages and currents to ensure a tight convex relaxation. To determine these bounds, the CI approach also includes a carefully designed pre-processing algorithm which uses generation and load forecasts and grid limits to iteratively calculate tight nodal bounds [46]. Notably, the CI model permits modeling of multi-phase unbalanced and meshed networks, without

requiring expensive post-processing steps or restrictive assumptions on the reciprocal unbalance of each phase, both which are common in other power flow models. The CI model has been shown to perform well on unbalanced networks with local generation, with maximum 1.2% optimality gap and 0.9% voltage error when compared to the AC-OPF for a number of use cases [46]. Due to its overall ability to model unbalanced grids and all of the aforementioned advantages including the computational simplicity of the linear model, the CI approach is adopted for the voltage regulation problem. The model is presented in Appendix B for reference.

4.2.2 Device Layer: Modeling DERs

We model three types of DERs: distributed PV units (only generation), prosumers (load and generation co-located), and storage. The models are detailed below.

Distributed Generation

We assume all renewable generation can be curtailed, and that all units are equipped with smart inverters capable of adjusting power factor [1]. We do not consider storage in this work. The full model for a PV unit at node j is:

$$\underline{P}_j = 0, \quad \overline{P}_j(t) = \alpha_{\text{PV}}(t) P_j^{\text{Cap}}, \quad \underline{Q}_j = -\overline{Q}_j \tag{4.1a}$$

$$P_j \tan(\cos^{-1}(-\text{pf})) \leq Q_j \leq P_j \tan(\cos^{-1}(\text{pf})) \tag{4.1b}$$

Prosumers

Prosumers are nodes where both load and generation are present. To model each device properly, we must introduce additional variables representing the load and generation powers:

$$P_j = P_j^G - P_j^L, \quad P_j^G \geq 0, P_j^L \geq 0 \tag{4.2a}$$

$$Q_j = Q_j^G - Q_j^L, \quad Q_j^L \geq 0 \tag{4.2b}$$

[1] Rule 21 interconnection rules for DERs in CAISO requires that all distributed generators be equipped with smart inverters, as of 2014.

$$\overline{P}_j = \overline{P}_j^G - \underline{P}_j^L, \quad \underline{P}_j = \underline{P}_j^G - \overline{P}_j^L \tag{4.2c}$$

$$\overline{Q}_j = \overline{Q}_j^G - \underline{Q}_j^L, \quad \underline{Q}_j = \underline{Q}_j^G - \overline{Q}_j^L \tag{4.2d}$$

The PV unit located at a prosumer node j will be represented by the equations in (4.1), with all P_j and Q_j variables replaced by the P_j^G and Q_j^G. Similarly, loads located at prosumer node j will replace P_j and Q_j variables with P_j^L and Q_j^L but with $\underline{P}_j^L = -\overline{P}_j$ and $\overline{P}_j^L = -\underline{P}_j$.

Storage

The storage model introduces a dependency on control action in one period to previous periods. The battery storage devices are modelled using the power charge and discharge, $P_j^{\text{sc}}(t)$ and $P_j^{\text{sd}}(t)$ respectively for node j and time t. These are nonnegative variables. The state of charge, $b_j(t)$, is calculated as an integral constraint using the actions of the previous period and the initial state of charge, $b_j^0 := b_j(t=0)$. We model charge and discharge efficiencies (η_j^C and η_j^D), self-discharge rate (η_j^{self}), and impose a minimum state of charge ($\underline{b}_j$) to ensure battery health. All batteries are assumed to operate at unity power factor, i.e. $Q_j = 0 \quad \forall t$, and do not contribute to voltage regulation services.

$$P_j = \frac{1}{\eta_j^{\text{D}}} P_j^{\text{sd}}(t) - \eta_j^{\text{C}} P_j^{\text{sc}}(t) \tag{4.3a}$$

$$0 \le P_j^{\text{sd}} \le \overline{P}_j^{\text{sd}}, \quad 0 \le P_j^{\text{sc}} \le \overline{P}_j^{\text{sc}} \tag{4.3b}$$

$$b_j(t) = (1 - \eta_j^{\text{self}}) b_j(t-1) + \eta_j^{\text{C}} P_j^{\text{sc}}(t) - \frac{1}{\eta_j^{\text{D}}} P_j^{\text{sd}}(t) \tag{4.3c}$$

$$\underline{b}_j \le b_j(t) \le \overline{b}_j \tag{4.3d}$$

4.2.3 Communication Layer: PAC-based Distributed Optimization

The CI-OPF problem is solved using a distributed optimization algorithm, based on the Proximal Atomic Coordination (PAC) approach [122]. The network-wide opti-

mization problem can be decomposed into a set of $\mathcal{J}$ coupled optimization problems of the form (4.4), each solved by an independent computational agent. The j-th agent solves its optimization problem to determine the power injections of the DER at node j which minimizes the local cost function f_j, subject to the usual power flow and DER limit constraints. Note that this local cost function can be suitably designed for different grid services. To address the coupling of the $\mathcal{J}$ subproblems coming from the network-wide power flow constraints, each agent will generate an *estimate* of the downstream nodal voltages; i.e. for an agent j with downstream nodes $i, \{(j,i)\} \in \mathcal{E}$, the agent j will generate estimates of the voltage V_i as denoted by V_i^j. To enforce agreement on the value of V_i, a set of coordination constraints are introduced as in (4.4b). Through iterative local optimization and communication of V_i and dual variables ν pertaining to the coordination constraints, the distributed algorithm converges to the global optimal solution.

$$\left\{P_j^*, Q_j^*\right\} = \underset{\{P_j,Q_j\}}{\operatorname{argmin}} \; f_j(V_j, P_j, Q_j) \tag{4.4a}$$

$$\text{subj. to: } V_j = \mathcal{V}_j\left(V_{o,\tau}, P_j, Q_j, V_i^j\right)$$

$$\{P_j, Q_j\} \in \varphi_j$$

$$V_i^j - V_i = 0 \tag{4.4b}$$

The Lagrangian function for each nodal optimization is as

$$\mathcal{L}_j\left(a_j, \mu_j, \nu\right) = \left\{ \begin{array}{c} f_j(a_j) + \mu_j\left(V_j - \mathcal{V}_j(V_{o,\tau}, P_j, Q_j, V_i^j)\right) \\ +\nu^T\left(V_i^j - V_i\right) \end{array} \right\}$$

where $a_j = \left[P_j, Q_j, V_j, V_i^j\right]$, μ is the dual variable corresponding to the equality constraints describing the power flow equations, and ν is the dual variable corresponding to the coordination constraints. It can be shown that $\mathcal{L}\left(a, \mu, \nu\right) \triangleq \sum_{j \in K} \mathcal{L}_j\left(a_j, \mu_j, \nu\right)$.

The PAC-based distributed voltage optimization is presented in Algorithm 3, where $\rho > 0$ is the step-size and $\gamma > 0$ is the over-relaxation term. The termination criteria can be set as a maximum number of iterations $k \leq \mathcal{K}$ or a threshold for change in primal and/or dual variables, ex. $|a_j[k] - a_j[k-1]| \leq \epsilon_{\mathcal{K}}$. The convergence

properties of PAC are demonstrated in [122], with algorithmic extensions presented in [49, 47]. In Algorithm 3 the primal variables in a_j are initialized to worst-case conditions (such as the lower limit of all variables), but initialization can be improved by using previous optimal dispatch values or predictions for the optimal dispatch.

Algorithm 3: PAC-based distributed algorithm for DER optimization

Data: DER limits at each node j: φ_j
Result: DER injection at each node j; P_j^*, Q_j^*
Initialize primal variable a_j, $P_j \leftarrow \overline{P}$, $Q_j \leftarrow \overline{Q}$, $V_j \leftarrow 1$, $V_i^j \leftarrow 1$;
Initialize equality dual variable, $\mu_j \leftarrow \rho\gamma\left(V_j - \mathcal{V}_j(V_{o,\tau}, P_j, Q_j)\right)$;
Initialize coordination dual variable, $\nu_j \leftarrow \rho\gamma\left(V_i^j - V_i\right)$;
Dual prediction, $\hat{\mu}_j \leftarrow \mu_j + \rho\gamma\left(V_j - \mathcal{V}_j(V_{o,\tau}, P_j, Q_j)\right)$;
Dual prediction, $\hat{\nu}_j \leftarrow \nu_j + \rho\gamma\left(V_i^j - V_i\right)$;
$k \leftarrow 1$;
while *termination criteria* **do**
  $a_j \leftarrow \underset{\{P_j,Q_j\}\in\varphi_j, V_j, V_i^j}{\operatorname{argmin}} \left\{ \begin{array}{c} \mathcal{L}_j\left(a_j, \hat{\mu}_j\left[k-1\right], \hat{\nu}\left[k-1\right]\right) \\ +\frac{1}{2\rho}\left\|a_j - a_j[k-1]\right\|_2^2 \end{array} \right\}$;
  $\mu_j \leftarrow \mu_j\left[k-1\right] + \rho\gamma\left(V_j - \mathcal{V}_j(V_{o,\tau}, P_j, Q_j)\right)$;
  $\hat{\mu}_j \leftarrow \mu_j + \rho\gamma\left(V_j - \mathcal{V}_j(V_{o,\tau}, P_j, Q_j)\right)$;
  Communicate V_j with upstream neighbours;
  $\nu_j \leftarrow \nu_j\left[k-1\right] + \rho\gamma\left(V_i^j - V_i\right)$;
  $\hat{\nu}_j \leftarrow \nu_j + \rho\gamma\left(V_i^j - V_i\right)$;
  Communicate $\hat{\nu}_j$ with neighbours;
  $k \leftarrow k-1$
end
$P_j^* \leftarrow P_j[k]$, $Q_j^* \leftarrow Q_j[k]$;

4.2.4 Emerging Grid Services

Upon the physics-aware distributed coordination framework we develop two grid services which are essential to grid operations in a low-carbon future: voltage regulation and load ramp mitigation. The first service, voltage regulation, uses DERs to provide services locally within the distribution grid. The second, load ramp mitigation, uses DERs in aggregate to provide a service to the bulk transmission grid. The following sections detail the proposed algorithms.

4.3 Voltage Regulation: A Hierarchical Coordination Approach for a DER-Rich Grid Edge

Secure and reliable operation of the power system requires the system frequency and voltages to be within certain acceptable limits. The increasing penetration of renewable energy resources and their intermittent nature results in voltage fluctuations throughout the grid and is making this task increasingly challenging. Particularly, voltage management has become an emerging challenge in distribution grids with high solar photovoltaic (PV) penetration. These distributed generation devices are intermittent in nature, and can cause reverse power flows, rapidly changing voltages, and risk of over-voltage events.

The traditional devices and methodologies used by utilities for voltage regulation, including capacitor banks (CBs), load tap changes (LTCs), and voltage regulators (VRs), will no longer be sufficient. These electro-mechanical devices were not designed to respond with the frequency or the fast timescales warranted by the high penetration of variable generation, such as PV. To maintain voltages within acceptable operating range under such high variability, these devices are actuated more frequently, degrading the lifespan of these expensive equipment from 30 years to as little as 5 to 10 years. This phenomenon has already been documented in regions with high solar adoption [98]; with the expected growth of solar PV globally, voltage regulation is a top priority for utility companies.

Dynamic Volt-Ampere Reactive (VAR) devices such as smart inverters are an attractive addition for achieving voltage control. Since DERs such as PVs are often accompanied by smart inverters, their proliferation enables many more opportunities for achieving voltage control. Current utility practices for voltage regulation must be adapted to include DERs, modernizing the way utilities manage grid voltages and improving grid performance. DERs are typically equipped with sensing, actuation, and control technologies, giving rise to a highly distributed and intelligent grid edge. Further, DERs equipped with power electronic control (such as smart inverters) can provide low-cost and fast timescale reactive power compensation throughout the dis-

tribution grid. Thus DERs provide two unique opportunities for voltage regulation: reducing the mechanical switching burden on traditional control devices, and tighter spatial-temporal voltage regulation to meet additional network objectives. The decision making across these two groups of devices, traditional mechanical devices and DERs, must be coordinated.

4.3.1 Prior work

Prior work can be grouped into three categories: (1) voltage regulation strategies for controlling either traditional devices [120, 143] or DERs [158, 153, 144, 156, 32, 145], but not both; (2) both together in a single large scale optimization [94, 93, 92, 80]; and (3) both together in a hierarchical approach [40, 128, 90, 44, 130, 42]. The hierarchical approach is motivated by the different characteristics of the two sets of control devices: traditional devices have discrete settings, actuate over slow timescales, and are costly to actuate; while DERs follow continuous setpoints, can actuate over fast timescales, and provide operational flexibility at a significantly lower cost.

In the first category, various methodologies for voltage regulation have been employed. For traditional device control, the tap setting for VRs can be determined by relaxing the discrete variables to continuous [120], and modeling the unbalanced power flow with a convex rank-constrained SDP formulation. The resulting optimization problem can be solved with traditional solvers (either centralized or distributed approaches). However, the approach may not scale for networks with a high number of discrete devices, with suboptimality concerns due to the continuous relaxation. In [143] the methodology for LTC tap optimization is extended to applications where the grid topology is known but the line parameters (resistance and reactance values) are unknown, by employing a data-driven approach using reinforcement learning. However, in both of these works there are no limits on the number of tap changes and the approach results in frequent tap operations. For DER control, proposed methods include gradient based convex optimization to determine DER set points using a local controller with stability guarantees [158], using distributed optimization [153, 144], or partitioning the network into clusters overseen by regional coordinators using a

Linearized DistFlow model [156]. The work in [32] also includes a procedure to select PV units to provide voltage control services. The approach in [145] considers both active power curtailment and reactive power control to meet voltage objectives, while minimizing active power curtailment using a chance-constrained approach. While many of these papers suggest that hierarchical architectures can be used, they do not carry out any design or simulations of the framework. It should be noted that the approach taken in [145] to estimate the admissible active power range and control DERs accordingly, is most conceptually similar to our proposed approach.

In the second category, methods are proposed to optimize over both traditional devices and DERs simultaneously. The resulting MIP is solved directly with additional uncertainty modeling for the PV and EV resources [94] and linearization strategies for lossless transformers [93], or relaxed with the discrete variables converted to continuous [80]. In [92] a distributed optimization approach is used to design a cooperative control scheme which implements a reactive power fair utilization ratio for both CBs and DERs, towards line loss minimization. However, these methods don't take advantage of the timescale separation of traditional devices and DERs which naturally gives rise to a hierarchical decision and control architecture: slow timescale control for traditional devices and a fast timescale control for DERs.

The third group of papers employ a hierarchical approach, of which our proposed method is a member. In [40] the MIP corresponding to shunt capacitor actuation and substation voltage setpoint is solved using a dynamic programming approach to search the entire state space, while DER optimization is modeled as a second order cone program (SOCP) and solved centrally. In [130] the MIP is relaxed to a continuous optimization for LTCs, distributed ADMM is used to optimize over DERs [44, 42], and [90] presents a network cluster approach to deal with voltage violations. Finally [128] considers a MISOCP to solve the hourly day-ahead dispatch of CBs and LTCs, and formulates a MILP to optimize EV navigation and charging strategies for voltage regulation at a 15-min timescale. In real time a feedback control strategy is proposed to mitigate voltage violations. In all of these proposed frameworks, intelligent coordination between the two levels of optimization is not present: each set

of devices are optimized independently with no mechanisms for cooperation across device types.

In this thesis, we present a hierarchical coordination approach for voltage regulation is proposed to coordinate the actions of slow-timescale utility-owned devices (i.e. load tap changers, LTCs) and fast timescale customer-owned devices (i.e. DERs). The lower level optimization of DERs is built upon the physics-aware distributed coordination substrate, and additional coordination mechanisms are introduced to support centralized optimization of LTC operations.

4.3.2 Voltage Regulation as a Bi-Level Optimization

In what follows we restrict our discussion to the set of traditional devices being an LTC at the Transmission-Distribution (T-D) substation (or point of common coupling, PCC), and the set of DERs being solar PV equipped with smart inverters and flexible loads. The LTC sets system-wide voltages by changing the slack bus voltage and the DERs adjust local voltages. These selected devices have the key characteristics of future voltage regulation practices: system-wide versus local voltage control, discrete versus continuous set points, slow versus fast actuation, and utility-owned versus privately-owned devices. The proposed framework can be readily extended to include other devices such as capacitor banks (local, discrete, slow, utility-owned) and storage (local, continuous, fast, both utility- or privately-owned).

The voltage regulation problem in the presence of DERs can be formulated as a mixed integer optimization over two sets of variables: the change in discrete tap setting for the LTC (ΔK_t) located at the transmission-distribution (T-D) substation which governs the feeder voltage $V_{o,t}$, and the real and reactive power injections by DERs ($\{P_t, Q_t\}$), at each time-step t. The MIP can be formally stated as:

$$\min_{\Delta K_t, P_t, Q_t} \sum_t |\Delta K_t| + f(V_t, P_t, Q_t) \tag{4.5a}$$

$$\text{subj. to: } V_t \in \mathcal{V}\left(V_{o,t}, P_t, Q_t\right) \qquad \forall t \tag{4.5b}$$

$$\Delta K_t \in \phi_t \qquad \forall t$$

$$\{P_t, Q_t\} \in \varphi_t \qquad \forall t$$

$$V_{j,t} \in [\underline{v}, \overline{v}] \qquad \forall j \in \mathcal{N},\ \forall t \tag{4.5c}$$

The problem in (4.5) aims to regulate the voltage magnitude within operating limits (4.5c)[2], while minimizing the number and magnitude of tap changes ΔK_t, and a network-wide objective $f(V_t, P_t, Q_t)$, subject to limits on LTC operations denoted by space ϕ_t and DER limits denoted by space φ_t. This MIP is challenging to solve due to the presence of discrete variables ΔK_t, and the nonconvex power flow equations represented in (4.5b).

We address the first challenge by observing the timescale separation present in the LTC and DER device actuation, where DERs can much more dynamically respond to system conditions as compared to the slow-actuating LTC. We convert the problem (4.5) into a bi-level optimization. The upper level optimizes the LTC setpoint every τ timestep, as denoted by $\tau = [1, ..., T]$. The lower level optimizes the DER setpoints more frequently, every t time-steps, where $t < \tau, t = [1, t+\Delta t, ...t+n\Delta t]$ where n is the number of subperiods in τ. For example, if the LTC were optimized every hour and DERs were optimized every 5-minutes, a single day of operation has $T = 24, n = 12$, and 288 DER optimizations. Figure 4-2 illustrates the hierarchical framework, where our contributions are in green.

Upper Level: LTC Optimization

The upper level optimization is given in (4.6), and deals with the discrete variables: it determines the minimum LTC tap operation required to maintain network voltages in (4.6b) within operating constraints (4.6c). The LTC sets the voltage at the distribution substation, $V_{o,\tau} = 1.0 + \Delta v_o K_\tau$, where Δv_o is the voltage per tap, $K_\tau \in [-\kappa, \kappa]$ is the tap setting at time τ, and κ is the number of taps (symmetrically about neutral

[2]Voltage standards detailing the allowable deviation from nominal voltage under normal grid conditions vary globally. North America follows ANSI C84.1 which allows ±5% deviation (so $\underline{v} = 0.95$ and $\overline{v} = 1.05$), while Europe follows IEC and European EN 50160 which allows ±10% deviation.

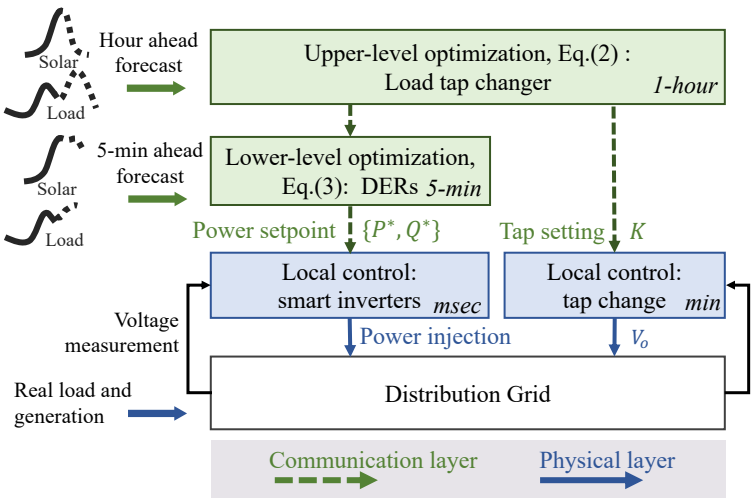


Figure 4-2: Proposed hierarchical architecture for voltage regulation in the presence of LTCs and DERs. We leverage the timescale separation of device actuation to present a bi-level optimization, as depicted in the green boxes.

state $K_0 = 0$).

$$\{K_1^*, ..., K_T^*\} = \underset{\Delta K_1 ... \Delta K_T}{\operatorname{argmin}} \sum_{\tau \in [1,T]} |\Delta K_\tau| \tag{4.6a}$$

$$\text{subj. to: } K_\tau + \Delta K_\tau \in [-\kappa, \kappa]$$

$$K_\tau = K_{\tau-1} + \Delta K_{\tau-1}, \; K_0 = 0$$

$$V_\tau \in \mathcal{V}\left(V_{o,\tau}, P, Q\right) \tag{4.6b}$$

$$V_j \in [\underline{v}, \overline{v}] \tag{4.6c}$$

Lower Level: DER Optimization

The lower level optimization problem is given in (4.7), and deals with the continuous variables: it determines the real and reactive power injection at every node in the network that minimizes a network-wide objective function, f, subject to power flow constraints $\mathcal{V}$ (4.7b), and DER constraints (4.7c).

$$\{P_t^*, Q_t^*\} = \underset{\{P,Q\}}{\operatorname{argmin}} \; f = \sum_j f_j(V_j, P_j, Q_j) \tag{4.7a}$$

$$\text{subj. to: } V \in \mathcal{V}\left(V_{o,\tau}, P, Q\right) \tag{4.7b}$$

$$\{P, Q\} \in \varphi_t \tag{4.7c}$$

The optimization problems (4.6) and (4.7) are coupled via the network voltages in (4.6b) and (4.7b), and the OLTC setpoint in (4.7b). To decouple the optimization problems and still achieve the objective in the original problem, we consider minimal data exchange within the bi-level framework. First we fix the substation voltage, V_o. We express the DER limits encoded in φ_t as inequality constraints on nodal P and Q injections, as $P_j \in [\underline{P_j}, \overline{P_j}]$ and $Q_j \in [\underline{Q_j}, \overline{Q_j}]$. We then note that these DER limits directly describe the limits on nodal voltages as described by the network power flow $\mathcal{V}$. The nodal voltages can be calculated as $V_{\mathcal{N}\setminus 0} = V_{o,\tau} + Z_{\mathcal{N}} I_{\mathcal{N}\setminus 0}$, where $V_{\mathcal{N}\setminus 0}$ and $I_{\mathcal{N}\setminus 0}$ denote the vector of nodal voltages and nodal current injections at all nodes except the T-D substation (node $j = 0$, slack bus), and $Z_{\mathcal{N}}$ denotes the network impedence submatrix with the columns and rows of the slack bus eliminated [46]. The nodal power injections can be translated into nodal current injections and voltage, with the bounds on nodal voltages at time t calculated, as in

$$\begin{aligned} \overline{V}_j &= \mathcal{V}_{\max}\left(V_o, \underline{P}_j, \overline{P}_j, \underline{Q}_j, \overline{Q}_j\right) \\ \underline{V}_j &= \mathcal{V}_{\min}\left(V_o, \underline{P}_j, \overline{P}_j, \underline{Q}_j, \overline{Q}_j\right) \end{aligned} \tag{4.8}$$

Since the nodal voltage bounds are calculated using the power physics, if a solution exists such that $P_j^* \in [\underline{P}_j, \overline{P}_j]$ and $Q_j^* \in [\underline{Q}_j, \overline{Q}_j]$, then $V_j^* \in [\underline{V}_j, \overline{V}_j]$. Next we consider the first stage optimization problem (4.6), with the voltage limits in (4.6b) calculated from (4.8). Solving the first stage (4.6) sets the optimal tap setting K_τ^* for time τ. Finally we solve (4.7) for $\{P_t^*, Q_t^*\}$ using the tap setting K_τ^* which sets $V_{o,\tau}$.

The proposed hierarchical framework motivated by timescale separation allows us to take advantage of different computational architectures at each level of optimization. Namely, the lower level problem in (4.7) is concerned with DERs that are distributed throughout the distribution grid, have private ownership, and whose power injections have relatively local effects. Meanwhile, the upper level problem in (4.6) is concerned with a single LTC at the T-D substation whose voltage setpoint

$V_{o,\tau}$ propagates through the entire feeder. We take advantage of these differences and propose the upper level problem is solved in a centralized fashion with network-wide information, and the lower level problem is solved in a distributed manner with peer-to-peer communication across DERs.

4.3.3 Distributed Voltage Optimization

Figure 4-3 shows a flowchart of the distributed scheme with two example nodes in the network: a load located at node j and a solar PV unit located at node k. A timeline of events is shown at the bottom of the figure. Each agent in the network[3] is equipped with sensing, computation, and communication capabilities. The agents will locally forecast the loads (P_L, Q_L and generation (P_{PV}) (Step 1) and run the pre-processing algorithm (Step 2) for the next time period. Next, the agents will begin the PAC-based distributed optimization step which consists of a local primal update (Step 3a), a peer-to-peer communication step sharing the real and imaginary voltage values (V^R, V^I respectively) with their neighbours (Step 3b), and update dual variables μ, ν (Step 3c) until convergence. We assume the timescales of optimization (i.e. time period from T_0 to T_1) are selected such that the algorithm will converge. The local devices then implement the optimal setpoint about which the primary controllers regulate. The example of a thermostat and a droop-based solar inverter are shown in Fig. 4-3. The main algorithm steps (Step 1 through 4) constitute the distributed voltage regulation scheme denoted by the green block in Fig. 4-2. The primary controllers (thermostat and droop-based solar inverter control) correspond to the blue local control block in Fig. 4-2.

4.3.4 Centralized Optimization for LTC Operation

The optimization problem in (4.6) is solved using a greedy algorithm which considers a single hour of operation τ and determines the smallest ΔK_τ that is needed to ensure

[3]For ease of exposition we assume a one-to-one mapping of nodes in the network to agents. In practice, an agent can represent one or more nodes, and will represent the aggregated capabilities across the nodes under its purview. The agent has full access to all information of the nodes and devices under its purview.

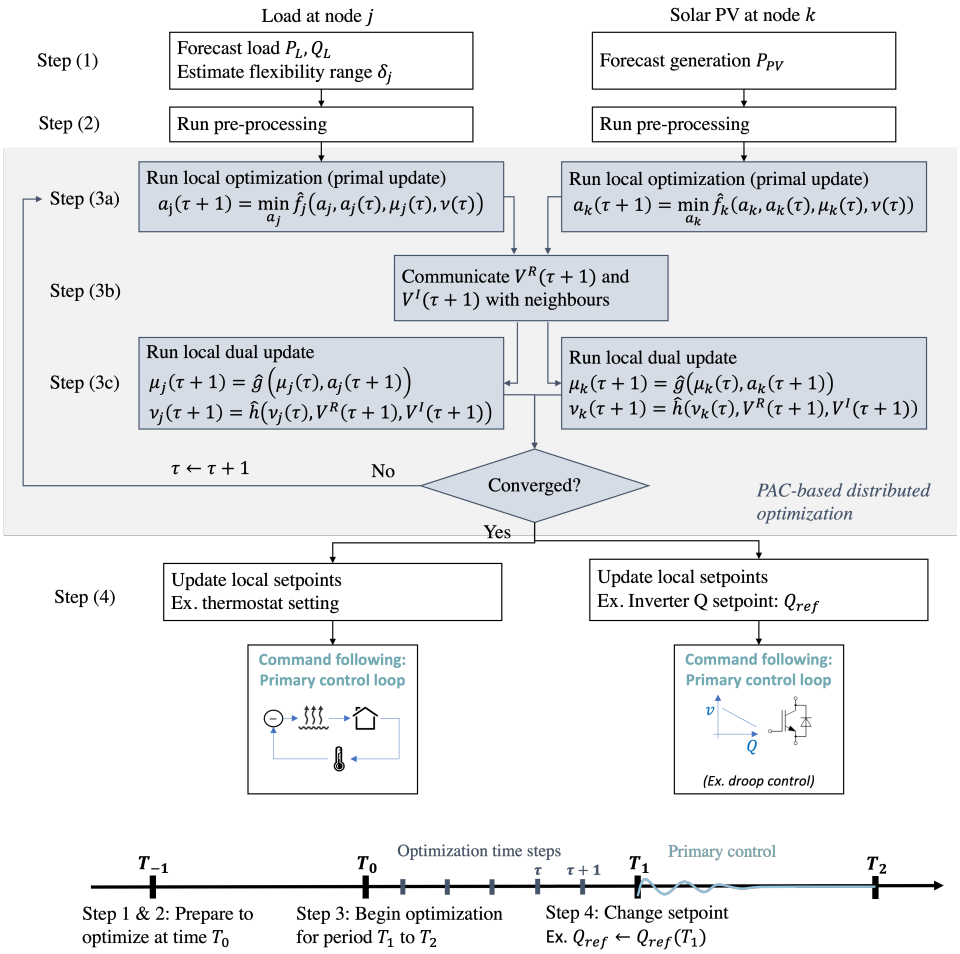


Figure 4-3: Distributed VVC scheme runs a PAC-based optimization to determine the device setpoints, around which the primary controllers regulate.

the voltage constraint (4.6b) is satisfied. The greedy algorithm runs every hour, for a single time step optimization, approximating the problem (4.6), as

$$\sum_{\tau\in[1,T]} \min_{\Delta K_\tau \in \phi_{\kappa_\tau}} |\Delta K_\tau| \approx \min_{\Delta K_1 \ldots \Delta K_T \in \phi_\kappa} \sum_{\tau\in[1,T]} |\Delta K_\tau| \tag{4.9}$$

where ϕ_{κ_τ} denotes the LTC feasible actuation space for time-step τ and ϕ_κ denotes the LTC feasible actuation space for all time-steps $[1, ..., T]$. The greedy algorithm determines the tap setting, as in

$$\Delta K_\tau = \begin{cases} \left\lceil \max_j \left(\underline{v} - \underline{V}_j \right) \right\rceil, \text{ undervoltage } \textit{risk} \\ - \left\lceil \max_j \left(\overline{V}_j - \overline{v} \right) \right\rceil, \text{ overvoltage } \textit{risk} \end{cases} \tag{4.10}$$

The evaluation of under- and overvoltage *risk* involves some tuning of hyperparameters in the greedy algorithm for the particular feeder and its load and DER conditions. These hyperparameters are part of the hierarchical coordination, discussed next.

4.3.5 Coordinating LTC and DER Actions

The voltage regulation framework described thus far is similar to the decision making paradigms proposed for transmission systems (across different sets of devices) and in literature for distribution systems. The original optimization problem in Eq. (4.5c) is split into two separate decisions which can result in a suboptimal solution. The introduction of additional coordination between the upper and lower levels of optimization can move the overall solution closer to optimality, as depicted in Fig. 4-4.

An additional coordination mechanism is proposed between the LTC decision in Eq.(4.10) and the subsequent DER optimization in Algorithm 3. The coordination mechanism will change the LTC tap setting only when the *risk* of voltage violations are *high*, and otherwise leverages the flexibility of DERs towards maintaining network voltages. First the LTC will evaluate the *risk* of voltage violations and set ΔK^*. The new network voltages after tap change can be approximated as $\overline{V}_{j,\tau} = \overline{V}_j + \Delta v_o K_\tau$ and

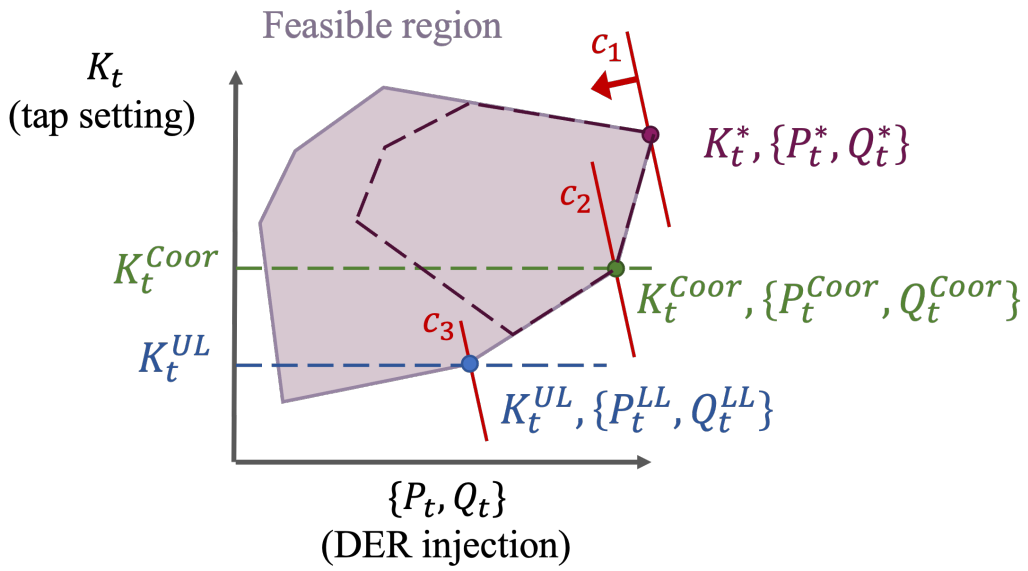


Figure 4-4: Consider an example with a single DER in the network and the LTC at the PCC. The linear equations of the CI model describes a polytope (indicated by the purple shaded area) showing the feasible space where voltages are within operating bounds, $V_t \in [0.95, 1.05]$. Note that in the illustration the integer points for the LTC are not shown for sake of simplicity in the diagram. Suppose the objective function is linear (for sake of illustration), where $f(V_t, P_t, Q_t)$ is a linear function, and that the integer points for the LTC coincide with the corner points of the polytope. This means the optimal solutions lie at one of the corners of the polytope (of the entire purple shaded area). When the optimization problem of Eq. 4.5c is solved the optimal solution reached is indicated by the purple dot, at $K_t^*, \{P_t^*, Q_t^*\}$ with objective value c_1. This result is that of a single centralized optimization over all possible device actuation, which gives the optimal solution. The direction of the red arrow indicates increasing cost, thus $c_1 < c_2 < c_3$ in this diagram (more generally, the costs may be less than or equal to each other). When the bi-level architecture without any additional coordination between upper and lower levels is used, the resulting solution is indicated by the blue dot, at $K_t^{UL}, \{P_t^{LL}, Q_t^{LL}\}$ with objective value c_3. This solution arises by first optimizing the tap operation and setting K_t^{UL}, then optimizing over the feasible DER operations which lie along the blue dashed line. Finally we consider the proposed approach of adding coordination between the upper and lower level optimization problems. The coordination restricts the operating range of the DER injections, as indicated by the new polytope boundaries in dashed purple. The new optimal solution with coordination is then indicated by the green dot, at $K_t^{Coor}, \{P_t^{Coor}, Q_t^{Coor}\}$ with objective value c_2. Notably the upper level optimization now considers the reduced DER operating range, resulting in a new optimal tap setting K_t^{Coor}. The lower level considers the reduced operating range (a subset of the original operating range) and the voltage at the PCC set by K_t^{Coor}, to arrive at the optimal DER injections $\{P_t^{Coor}, Q_t^{Coor}\}$.

$\underline{V}_{j,\tau} = \overline{V}_j + \Delta v_o K_\tau$. Finally, the LTC agent will limit the operating range of DERs in order to ensure the essential network voltage constraint in Eq.(4.6c) is satisfied. In doing so, the coordination mechanism improves upon the LTC objective in Eq.(4.9). Figure 4-5 shows the coordination scheme.

Mathematically, the upper level optimization problem with coordination can be described as:

$$\min_{\Delta K, \underline{V}', \overline{V}'} \ |\Delta K| \tag{4.11a}$$

$$\text{subj. to: } \Delta K \in \phi_\kappa$$

$$\underline{V}' \in \left[\max(\underline{V}, \underline{v}), \overline{V}'\right] \tag{4.11b}$$

$$\overline{V}' \in \left[\underline{V}', \min(\overline{V}, \overline{v})\right] \tag{4.11c}$$

The solution of the above optimization is the change in tap setting of the LTC ΔK^*, and the truncated feasible space $V \in [\underline{V}', \overline{V}']$. The truncated voltage bounds directly restrict the DER injections, $\{P, Q\} \in \varphi_t'$, $\varphi_t' \subset \varphi_t$, through interaction with the power flow constraints $V \in \mathcal{V}(V_o, P, Q)$. By adding the truncated voltage constraint directly in the DER optimization (4.7), we eliminate the need to explicitly calculate φ_t'.

The LTC optimization with coordination is presented in Algorithm 4 for the undervoltage case, and a schematic of the same is shown in Fig. 4-6. The algorithm has two tunable parameters, ϵ and σ. The parameter $\epsilon \in (0, 1]$ is a proxy for the likelihood of undervoltage occurring and represents a willingness to actuate the LTC. The parameter $\sigma \in (0, 1]$ represents a willingness to risk undervoltage, equivalently the tradeoff to relying on more DER flexibility. Take for example $\epsilon = 0.9$ and $\sigma = 0.5$: if there exists a node whose voltage bounds lie below $\underline{v}$ for more than $\epsilon = 90\%$ of the total voltage range, the LTC will be called upon. To calculate the new tap setting, we require that at least $\sigma = 50\%$ of all voltage ranges be above $\underline{v}$. Lower values of ϵ correspond to higher likelihood of actuating the LTC, and higher values of σ correspond to a higher undervoltage risk aversion (and lower reliance on DERs to restrict operations within φ'). These parameters must be tuned for each network

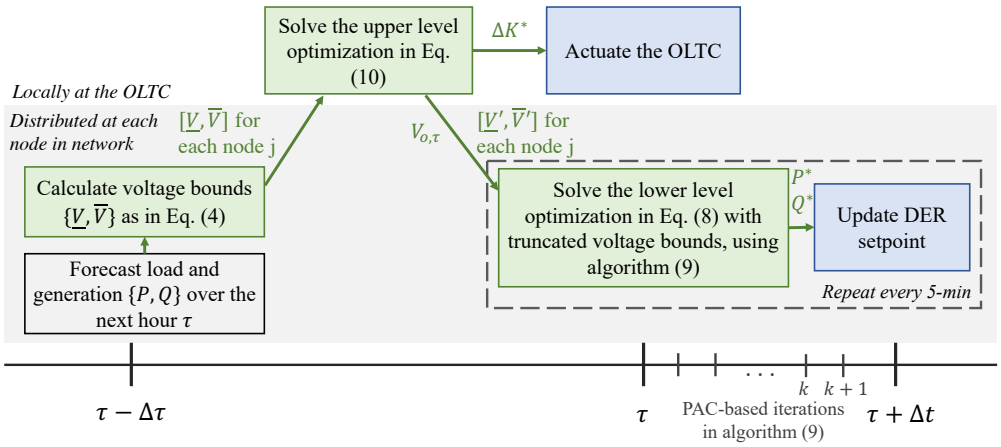


Figure 4-5: A schematic of the proposed coordination between the upper level optimization for LTC setpoint and lower level optimization for DERs. The output of the coordination is the adjusted voltage bounds $[\underline{V}', \overline{V}']$ which are communicated by the central LTC agent to each distributed DER agent. The timescales τ for each upper level optimization and Δt for the lower level optimization with iterations k are depicted.

during offline simulation studies, and may take into account additional factors such as load/generation forecast error and DER reliability (i.e. how reliably a DER will remain within φ' when requested).

Algorithm 4: Upper-level optimization for LTC, with coordination

Data: $[\underline{V}, \overline{V}]$

Result: $\Delta K^*, \left[\underline{V}', \overline{V}'\right]$

if $\exists\ j\ s.t.\ \frac{\underline{v}-\underline{V}_j}{\overline{V}_j-\underline{V}_j} > \epsilon$ **then**
  High risk of undervoltage. Calculate required tap change;
  $\Delta V \leftarrow \max_j \left(\underline{v} - \underline{V}_j + \sigma(\overline{V}_j - \underline{V}_j)\right)$;
  $\Delta K^* \leftarrow \lceil \frac{\Delta V}{\Delta v_o} \rceil$;
  Estimate new voltages;
  $\underline{V}_j^{\text{est}} \leftarrow \underline{V}_j + \Delta v_o \Delta K^*$;
  $\overline{V}_j^{\text{est}} \leftarrow \overline{V}_j + \Delta v_o \Delta K^*$;
else
  $\Delta K^* \leftarrow 0$;
  $\underline{V}_j^{\text{est}} \leftarrow \underline{V}_j$;
  $\overline{V}_j^{\text{est}} \leftarrow \overline{V}_j$;
end
$\underline{V}_j' \leftarrow \max\left(\underline{V}_j^{\text{est}}, \underline{v}\right)$

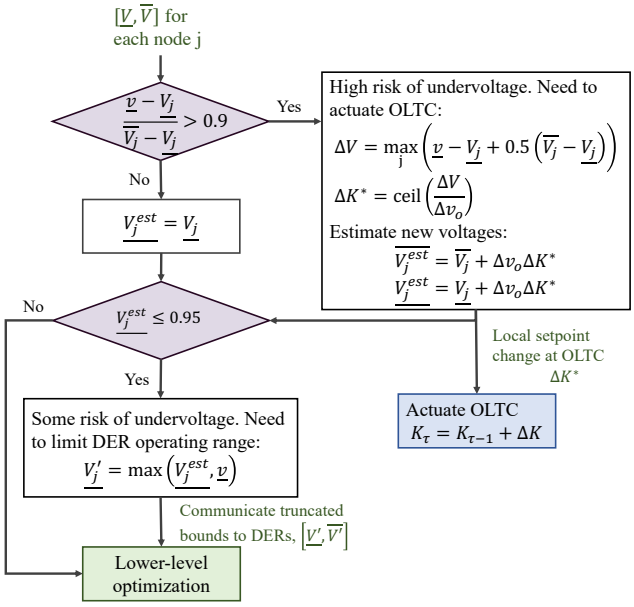


Figure 4-6: A schematic of the algorithm to solve the upper level optimization with coordination, with parameters in Algorithm 4 as $\epsilon = 0.9, \sigma = 0.5$.

4.3.6 Simulation Results on Utility Feeder

The proposed voltage regulation framework was validated on offline simulations of a real utility distribution feeder. The feeder has 56 nodes with total maximum load of 4.0833MW and 1.56MVAr. The distribution substation has a rating of 18MVA, and is equipped with an LTC. The LTC is on a three-phase step-down transformer with 32 taps and 0.00625pu voltage per tap. The LTC is initialized at the neutral state, $K_0 = 0$. These parameters characterize the LTC feasible space ϕ_t, where $\kappa = 16, \Delta v_o = 0.00625$. The feeder has a total of 750kW installed PV, distributed across 10 of the nodes (roughly 20% penetration of nameplate capacity to maximum real power load). We assume each PV unit is equipped with a smart inverter with power electronic control, and has the requisite primary control scheme that responds to setpoint changes. For the lower level optimization, we consider a common voltage regulation objective of Conservation Voltage Reduction (CVR), whereby we try and regulate the network voltages about a particular setpoint less than unity, i.e. $f_j = (V_j - v)^2$, where typically $v \in (0.95, 1.0)$. We set $v = 0.97$ as per utility preference. The LTC is optimized every hour ($T = 24$) and the DERs are optimized every 5-min

$(n = 12)$.

Simulation Results

Three voltage regulation cases are tested on the feeder: (i) no upper level optimization; (ii) no coordination, where in the upper level determines only ΔK^*; and (iii) with coordination (proposed method) where the upper level determines both ΔK^* and the reduced voltage bounds $[\underline{V}', \overline{V}']$. Case (ii) can be thought of as the framework typically proposed in literature [40, 128, 90, 44, 130, 42]. Figure 4-7 shows the resulting LTC operation, with the tap setting for each hour plotted on the graph, and Figure 4-8 plots the network voltages for each case. Without the upper level optimization we assume the LTC is not utilized (red). The network voltages frequently violate the lower limit $\underline{v}$, coming down as low as 0.925pu, in the beginning and end of the day when PV generation is not available to provide voltage support. The network voltages cannot be managed with only DER support. The next case (case ii) adds the LTC to provide the requisite support. The LTC is actuated frequently (blue) during the late evening hours when voltages are lowest. The LTC is actuated in 3 consecutive hours from 21 to 23 in order to maintain network voltages above $\underline{v}$. However, in this case the risk of very low network voltages during hour 22 results in an overvoltage event, driven by the high tap setting, $K = 9 \rightarrow V_{o,22} = 1.05625$. In comparison, the proposed method (case iii) with coordination between the LTC and DER actions only has a single tap change (green), and regulates the voltages between $\underline{v}$ and $\overline{v}$ without any violations. Figure 4-9 shows the total voltage adjustments for case (iii) when there is an undervoltage risk (blue) and an overvoltage risk (red). Notably, the voltage range adjustments for undervoltage during hours 0 to 2 reduces the required tap setting, where $K = 4$ in case (iii) as compared to $K = 5$ in case (ii), which allows the LTC to maintain voltages during the day without any PV nodal voltages reaching $\overline{v}$. Additional voltage adjustments happen throughout the day, with significant adjustments at hour 15, where the undervoltage events begin in case (i). The reductions due to overvoltage risk are significant through the peak PV generation hours, and successfully prevent voltages from reaching or exceeding $\overline{v}$.

The proposed voltage regulation framework is able to regulate network voltages by limiting the DER operating range, without excessive curtailment of PV generation during periods of overvoltage risk. Figure 4-10 plots the percentage of real and reactive load served by the distributed PV, as well as total network losses as a percentage of total load served, for case (iii). The results show that upwards of 20% of real power load and 5% of reactive power load can be met by distributed PV, as expected for the PV penetration level. Table 4.1 provides a summary of metrics across cases (i) thru (iii). The proposed method with coordination is better able to utilize the reactive power capabilities of the PV units as compared to without coordination, while also reducing reactive power load at the T-D substation and eliminating both under and overvoltage events. Further, the proposed method is able to reduce total network losses, from from 1.34MWh in case (i) to 1.32MWh for case (iii). The hourly losses as a percentage of load served is also plotted on Fig. 4-10 (yellow), with roughly 2%-3% losses throughout the day, and a slight increase at hour 9 where the a higher percentage of the load must be served by the bulk system at the Transmission-Distribution substation (or point of common coupling, PCC).

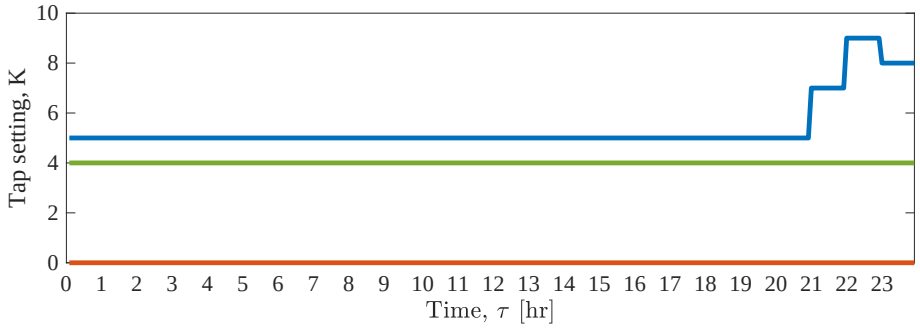


Figure 4-7: Tap settings over a day of operations for: (i) no upper-level optimization, red; (ii) no coordination, blue; (iii) coordination (proposed method), green.

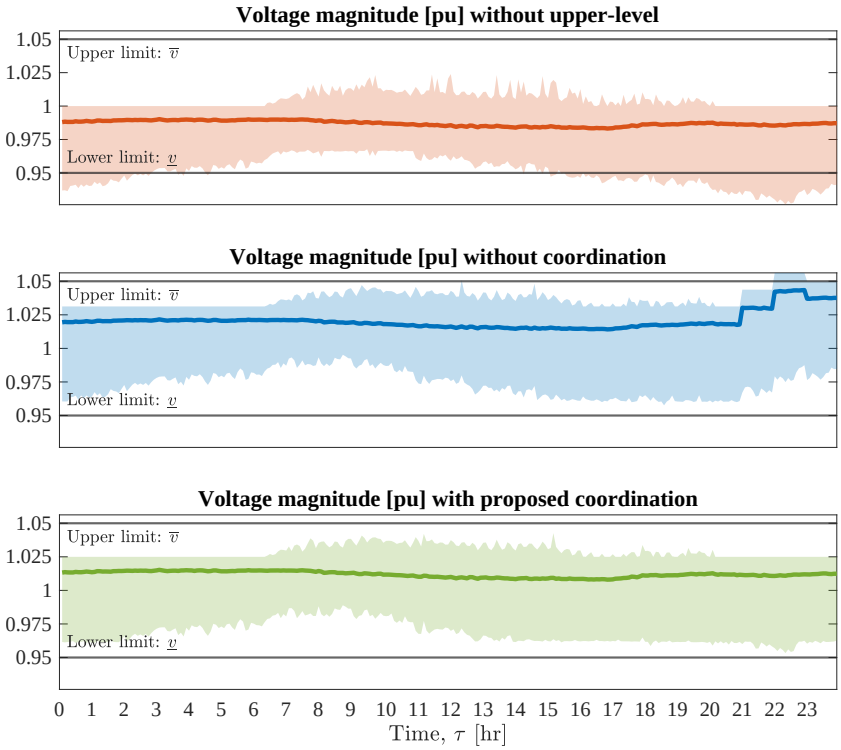


Figure 4-8: Network voltages over a day of operations for: (i) no upper-level optimization, top; (ii) no coordination, middle; (iii) coordination (proposed method), bottom. The average network voltage is shown in a dark solid line, and the range of network voltages as a shaded region. The upper and lower operational limits on voltage $(\overline{v}, \underline{v})$ are marked with solid black lines. Deviations outside of these limits (as in the top graph) are not permissible for grid operations.

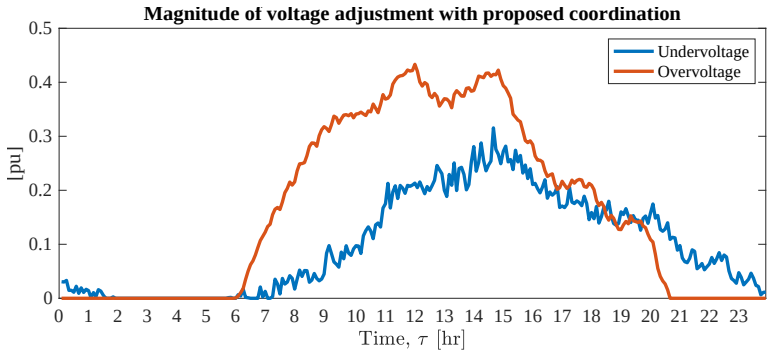


Figure 4-9: Magnitude of voltage adjustment over 24 hours of operation with proposed method, in response to risk of undervoltage (blue) and overvoltage (red). The magnitude is calculated as: $\underline{V}_{\text{mag}} = \sum_j \underline{V}'_j - \underline{V}_j$ and $\overline{V}_{\text{mag}} = \sum_j \overline{V}_j - \overline{V}'_j$, for under and overvoltage respectively. It should be noted that for cases (i) and (ii) the total voltage adjustment would be zero, as no coordination is present.

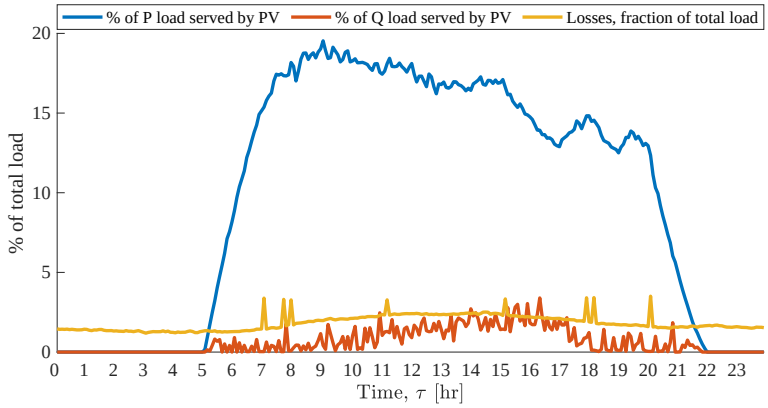


Figure 4-10: Results for proposed method over a day of operations: percentage of real and reactive power load served by distributed PV, and network losses as a percentage of load served.

Table 4.1: Summary of metrics from offline simulation studies on real utility feeder

	Case (i)	Case (ii)	Case (iii)
Number of tap changes	–	4	1
Norm of DER voltage adjustment	–	–	2.196 pu
Q injection at feeder	24.0 MVAr-h	27.8 MVAr-h	27.2 MVAr-h
Q injection from DERs	3.018 MVAr-h	0.156 MVAr-h	0.230 MVAr-h
Number of voltage violations	Under: 429 Over: 0	Under: 0 Over: 446	Under: 0 Over: 0
Norm of voltage violations	Under: 0.22 Over: 0	Under: 0 Over: 0.08	Under: 0 Over: 0
Network losses	1.34 MWh	1.35 MWh	1.32 MWh

4.4 Load Ramp Mitigation: Distributed Coordination of Storage

The push towards decarbonization of the electric grid has seen an explosive growth in renewable energy installation. The state of California generates approximately 25% of its electricity demand from solar resources, resulting in the characteristic "duck curve" in net system load [26]. This new operating condition, where dispatchable bulk resources must quickly meet the large and rapid change in electricity demand, introduces challenges to grid operators.

The deepening duck curve is no longer an infrequent event. On April 30, 2022, the California ISO (CAISO) reported a new record of 99.87% load served by renewable resources[4][28]. Almost 2 hours after this event, the CAISO system experienced a steep increase in net load (system load minus solar and wind), with a 3-hour ramp of almost 15.9GW, a fairly typical event for the CAISO grid[5]. Figure 4-11 shows the net demand in the CAISO grid on March 11, 2022 when the ramp record was set, and April 3, 2023 a fairly normal day of operations, with a ramp of 16.8GW[6].

Dispatchable generators are simple-cycle combustion turbines that typically operate on natural gas or petroleum liquids, and are capable of quickly ramping electricity generation (either up or down). These turbine units have ramp times of minutes to hours and minimum run times in the order of minutes. In comparison, combined-cycle combustion turbines and steam turbines take hours to ramp and have minimum run times of hours to days. Looking at the two example days from the CAISO grid, the majority of load ramp was supplied by natural gas and electricity imports, and in the 2023 case, some battery storage. This is shown in Fig. 4-12.

Although simple-cycle combustion turbines are *capable* of ramping quickly, it is not economical to do so. These generator units, often called peaker plants, are designed to run less than 10% of the time, where traditional plants and combined cycle plants must run 30% to 70% of the time to remain cost effective. This reduced operation pushes the electricity cost of a peaker plant from $0.08 per kWh to $0.15 per kWh[7][132]. The high costs of maintaining and calling upon peaker plants to supply the afternoon load ramp and the associated carbon emissions of these fossil-based resources makes this solution uneconomical and in opposition to green energy directives. While storage offers flexibility and low-inertia (fast ramping) capabilities, as seen by the 2023 example in Fig. 4-12, it remains too expensive for utility-scale system-wide adoption.

[4]This near 100% renewable electricity was sustained for 2 minutes, and was composed of 66% solar, 25% wind, and some geothermal, biomass, biogas and hydro.

[5]As of Feb 2023 the steepest 3-hour average ramp was 18.2GW on March 11, 2022. In comparison, the steepest 3-hour average ramp in 2015 was 10GW.

[6]Figures taken from CAISO 'Today's Outlook' widget. Available online as of 2023: http://www.caiso.com/TodaysOutlook/Pages/default.aspx

[7]These prices are estimates for the California grid as of 2014.

Instead, we propose system operators look towards the distribution grid to provide some support. Small-scale consumer owned DERs, in particular distributed storage devices, can provide support to the bulk grid. This load ramp mitigation service is provided by coordinating the actions of variable power factor setting of solar inverters, flexible loads to reduce consumption, and distributed storage devices such as community batteries. These DERs are coordinated across a large region of the distribution grid using the PAC-based distributed CI-OPF algorithm, upon the physics-aware distributed coordination substrate.

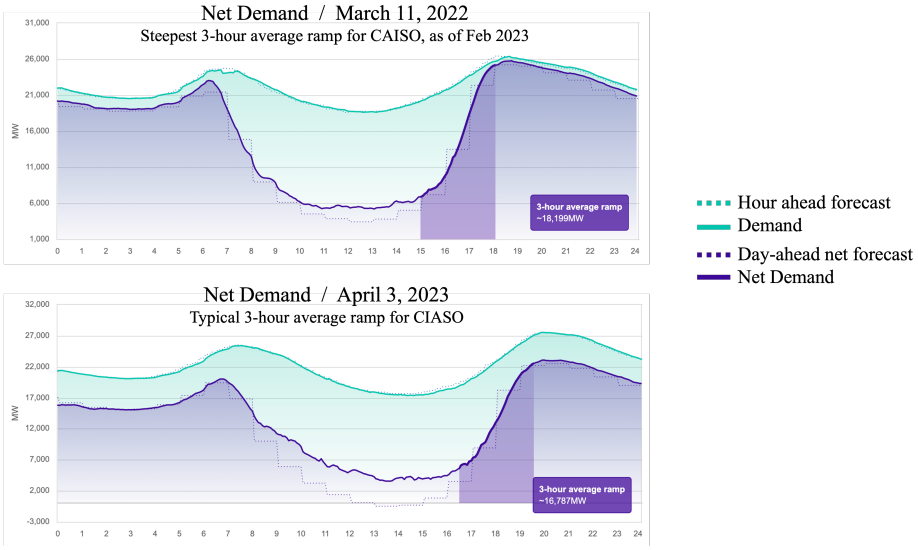


Figure 4-11: System demand and net demand (minus wind and solar) compared to forecasted demand of the CAISO grid. The plot is in 5-minute increments, and illustrates the duck curve phenomenon and 3-hour ramp rate. Figure is taken from the CAISO 'Today's Outlook' widget.

4.4.1 Simulation Results: Case Study of San Francisco

The proposed distributed decision making is tested on a case study of San Francisco, California, using the IEEE-34 node network as a proxy for the distribution grid. The load data from the IEEE datasheet serves as the daily average load, and the 24-hr load profiles are obtained from the ODEI dataset from NREL for the Typical Meteorological Year [105]. All loads are assumed to have a constant power factor of 0.95. The network loads are classified as residential or commercial loads based

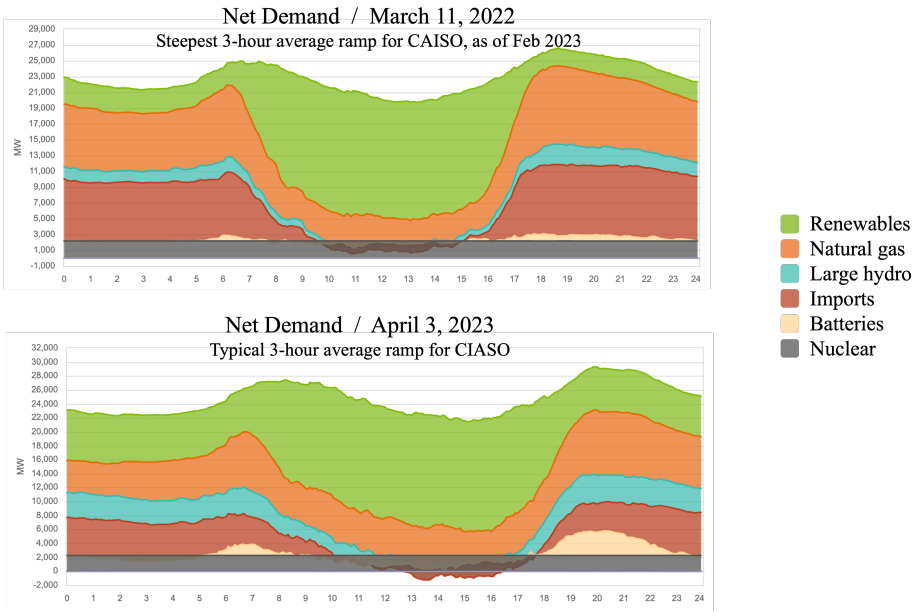


Figure 4-12: Breakdown of generation sources for the CAISO grid, corresponding to the duck curve illustrations. Top: Ramp onset is 3pm. Bottom: Ramp onset is 4:30pm. Figure is taken from the CAISO 'Today's Outlook' widget.

on the size of the load, by matching the load levels of the IEEE-34 network with the TMY data. Commercial loads include retail space, small and medium office buildings, primary school, medium and large restaurants, and a hospital. The network is modified to include DERs which include clusters of rooftop photovoltaic (PV) units, flexible loads, and three battery storage units. The penetration of PV is 38%, as measured by the ratio of nameplate capacity to average system load. This high DER penetration scenario is a reasonable projection given the RPS initiatives in California. Each PV unit is assumed to be equipped with an inverter with corresponding power electronic control, which can be operated at variable power factor in the range of 0.8 to 1. PV curtailment is not considered. The flexible loads are modelled as typical residential cooling loads for California [111]. Variations in nodal demand response are obtained by shifting the baseline profile obtained from [111] in time and space, with both following zero-mean Gaussian distributions with variances of 0.075 and 0.1 respectively. The three battery units are a 450kW-120kWh community unit[8] at node 6, a cluster of 40 Tesla Powerwall+ batteries (each 13.5kW-5kWh) at node 19, and a

[8]Modeled on the Ellenbrook unit from the PowerBank community storage trial, Australia [129]

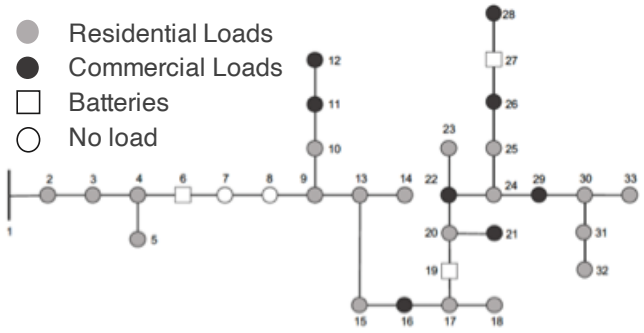


Figure 4-13: Topology of IEEE-34 network. PV units and flexible loads are present throughout the network.

800kW-185kWh hospital unit at node 27. Any load not met by local generation (or storage) is assumed to be served by the bulk grid at the point of common coupling (PCC), at node 1 in the network ($j = 1$). Figure 4-13 shows the network topology.

Scenarios Considered

We consider three scenarios, as below:

- Scenario A: *Baseline*. All PV inverters operate at unity pf, and batteries and flexible loads are not present.
- Scenario B: *Local control*. Each DER owner operates its devices and manages its loads.
- Scenario C: *Distributed control*. All devices and loads are coordinated using the proposed distributed coordination with PAC-based optimization.

Scenario A quite trivially is the characteristic duck curve, where the high PV generation at unity power factor results in large ramping requirements of transmission-level generators. Scenario B is a local approach, where each agent will minimize its peak load throughout the day. This serves as a very rough approximation of reducing the ramping requirements of bulk resources, by noticing that the largest ramp typically coincides with the peak demand in the evening. This action can be motivated by the fact that consumers can be charged based on their peak energy

consumption. To leverage the capabilities of the storage devices, neighbouring nodes are clustered with the battery. Residential loads (and corresponding DERs) at nodes 3, 4, and 5, share the community battery at node 6. Residential loads at node 20 and the primary school at node 21 share the cluster of Tesla Powerwall+ batteries at node 19. The hospital at node 26 is assumed to own and operate the battery at node 27. All remaining nodes are treated as independent agents. After clustering, there are a total of 26 agents in the network, each managing its own consumption and generation, to minimize its peak load. The multi-period optimization problem solved by each agent is a simple power balance, where any load in excess of local generation is assumed to be served by the bulk system. The power physics between nodes within a cluster is not modelled. The objective function is $f_{\text{local}}(y) = \max_t \{-P(t)\}$, where $y = \left[P, Q, P^L, P^G, Q^L, Q^G, P^{\text{sd}}, P^{\text{sc}}\right]$.

Scenario C is the proposed distributed coordination approach which employs the PAC-based optimization algorithm to coordinate device actions in a physics-aware framework. This approach accommodates system-level constraints including grid power physics, and minimizes the ramping requirements at the PCC. In Scenario C, the individual device agents solve the CI-based multi-phase unbalanced OPF, to minimize the objective function $f(x)$ which minimizes the difference in power supplied by the PCC from one hour to the next. The function is $f_{\text{distributed}}(x) = \left\| \sum_{\phi \in \mathcal{P}} P_1^{\phi}(t) - P_1^{\phi}(t-1) \right\|$.

Results and Discussion

Each scenario was simulated on the IEEE-34 node network. Figure 4-14 shows the net load served by the bulk system. The load curve for Scenario A shows the characteristic high ramps down and up when solar generation begins and ends. Scenario B provides minimal improvement, reducing the load for most hours of the day, while Scenario C effectively leverages the DERs to reduce the ramping requirements. Note that the objective function reduces the hour-to-hour change in load, and to do so, increases load during the hours of peak PV generation (roughly 10am to 4pm) to charge the batteries, which are then discharged in the late evening to reduce the net

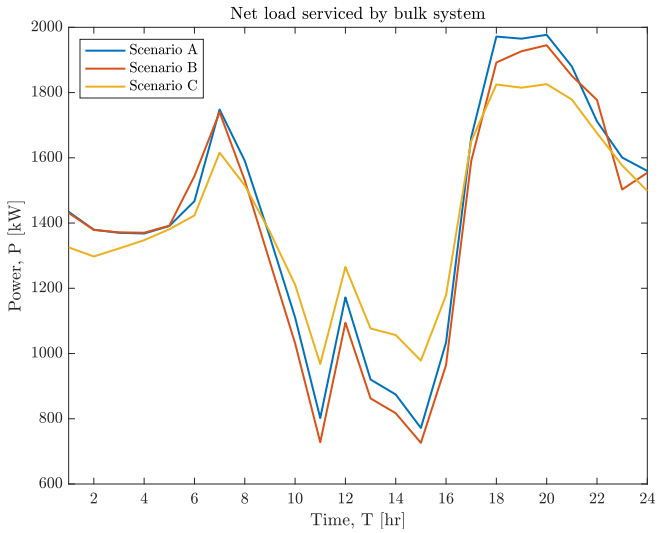


Figure 4-14: Net load serviced by the bulk system, for each scenario. These plots do not include power loss over lines.

load. The magnitude of hour-to-hour ramping is shown in Figure 4-15. Notably, the local optimization of Scenario B, which minimizes the peak load throughout the day as a proxy for minimizing the load ramp in the evening, is not able to suitably reduce the ramping requirement. The distributed optimization approach, on the other hand, is able to leverage system-wide information and coordinate the DERs to provide grid-level support, as needed by the bulk system. This coordinated approach is able to reduce ramping requirements throughout the day, with a 23% reduction in ramping requirements at the 4pm peak. Table 4.2 presents the total ramping reduction for Scenarios B and C, as compared to the baseline in A, and the computational run times. As expected, the proposed distributed coordination significantly outperforms the local approach, with 28% reduction in ramping required. The local approach is unable to provide any reduction. Unsurprisingly, the local approach takes less time to reach a decision (albeit an inferior one), while the distributed approach takes considerably longer (completing 1000 iterations). However, computational time of 20s is still well within the acceptable time-frame for decision making, and enables DERs to provide bulk-level support.

We next investigate the impact of the battery units' initial state of charge. We run each Scenario for (1) Minimum SOC where initial capacity is at 45, 0, and 160

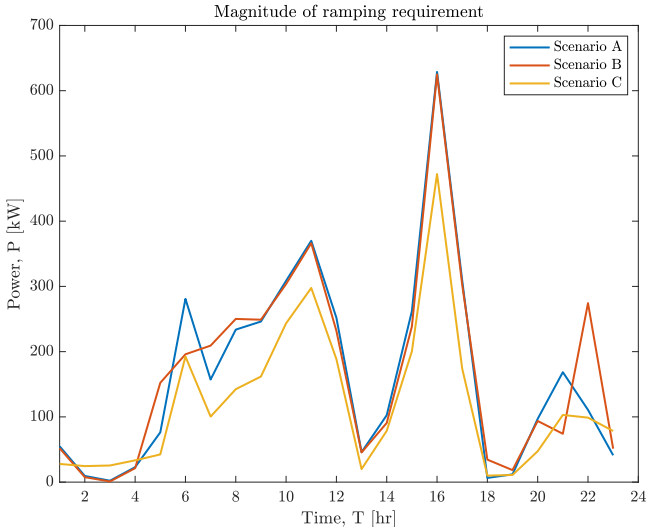
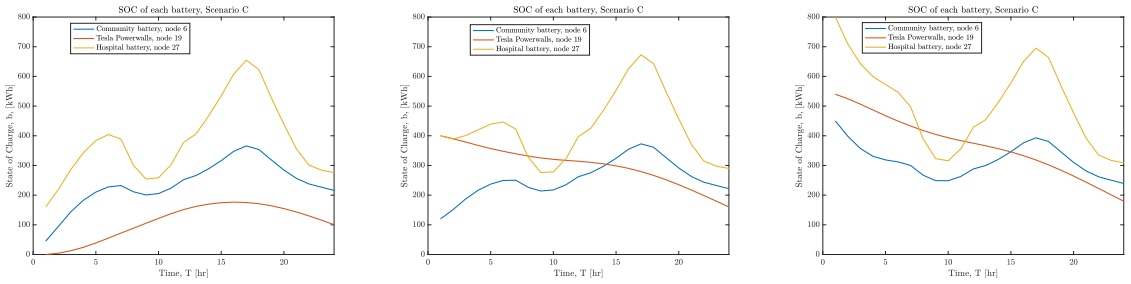


Figure 4-15: Magnitude of ramping required by bulk system generators for each scenario.

	A (Baseline)	B (Local)	C (Proposed)
Total ramping need (kW)	3923.7	3941.0	2839.4
Ramping reduction	-	-0.44%	27.63%
Mean run time per agent	-	0.0947s	16.97s

Table 4.2: Summary of results for Scenarios A thru C. Scenario A is baseline with no decision making, so no ramping reduction or computational time to report.



(a) Minimum initial SOC (b) Mid initial SOC (c) Full initial SOC

Figure 4-16: State of charge of each storage device for Scenario C, with different initial state of charge for each battery.

kWh; (2) Mid SOC where initial capacity is at 120, 400, and 400 kWh; and (3) Full SOC where initial capacity is at 450, 540, and 800 kWh, respectively for batteries at nodes 6, 19, and 27. Figure 4-16 plots the SOC of each battery unit for the three cases. The usage pattern in each of the cases is very similar, with charging in the early morning and afternoon to build up storage capacity for the evening. Interestingly, the final SOC are non-zero for the three cases, and are quite similar, suggesting the batteries may be able to retain a higher minimum charge to be used as backup power in the case of emergencies. The batteries provide flexibility to increase or decrease load throughout the day, as required by the grid.

4.5 Concluding Remarks

A physics-aware distributed coordination architecture was presented. The proposed framework leverages intelligence of grid-edge devices towards meeting system-level objectives while satisfying device and grid constraints. The distributed coordination reduces the communication and computational burden as compared to using a central decision making agent, thus allowing it to be used in future distribution grids with high DER penetration. Two services are developed upon the physics-aware distributed decision-making architecture.

The first service was voltage regulation wherein a hierarchical coordination approach was proposed for distribution grids with high DER penetration. The key challenges in voltage regulation were addressed by taking advantage of the timescale

separation between device actuation. Under this principle, a bi-level optimization problem was formulated to coordinate LTCs and distributed PV actions, separately optimizing over discrete and continuous variables. At the lower level optimization over DERs a distributed PAC-based optimization approach was employed, solving a linearized optimal power flow problem with limited data exchange between neighbouring PV units. The upper level optimization determines LTC tap settings, with additional coordination between DER actions and LTC. This proposed coordination improves upon the state of art by fully integrating the flexibility of DERs into the optimization process by restricting the DER operating regions to maintain network voltages. The proposed method is tested on a real utility distribution feeder with moderate PV penetration (20% installed capacity to load). The proposed method is capable of (i) reducing the number of LTC tap changes; (ii) eliminating all voltage violations; (iii) effectively utilizing both real and reactive power capabilities of PV units; and (iv) reducing network losses.

The second service was load ramp mitigation wherein a distributed coordination of storage devices was proposed to aggregate and dispatch services for the bulk grid. The case study on a modified IEEE-34 node network shows how distributed techniques can leverage information from different resources to successfully mitigate the duck curve, reducing ramping requirements of bulk system generators by up to 23%. This framework can be extended to include electric vehicles, by modeling the vehicle-to-grid capabilities and corresponding cost of battery cycling and lifetime degradation in the optimization problem. Using such an approach throughout the distribution grid can reduce the challenges of the new operating condition resulting from high solar and renewable penetration, reduce system costs, and improve renewable integration.

Chapter 5

Additional Research Contributions

I completed extensive research on several other research topics, in addition to what was outlined in Chapters 3 and 4. These topics include retail electricity market design, accelerated distributed optimization methods, and energy management of interconnected energy hubs. What's presented below is the outcome of extensive collaborations on the project topics, either in part of full. A note of the collaborators involved will be made at the beginning of each research project.

5.1 Retail Electricity Markets

The transformation of the power sector – the emergence of intermittent renewable generation, battery storage, and the rapid growth of small-scale customer-owned DERs – necessitates a change in how the electrical system is operated and a revitalization of the corresponding market structures and business models governing electricity economics. The centralized paradigm of current electricity markets may no longer be adequate to support the energy transition and clean energy targets, while maintaining power system stability, security, and affordability. Rather, a distributed paradigm must be invoked in the economic substrate of the power grid as well, which leads to new retail market mechanisms to efficiently operate assets and support investments within a distribution system. Emerging market structures must encourage flexibility and value the grid services provided by resources, as opposed to traditional cost recov-

ery mechanisms. Markets of the future should take advantage of grid modernization initiatives and inform timely integration of grid digitalization, energy management systems, communication networks, and the creation and widespread adoption of new standards and protocols.

5.1.1 Precursors to Retail Markets

A portion of this research was the outcome of R. Haider's SM thesis. Research in collaboration with David D'Achiardi

The electricity deregulation movement of the 1990s divided the vertically-integrated value chain along the power system into generation, transmission, distribution, and electricity markets. More recently, a similar deregulation movement facilitated competition within retail power sales, triggering the emergence and growth of competitive retail suppliers and Community Choice Aggregations. Despite these steps, distributed electricity markets do not currently exist. Distribution-level markets displace agents on all revenue streams associated with the vertically-integrated utility. For this reason, the design of retail market precursors varies greatly. Existing retail compensation schemes rely on static or averaged tariff rates which do not incorporate locational or temporal pricing. As a result, these schemes fail to provide adequate compensation for DERs, whose grid services are inherently variable in location and time. Tariff structures based on time-variable pricing (TVP) have the ability to provide cost-reflective price signals and enable more efficient operation of DERs; however, retail customer enrollment in these programs has been limited. Other retail compensation schemes typically overcompensate resources at the expense of higher future retail prices. In the absence of retail markets, wholesale electricity market (WEM) participation has opened up to DERs, through aggregators and new market models, and will continue to grow in the US as per FERC Order 2222 [45]. The implications of this wholesale participation as compared to retail markets will be discussed next.

In [110] a description of the current US electricity market and settlements are provided. Both [159] and [100] discuss the transformation of the power sector, and [71] suggests new market structures and business models are needed to encourage

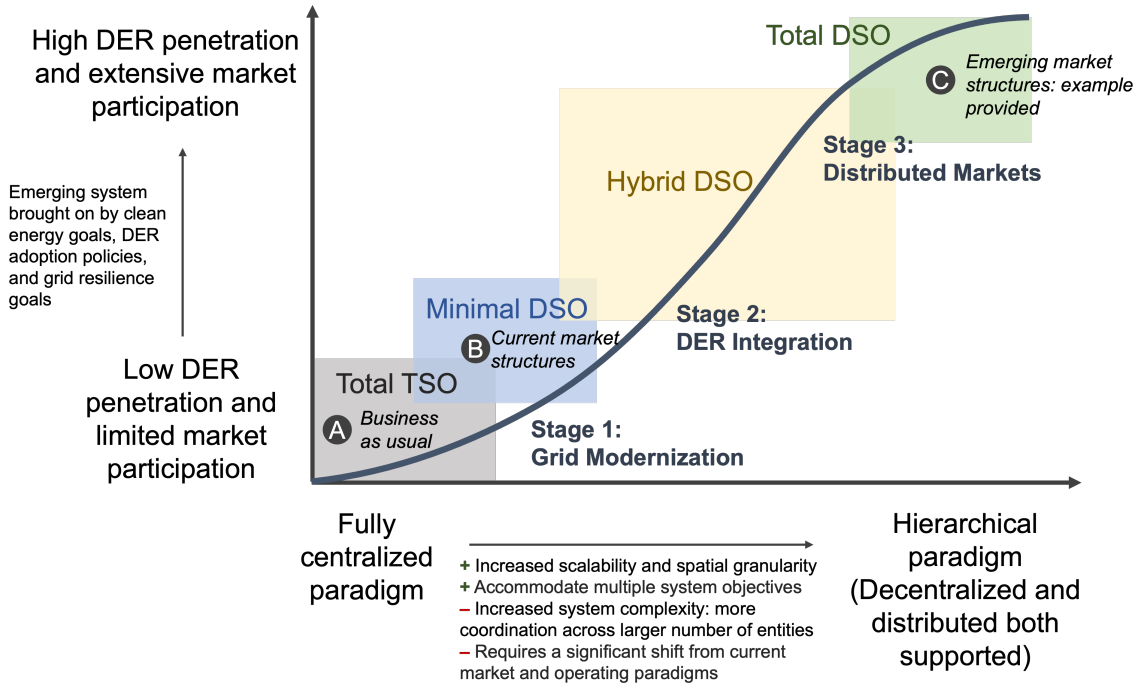


Figure 5-1: Progression of market structures as DER penetration increases. Adapted from [96, 81], with additional input from [63, 65].

flexibility and value services for a renewable-rich energy system.

5.1.2 Emerging Market Structures

Different market models have been proposed in literature, and can be classified as per Figure 5-1. The operating and market participation paradigms ranging from the Total TSO to Total DSO classifications vary significantly in market clearing, dispatch, and control, as well as implementation concerns like scalability, plug-and-play features, and communication requirements. A brief discussion of these overarching models is presented below.

Centralized paradigm

The centralized paradigm considers a single market (i.e. the wholesale electricity market) which receives and clears the bids of all market participants, including large-scale bulk resources and small-scale DERs. Under the single market structure, the TSO dispatches resources and sets prices, communicating setpoint decisions with the

DSO. The centralized paradigm includes both Total TSO and Minimal DSO models. The Total TSO structure will model the entire electrical grid – including the distribution lines down to individual DERS – providing highly granular spatial and temporal pricing, along with dispatch setpoints for individual DERs. The Minimal DSO model would dispatch DERs as if they were located at the T-D substation. Although local market structures, such as distribution level markets and peer-to-peer structures, can be integrated into the centralized paradigm, the primary market price signal and dispatch would be at the wholesale level, and sent by the TSO. The centralized paradigm includes "business-as-usual" markets wherein DERs are not permitted to participate or have limited participation mechanisms, and current market structures which attempt to integrate DERs into the WEM (such as California's DERP model, and FERC 2222 in the U.S.).

Hierarchical paradigms

Hierarchical paradigms consider a separation of transmission and distribution resources, in that dispatch decisions for DERs are completed by the DSO. Under a single market framework, the DSO will aggregate DER flexibility into a single bid at the T-D substation which is cleared by the TSO in the wholesale market. The TSO then sends the market dispatch to the DSO, which then disaggregates the dispatch request to the constituent DERs. Emerging hierarchical market designs typically include the establishment of a DSO which significantly expands on the primarily asset-centric role of current DSOs (such as those in Europe), distribution utilities, and network operators [134]. Additional roles may include oversight over DER-participation, aggregation and bidding functions, and additional market oversight of distribution level markets.

Implementation Challenges and Opportunities of Market Paradigms

Market structures along all ends of the spectrum face implementation challenges. While centralized paradigms are a natural extension of current market structures, they may not be scalable to the high penetrations of DERs expected to be installed

in the distribution grid. Perhaps the most pertinent challenge remains that of tier bypassing: DERs which reside in the distribution grid are directly dispatched by wholesale markets in response to the bulk grid objectives without considering local constraints [81]. In a centralized paradigm the grid physics of the distribution grid, including necessary voltage constraints, are not accounted for in market dispatch. There remain open questions around how a DSO or distribution utility may respond to market dispatch signals which violate local constraints or capabilities – if DERs are re-dispatched at the local level are they still subject to a penalty for reneging on the market dispatch? Further, inefficiencies of centralized market structures and precursors to retail market designs must be noted. These include: (i) relatively static pricing; (ii) voluntary enrollment (ex. time-of-use retail tariff programs); (iii) competing retail and wholesale programs (dual-participation not typically permitted); (iv) prohibitive technical requirements (ex. metering and telemetry requirements); and (v) prohibitive regulatory requirements (ex. interconnection rules, communication requirements, and 24/7 participation rules) [65]. Although the Minimal DSO model attempts to address issues of static pricing by exposing DERs and their aggregations to wholesale prices, the real world implementations of this model leave much to be desired, with unclear regulation around (iii)-(v), and no locational granularity in electricity pricing. Further, the growth of third-party DER aggregators introduces additional questions around cost allocation of the equipment, registration processes, and any market penalties between the aggregator entity and the individual DERs. Depending on the participation rules and eligible market products for DERs, the DER owners must be concerned with both the high cost of market entry where economies of scale do not apply, and potential limited profitability for DER services due to the misalignment of different tariff structures between wholesale and retail programs.

At the other end of the spectrum, the hierarchical paradigm is a significant shift from current market structures and would require a significant re-imagining of the roles and responsibilities of various market agents, and significant regulatory reform. However, the hierarchical paradigm has better scalability across high DER penetration levels, and when appropriately designed can enable plug-and-play characteristics

which are highly desirable in future operating frameworks. Furthermore, the hierarchical paradigm lends itself more naturally to the establishment of distribution-level retail electricity markets. Retail markets can support the creation of new DER services which take advantage of the operational flexibility of DERs by developing distribution-specific market derivatives and prices with a high spatial temporal granularity [63, 65, 72]. However, retail markets displace agents on all revenue streams associated with the vertically-integrated utility. Due to this complexity, retail markets remain a theoretical market framework. Discussions of such market structures remain in a nascent stage, and the design of such markets, the operational changes, and the regulatory requirements are all open research and implementation questions.

5.1.3 Proposal for a Local Electricity Market

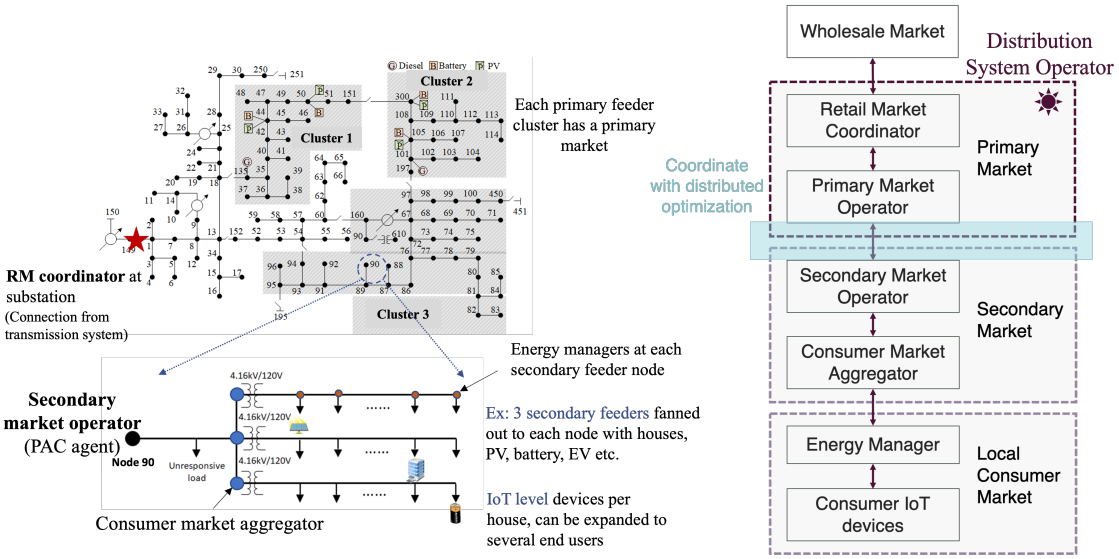
A portion of this research was the outcome of R. Haider's SM thesis. Extensive work has continued on this project. Research in collaboration with Vineet Nair at MIT; and Venkatesh Venkataramanan at NREL.

A local electricity market design is proposed in [63, 65, 72] in which a retail market structure is introduced in the distribution grid to coordinates the flexibility of DERs. We leverage the concept of *transactive energy*: Defined by NIST as "a system of economic and control mechanisms that allows the dynamic balance of supply and demand across the entire electrical infrastructure using value as a key operational parameter" [109], transactive energy bridges the gap between physical power flow in the grid and market derivatives. Such a retail market has its foundation in an advanced distributed optimization algorithm, which enables local and private bidding transactions, to achieve network-level objectives. In particular, we propose a DSO-centric retail market that determines the appropriate incentives for DERs to participate in the market [63]. These monetary incentives take the form of distribution-level Locational Marginal Prices (d-LMPs) to participants at the distribution primary feeders, similar to the notion of LMPs employed as pricing signals in the wholesale energy market by Independent System Operators (ISOs) at the transmission level [36]. The d-LMPs are determined using the distributed PAC optimization algorithm [122, 63],

as a core component. All underlying grid physics and constraints in the distribution system are incorporated in deriving the d-LMPs. As a result, they have the potential to fully exploit the emerging flexibility of the distribution system, and reduce operational costs across the power supply chain.

In the proposed market structure the DSO provides regulatory oversight of the retail market and DER integration (ex. interconnection, participation rules) similar to an ISO at the wholesale level. The billing and financial transactions pertaining to the local electricity market are completed by a Primary System Operator (PMO) and Secondary System Operator (SMO) at the primary and secondary levels of the distribution grid respectively. This tiered structure fully embraces the hierarchical paradigm of distributed markets. The various roles and responsibilities of the DSO, including maintaining system reliability, facilitating transactions between agents and aggregators, enabling energy procurement, and market clearing and scheduling, are distributed across the two market operators (PMO and SMO) and the DSO. The upper market tier operated by the PMO focuses on managing distribution grid physics and interactions with the wholesale market. The lower market tier operated by the SMO focuses on addressing consumer preferences, the reliable performance of DERs through introduction of a commitment score, and monitoring of security breaches such as cyber or IoT attacks on DERs. Within this structure, each of these market operators are non-profit entities, providing market oversight to ensure fair retail market access for participating DERs and aggregators. Figure 5-2 shows the market structure.

A DSO-centric market has several advantages, the first of which is that it enables DERs to directly participate in a local setting, and have the market location be collocated electrically, such as at a primary feeder node or a secondary feeder node. Second, the separation of wholesale and retail markets also permits a separation in market objectives. The wholesale market is concerned with the cost and reliability of electricity delivery in the transmission system, which corresponds to solving the unit commitment and security constrained economic dispatch problems subject to generator and line flow limits. The distribution market which presides closer to end-use



(a) Diagram of a distribution feeder wherein there are 3 primary feeder clusters, each with their own primary market overseen by the Primary Market Operator (PMO). Each node in the primary feeder cluster is a Secondary Market Operator (SMO) and oversees the participation of customer DERs and aggregation of flexibility bids into the primary market.

(b) Proposed structure of the hierarchical local electricity market, overseen by the DSO. The market structure is electrically collocated with the underlying distribution system.

Figure 5-2: Proposed Local Electricity Market at the retail level.

customers would also be concerned with cost and reliability, but may include additional objectives and constraints. These may include: (i) minimizing the disutility of flexibility (ex. inconvenience caused by curtailed or shifted load); (ii) maximizing the aggregated DER flexibility (ex. for bidding into wholesale market); (iii) minimizing deviation of DER performance from market dispatch (ex. performance penalties, including commitment or performance scores); (iv) and enforcing electricity affordability for all customers (ex. retail price caps). Fundamentally, these two features – local setting with electrical collocation and inclusion of distribution-specific objectives – fully address the issue of tier bypassing, the primary concern for grid operators in the centralized paradigm.

Details of the Proposed Market

The retail market leverages the physics-aware distributed coordination architecture proposed in Chapter 4. While the contributions presented therein focused on developing the grid services towards enabling DER flexibility, the contributions here are

building the framework for financial transactions which will encourage DER owners to provide those services. As such, the market structure is built upon the OPF problem, where the power physics of the primary distribution system and DER constraints are modelled. Notably, an AC power physics model is used, wherein both real and reactive power are considered in the proposed market framework. Thus, the proposed market solution covers simultaneous real and reactive power deployment, and the corresponding energy price settlements. A more detailed discussion of the reactive power prices is presented later.

The objective function $f(x)$ of the optimization problem is chosen to carefully reflect the energy market structure. The objective function is chosen to be a combination of Social Welfare and line losses and is given by:

$$\begin{aligned} f\left(x\right) = & \sum_{j \in \mathcal{N}_G} \left(a_j^P P_j^2 + b_j^P P_j + a_j^Q Q_j^2 + b_j^Q Q_j \right) \\ & - \sum_{j \in \mathcal{N}_L} \left(\alpha_j^P (P_j - \underline{P_j})^2 + \alpha_j^Q (Q_j - \underline{Q_j})^2 \right) \\ & + \zeta \, \mathrm{Re}\left(I^H Z I \right) \end{aligned} \tag{5.1}$$

where the first two terms represents generator cost and customer load disutility, and the third term represents line losses, in which superscript H denotes the Hermitian. In (5.1), $\mathcal{N}_G$ and $\mathcal{N}_L$ are the set of generator and load nodes respectively, ζ describes the tradeoff between economic and energy efficiency objectives, $a_j^P, b_j^P, a_j^Q, b_j^Q$ are generating cost coefficients, and α_j^P, α_j^Q are load disutility coefficients. All nodes $j \in \mathcal{N}_G$ and $j \in \mathcal{N}_L$ represent active distributed generation and demand response participants in the distribution grid, respectively. Node 1 is treated as the point of common coupling with the transmission grid (i.e. transmission-distribution substation), and reflects the wholesale price of electricity, which is the LMP λ^P from the wholesale electricity market. Thus for $j = 1$, $a_j^P = 0$ and $b_j^P = \lambda^P$. The cost coefficients related to reactive power are chosen as $a_j^Q = 0, \quad b_j^Q = 0.1 b_j^P$. The motivation for these choices comes from [43] which cites that reactive power prices are often one-tenth of that of real power. It should be noted that the weighted combination of the social

welfare and line losses is not standard practice, but is included here as line losses are more significant for an optimal functioning of a distribution grid compared to a transmission grid. The resulting solutions of the CI-OPF problem forms the backbone of our proposed reactive power market.

Validation of the Retail Market Design

The proposed market design has been validated through multiple case studies including a modified case study of an IEEE-123 bus primary feeder test case with multiple secondary feeders at each bus. We validate the performance of the market compared to existing retail practices, including demand response with no-export rules and net metering. The simulations are on a modified IEEE-123 node network using data from ISO-NE, the utility Eversource, and weather data for Boston, MA. The results for the primary market show significant customer cost savings, with a decrease in electricity prices from 0.114 \$/kWh down to 0.0291 \$/kWh [65]. The results for the hierarchical structure show that the LEM can coordinate and aggregate local DERs more effectively when an SMO is present to aggregate the flexibility bids [72]. This enables an optimal combination of local power and power drawn from the bulk grid, and reduces distribution level costs and d-LMPs. The incorporation of a commitment score helps to maintain better reliability while still extracting flexibility from customers and DERs. Finally, the time-varying local retail tariffs lead to more efficient market scheduling and lower final costs for end-users, while ensuring that DERs and consumers are correctly compensated for the flexibility services they provide to the grid. In what follows, a focus on the primary market is presented, as it pertains most directly to cost reductions for customers.

The proposed market operation is benchmarked against four operating models, wherein the utility purchases power from the WEM at the wholesale price and sells to customers at a fixed retail price, as is currently done in the US. The first of these is a 'Traditional' model where there is no DER utilization. The 'No Export' model realizes DR as continuous DER operation, rather than only during specific call windows which are a limitation of the programs, and retain the no-export rule

for behind-the-meter resources from wholesale market participation models. Retail compensation schemes of net energy metering (NEM) and net energy billing (NEB) are used in the 'Retail_M' and 'Retail_B' models.

Traditional: There is no DER utilization within the network. All load is serviced by the utility.

No Export: DG resources are used to offset local load and cannot export excess generation to the grid (excess generation is curtailed). All load from customers without DGs and any excess load of DG owners is serviced by the utility.

Retail_M: DG resources are used to offset local load and can export excess generation to the grid. Compensation for DGs follows NEM, at a fixed retail purchase rate. Any excess network load is serviced by the utility.

Retail_B: DG resources are used to offset local load and can export excess generation to the grid. Compensation for DGs follows NEB, at a fixed retail purchase rate. Any excess network load is serviced by the utility.

We use several metrics to validate the market performance using different stakeholder perspectives. They include the revenue for a DER owner, the cost for a customer consuming electricity, and the net revenue for the DSO. These metrics are calculated as follows. Real and reactive power are denoted as P and Q, with superscripts G and L for generation and consumption respectively. Subscript j denotes the j-th agent participating in the market. The wholesale LMP is denoted as λ^{P}, retail electricity prices as $\mu^{\mathrm{P}}_{\text{retail}}$, retail purchase rate as $\mu^{\mathrm{P}}_{\text{retail-G}}$, and the d-LMP for an agent j as μ^{P}_j and μ^{Q}_j. The baseline load for an agent j is denoted as P^{L0}_j and Q^{L0}_j. With this notation, we define the following quantities.

Payment made to WEM, for purchasing power:

$\mathcal{C}_{\text{WEM}} = \lambda^{\mathrm{P}} \sum_j P^{\mathrm{L}}_j$

Revenue earned from loads without proposed market:

$\mathcal{R}^{\text{base}}_{\text{load}} = \sum_j \mu^{\mathrm{P}}_{\text{retail}} P^{\mathrm{L}}_j$

Revenue earned from loads with proposed market:

$\mathcal{R}^{\text{market}}_{\text{load}} = \sum_j \left(\mu^{\mathrm{P}}_j P^{\mathrm{L}}_j + \mu^{\mathrm{Q}}_j Q^{\mathrm{L}}_j \right)$

Remuneration to distributed generators:

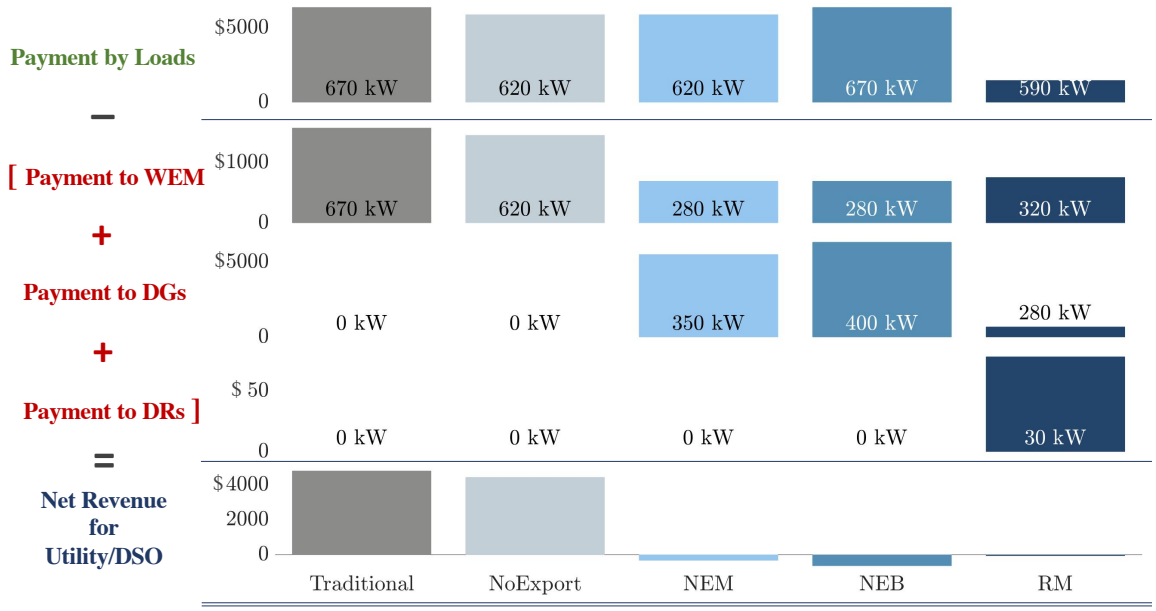


Figure 5-3: A breakdown of the net revenue calculation under all scenarios, with all revenue and cost in USD. The kW value used to calculate the amount is annotated on the bars - these are the load serviced by the utility, power purchased from the WEM, load serviced by the DGs, and load curtailment by DRs.

$\mathcal{R}_{\text{gen}} = \sum_j \left(\mu^{\text{P}} P_j^{\text{G}} + \mu^{\text{Q}} Q_j^{\text{G}}\right)$

where for traditional and no export cases $\mu^{\text{P}} = 0$ and $\mu^{\text{Q}} = 0$, for Retail_M and Retail_B cases $\mu^{\text{P}} = \mu^{\text{P}}_{\text{retail-G}}$ and $\mu^{\text{Q}} = 0$, and for the proposed market $\mu^{\text{P}} = \mu_j^{\text{P}}$ and $\mu^{\text{Q}} = \mu_j^{\text{Q}}$.

Remuneration to flexible loads:

$\mathcal{R}_{\text{flex}} = \sum_j \left(\mu_j^{\text{P}} \left(P_j^{\text{L0}} - P_j^{\text{L}}\right) + \mu_j^{\text{Q}} \left(Q_j^{\text{L0}} - Q_j^{\text{L}}\right)\right)$

The metrics are then defined as:

Revenue for DER owner: $\mathcal{R}_{\text{flex}}$ and $\mathcal{R}_{\text{gen}}$

Cost for consumer: $\mathcal{R}^{\text{x}}_{\text{load}}$

Net revenue: $\mathcal{P} = \mathcal{R}^{\text{x}}_{\text{load}} - \mathcal{R}_{\text{flex}} - \mathcal{R}_{\text{gen}} - \mathcal{C}_{\text{WEM}}$

A detailed comparison of the cost of market operation is shown in Fig. 5-3. Both the Traditional and No Export scenarios result in large profits for the utility, due to the large difference between the retail and wholesale prices of electricity. Both

Retail_M and Retail_B result in a loss for the utility. While these retail compensation structures are currently used in US electricity markets, the high retail purchase rate means the utility is not only overcompensating the DGs, but that under high penetration of these DG resources, this participation model becomes uneconomical. One option is to provide lower purchase rates, however deciding the value of the energy service being provided is challenging. Another option is to enable participation at the wholesale level, but this continues to be challenging for small resources, even through aggregator models. Most notably in Fig. 5-3, all the quantities for the retail market scenario are significantly lower than of the Traditional and No Export case, and only comparable to the Retail_M/Retail_B cases for the cost of electricity from the WEM. Despite serving the same load, the proposed retail market is able to do this at a much lower retail cost: an average of 0.0291 $/kWh, compared to the current utility retail price of 0.114 $/kWh. Rather than simply making a large profit, the proposed DSO is building social equity and redistributing wealth through socialization of the profit. With the retail market, the true value of energy is recovered, which results in lower electricity costs for consumers and lower compensation for DERs, while ensuring power balance and economic efficiency in the market.

The lower revenue stream is not endemic to the proposed retail market model, but rather a reality of the modern electricity grid with negative electricity prices (frequently occurring in US states of California and Texas, and Germany), high renewable curtailment, and unprecedented ramp rates as in the famous California 'duck curve'. This is already manifesting in systems with high renewable penetration at the transmission level: the share prices for the three largest utilities in Germany have dropped by 45% to 66% between 2010 and 2016, and other utilities in Western Europe have similarly lost market value over the past decade [100]. While this may seem concerning for the future of the utility as we know it, it is a reality stemming from the misalignment of incentives present in the current utility business model. The utility model and corresponding rate structures must be redesigned to shift away from a commoditized market with capital expenditures and energy sales as the main revenue stream, and towards performance-based ratemaking (PBR) where utility revenue is

instead based on achieving performance metrics and other non-investment factors. The shift in business model will be revisited when discussing policy implications of the proposed retail market.

5.1.4 Reactive Power Markets

Research in collaboration with Adam Potter as an Undergraduate Researcher at MIT; and Giulio Ferro and Michela Robba at the University of Genoa

The topic of reactive power compensation was visited in Chapter 4 in the design of a hierarchical voltage regulation service which coordinated distributed DERs and utility-owned devices. The topic of voltage regulation is tied intimately to grid stability and power quality, where reactive power is the key lever with which to maintain them. The paradigm shift in reactive power support from centralized large-scale traditional generators to distributed small-scale consumer owned DERs is illustrated in Fig. 5-4.

The operational change wherein DERs provide reactive power compensation implies that they must be compensated for through an appropriate market mechanism. A market structure that compensates DGs for their services, particularly for reactive power, is structurally different from the current practice [43]. On top of the retail market structure presented earlier, we can add a reactive power market, wherein we consider the deployment of DER resources to provide reactive power, the financial remuneration for the corresponding grid services, and the role of the market in an unbalanced distribution grid with increasing DER penetration [118].

Reactive Power Pricing in the U.S.

Currently, reactive power markets do not exist in the US – neither energy or ancillary markets. In Europe, approaches to ancillary services vary but none are inclusive to DGs despite recent efforts to reduce minimum power bids from 10 MW to 1 MW in several countries [119]. For both regions, and many others, transitioning ancillary service structure to be inclusive of DGs is a proven priority for regulatory agents. In

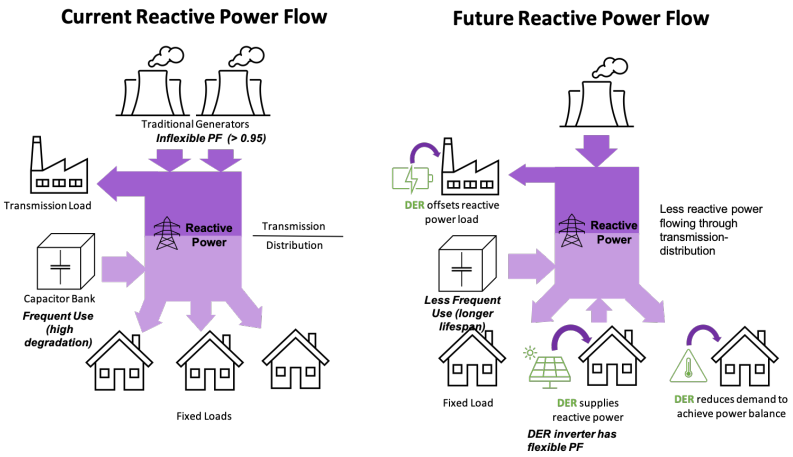


Figure 5-4: Balancing reactive power in an electric grid is necessary to stabilize voltage and ensure power transfer. As grids decentralize, DERs can be utilized to provide reactive power support throughout the distribution system, with more flexible power factors and distributed control units. They can replace traditional reactive power suppliers, which are limited in PF and expensive to operate. However, to provide reactive power DERs must be incentivized to do so.

a common approach to reactive power compensation, and the approach used in the US, any payments to generators that provide reactive power are static rates, designed to recover the capital costs of installing the equipment or based on the LOC (lost opportunity cost) of not selling real power. These payments can consist of capacity, or fixed payments, based on generator's reactive power capabilities; and utilization, or variable, payments based on the delivered kVARs of reactive power. These static rates do not reflect the temporal shifts in supply and demand, or spatial variation across the grid, both of which are pertinent characteristics of DGs. Further, these payments are typically cleared on a case-by-case basis for each generator, a process which is generally exclusive to larger generators. This approach of static case-by-case rates will be insufficient in accommodating a deep penetration of DGs. These DGs offer significant operational flexibility at little to no cost, and can be leveraged to provide much needed grid services through the clean energy transition. To do so, a market mechanism which prices their reactive power capabilities and their spatial and temporal variations is vital. This is possible with approaches based on OPF, which allow the simultaneous determination of prices corresponding to both real and reactive

power injections, enables an increasing penetration of DERs, and accommodates grid physics [63]. One such approach is the market structure proposed here.

Priorities for Emerging Grid

An analysis of the state-of-art for reactive power management and pricing in the US and an overview of both industry practice and technical solutions is presented in [118]. The state-of-art in reactive power management and pricing are insufficient in accommodating a deep penetration of DERs, which is likely to occur in the future grid. In order to leverage the flexibility and low cost of DERs, a new set of requirements and objectives are needed. For the emerging low-carbon grid with high DER penetration, we propose the following set of new priorities for industry practice:

- Utilize the significant reactive power capabilities of a large penetration of DERs and their wider PF operating range
- Support efficient grid operation by minimizing grid losses using local DERs
- Provide fair remuneration to DERs for their grid services
- Incentivize long-term growth of DER adoption through market inclusion and corresponding reliable revenue streams

The priorities of reactive power compensation for the emerging grid as summarized in Table 5.1.

Proposed Reactive Power Pricing

The retail market structure remains the same as in the previous sections, where the market cost is one of social welfare, wherein the total generation cost, customer load disutility, and line losses are minimized. The market prices are the marginal costs of generators/consumers as determined by the Lagrange multipliers in the CI-OPF model. This is the same as before, where we considered the real power d-LMPs.

Each market clearing sets the reactive power dispatch, $Q_{j,t}$ for node j and time t, and produces the corresponding dual variable $\mu^{Q}_{j,t}$. This dual variable sets the basis for

Current Practice (State-of-Art)		Proposed Practice for Emerging Grid (Proposed)	
Methodology for Q pricing	Limitations of state-of-art	Methodology for Q pricing	Features of proposed method
Fixed pricing	Emerging grid is more dynamic: fixed prices cannot provide adequate signals for devices to respond to	Dynamic pricing	Time-varying and spatially varying Q prices provide signal for DERs to respond to system needs
Cost-based pricing	Not inclusive to DERs which don't incur high costs for producing Q	Service-based pricing	Q price is determined by an OPF, thereby modeling the real-time value of Q injection to the grid
Case-by-case clearing	Intractable for a large number of devices (i.e. DER-rich grid)	Coordinated and distributed clearing	Maintain tractability of large-scale OPF by using distributed optimization
Price deadbands where no payment is made (0.95 to 1 PF)	Deadbands are anti-competitive and not recommended for inclusive markets	No price deadbands	Provide compensation through the full range of PFs enabled by power electronic control

Table 5.1: Summary of the state-of-art industry practice on reactive power management and pricing. We propose a reactive power market for the emerging grid, with new methodologies and features.

the time-varying reactive power price $\bar{\mu}_j^Q$, which we denote as the d-LMP, determined for node j as:

$$\bar{\mu}_j^Q = \frac{\sum_{t \in T} \mu_{j,t}^Q Q_{j,t}}{\sum_{t \in T} Q_{j,t}} \tag{5.2}$$

This d-LMP varies daily, calculated as a weighted average of the dual variable from the OPF problem in (B.1). Thus for a 5-minute retail market clearing period, the set T includes 288 points. Such a time-varying price allows DGs to adjust their generation and/or PF settings and DRs to shift their consumption behavior, all in a coordinated manner so that the DSO can accurately recover costs. Rather than exposing end-

use customers and DER owners to the complete dynamics of the electricity system by using $\mu^Q_{j,t}$ as a real-time d-LMP, as wholesale market participation models (like FERC 2222) would do with a corresponding λ^Q, the averaging procedure in (5.2) allows the price volatility to be contained while resulting in the same payout as the corresponding real-time prices.

We note that the reactive power price at individual nodes is affected by the line loss component of the objective function. The first two terms of the objective function represent the generator cost and customer load disutility, capturing the P and Q prices at a subset of nodes, i.e. generators and flexible loads. The third term represents line losses ($\text{Re}\left(I^H Z I\right)$) and captures a global view of the network, introducing Q prices at every node. It is important to note that by virtue of the Ohm's law relationship (modelled directly in the CI-OPF), the minimization of losses in the objective function implicitly minimizes the voltage drop across the whole network.

Market Simulation Results

A numerical case study of the US New England grid modeled as a modified IEEE-123 bus is conducted in [118]. The salient results from the simulation study are presented below. An interested reader can find all details of simulation setup and results in [118]. We demonstrate the following features:

- *Our market can accommodate a large amount of distributed solar:* The simulation results show how reactive power support can be realized even with a high penetration of distributed solar generation. We show in particular that the revenue stream for reactive support to the distributed solar remains steady even as DG penetration increases from 5% to 160%, which indicates the feasibility of the proposed market at high penetrations, and even at low penetrations.

- *Our market uses all available distributed solar efficiently:* We utilize the full flexibility of distributed solar to operate over a range of PF, from 0.6 to 0.95, to meet over 45% of reactive power load, even with the reactive power price per unit kVAR remaining the same. The corresponding increase in the requisite reactive

power injection is appropriately compensated by an increasing percentage of the revenue from reactive power support compared to total revenue: distributed solar participating in reactive power support can enjoy Q-revenue streams of up to 10% of total revenue.

- *Our market supports grid voltages:* The reactive power support from distributed generators spatially distributed throughout the grid enables grid voltages to remain within allowable operating range even at high distributed solar penetration. Further, the reactive power support enhances grid operation by increasing the mean voltage across the network (by up to 2.2%) and providing necessary support during periods of high demand.

- *Our market is stable and provides reliable revenue:* A detailed assessment of the price variations across all distributed solar in the IEEE 123-bus over a week shows that the price variations primarily follow load fluctuations and are otherwise minimal. The limited price volatility and the stable revenue streams under varying distributed solar penetration can instill confidence in the market, through the energy transition. This could encourage additional enrollment into the retail market, and also help drive investment decisions for additional DER adoption.

5.1.5 Policy Implications and Interactions with Policymakers

The development and realization of new market structures is predicated on significant interactions between technologists, researchers, system regulators, utilities, electricity system operators, policy makers, and other interest groups including DER advocates. Within the realm of electricity markets a concerted effort must be made to translate technological contributions to policy and regulatory reform which support emerging market structures. Below is a discussion of potential policy implications and recommendations for realizing a retail market structure. This section is a key contribution to the IEEE Power and Energy Society (PES) Technical report on Behind-the-Meter DERs [154].

The creation of a DSO entity is necessary to support emerging market structures and BTM-DER integration.

The DSO would oversee the task of DER integration and compensation, oversee the operation of retail markets, and coordinate with the regional ISO to address system-level questions. In regions where DSOs currently exist (such as those in Europe) the DSO must evolve to take on more roles and responsibilities beyond managing distribution system wires and towards market oversight, DER integration, and coordination with ISOs [65]. There must be clear guidelines for the functional roles and responsibilities of a DSO to meet the needs of modern distribution systems. Open regulatory questions include:

- Guidelines for participation of DERs in either or both the retail and wholesale markets (such as through aggregators enabled by FERC 2222 in the U.S.), as well as current retail programs (such as net metering for solar or tariff programs for batteries);

- Development of ancillary markets at the distribution level to provide local resiliency and support (such as during storm-related outage events) would require additional regulatory development. Guidelines for DERs participating in both energy and ancillary markets must be developed to support value stacking for DER investors, but prevent 'double dipping' concerns [139]. Suitable interactions with the wholesale ancillary markets must be developed in concert with existing market models (such as U.S. FERC 2222 proceedings);

- Regulatory oversight is needed to determine the types of grid support functions that can be performed by DERs (ex. voltage, reactive power, regulation) and adapted based on the resource type, location, and market participation;

- Coordination between the DSO and ISO is needed for operational and planning decisions. Based on the DSO model describing its functional roles and responsibilities, different coordination activities may be required. Some such activities include: (i) analyzing the reliability impacts of DERs at both the bulk and

distribution level, and the associated data collection and cost allocation of such studies; (ii) capacity procurement, minimum system load, and resource adequacy planning; (iii) interconnection procedures and managing interconnection queues for distribution-level devices which may participate at the wholesale market level; (iv) informing grid protection standards and grid-connected inverter standards including standards for voltage ride-through, reconnection behaviour and system restart with inverter-based DERs, response to frequency disturbances, and low system load behaviour [13].

The timelines for market reform must align with grid modernization efforts and investment timelines.

Significant investments must be made to upgrade distribution grid infrastructure to support resilient, secure, and flexible operations with high BTM-DER penetrations. Emerging market structures which enable DERs to provide flexibility services also require numerous upgrades to energy management systems (EMS), adoption of distributed energy resource management systems (DERMS) grid-wide communication capabilities, sensors, telemetry and metering infrastructure, and cyber-secure operating protocols. Although grid modernization and market reform may fall under the purview of different decision making authorities, they are codependent initiatives which require alignment in goals and implementation timeline to fully realize the potential of BTM-DERs.

Establishing standardized communication protocols and data collection requirements is critical.

Emerging electricity markets require hardware, software, and situational awareness on a real-time basis to support market transactions, send and receive dispatch signals, and to coordinate real-time market participation. There remain many open questions around the communication requirements for DERs participating in existing and emerging market structures, and the ability for DSO-like entities to receive and respond to dispatch signals.

- Communication: The IEEE 2030.5 Smart Energy Profile (SEP) 2.0 standard [9] provides a framework for monitoring and control of DER assets, and has been suggested as the standardized communication protocol for DER aggregation programs. California has adopted the IEEE 2030.5 standard within its Rule 21 protocol, which includes communication protocols for smart inverters. While this is a start, the limited scope of adoption of the standard remains a concern.

- Data collection: The increasing penetration of BTM-DERs requires a significant data collection and modelling effort to support DER visibility at a high-resolution [8]. The data is needed to support integration and reliability studies which inform operating and interconnection standards. These in turn impact the grid services that DERs can provide and the corresponding market derivatives to acquire, price, and compensate DERs for their services.

A concerted effort to promote the design of appropriate performance-based ratemaking (PBR) schemes and widespread adoption must be a priority.

Performance-based ratemaking (PBR) has emerged as the future of the utility model and corresponding rate structures: shifting away from a commoditized market with capital expenditures and energy sales as the main revenue stream, and towards PBR where utility revenue is instead based on achieving performance metrics and other non-investment factors. Interestingly, PBR is not a new concept. A report from the U.S. Lawrence Berkeley National Laboratory highlighted the underpinning concept of PBR in a 1995 study: "Current regulation, which is typically a form of cost-of-service, rate-of-return (COS/ROR) regulation, does not reward utilities for exemplary performance and can be complex and costly to conduct for a utility that provides a mix of monopoly and competitive services. [...] PBR can provide utilities with a greater incentive to make productivity-improving actions, and greater ability to price flexibly and reduce regulatory costs." [30]

The re-emergence of PBR in recent years comes from the growth of BTM-DERs and the need for coordinated flexibility services to operate a 100% carbon-free grid.

With emerging market structures – including aggregator participation at the wholesale level and retail markets at the local level – the compensation mechanisms for resources, in particular DERs like generators and storage, will become increasingly complex. The development and adoption of PBR can help realign utility investments and revenue streams with clean energy goals, supporting utility investment into DER integration and non-wires alternatives, and support efficient grid utilization.

Performance-based ratemaking can help realign utility revenue streams with state RPS and energy goals, support investment into non-wire alternatives (NWA) such as leveraging DR flexibility and storage integration, and more efficient grid utilization [65, 87, 30]. The discussions around PBR bring to the forefront the need to measure and enforce, through profitability, goals for customer service, reliability, storm response, compliance with environmental policy goals and initiatives, and interconnection of DERs in the distribution system. While PBR has been adopted in the U.S. by the state of Hawaii [10], and is being investigated in states such as New York, Massachusetts, and Maine, the slow adoption of PBR would delay the implementation of more complex (and necessary) DER compensation mechanisms and structured retail markets.

The future design of electricity markets must reflect all the priorities of equitable decarbonization and sustainability.

Investigation into rate equity across socio-economic and demographic classes, and suitable policies to ensure fairness, electricity access, and reliability of service is necessary. Some areas include:

- Electricity rates: The adoption of temporally and locationally varying prices introduces new questions. Concerns about exposing customers to the price volatility of real-time rates can be addressed by employing hybrid time-variable pricing models such as Block-and-Index pricing to enable risk hedging and other solutions. Concerns around fairness of locational prices must be addressed with both technological and regulatory solutions. New rate design must reduce energy burden for communities with disproportionately higher energy burden (such as low

income households) [25];

- Market access: Barriers to market participation must be mitigated, which include costs to install telemetry and metering equipment (including sub-metering and parallel metering), reporting requirements, and participation models. Resources like community solar and community battery installations which democratize access to clean energy DERs must not face undue barriers to market participation;

- Cost allocation and cost recovery: Regulatory standards must be updated to include guidelines for the jurisdiction of cost allocation among the different market participants including, among others, ISO, DSO, utilities, aggregators, and DER owners. In addition to the costs of system planning and studies (already discussed in (1)), the costs of new meters and telemetry equipment, registration costs, and any penalties for underperformance are some areas where cost allocation guidance is needed.

5.1.6 Concluding Remarks

This section proposed a retail market structure using a distributed optimization algorithm capable of solving for the optimal dispatch and real-time d-LMPs, while leveraging grid-edge intelligence and peer-to-peer communication [63, 72]. Such an approach enables market participation for small scale resources, allows them to operate more dynamically, and eliminates the no-export rules for behind-the-meter DERs, so they can generate, reduce load, or even increase load as needed by the network. As DER penetration continues to increase, technology costs reduce, and subsidies for these resources are removed, new incentives for DER participation in markets is required. This can be achieved through new revenue streams from retail markets. Furthermore, our retail market results in customer cost savings from such a market, through a decrease in electricity prices from 0.114 \$/kWh down to 0.0291 \$/kWh [65].

The retail market also includes a market derivative for reactive power, wherein DERs providing reactive power support (say for voltage regulation purposes) are

compensated for their injections [118]. The reactive power market incentivizes DERs to enroll and participate in reactive power support by offering a reliable additional source of income through daily reactive power distributed locational marginal prices (Q d-LMPs). The results from our numerical study show that this revenue stream remains dependable even as DG penetration rapidly grows. By building confidence with stable prices, additional DER enrollment may perhaps by encouraged into the retail market and help drive investment decisions for additional DER adoption. Our results appear promising in this regard and permit better DER integration and utilization to meet distribution grid objectives, and can enable deeper DER penetration.

5.2 Accelerated Distributed Optimization Methods

In Chapter 4 the PAC distributed optimization algorithm was proposed as the communication layer in the physics-aware coordination architecture. The PAC algorithm was used to solve the CI-OPF problem, and applied to the voltage regulation and load ramp mitigation services. However, the PAC algorithm is a first order algorithm which needs many iterations to converge, and each iteration is prone to communication failures such as packet drop or cyber-attacks. To address the slow convergence, accelerated methods have been proposed which leverage momentum terms (i.e. Nesterov acceleration) [49, 66] and second-order information (i.e. Newton's method) [47]. This section provides a second presentation of the PAC algorithm and the decomposition ('atomization') of a global problem into a distributed one. Then two proposed accelerated methods are presented.

5.2.1 Proximal Atomic Coordination Algorithm

Consider a global optimization problem composed of equality and inequality constraints, which may be coupled in time

$$\min_{x} \sum_{i=1}^{S} f_i(x)$$

$$\text{s. t.} \quad Gx = b, \quad Hx \leq d \tag{5.3}$$

where $\sum_{i \in S} f_i(x)$ represents the total objective function. Problem (5.3) can be distributed into $K = \{1, ..., k, ...K\}$ separate coupled optimization problems, denoted as *atoms*. We use a decomposition profile which separates the vector of variables x into two sets: $\mathsf{L} = \{L_j, \quad \forall j \in K\}$ and $\mathsf{O} = \{O_j, \quad \forall j \in K\}$, which represent the partition of decision variables "owned" and "copied" by atom j. The set of total variables (owned and copied) by an atom is denoted as T. The decomposition profile also separates the constraints into sets owned by each atom, as $\mathsf{C} = \{C_j, \quad \forall j \in K\}$. The notion of variables copies are used to satisfy the coupling in constraints and/or objective function. In the context of the CI-OPF problem, the power physics of the grid (Ohm's Law) result in these coupling constraints. Note that the CI does not have coupling introduced by inequality constraints, however the decomposition can be trivially extended to coupled inequality constraints. Using the decomposition profile, we obtain:

$$\begin{aligned} &\min_{a_j} \quad \textstyle\sum_{j \in K} f_j(a_j) \\ \text{subj. to:} \quad &\begin{cases} G_j a_j = b_j, & \text{for all } j \in K \\ H_j a_j \leq d_j, & \text{for all } j \in K \\ B_j a = 0, & \text{for all } j \in K \end{cases} \end{aligned} \tag{5.4}$$

where a_j is atom j's variables (both owned and copied), $f_j(a_j)$ is the atomic objective function, and G_j, b_j, H_j, and d_j represent the submatrix or subvector of G, b, H, and d respectively. Finally, B is in incidence matrix over the owned and copied atomic variables, defined as

$$B_i^m \triangleq \begin{cases} -1, & \text{if } i \text{ is 'owned' and } m \text{ a related 'copy'} \\ 1, & \text{if } m \text{ is 'owned' and } i \text{ a related 'copy'} \\ 0, & \text{otherwise} \end{cases}$$

We then use B_j (B^j) to denote the relevant incoming (out-going) edges of the directed graph for atom-j. To fully parallelize the optimization, we introduce *coordination*

constraints, which must be satisfied for every atom. These require all atomic copied variables in a given jth atom to equal the value of their corresponding owned in ith atom, $i \neq j$:

$$B_j a = 0 \quad \forall j \in K \tag{5.5}$$

A detailed analysis of convergence rate, communication and computational complexity, and privacy are provided in [122].

5.2.2 Common Accelerated Methods: Momentum and Heavy Ball

Optimization algorithms rely on successive iterations which typically utilize information of the gradient of the loss function to guide the algorithm towards the optimal solution (where the gradient is zero).

Basic gradient descent:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \tag{5.6}$$

However, gradient-based methods are very slow to converge. This is well known, and various accelerated methods have been proposed to reduce convergence time and the number of iterations. The distributed ADMM algorithm [24] is a popular optimization method for which several papers [59, 54, 73] have attempted to obtain accelerated convergence in ADMM by considering the Nesterov's [108] and the Heavy Ball methods [117]. Other approaches for accelerated convergence have been examined in [73, 56] by either using second-order methods and adaptive techniques for adjusting the penalty parameter. Other methods such as [99] and [37] imply parallel solver with a matrix-splitting technique and a distributed quasi-Newton method, respectively.

Polyak Heavy-ball method: The approach is presented in [117], as below:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{x-1}) \tag{5.7}$$

which can be rewritten using PAC notation as:

$$x_{k+1} = \hat{x}_k - \alpha \nabla f(\hat{x}_k) \tag{5.8}$$

$$\hat{x}_{k+1} = x_{k+1} + \beta(\hat{x}_k - \hat{x}_{k-1}) \tag{5.9}$$

where we can further consider the primal optimization of PAC:

$$x_{k+1} = argmin \left(\mathcal{L}(x, \mu_k, \nu_k) + \frac{1}{2\rho} \|x - \hat{x}_k\|_2^2 \right) \tag{5.10}$$

$$\hat{x}_{k+1} = x_{k+1} + \beta \left(\hat{x}_k - \hat{x}_{x-1} \right) \tag{5.11}$$

Nesterov acceleration with momentum: The approach is presented in [108], as below:

$$y_{k+1} = x_k - \frac{1}{L} \nabla f(x_k) \tag{5.12}$$

$$x_{k+1} = y_{k+1} + \beta \left(y_{k+1} - y_k \right) \tag{5.13}$$

which can be rewritten using PAC notation as:

$$x_{k+1} = \hat{x}_k - \frac{1}{L} \nabla f(\hat{x}_k) \tag{5.14}$$

$$\hat{x}_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k) \tag{5.15}$$

where we can further consider the primal optimization of PAC:

$$x_{k+1} = argmin \left(\mathcal{L}(x, \mu_k, \nu_k) + \frac{1}{2\rho} \|x - \hat{x}_k\|_2^2 \right) \tag{5.16}$$

$$\hat{x}_{k+1} = x_{k+1} + \beta \left(x_{k+1} - x_k \right) \tag{5.17}$$

Both these methods find the gradient at the accelerated point, but the update for the accelerated primal variable is different. The Polyak heavy-ball method uses the past two accelerated values $(\hat{x}_{k+1}, \hat{x}_k)$, while Nesterov's momentum term uses the current and past result of the primal optimization (x_{k+1}, x_k).

In contrast to the above mentioned papers, the PAC approach presented [122] includes elements that enhance privacy in the exchange of dual variables. In what follows, a Nesterov-accelerated PAC algorithm, NST-PAC, is proposed which increases the algorithm convergence speed and introduces additional privacy preservation features in the primal variables via time-varying parameters. As a result, a combined realization of data-privacy and accelerated convergence is made possible, aspect that is not considered in the most recent literature.

5.2.3 Nesterov-Accelerated PAC Algorithm: NST-PAC

Research in collaboration with Giulio Ferro and Michela Robba at the University of Genoa

In this section, an accelerated variant of the PAC algorithm is presented, called NST-PAC [49, 66]. The NST-PAC includes time-varying gains and Nesterov type acceleration [108]. It is a primal-dual method with ℓ_2 and proximal regularization, Nesterov type acceleration for both primal and dual variables, and privacy-preserving features. We begin by forming the atomic Lagrangian function:

$$\begin{aligned}\mathcal{L}(a, \mu, \nu) &= \sum_{j \in K} \left[f_j(a_j) + \mu_j^T (G_j a_j - b_j) + \nu_j^T B_j a \right] \\ &= \sum_{j \in K} \left[f_j(a_j) + \mu_j^T (G_j a_j - b_j) + \nu^T B^j a_j \right] \\ &\triangleq \sum_{j \in K} \mathcal{L}_j(a_j, \mu_j, \nu). \end{aligned} \tag{5.18}$$

The algorithm is carried out as below:

$$a_j[\tau + 1] = \underset{a_j \in \mathbb{R}^{|T_j|}}{\operatorname{argmin}} \{ \mathcal{L}_j(a_j, \hat{\mu}_j[\tau], \hat{\nu}[\tau]) + \frac{\rho_j \gamma_j}{2} \|G_j a_j - b_j\|_2^2$$

$$+\frac{\rho_j\gamma_j}{2}\left\|B_j a_j\right\|_2^2+\frac{1}{2\rho_j}\left\|a_j-a_j\left[\tau\right]\right\|_2^2\Bigg\}, \tag{5.19}$$

$$\hat{a}_j\left[\tau+1\right]=a_j\left[\tau+1\right]+\alpha_j[\tau+1](a_j\left[\tau+1\right]-a_j\left[\tau\right]) \tag{5.20}$$

$$\mu_j\left[\tau+1\right]=\hat{\mu}_j\left[\tau\right]+\rho_j\gamma_j(G_j\hat{a}_j\left[\tau+1\right]-b_j) \tag{5.21}$$

$$\hat{\mu}_j\left[\tau+1\right]=\mu_j\left[\tau+1\right]+\phi_j[\tau+1](\mu_j\left[\tau+1\right]-\mu_j\left[\tau\right]) \tag{5.22}$$

$$\text{Communicate } \hat{a}_j \text{for all } j\in[K] \text{ with neighbors} \tag{5.23}$$

$$\nu_j\left[\tau+1\right]=\hat{\nu}_j\left[\tau\right]+\rho_j\gamma_j B_j\hat{a}_j\left[\tau+1\right] \tag{5.24}$$

$$\hat{\nu}_j\left[\tau+1\right]=\nu_j\left[\tau+1\right]+\theta_j[\tau+1](\nu_j\left[\tau+1\right]-\nu_j\left[\tau\right]) \tag{5.25}$$

$$\text{Communicate } \hat{\nu}_j \text{ for all } j\in[K] \text{ with neighbors} \tag{5.26}$$

where ρ_j, γ_j are atom-varying over-relaxation and step-size parameters, respectively. The proposed NST-PAC uses $\ell 2$ regularization terms rather than the prox-linear variant in PAC. Further, both primal and dual variables are accelerated using Nesterov type acceleration, to speed up convergence. Further, we extend the privacy-preserving feature of the PAC algorithm to both the primal and dual variables, by using three iteration-varying and atom specific parameters for the accelerated terms, $\alpha_j\left[\tau\right]>\alpha_j^{\min}$, $\phi_j\left[\tau\right]>\phi_j^{\min}$ and $\theta_j\left[\tau\right]>\theta_j^{\min}$. In the original algorithm privacy is kept only for dual variables.

Convergence Properties of NST-PAC

We now present the convergence properties of the NST-PAC in (5.19)-(5.26). A continuous time equivalent of the NST-PAC is first derived (Theorem 1) following which its convergence is established (Theorem 2).

Theorem 1. *The continuous time limit of the accelerated NST-PAC algorithm without time varying parameters is given by the following ODE:*

$$\left(I_{N^T}-\mathcal{V}(\varrho^2\mathrm{A}\Gamma)\right)\ddot{\mathcal{A}}\;+\;\nabla^2 f\left(\mathcal{A}\right)\dot{\mathcal{A}}+\Big[\tilde{G}^T(I_{N^O}-\Phi^{\tilde{G}})^{-1}\Gamma^{\tilde{G}}\tilde{G}$$
$$+B^T(I_{N^O}-\Theta^B)^{-1}\Gamma^B B\Big]\mathcal{A}=0 \tag{5.27}$$

with $\mathcal{V}(X) = \tilde{G}^T X^{\tilde{G}} \tilde{G} + B^T X^B B$*, and where* X *is a matrix of appropriate dimension. The vectorized atom-varying over-relaxation terms are* Γ*, with* $\Gamma^{\tilde{G}} \triangleq \{\gamma_1 I_{N_1^C}, \ldots, \gamma_{|K|} I_{N_{|K|}^C}\}$*; likewise for* A*,* Φ*,* Θ*, and* ϱ *representing the vectorized* α*,* ϕ*,* θ*, and* ρ *parameters respectively.*

Proof. We use the approach of [53]. We take $t \approx \delta k$. Use implicit discretization such that $a_j[k+1] = \mathcal{A}_j(t), \mu_j[k+1] = \mathcal{M}_j(t), \nu_j[k+1] = \mathcal{N}_j(t)$; the same for the hats. The atomic vectors are denoted as $\mathcal{A}(t) = \left[\mathcal{A}_1(t) \ldots \mathcal{A}_{|\mathcal{B}|}(t)\right]^T$, $\mathcal{M}(t) = \left[\mathcal{M}_1(t) \ldots \mathcal{M}_{|\mathcal{B}|}(t)\right]^T$, and $\mathcal{N}(t) = \left[\mathcal{N}_1(t) \ldots \mathcal{N}_{|\mathcal{B}|}(t)\right]^T$. For a general variable x with continuous form $\mathcal{X}(t)$, Taylor's theorem gives us $x[k] = \mathcal{X}(t-\delta) = \mathcal{X}(t) - \delta\dot{\mathcal{X}}(t) + \frac{\delta^2}{2}\ddot{\mathcal{X}}(t)$.

We apply the KKT conditions for first order optimality to the PAC algorithm with Nesterov addition, (5.19)-(5.26). The primal update equation of (5.19) in continuous time is (for brevity we consider $b_j = 0$):

$$\begin{aligned} 0 = {} & \partial_{a_j} \mathcal{L}_j \left(\mathcal{A}_j(t), \mathcal{M}_j(t), \mathcal{N}_j(t)\right) + \left(I_{N^T} - \rho_j^2 \mathcal{V}(\gamma_j \alpha_j)\right) \dot{\mathcal{A}}_j(t) \\ & + \frac{\rho_j^3}{2} \mathcal{V}(\gamma_j \alpha_j) \ddot{A}_j(t) - \rho_j \mathcal{V}(\gamma_j) \mathcal{A}_j(t) \\ & + \rho_j \gamma_j \left(\tilde{G}_j \mathcal{A}_j(t) \tilde{G}_j^T + B_j \mathcal{A}_j B_j^T\right) \end{aligned} \tag{5.28}$$

where we have applied Taylor's theorem to the predictor term:

$$0 = \hat{\mathcal{A}}_j(t) - \mathcal{A}_j(t) - \delta\alpha_j \dot{\mathcal{A}}_j(t) + \frac{\delta^2}{2}\alpha_j \ddot{\mathcal{A}}_j(t)$$

Applying Taylor's theorem to the dual update equations:

$$\begin{aligned} 0 &= \mathcal{M}_j(t) - \left[\hat{\mathcal{M}}_j(t) - \delta\dot{\hat{\mathcal{M}}}_j(t) + \frac{\delta^2}{2}\ddot{\hat{\mathcal{M}}}_j(t)\right] - \rho_j \gamma_j \tilde{G}_j \mathcal{A}_j(t) \\ 0 &= \mathcal{N}_j(t) - \left[\hat{\mathcal{N}}_j(t) - \delta\dot{\hat{\mathcal{N}}}_j(t) + \frac{\delta^2}{2}\ddot{\hat{\mathcal{N}}}_j(t)\right] - \rho_j \gamma_j B_j \mathcal{A}(t) \end{aligned}$$

Applying Taylor's theorem to the dual Nesterov equations:

$$\hat{\mathcal{M}}_j(t) = \mathcal{M}_j(t) + \delta\phi_j \dot{\mathcal{M}}_j(t) - \frac{\delta^2}{2}\phi_j \ddot{\mathcal{M}}_j(t)$$

$$\hat{\mathcal{N}}_j(t) = \mathcal{N}_j(t) + \ \delta\theta_j\dot{\mathcal{N}}_j(t) - \frac{\delta^2}{2}\theta_j\ddot{\mathcal{N}}_j(t)$$

Solving the system of equations for the dual variables with $\rho_j = \delta$ and the continuous time limit $\delta \to 0$:

$$\dot{\mathcal{M}}(t) = (I_{N^O} - \Phi)^{-1}\Gamma^{\tilde{G}}\tilde{G}\mathcal{A}(t) \tag{5.29}$$

$$\dot{\mathcal{N}}(t) = (I_{N^O} - \Theta)^{-1}\Gamma^{B}B\mathcal{A}(t) \tag{5.30}$$

We then take the time derivative of (5.28), substitute with (5.29)-(5.30), and discard higher over derivatives, to recover the ODE for NST-PAC without time varying parameters. □

The following assumptions are needed to prove the asymptotic stability of the optimal solution.

Assumption 1. *There exists a non-trivial optimal solution to problem* (5.3)*,* $x^* \in \mathbb{R}^N$*. The optimal solution to the distributed problem* (5.4) *is* $a^* \in \mathbb{R}^{|\mathsf{T}|}$*, and is related to* $y^* \in \mathbb{R}^N$ *via:* $y^* = \Pi^{\mathsf{L}}a^*$

Assumption 2. *Let the PAC parameters satisfy:*

$$1 > \max_j\{\rho_j^2\alpha_j\gamma_j\}\lambda_{\max}(\mathcal{V})$$

Assumption 3. *Let* $0 \leq \phi_j < 1$ *and* $0 \leq \theta_j < 1, \quad \forall j$

Assumption 4. *Each objective summand* $f_j\ \forall j \in K$ *is differentiable and* χ*-strongly convex.*

Theorem 2. *Let* $\mathcal{A}^*$ *be the optimal solution of problem* (5.4)*, then* $\mathcal{A}^*$ *is an asymptotically stable critical point of the NST-PAC algorithm.*

Proof. For the sake of simplicity, we take $\theta = \phi$. Since $\mathcal{A}^*$ is the optimal solution, then $\mathcal{V}(\varrho^2\mathrm{A}\Gamma)\mathcal{A}^* = 0$ (which holds for all feasible points, see Theorem 4.7 in [122]).

Define a Lyapunov candidate function

$$V = \int_0^{\dot{\mathcal{A}}} \left(I_{N^T} - \mathcal{V}(\varrho^2 \mathrm{A}\Gamma)\right) y dy + \int_0^{\mathcal{A}} (I_{N^O} - \Theta)^{-1} \mathcal{V}(\Gamma) y dy \tag{5.31}$$

whose derivative is

$$\dot{V} = -\dot{\mathcal{A}}^T \left[\nabla^2 f\left(\mathcal{A}\right)\right] \dot{\mathcal{A}} \tag{5.32}$$

We have $V \succ 0$ and $\dot{V} \prec 0$ guaranteed by Assumptions 2 and 4. To retain privacy and asymptotic stability, each atom j needs to choose its own private lower bounds for α_j, ϕ_j, and θ_j. □

Remark: It is important to note that (5.27) represents a mass-spring-damper system. The accelerated nature of the algorithm compared to PAC is evident from the decrease in the inertia term $(I_{N^T} - \mathcal{V}(\varrho^2 \mathrm{A}\Gamma))$ and by noting that $\alpha_j = 0 \ \forall j$ in the PAC approach in [122].

Simulation Results

We now evaluate the performance of the proposed architecture from the point of view of the distributed optimization approach, NST-PAC. We benchmark this performance through comparison with other popular methods such as dADMM [91], and PAC [122]

The algorithms were applied to the problem of energy management in a network of interconnected energy hubs (discussed in more detail in Section 5.3. The overall idea of the energy management is similar to that of OPF, but extends beyond just the electric grid. Instead a network of interconnected energy hubs including buildings, polygeneration microgrid, and electric vehicles are coordinated with the electric grid and district heating network. The energy management problem is supervised by an Energy Community Manager (ECM). The size of the underlying optimization problem then depends on the size of the electric grid and district heating network. Two test cases were simulated: (1) the IEEE-13 bus electric grid and a 9-node district heating network; and (2) the IEEE-123 bus electric grid and a 30-node district heating network.

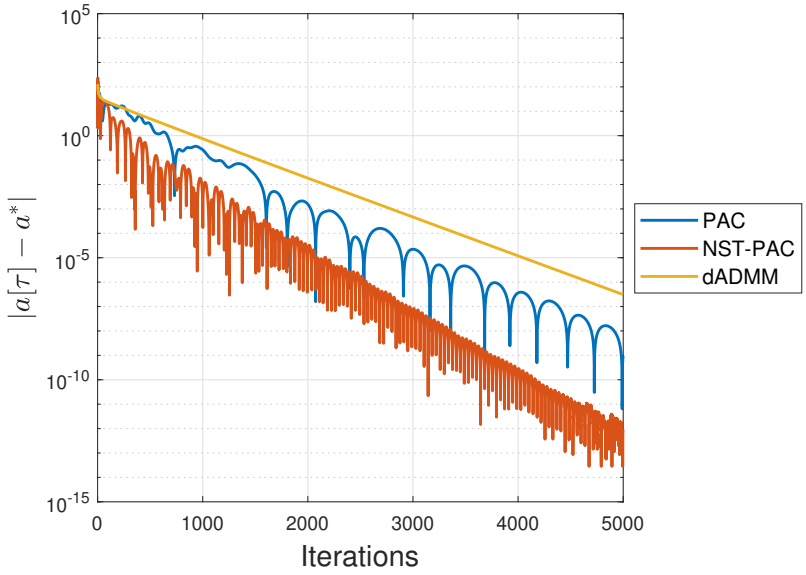


Figure 5-5: Comparison between PAC, NST-PAC, and dADMM

Figure 5-5 illustrates the speed of convergence of the three algorithms using the metric, distance to the optimal solution $|a[\tau] - a^*|_2$, for case (1). It was observed that over 5000 iterations, the NST-PAC metric is 10^{-10}, while dADMM and PAC resulted in the values 10^{-6} and 10^{-7}. This shows the accelerated nature of the NST-PAC. The added advantage of the NST-PAC is its enhanced privacy-preservation, as it fully keeps the primal and dual variables private, while PAC ensures the privacy of only the dual variables, and dADMM has neither of these privacy properties. The scalability of the proposed method was evaluated by testing on case (2), wherein the algorithm converged in 15000 iterations, taking an overall run time of 89.36 seconds.

5.2.4 A Second Order Dual Update Approach to Distributed Optimization

Research in collaboration with Giulio Ferro, Michela Robba, and Federico Delfino at the University of Genoa

While simulation results in the previous section showed the improved performance of NST-PAC as compared to PAC and ADMM, the convergence still remains slow. To this end, second-order update methods are proposed to enable fast convergence of

distributed algorithms [47].

Recent literature has applied second-order methods to distributed optimization, typically with a focus on the primal problem. In [133], a distributed Newton method for global consensus optimization problems is stated and compared with state-of-the-art algorithms like ADMM. A similar approach is given in [141] where a distributed reinforced Newton method for unconstrained consensus problems is presented. It is well known that depending on the shape of the cost function (especially for machine learning problems), the use of second-order information about the cost function, i.e., the Hessian, can be computationally expensive and prohibits application of second-order methods. Recent works have tried to address this by proposing approximate Newton steps [149], or approximating the inverse Hessian using matrix splitting techniques [152].

With respect to the papers above, the AL-SODU algorithm proposed below is suited for constrained optimization problems, and the second-order update is performed on the dual variable update, which is responsible for the slow convergence of algorithms based on the augmented Lagrangian (see PAC [122], ADMM [53], and NST-PAC [49]). It must be noted that in general optimization problems, the direction of steepest descent is represented in the primal update – thus, when the primal update step can be explicitly resolved (as is the case of quadratic problems considered here) the use of second-order methods for the primal update is less effective in accelerating algorithm performance. For this reason, we focus on the dual update step which follows a dual ascent approach.

Construction of Second Order Dual Update

The use of a Second Order dual Update (SODU) allows for faster convergence to the optimal solution, enabling the application of distribution optimization to large-scale problems such as those encountered in energy systems. To introduce SODU, we first consider a general convex optimization problem of the form:

$$\min_{x} \quad f(x)$$

$$s.t \;\; Ax = b \tag{5.33}$$

where A and b are the constraint matrix and vector, and $f(x)$ is a general convex objective function. The corresponding Lagrangian function of (5.33) is:

$$\mathcal{L}(x, \nu) = f(x) + \nu^T (Ax - b) \tag{5.34}$$

where the constraint has been dualized with dual variable ν. The resulting Lagrangian function can be solved as an unconstrained optimization problem with various iterative primal-dual and dual-ascent algorithms. Current state of the art optimization algorithms generally carry out the dual update as a gradient step of the dual function as:

$$\nu[\tau + 1] = \nu[\tau] + \rho(Ax[\tau + 1] - b) \tag{5.35}$$

where ρ is the step size.

We propose an alternative dual update which employs Newton's method to solve the Lagrangian system in (5.34) [22]:

$$\begin{aligned} \nu[\tau + 1] = \nu[\tau] + \left[A^T (\nabla_{xx} f(x[\tau + 1]))^{-1} A\right]^{-1} \\ \left[-A^T (\nabla_{xx} f(x[\tau + 1]))^{-1} \nabla_x \mathcal{L}(x[\tau + 1], \nu[\tau]) \right. \\ \left. + Ax[\tau + 1] - b\right] \end{aligned} \tag{5.36}$$

To derive the dual update in (5.36) we consider the nonlinear system of equations in x and ν:

$$\nabla f(x) + \nabla \nu^T (Ax - b) = \nabla \mathcal{L}(x, \nu) = 0 \tag{5.37}$$

We then apply Newton's method to solve (5.37):

$$\nabla^2 \mathcal{L}(x, \nu) = -\nabla \mathcal{L}(x, \nu) \tag{5.38}$$

Therefore we obtain

$$\nabla_{xx} f(x)\,\Delta x + A\Delta\nu = -\nabla_x \mathcal{L}(x,\nu) \tag{5.39}$$

$$A^T \Delta x = -Ax + b \tag{5.40}$$

By multiplying (5.39) by $A(\nabla_{xx} f(x))^{-1}$, and using (5.40) we have:

$$\begin{aligned} &-Ax + b + A^T(\nabla_{xx} f(x))^{-1} A\Delta\nu \\ &= -A^T(\nabla_{xx} f(x))^{-1}\nabla_x \mathcal{L}(x,\nu) \end{aligned} \tag{5.41}$$

Finally, since the dual update is given by

$$\nu[\tau+1] = \nu[\tau] + \Delta\nu \tag{5.42}$$

by solving (5.41) with respect to $\Delta\nu$ we obtain (5.36).

Statement of the Algorithm

This section will state the AL-SODU algorithm which can be used to solve convex optimization problems of form of (5.2.4). Note the use of positive slack variables z converts inequality constraints to equality constraints. The resulting optimization problem will be of the form:

$$\begin{aligned} &\min_{y,z} \quad f(y) \\ s.t \quad &\begin{bmatrix} \hat{A} & I \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = b \\ &z \geq 0 \end{aligned} \tag{5.43}$$

where y is the vector of decision variables and $\hat{A}$ is the constraint matrix of the original optimization problem (5.3) composed of both G and H, equality and inequality constraints.

The AL-SODU algorithm is a dual ascent algorithm which employs a second order dual update (SODU). We begin by forming the augmented Lagrangian by introducing a proximal augmentation term to ensure strong convexity:

$$\mathcal{L}^A\left(x,\nu,x\left[\tau\right]\right)=\mathcal{L}\left(x,\nu\right)+\frac{1}{2\rho}\left\|x-x\left[\tau\right]\right\|_2^2 \tag{5.44}$$

with

$$\mathcal{L}\left(x,\nu\right)=f\left(x\right)+\nu^T\left(Ax-b\right) \tag{5.45}$$

The AL-SODU algorithm can then be stated as

$$x\left[\tau+1\right]=\underset{x\in\mathcal{X}}{\operatorname{argmin}}\left\{f\left(x\right)+\nu^T\left[\tau\right]\left(Ax-b\right)+\frac{1}{2\rho}\left\|x-x\left[\tau\right]\right\|_2^2\right\}, \tag{5.46}$$

$$\begin{aligned}\nu\left[\tau+1\right]=\nu\left[\tau\right]+&\left[A^T\left(\nabla_{xx}\mathcal{L}^A\left(x\left[\tau+1\right]\right)\right)^{-1}A\right]^{-1}\\&\left[-A^T\left(\nabla_{xx}f\left(x\left[\tau+1\right]\right)\right)^{-1}\nabla_x\mathcal{L}^A\left(x\left[\tau+1\right],\nu\left[\tau\right]\right)\right.\\&\left.+Ax\left[\tau+1\right]-b\right]\end{aligned} \tag{5.47}$$

where $x=\begin{bmatrix}y & z\end{bmatrix}^T$ is the whole decision vector including slack variables and $A=\begin{bmatrix}\hat{A} & I\end{bmatrix}$. The positivity constraint in (5.2.4) is subsumed in the primal update of (5.46), where the slack primal variables $x\left[\tau+1\right]$ belongs to the set $\mathcal{X}$, where $\mathcal{X}$ is the positive orthant $\mathbb{R}^+$. The contribution in this algorithm is the dual update in (5.36) which uses second order information from the Hessian preforming a Newton-like iteration, rather than the classical gradient iteration of the dual function [22]. The use of SODU accelerates the convergence of the algorithm, as presented next.

Comparison with Other Algorithms

In order to show the effectiveness of the AL-SODU algorithm we compare it against three possible other variants of the algorithm:

- AL-FODU: which replaces Eq. (5.47) with a first order dual update as in classical AL algorithms [22];
- AL-HB: which includes a Polyak's Heavy ball like predictor corrector step for

the primal and dual update of the AL-FODU algorithm as in [55]

- AL-NST: which includes a Nesterov like predictor corrector step for the AL-FODU algorithm as in [59]

All the algorithms' specifications are reported in detail below.

AL-FODU algorithm: This algorithm employs a first order dual update for dual variables, like most primal-dual algorithms. The algorithm steps are given by:

$$\hat{x}_i[\tau+1] = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\hat{x}_i{}^T Q_i \hat{x}_i + q^T \hat{x}_i + \left[\nu^T[\tau](Ax-b)\right]_i + \frac{1}{2\rho}\|\hat{x}_i - x_i[\tau]\|_2^2 \right\} \quad i \in \mathcal{A} \tag{5.48}$$

$$\text{Communicate } \hat{x}_i[\tau+1] \quad \forall i \in \mathcal{A} \text{ to the Coordinating agent} \tag{5.49}$$

$$x[\tau+1] = \underset{x\in\mathcal{X}}{\operatorname{argmin}} \left\{ \|x - \hat{x}[\tau+1]\|_2^2 \right\} \tag{5.50}$$

$$\nu[\tau+1] = \nu[\tau] + \rho(Ax[\tau+1] - b) \tag{5.51}$$

$$\text{Communicate } \nu[\tau+1] \quad \text{to} \quad \mathcal{A} \tag{5.52}$$

AL-HB algorithm: This algorithm employs an accelerated primal dual variables update based on Polyak's heavy ball method. This acceleration exploits the value of the primal/dual variables at the previous iteration to perform a predictor-corrector update (with ϕ as fixed parameter). The algorithm steps are given by:

$$\hat{x}_i[\tau+1] = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\hat{x}_i{}^T Q_i \hat{x}_i + q^T \hat{x}_i + \left[\nu^T[\tau](Ax-b)\right]_i + \frac{1}{2\rho}\|\hat{x}_i - x_i[\tau]\|_2^2 \right\} \quad i \in \mathcal{A} \tag{5.53}$$

$$\text{Communicate } \hat{x}_i[\tau+1] \quad \forall i \in \mathcal{A} \text{ to the Coordinating agent} \tag{5.54}$$

$$\overline{x}[\tau+1] = \underset{\overline{x}\in\mathcal{X}}{\operatorname{argmin}} \left\{ \|\overline{x} - \hat{x}[\tau+1]\|_2^2 \right\} \tag{5.55}$$

$$x[\tau+1] = \overline{x}[\tau+1] + \phi(\overline{x}[\tau+1] - \overline{x}[\tau]) \tag{5.56}$$

$$\overline{\nu}[\tau+1] = \nu[\tau] + \rho(Ax[\tau+1] - b) \tag{5.57}$$

$$\nu[\tau+1] = \overline{\nu}[\tau+1] + \phi(\overline{\nu}[\tau+1] - \overline{\nu}[\tau]) \tag{5.58}$$

$$\text{Communicate } \nu\left[\tau+1\right] \quad \text{to} \quad \mathcal{A} \tag{5.59}$$

AL-NST algorithm: This algorithm employs an accelerated primal dual variables update based on Nesterov's method (with $\phi\left[\tau\right]$ as an iteration varying parameter). The algorithm steps are given by:

$$\begin{aligned}\hat{x}_i\left[\tau+1\right] = \ & \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2}\hat{x}_i{}^T Q_i \hat{x}_i + q^T \hat{x}_i + \left[\nu^T\left[\tau\right]\left(Ax-b\right)\right]_i \right. \\ & \left. + \frac{1}{2\rho}\left\|\hat{x}_i - x_i\left[\tau\right]\right\|_2^2 \right\} \quad i \in \mathcal{A}\end{aligned} \tag{5.60}$$

$$\text{Communicate } \hat{x}_i\left[\tau+1\right] \quad \forall i \in \mathcal{A} \text{ to the Coordinating agent} \tag{5.61}$$

$$\phi\left[\tau+1\right] = \frac{\tau}{\tau+r} \tag{5.62}$$

$$\overline{x}\left[\tau+1\right] = \underset{\overline{x}\in\mathcal{X}}{\operatorname{argmin}} \left\{ \left\|\overline{x} - \hat{x}\left[\tau+1\right]\right\|_2^2 \right\} \tag{5.63}$$

$$x\left[\tau+1\right] = \overline{x}\left[\tau+1\right] + \phi\left(\overline{x}\left[\tau+1\right] - \overline{x}\left[\tau\right]\right) \tag{5.64}$$

$$\overline{\nu}\left[\tau+1\right] = \nu\left[\tau\right] + \rho\left(Ax\left[\tau+1\right] - b\right) \tag{5.65}$$

$$\nu\left[\tau+1\right] = \overline{\nu}\left[\tau+1\right] + \phi\left(\overline{\nu}\left[\tau+1\right] - \overline{\nu}\left[\tau\right]\right) \tag{5.66}$$

$$\text{Communicate } \nu\left[\tau+1\right] \quad \text{to} \quad \mathcal{A} \tag{5.67}$$

with $r \geq 3$ chosen a priori.

Simulation Results

The AL-SODU is applied to the the real-time scheduling problem for the energy management of interconnected energy hubs (as presented in the previous Section, and discussed in more detail in Section 5.3). Recall that the energy management problem is supervised by an Energy Community Manager (ECM), which is denoted as the 'Coordinating Agent' in the above algorithms.

The convergence comparison is given in Figure 5-6 in which for each algorithm is plotted the precision in terms of the optimality gap [%] the percentage distance between the algorithm solution and the solution obtained by the IPOPT solver [135] using the YALMIP interface [88]. The plot shows that the AL-SODU algorithm com-

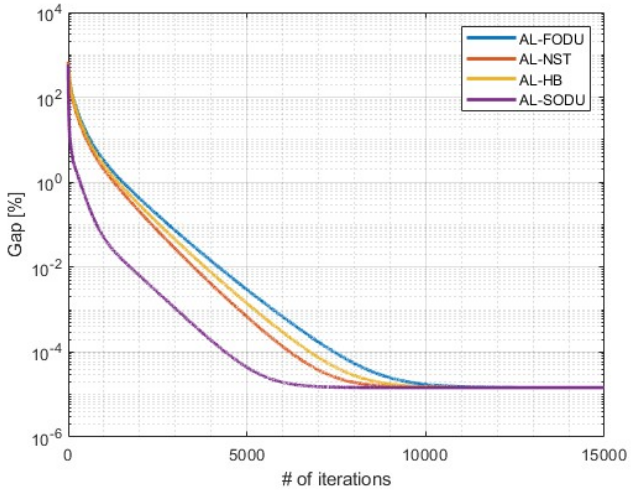


Figure 5-6: Comparison between: AL-FODU, AL-HB, AL-NST and AL-SODU algorithms.

pared to the other algorithms, takes significantly fewer iterations to achieve the same precision (namely 7000 iterations for the AL-SODU versus 10000, a 30% reduction in number of iterations). Thus the proposed AL-SODU is able to effectively combine the benefits of distributed computations with the speedup afforded by second order methods.

5.2.5 Concluding Remarks

In this section two accelerated distributed optimization algorithms were presented. The first algorithm, NST-PAC, introduces Nesterov momentum term to the PAC algorithm to provide accelerated convergence. The convergence is established for the continuous time equivalent of the NST-PAC algorithm. The second algorithm, AL-SODU, is based on a second order dual update approach and can improve the performances of current first order algorithms that are widely used in literature. The proposed AL-SODU algorithm is a dual descent algorithm using a augmented Lagrangian primal step, and second order Newton method update for the dual step. Simulation results shown that AL-SODU considerably outperforms other distributed algorithms, with 30% fewer iterations as compared to state-of-art first order and accelerated dual updates (including heavy ball and Nesterov).

5.3 Managing Interconnected Energy Hubs

Research in collaboration with Giulio Ferro and Michela Robba, at the University of Genoa

The emergence of distributed energy resources at the edge and closer to the consumer has implications on the energy sector, not only at the grid level, but at a city-level. A novel concept that has been recently been introduced in the context of smart energy management in cities is that of an Energy Community (EC). An EC denotes a set of residential or small commercial agents, each acting as a prosumer, and includes generation (electric and thermal), storage units such as batteries, and flexible loads [102]. While the EC concept is still emerging as a theoretically proposed framework (much like the smart grid concept) the EC may become more prevalent with increased electrification and smart city efforts. Notably, ECs extend beyond just the electric grid and considers energy systems in a broader sense, by including thermal energy as well. This may include the heating of homes and buildings or water heating. Further, the framework of EC is broadly applicable to interconnected networks: while smart cities are a nice example, the EC may be more applicable for facilities like water sanitation and distribution which presents a significant electric load on the power grid and can provide some flexibility in operations, but must satisfy constraints of the water network as well.

In this section, we consider an EC which consists of different Energy Hubs (EHs), including microgrids, smart buildings, and charging parks in a smart city (see Figure 5-7 for a schematic) and an ECM managing all EHs. We propose the EC is supervised by an Energy Community Manager (ECM), which oversees a peer-to-peer (P2P) market structure and the energy exchanges among the various EHs [49]. In particular, we will consider how the ECM will ensure a fair sharing of costs and benefits across the entire EC, while minimizing the requisite objective function of each hub and ensuring that all underlying constraints are met. As each of the EHs is a complex cyber-physical system in itself, what is needed is an optimization approach that ensures data privacy and security while preserving plug and play modularity, i.e., no system

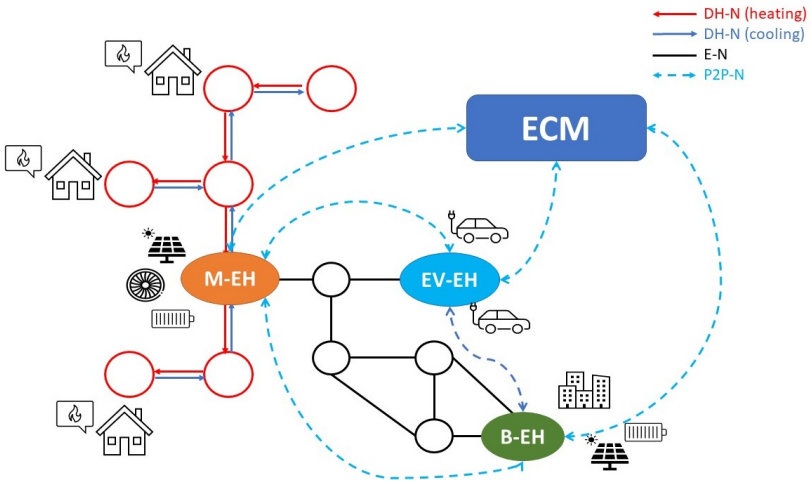


Figure 5-7: The overall systems architecture with an ECM managing an EC consisting of EHs including a microgrid (M-EH), buildings (B-EH), and electric vehicle carport (EV-EH). An electric network (E-N) interconnects the first two energy hubs while a district heating network (DH-N) interconnects the last two. Overall, the energy transactions between the EHs is managed using a P2P network, and the ECM provides overall oversight using a global optimization problem that is solved using a distributed approach.

reconfiguration is needed when a player leaves or is added in the network, and optimal performance of the entire P2P network.

ECM Architecture

The detailed models of each component of the EC is presented in [49]. A summary of the salient features of the EC is presented here.

Building Energy Hub, B-EH: The B-EH is assumed to consist of electrical storage systems, deferrable loads renewable generation, and HVAC units in buildings with multiple interconnected rooms. The goal of the B-EH is to minimize the objective function characterized by: a) economic costs of purchasing/selling electrical energy among EHs interconnected through an electrical network (E-N, where prices are chosen through bilateral agreements between market participants); and b) ensuring certain comfort levels. The optimization constraints include the power exchange between the B-EH and the E-N, constraints on loads, temperature, and storage state-of-charge levels.

Polygeneration Microgrid Energy Hub (M-EH): The M-EHs are characterized by a storage system, CHP units, boilers, biomass-based plants, etc., electrical and thermal demands, renewables, and shiftable loads [33]. The goal of the M-EH is similar to the B-EH to minimize: a) the economic cost of electricity exchange; and b) ensuring comfort levels. In addition, the microgrid will also minimize the cost of gas used in combined heat and power (CHP) facility and boilers. The optimization constraints include the similar electric power constraints as in B-EH, with the addition of local thermal balance constraints.

Electric Vehicles-Energy Hub (EV-EH): The EV-EH consists of several charging stations, which may buy power directly from the grid or from other EHs and sell power to other EHs. The goal of the EV-EH is to minimize the electric power purchased from other hubs (to minimize cost) and a penalty term to ensure vehicles charge by their specified deadlines. The optimization constraints include the electric power balance and state-of-charge management.

Energy Community Manager (ECM): The objective of the ECM is to minimize operating costs across all EHs subject to their respective constraints, while accommodating the underlying physical constraints of both the E-N and DH-N. The E-N is modelled using the CI power flow constraints. The DH-N is modelled as a directed graph with nodes corresponding to heat sources, consumers, or connection nodes, and the links representing pipes, using well-known models [114, 116]. All energy hubs are assumed to be prosumer nodes in the E-N, and interconnected through the DH-N based on thermal load and supply.

Solution methodology: The ECM optimization problem is solved in a distributed fashion, with each EH minimizing its own cost subject to the local constraints. A P2P market is constructed to enable energy sharing and price determination. All of the EHs are assumed to participate in the proposed P2P market network, with the information flow (possibility of trading) occurring between neighbouring EHs (as determined by the E-H and DH-N networks). The market dispatch is solved in a distributed fashion, using the proposed accelerated NST-PAC algorithm.

Simulation Results

Two test cases are simulated: (1) the IEEE-13 bus electric grid and a 9-node district heating network; and (2) the IEEE-123 bus electric grid and a 30-node district heating network. The salient results from the simulation study are presented below. An interested reader can find all details of simulation setup and results in [49]. Three different scenarios have been considered:

- Scenario 1. This corresponds to the current state of the art in practice. All three EHs are present (minimizing their own objective function), but no EC, ECM, or energy exchange between the energy hubs is assumed to be present. Any interconnection between EHs and the underlying E-N and DN-H are not modelled.

- Scenario 2. This corresponds to the inclusion of E-N and DH-N, similar to recent research approaches proposed in [103, 136], but includes a DH-N. No ECM or a systems-wide objective is however included to supervise the operation of the EHs.

- Scenario 3. The overall systems architecture proposed in this section, with the ECM overseeing the network-wide optimization problem and P2P market.

Table 5.2 shows the resulting costs for each scenario across each agent in the EC, for case (1) and (2). The first two rows correspond to the overall costs of the ECM, which are followed by costs corresponding to the B-EH, M-EH, and EV-EH in the simulation. Lower values are better for each cost metric. The table also shows the percentage reduction in cost using our method (Scenario 3) compared to the other two scenarios, in the last two columns, denoted as $\Delta 1$ and $\Delta 2$. The table shows the advantage of our method, as shown in the first two rows, especially the last two columns. We note that our proposed method results in a 10% reduction in C^P and a 69% reduction in C^C compared to the current practice. In comparison, the advantage from Scenario 2, the nearest approach reported in the literature that has no ECM, results in a corresponding reductions of 2% and 10%. This illustrates the

Table 5.2: Cost results in the three Scenarios

	Scenario 1	**Scenario 2**	**Scenario 3**	$\Delta 1$	$\Delta 2$
Test case 1 (IEEE-13 bus E-N, 9 node DH-N)					
Cost of electricity purchased from external grid, C^P	2219.37	2180.21	2007.88	-9.7	-1.76
Tracking comfort levels in the DH-N, C^C	1100.81	665.41	346.75	-68.49	-39.54
Cost of B-EH Number 1	1640.07	1560.4	1664.27	1.47	-4.87
Cost of B-EH Number 2	1353.95	1353.38	1387.89	2.50	-0.042
Cost of M-EH Number 1	1212.2	1286.01	1396.55	15.20	6.1
Cost of M-EH Number 2	1345.02	1348.26	1370.01	1.85	0.22
Cost of EV-EH Number 1	102.19	104.33	116.87	13.72	1.95
Cost of EV-EH Number 2	122.90	124.68	135.75	10.45	1.63
Test case 2 (IEEE-123 bus E-N, 30 node DH-N)					
Cost of electricity purchased from external grid, C^P	28896.3	27610.4	27257.8	-5.67	-4.45
Tracking comfort levels in the DH-N, C^C	4743.7	3100.1	2228.65	-53.03	-34.65

advantage of the proposed systems architecture involving an ECM. The table also shows the breakdown of the costs experienced by all EHs. A sum of all Δ_1 across all EHs indicates that Scenario 3 results in a net increase of 46% over Scenario 2. This is mainly because MG1 (polygenerative) substantially increases thermal energy production via CHP in order to help sustaining the comfort temperature of the DHN building clusters, as shown in Figure 5, which in turn leads to a 69% reduction in the overall comfort cost in the second row. This demonstrates that sub-optimal performances of the EHs help to increase the overall welfare of the city networks. In contrast, Scenario 2 does not yield such an optimal performance even with the net increase of 10% in energy costs across all hubs. This indicates that just the inclusion of the interconnections in the E-N and DH-N alone is not sufficient to yield a city-wide optimization and that an ECM was a necessary inclusion. The drawback of Scenario 1 is obvious, as it includes none of the realistic network constraints in the EH-operation.

Concluding Remarks

This section proposes a systems architecture for an EC for the optimal management of interconnected EHs. Three classes of EHs have been considered (B-EH, M-EH and EV-EH) within a P2P market that allows the energy exchange among the EHs. The EHs are coupled through an E-N and a DH-N. We introduce in the system architecture

the ECM that ensures the overall optimal management of the P2P market in the EC, including the needs of the EHs, the E-N and DH-N. To solve the overall resulting constrained optimization problem, we propose and use the distributed algorithm NST-PAC which provides fast solution through Nesterov's acceleration term and ensures data privacy between market agents. The overall ECM, EHs, and the E-N and DH-N networks are evaluated using two case studies: an IEEE 13-bus as the E-N and a 9 nodes DH-N and IEEE 123-bus as the E-N and a 30 nodes DH-N . Results show the advantage of the systems architecture compared to the state of the art in practice and recent approaches from literature [49].

Chapter 6

Towards Implementation of Proposed Algorithms

6.1 Situational Awareness

Power systems operators have historically placed a greater emphasis on transmission grid monitoring and real-time data collection, as the critical backbone of the electrical grid. However, this paradigm of information and visibility only at the bulk level will no longer work for emerging systems. The increasing intelligence of the grid's edge – through adoption of customer-owned devices in the distribution grid which can now inject power, as opposed to the passive paradigm of the past – necessitates more situational awareness. This situational awareness can provide distribution grid operators with visibility and real-time information into the distribution grid, to support the control mechanisms which keep the grid operable. Further, with enhanced situational awareness, the distribution system can improve in efficiency, reliability, and resiliency.

6.1.1 Data Acquisition and Monitoring Systems

Modern distribution system operators (including distribution utilities and network operators) need access to real-time data for a range of functions including fault and outage detection, remote asset health monitoring, load management applications, and

improving DER visibility. Currently, SCADA (Supervisory Control and Data Acquisition) systems provide limited communication and control in distribution systems.

Advanced Distribution Management Systems (ADMS): These software platforms integrate multiple software applications including SCADA, outage management systems, and distribution management. The ADMS may also be integrated with a distributed energy resource management systems (DERMS), or kept alongside a DERMS implementation. These software platforms look to integrate multiple data sources (from SCADA, smart meters, other sensors) with computational and control functionalities to support grid operations, monitoring, and control. The design and implementation of these platforms varies by vendor. DERMS applications may include voltage control, load forecasting, DER situational awareness, and DER fleet management. Pertaining to the proposed algorithms in this thesis, significant effort must be made to integrate data sources with network-wide communication infrastructure. The proposed dynamic reconfiguration framework may require integration of ADMS and DERMS platforms, where the ADMS platform focuses on communication with and monitoring of switches, and the DERMS platform focuses on DER dispatch and control.

Advanced metering infrastructure (AMI): The proliferation of AMI systems aids in grid operations and certain electricity market functions [52]. A network of AMI devices provides multiple benefits, from simply collecting meter data for billing purposes, to providing customers with granular information about their usage patterns, grid operators with outage information, and voltage monitoring to ensure safe operable voltage ranges and assess performance of programs like CVR . The deployment of AMI infrastructure typically include three components [7]:

1. Smart meters installed at customer locations to collect electricity consumption data. The interval of data collection can vary. Most countries in the EU currently support 60-min data granularity. Deployments in the US range from 5-, 15-, 30-, and 60-min intervals.

2. Communications networks which are upgraded to transmit the interval load data

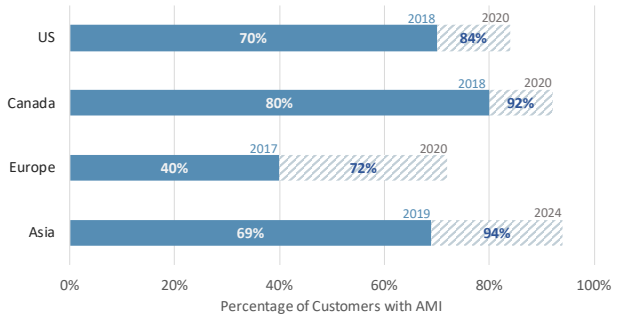


Figure 6-1: Smart meter deployment in different regions of the world. Solid bars indicate the percentage that has already been deployed,. The dashed area represents the forecasted deployment by the end of 2020 or 2024. Data from Asia includes China, Japan, South Korea, and India.

from the network of meters to the utility. The communication technologies used for AMI networks includes both wired and wireless technologies

3. Meter data management system (MDMS) which resides on the utility side to store and process the interval load data. The MDMS can be connected with other information and control systems including billing, customer information systems, locational systems (ex. GIS), outage management systems, and distribution management systems.

Smart meter deployment has largely been the focus of digitization of the distribution grid. An average of 65% of customers in North America, Europe, and Asia are already using smart meters, with a projected increase to 85%, as shown in Fig. 6-1. While the EU lags behind other regions in AMI deployment, recent regulations continue to push for a customer-centric approach to grid enhancement, where smart meter installment requested by the customer must happen within 4 months of request [19]. In the US, the focus of AMI infrastructure has been for residential customers, to support billing, customer information, and outage management systems, and to enable time-varying price tariff structures [7].

Sensors: The deployment of sensors in the electric grid provides invaluable operational data in near real-time. Various sensor types are used including, but not limited to, voltage/temperature sensors, outage detection sensors, transformer monitoring sensors, and dynamic line rating sensors. In the distribution grid Micro PMUs

(μPMU, phasor measurement units) are being deployed throughout the distribution grid. These devices provide 120 measurement samples per second, and can measure grid voltages, currents, frequency, and power (real, reactive, and apparent) power. Literature has also looked at using μPMU data to detect or estimate grid topology, phase topology, and develop metrics for distribution grid observability.

6.1.2 DER Visibility

The increasing penetration of DERs requires a significant data collection and modelling effort to support DER visibility at a high-resolution. The lack of visibility in DER data is an operational and planning concern, for grid operations, reliability, modernization strategies, and utility investment decisions to ensure low-cost DER integration. In light of this, the North American Electric Reliability Corporation (NERC) established the System Planning Impacts from DER Working Group (SPIDERWG) in 2018, to address key aspects related to system planning, modeling, and reliability impacts of DERs to the bulk power system [60]. More generally, the data is needed to support integration and reliability studies which inform operating and interconnection standards. These standards must reflect goals of distribution grid modernization and full DER integration by supporting various DER services (such as voltage management and reactive power support). With regards to the proposed algorithms and distributed decision making framework, significant effort must be made in informing interconnection and operating standards with accurate data and models. These standards will then govern how DERs can participate in receiving and responding to dispatch signals (whether from a centralized decision-maker as in 3 or in a distributed framework as in 4).

6.2 Communication Infrastructure

An important point of interest in the implementation of the proposed algorithms is the communication infrastructure. In order to implement the proposed dynamic reconfiguration architecture, communication from a central operator to each switch

device is needed. Although some utilities may already have such capabilities (see discussion in Section 3.1.3), not all distribution systems have this infrastructure in place. Similarly, the proposed distributed voltage optimization algorithm requires significant infrastructure upgrades, to enable peer-to-peer coordination.

6.2.1 Communication Latency

The successful implementation of distributed optimization relies heavily on the communication delay or latency present in the communication infrastructure can impact the performance of the voltage regulation strategy if information is not received in a timely manner. Table 6.1 presents a summary of communication technologies and their corresponding latencies. Presumably, a distributed voltage control approach that requires iterative information sharing between neighboring agents may not be realizable with slow communication; the communication latency may be prohibitive in time to reach an optimal decision. However, the emergence of technologies for critical IoT systems – including smart power systems – can enable faster data sharing with high reliability guarantees.

Table 6.1: Summary of communication latency and technology readiness levels of different communications used in (or projected for) power grid applications.

	Maximum latency	**Technology readiness**
Slow communication [77]	100 ms	Past: transmission grid transient stability
Critical IoT Connectivity [38]	50 ms (99.9% reliability)	Current: Piloting in many industries
Ultra-reliable low-latency communication (URLLC) [38]	1 ms (99.9% reliability)	Future: 5G New Radio standard release target, one-way latency

6.2.2 Communication Standards

The IEEE 2030.5 Smart Energy Profile (SEP) 2.0 standard provides a framework for monitoring and control of DER assets, and has been suggested as the standardized communication protocol for DER aggregation programs [9]. California has adopted

the IEEE 2030.5 standard within its Rule 21 protocol, which includes communication protocols for smart inverters. This standard requires DERs to have monitoring and reporting capabilities and grid support functionalities such as Volt-VAr Control (VVC).

6.3 Towards Implementation of Grid-SiPhyR

The sections above present a discussion on state-of-art and emerging technologies for the distribution grid, and the technical considerations for optimization and control of future grids. Some additional comments must be made with regards to the proposed Grid-SiPhyR framework in Chapter 3. Dynamic reconfiguration was shown to improve grid efficiency, operability, and clean energy targets, by reducing line losses, improving voltage profiles, and increasing PV utilization. To enable such functionality, a physics-informed machine learning framework was proposed to allow for dynamic (fast and frequent) decision making. There remain some key implementation challenges to the proposed method.

Grid protection: A major concern with high DER penetration is the reliable operation of grid protection devices with bi-directional power flows and power flow reversal in key lines. Frequent topology changes in the grid – such as those proposed by dynamic reconfiguration – can also make traditional grid protection devices and schemes ineffective. Under the aforementioned conditions, there are challenges in detecting fault currents, a necessary functionality for safe and reliable grid operations. New protection schemes must be developed to support future distribution grid operations with high DER penetration. The coordination of multiple protective devices, such as those located in series, is also needed.

Operator trust in ML methods: Machine learning methods suffer from the black box problem, wherein trained models may perform accurately against test data, but give little to no insight into how the decision was formed. For power systems operators who are concerned with keeping the safety-critical large scale system operational with high reliability, such black box methods remain impracticable. To build opera-

tor trust in ML-based methods, there is a need to develop *trustworthy ML* solutions which are explainable, fair, privacy-preserving, causal, and robust[1]. The proposed physics-informed machine learning framework not only improves prediction accuracy, but moves towards an explainable framework by embedding system physics into the decision loop. Further, the proposed physics-informed rounding enhances visibility into how switching decisions are made. Finally, the proposed framework need not be directly integrated with Distribution Automation Control. Instead, the switching decisions can be proposed to an operator with additional information on why the switch is desirable (perhaps against the efficiency, operability, and clean energy metrics). The operator can then make an informed decision on whether to reconfigure the network or not.

6.4 Grid Modernization and DER Integration Projects

This section presents grid modernization project cost data and design decisions. These can be used towards understanding the state-of-art in distribution grid modernization and digitalization efforts, and the associated costs for moving towards smart grid functionalities, such as those proposed in this thesis.

DER Integration project cost estimates in Massachusetts

Eversource, an energy provider in the New England area, publishes cost estimates for typical distribution system and substation modification projects related to DER integration [14]. As of July 2022 the associated project costs in Massachusetts are estimated within the range of $1.65 to $1.94 million USD, *excluding distribution line costs*. Of this project cost, recloser installation and upgrades was estimated between $100,000-150,000. The cost to install or remove a capacitor bank (rated at 600-1800 kVAR) was estimated around $100,000.

[1]See more at: https://www.trustworthyml.org/

Smart grids project in Washington State

The Pacific Northwest Smart Grid Pacific Northwest Smart Grid Demonstration Project was a $178 million project led by Battelle in 2010 [67]. As a part of this project, Avista Utilities led a $42 million Spokane Smart Circuits project to automate and upgrade portions of their electric distribution system with a distribution management system, advanced metering, enhanced utility communication abilities, and intelligent end devices. The scope of the project included 72 distribution circuits and 17 substations serving over 125,000 electric and gas customers in the cities of Spokane and Pullman [1]. A component of this project was upgrading switch devices towards dynamic reconfiguration.

Reliability project in Pennsylvania

In 2008 the Philadelphia Electric Company (PECO) led a $2 million project to upgrade the distribution grid to improve electric service reliability [2]. PECO installed automated computer-controlled 'smart' switches along 36 electric distribution circuits. Each switch (called a recloser) cost between $50,000 to $65,000, and required roughly five days to install.

Smart grid communication infrastructure in Japan

The Japanese smart grids model has adopted three systems for connectivity and communication: cellular in rural areas, wireless 920MHz radio frequency for urban areas, and programmable controllers for households in high-rise buildings [3]. Further, energy management systems for communities, buildings, and households are being adopted as part of the smart grid model, with utilities supplying the system hardware. This has supported the growth of both automated and customer-led energy usage reductions. While much of these technologies have been rolled out, originally planned for 2020 Tokyo Olympics [4], the technology has yet to undergo "stress tests" with high customer engagement *and full realization of smart grid capabilities.*

6.5 Pathway Forward

Grid modernization initiatives are underway and many necessary discussions around situational awareness, distributed intelligence, and DER services are being had. With the requisite infrastructure and processes in place, the frameworks and algorithms proposed in this thesis can be implemented towards maintaining high reliability, efficiency, and service quality in a future high renewable and low-carbon power system.

The above discussions indicate that while some of the infrastructure required to adopt the proposed frameworks and implement the proposed algorithms is present, the current level of digitalization of distribution networks is still far from what is required. There are a number of recent studies that have underscored the need for a concerted push in infrastructure funding in improving the grid-edge (see for example, [104]). Likewise, the revised EU CEP is requiring of smart meters, (1) two-way communication between the metering system and external networks; (2) support for advanced tariff systems; and (3) data of at least 15-minute granularity [19]. Certainly, the implementation of the proposed physics-aware optimization schemes will likely require some additional hardware, software, and cybersecurity measures, but progress towards enhancing situational awareness and communication capabilities through data acquisition, monitoring, and automatic control suggests that these additional implementations are becoming more and more practicable.

Despite these advances, the pace of technological upgrades and operational reform remains too slow compared to what is needed to meet carbon reduction goals and climate targets. Research and technology transfer must focus on developing frameworks which can be deployed in the field today using the equipment currently installed in the distribution grid and augmenting existing operational and decision making approaches, with clear pathways to transition into more advanced frameworks in the future.

With regards to this thesis, the Grid-SiPhyR framework presented in Chapter 3 can be used first in offline studies which have high computational burden and inform operational and planning decisions involving switch installation and/or actuation,

and DER dispatch. This can promote industry and operator trust in machine learning methods while providing the opportunity to iterate on framework design before eventually piloting Grid-SiPhyR for online operational decision making. Likewise, the physics-aware framework presented in Chapter 4 can be deployed in the field today with some modifications. An ideal partner would be a utility company with many DERs installed in the grid (irrespective of ownership, so including both utility owned and private customer ownership), some amount of coordination and communication capabilities (even with a centralized decision maker and distributed devices), and grid modernization plans which can eventually support distributed intelligence. In an early pilot, clusters of devices can be modelled and controlled by a single decision making agent (which acts as a single agent in the distributed PAC algorithm), requiring fewer computational units and a more sparse communication graph. All utility owned devices can be centrally managed, including utility-owned solar and storage, in coordination with the device clusters. This hybrid approach is not a vastly new paradigm for the electricity system, which has historically relied on hierarchical control architectures, and a mix of both central and local optimization and control. In doing so, the hybrid approach can also support the full integration of aggregators, virtual power plants composed of DERs, and other non-utility owned DERMS platforms. This reduces the burden on utility companies to singularly invest in and develop the platforms needed to integrate DERs, while organically moving towards a distributed architecture for resource management and optimization with heterogeneous agents. A hybrid approach is also amendable to the Energy Hubs proposed in Section 5.3, which takes a smart cities approach to managing multiple agents and energy consumption/production devices.

Chapter 7

Conclusion and Future Work

This thesis presented two physics-aware optimization frameworks for operating future low-carbon power systems and investigated different algorithmic paradigms. First data-driven methods were employed towards end-to-end learning how to optimize for dynamic applications. Second the centralized approach of current grid operations was challenged by employing distributed methods for resource coordination and optimal dispatch solutions. Both frameworks were applied to different combinatorial problems: machine learning for dynamic reconfiguration, and hierarchical coordinated distributed optimization for voltage regulation. Herein two paradigms were investigated: a single one-shot optimization over both discrete and continuous variables; and bi-level optimization to separate decisions over discrete versus continuous variables. The proposed algorithms present new operating paradigms that enable the operational flexibility necessary for realizing a safe and reliable low-carbon power system. Building upon the results of this thesis, are multiple directions of future research which can yield useful results, some of which are described next.

7.1 Future Work – Physics-informed ML for Power Systems

A natural evolution of Grid-SiPhyR is to replace the multi-layer perceptron with a graph neural network (GNN). A GNN structure incorporates the underlying physical structure of the power grid, and presents another element of physics-informed neural design. Through successive convolutional layers, the GNN can extract pertinent compositional features from the graph nodes and edges. The output layer would then make predictions from the extracted features. The probability of switch statuses is one of the predicted variables from a GNN-based Grid-SiPhyR, and would pass through the physics-informed rounding as proposed in this thesis. The GNN provides the modeling flexibility to support locally- and globally-informed predictions: the output layer may use features only from a set of immediate neighbours or use network-wide information. This embeds the locality of power physics such as voltages and reactive power injections within the neural architecture. Further, the GNN-based Grid-SiPhyR may be extended to the cases of dynamic reconfiguration and DER dispatch when the set of devices change. Chapter 3 investigated the impact of different load profiles and different solar generator locations and capacities on prediction performance. It did not consider cases where some switches may not be available (due to scheduled maintenance or equipment failure) or a line fault has occurred on the grid. The GNN may provide the generalizability to these conditions, and to unseen grid topologies [74].

Grid-SiPhyR presents a physics-informed approach for end to end learning to optimize for combinatorial problems that can be extended to other applications in power systems. The unit commitment (UC) problem in electricity market dispatch asks the following question: "Given a list of generators and their operating constraints and costs, what is the least cost dispatch solution which satisfies loads in the network?" Generator constraints include time to ramp electricity generation up (or down) and minimum on-time, along with the associated operating costs. The UC problem must be solved subject to the power physics constraints. Similar to the dynamic reconfig-

uration problem, the UC problem considers a safety critical application with integer variables, and the additional challenge of the full nonlinear non-convex AC power flow. Where the dynamic reconfiguration problem presented a set of equations which fully described the rounding function, the UC problem (and its extensions) do not present a unique rounding function. Instead, a family of functions exist which produce feasible solutions. The framework can be further extended to deal with general integer constraints (currently only considers binary constraints). These two extensions would enrich the application of the Grid-SiPhyR framework to a broader class of problems, and provide ways to broaden the framework to other domains such as robotics, route planning, and device actuation. Additional extensions to the SiPhyR framework were also discussed in Section 3.8.

The committee machine approach employed in Chapter 3 was a simple averaging across all predictors. There are three main directions of future work which can be taken to improve performance of the committee machine prediction: investigating impact of initialization, employing different ensemble models, and forming different ensemble combinations. Different initialization techniques can be used to test the impact of initialization method on model prediction performance and model variance. The goal of this direction would be to both improve prediction performance and reduce model variance. This would also reduce the complexity of the next two tasks. The second approach is to investigate possible variations in the models, including networks with different depth and width (number of layers or nodes), different architectures (using dropout, different activation functions, or convolutional/recurrent layers), and models trained under different conditions (including learning rates and regularization). In particular, the regularization may help in guiding the neural network towards global minima or preventing the network from getting stuck at a local minima; it is well understood that model performance is highly variable due to the existence of numerous sub-optimal local minima [115]. Third, the ensemble combination may also be improved by using model blending: a weighted average ensemble which selects models based on prediction performance and can acknowledge the trade-off between optimality and feasibility [82].

7.2 Future Work – Physics-Aware Distributed Coordination

The proposed architecture relies heavily on the PAC-based distributed optimization. The PAC algorithm is a first order algorithm which needs many iterations to converge, with each iteration prone to communication failures such as packet drop or cyber-attacks. Accelerated methods have been proposed for the PAC algorithm, including a second-order dual update with considerably faster performance [49, 47], as presented in Chapter 5. The further development of accelerated optimization methods and robustness of the algorithm against cyber-attack or communication failure can be investigated.

The power physics models used throughout this thesis assumed PQ load models – constant real power and constant reactive power models – to represent the loads connected in the distribution grid. However, the PQ model does not capture the sensitivity of loads to voltages, which is of particular importance to voltage regulation in the distribution grid. The extension of the algorithms presented in Chapter 4 for voltage dependent loads is a future research direction. Characterizing time-varying loads and accurately modeling them is essential to developing and testing voltage control methodologies. Offline testing with representative load data can accelerate online testing and deployment, reducing barriers to adopting new methodologies in the field [125].

The hierarchical coordination mechanism proposed in Chapter 4 relies on a heuristic design process, to set the two tunable parameters ϵ and σ. The value of these design parameters may also be sensitive to system conditions. A sensitivity analysis is needed. Future work should also evaluate data-driven methods for optimally adjusting DER operating regions – either by determining optimal (possibly time-varying) parameter values, or proposing new methods for evaluating under- or overvoltage risk. This research direction is perhaps another instance of physics-informed algorithm development, wherein the blending of analytical approaches using power physics and circuit models with data-driven methods can provide better performance with some

feasibility guarantees.

7.3 Future Work – Implementation

Considerable work remains in transferring technologies and algorithms from the research domain towards industry implementation and adoption, as pointed out in Chapter 6. Concerted effort must be made to gain operator trust in ML-based methods through collaborative research activities with industry partners. The development and deployment of distributed optimization in field begins with hardware-in-the-loop demonstrations at the lab scale, prior to field pilots and widespread deployment. This may include demonstrations with Real Time Digital Simulator (RTDS) environment, real power hardware (ex. inverters, load banks, batteries), and a network of distributed computational units like single-board computers (ex. BeagleBone Blacks, Raspberry Pi, etc.) equipped with communication capabilities. While the requisite hardware and communication technologies may not exist in today's utility systems, the development of these technologies must move forward towards showing the technical potential as this infrastructure becomes economically viable at the large scale.

Appendix A

Datasets: Dynamic Reconfiguration of Distribution Grids

A.1 BW-33 (Baran-Wu) Distribution Grid

Grid data available in [20] includes the grid topology and line parameters, as presented in Table A.1, and location of loads and their nominal power demand (P and Q) for a single period as presented in Table A.2. As discussed in Chapter 3 each grid was modified to include distributed generation. The location and nameplate capacity of each solar facility is provided in Table A.3. Figure A-1 shows a sample 24-hour generation profile of a solar facility in the BW-33 dataset, at hourly intervals.

Table A.1: BW-33 grid topology data and line parameters

Branch No.	Upstream Node	Downstream Node	R [ohm]	X [ohm]
1	1	2	0.0922	0.0470
2	2	3	0.4930	0.2511
3	3	4	0.3660	0.1864
4	4	5	0.3811	0.1941
5	5	6	0.8190	0.707
6	6	7	0.1872	0.6188
7	7	8	0.7114	0.2351

8	8	9	1.030	0.7400
9	9	10	1.0440	0.7400
10	10	11	0.1966	0.0650
11	11	12	0.3744	0.1238
12	12	13	1.4680	1.1550
13	13	14	0.5416	0.7129
14	14	15	0.5910	0.5260
15	15	16	0.7463	0.5450
16	16	17	1.2890	1.7210
17	17	18	0.7320	0.5740
18	2	19	0.1640	0.1565
19	19	20	1.5042	1.3554
20	20	21	0.4095	0.4784
21	21	22	0.7089	0.9373
22	3	23	0.4512	0.3083
23	23	24	0.8980	0.7091
24	24	25	0.8960	0.7011
25	6	26	0.2030	0.1034
26	26	27	0.2842	0.1447
27	27	28	1.0590	0.9337
28	28	29	0.8042	0.7006
29	29	30	0.5075	0.2585
30	30	31	0.9744	0.9630
31	31	32	0.3105	0.3619
32	32	33	0.3410	0.5302
33	8	21	2.00	2.00
34	9	15	2.00	2.00
35	12	22	2.00	2.00
36	18	33	0.500	0.500

37	25	29	0.500	0.500

Table A.2: BW-33 grid load data

j	P^L [kW]	Q^L [kVAR]	j	P^L [kW]	Q^L [kVAR]	j	P^L [kW]	Q^L [kVAR]
2	100	60	13	60	35	24	420	200
3	90	40	14	120	80	25	420	200
4	120	80	15	60	10	26	60	25
5	60	30	16	60	20	27	60	25
6	60	20	17	60	20	28	60	20
7	200	100	18	90	40	29	120	70
8	200	100	19	90	40	30	200	600
9	60	20	20	90	40	31	150	70
10	60	20	21	90	40	32	210	100
11	45	30	22	90	40	33	60	40
12	60	35	23	90	50			

Table A.3: Locations and capacity of community solar facilities under each DD. The generating capacity is in kW

DD-U		DD-I		DD-II		DD-III		DD-II+III	
j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$
4	60	2	185	12	300	29	300	12	170
7	100	4	160	14	160	30	160	14	100
12	120	7	200	15	200	31	200	15	180
14	80	19	185	16	280	32	280	29	160
19	110	23	210					31	190
23	80							32	140
21	110								
26	70								

29	60				
32	150				

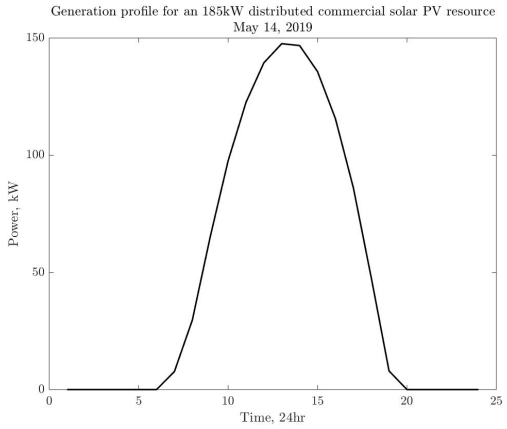


Figure A-1: Sample solar PV generation profile for community PV facility, queried for May 14, 2019.

A.2 TPC-94 Distribution Grid

Grid data available in [126] includes the grid topology and line parameters, as presented in Table A.4, and location of loads and their nominal power demand (P and Q) for a single period as presented in Table A.5. The location and type of commercial loads are shown in Table A.6. Two sets of solar location and capacity data are created: S1 corresponding to the training data (presented in Table A.7) and 'Solar Error' corresponding to a set of test data (presented in Table A.8).

Table A.4: TPC-94 grid topology data and line parameters

Branch No.	Upstream Node	Downstream Node	R [ohm]	X [ohm]
1	1	12	0.0085	0.0290
2	12	13	0.0092	0.0188
3	13	14	0.0103	0.0212
4	14	15	0.0040	0.0082

5	15	16	0.0092	0.0188
6	16	17	0.0017	0.0035
7	17	18	0.0018	0.0060
8	18	19	0.0046	0.0094
9	18	20	0.0103	0.0212
10	18	21	0.0046	0.0094
11	2	22	0.0034	0.0071
12	22	23	0.0149	0.0303
13	23	24	0.0011	0.0024
14	23	25	0.0034	0.0071
15	3	26	0.0050	0.0169
16	26	27	0.0023	0.0047
17	27	28	0.0023	0.0047
18	28	29	0.0069	0.0141
19	29	30	0.0017	0.0035
20	30	31	0.0074	0.0153
21	31	32	0.0103	0.0212
22	32	33	0.0069	0.0141
23	32	34	0.0086	0.0176
24	34	35	0.0057	0.0118
25	4	36	0.0025	0.0084
26	36	37	0.0046	0.0094
27	37	38	0.0109	0.0223
28	38	39	0.0021	0.0072
29	39	40	0.0057	0.0118
30	5	41	0.0086	0.0173
31	41	42	0.0057	0.0118
32	42	43	0.0057	0.0118
33	43	44	0.0011	0.0024
34	44	45	0.0074	0.0153

35	45	46	0.0023	0.0047
36	46	47	0.0218	0.0447
37	47	48	0.0017	0.0035
38	48	49	0.0017	0.0035
39	49	50	0.0034	0.0071
40	50	51	0.0092	0.0188
41	49	52	0.0086	0.0176
42	52	53	0.0092	0.0188
43	6	54	0.0021	0.0072
44	54	55	0.0017	0.0035
45	55	56	0.0057	0.0118
46	56	57	0.0103	0.0212
47	7	58	0.0106	0.0362
48	58	59	0.0029	0.0059
49	59	60	0.0029	0.0059
50	60	61	0.0017	0.0035
51	61	62	0.0034	0.0071
52	62	63	0.0017	0.0035
53	63	64	0.0034	0.0071
54	64	65	0.0023	0.0047
55	65	66	0.0057	0.0118
56	8	67	0.0099	0.0338
57	67	68	0.0235	0.0482
58	68	69	0.0023	0.0047
59	69	70	0.0018	0.0060
60	70	71	0.0017	0.0035
61	71	72	0.0011	0.0024
62	72	73	0.0046	0.0094
63	73	74	0.0103	0.0212
64	74	75	0.0011	0.0036

65	9	76	0.0021	0.0072
66	76	77	0.0074	0.0153
67	77	78	0.0053	0.0181
68	78	79	0.0096	0.0326
69	79	80	0.0021	0.0072
70	80	81	0.0032	0.0109
71	81	82	0.0025	0.0084
72	82	83	0.0011	0.0023
73	10	84	0.0142	0.0483
74	84	85	0.0014	0.0048
75	85	86	0.0025	0.0084
76	86	87	0.0021	0.0072
77	11	88	0.0110	0.0374
78	88	89	0.0057	0.0193
79	89	90	0.0021	0.0072
80	90	91	0.0057	0.0115
81	91	92	0.0057	0.0115
82	92	93	0.0040	0.0082
83	93	94	0.0137	0.0282
84	16	66	0.0057	0.0118
85	18	71	0.0057	0.0118
86	22	54	0.0057	0.0118
87	23	83	0.0149	0.0306
88	24	87	0.0200	0.0411
89	25	29	0.0235	0.0473
90	27	37	0.0040	0.0082
91	31	94	0.0034	0.0071
92	39	43	0.0023	0.0047
93	40	50	0.0034	0.0071
94	45	57	0.0011	0.0024

95	51	53	0.0086	0.0176
96	64	75	0.0017	0.0035
97	70	78	0.0077	0.0158

Table A.5: TPC-94 grid load data

j	P^L [kW]	Q^L [kVAR]	j	P^L [kW]	Q^L [kVAR]	j	P^L [kW]	Q^L [kVAR]
13	100	50	14	300	200	15	350	250
16	220	100	17	1100	800	18	400	320
19	300	200	20	300	230	21	300	260
23	1200	800	24	800	600	25	700	500
27	300	150	28	500	350	29	700	400
30	1200	1000	31	300	300	32	400	350
33	50	20	34	50	20	35	50	10
36	50	30	37	100	60	38	100	70
39	1800	1300	40	200	120	42	1800	1600
43	200	150	44	200	100	45	800	600
46	100	60	47	100	60	48	20	10
49	20	10	50	20	10	51	20	10
52	200	160	53	50	30	55	30	20
56	800	700	57	200	150	61	200	160
62	800	600	63	500	300	64	500	350
65	500	300	66	200	80	68	30	20
69	600	420	71	20	10	72	20	10
73	200	130	74	300	240	75	300	200
77	50	30	79	400	360	82	2000	1500
83	200	150	86	1200	950	87	300	180
89	400	360	90	2000	1300	91	200	140
92	500	360	93	100	30	94	400	360

Table A.6: Locations and type of commercial loads. The labels (1) thru (5) represent commercial loads of the following profiles, respectively: hospital, medium office building, quick service restaurant, stand-alone retail space, and warehouse.

A		B		C		D		E		F		G		H		I		J		K	
j	Label	j	Label	j	Label	j	Label	j	Label	j	Label	j	Label	j	Label	j	Label	j	Label	j	Label
13	4	23	2	28	2	36	5	42	4	55	3	61	2	74	2	77	5	86	3	90	5
16	2			30	1	37	4	47	4							82	2	87	2	93	4
17	1			34	5	39	1	52	2												
				35	5	40	2														

Table A.7: Locations and capacity of the solar facilities per feeder in the TPC-94 grid in training data. The nameplate generating capacity is in kW.

A		C		D		E		F		G		I		J	
j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$
5	150	20	585	40	250	42	450	43	305	48	240	65	180	74	250
8	200	21	340					45	250	50	120	69	340		
10	330	22	750					46	320	53	430	70	250		
		24	400												
		28	350												

Table A.8: Locations and capacity of the solar facilities per feeder in the TPC-94 grid in 'Solar Error' test data. The nameplate generating capacity is in kW.

A		C		D		E		F		G		I		J		K	
j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$	j	$\overline{P}^G$
3	250	20	585	29	350	36	250	43	325	48	240	65	230	74	250	78	120
6	310	21	380			42	450	45	310	50	120	69	380			81	150
8	200	22	750					46	290	53	430	70	230				
9	280	24	450														
		28	350														

Appendix B

Current Injection Model for Optimal Power Flow

The Current Injection (CI) model [46, 48] is a model for the power physics of unbalanced and meshed AC systems. It is an ideal candidate for unbalanced distribution grids with unbalanced lines, and single-phase loads and generation: unlike other proposed models in literature, the model is valid for the full range of angle imbalances. The CI model uses nodal variables, similar to the Bus Injection model, and represents all loads and generators as nodal current injections, with all power, current, and voltage phasors represented in Cartesian coordinates. The 3-phase impedance matrix is used to describe the self and mutual inductance between phases to model the coupling of phases that are common to a distribution grid. This appendix section provides an overview of the model.

B.1 Statement of the CI Model

Denoting the column vector of line currents as I_{flow}, nodal voltages as V, and nodal current injections as I, the optimal power flow formulation (CI-OPF) is written as:

$$\min_{x} f(x) \tag{B.1a}$$

$$AV = ZI_{\text{flow}} \tag{B.1b}$$

$$I^R = \text{Re}\left(A^T I_{\text{flow}}\right) \tag{B.1c}$$

$$I^I = \text{Im}\left(A^T I_{\text{flow}}\right) \tag{B.1d}$$

$$\underline{V}_j^{\phi,R} \leq V_j^{\phi,R} \leq \overline{V}_j^{\phi,R} \qquad \forall j \in \mathcal{N}, \phi \in \mathcal{P} \tag{B.1e}$$

$$\underline{V} j^{\phi,I} \leq V_j^{\phi,I} \leq \overline{V}_j^{\phi,I} \qquad \forall j \in \mathcal{N}, \phi \in \mathcal{P} \tag{B.1f}$$

$$\underline{P}_j^{\phi} \leq P_j^{\phi} \leq \overline{P}_j^{\phi} \qquad \forall j \in \mathcal{N}, \phi \in \mathcal{P} \tag{B.1g}$$

$$\underline{Q}_j^{\phi} \leq Q_j^{\phi} \leq \overline{Q}_j^{\phi} \qquad \forall j \in \mathcal{N}, \phi \in \mathcal{P} \tag{B.1h}$$

$$P_j^{\phi} = V_j^{\phi,R} I_j^{\phi,R} + V_j^{\phi,I} I_j^{\phi,I} \qquad \forall j \in \mathcal{N}, \phi \in \mathcal{P} \tag{B.1i}$$

$$Q_j^{\phi} = -V_j^{\phi,R} I_j^{\phi,I} + V_j^{\phi,I} I_j^{\phi,R} \qquad \forall j \in \mathcal{N}, \phi \in \mathcal{P} \tag{B.1j}$$

where $x = \left[I^R\ I^I\ V^R\ V^I\ P\ Q\ I^R_{\text{flow}}\ I^I_{\text{flow}}\right]$ is the decision vector for the CI-OPF problem. The real and imaginary components of a complex number x are denoted as x^R and x^I; overbar $\overline{x}$ and underbar $\underline{x}$ denote the upper and lower limits of a variable x; $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the real and imaginary components of a complex number.

Constraint (B.1b) describes Ohm's law, and (B.1c)-(B.1d) describe Kirchhoff's Current Law. The objective function (B.1a) is the performance index to be minimized, which can be, for example, to minimize cost for power production, or line losses. Note the use active sign convention means positive P and Q values are generators, while negative values are loads.

B.2 Convexification of the CI-OPF

The OPF problem in (B.1) fully describes the power physics of an unbalanced network of either radial or meshed topology. However, constraints (B.1i)-(B.1j) render the problem nonconvex. To convexify these bilinear constraints, McCormick Envelopes (MCE) [97] are employed. MCE describe the convex hull of a bilinear product $w = xy$ by utilizing the bounds on x and y. This is denoted as $\text{MCE}(w, \underline{x}, \overline{x}, \underline{y}, \overline{y}) = \{w = xy : x \in [\underline{x}, \overline{x}], y \in [\underline{y}, \overline{y}]\}$, and formally defined as:

$$M(w,\underline{x},\overline{x},\underline{y},\overline{y}) = \begin{cases} w \geq \underline{x}y + x\underline{y} - \underline{x}\underline{y} \\ w \geq \overline{x}y + x\overline{y} - \overline{x}\overline{y} \\ w \leq \underline{x}y + x\overline{y} - \overline{x}\underline{y} \\ w \leq \overline{x}y + x\underline{y} - \underline{x}\overline{y} \end{cases} \tag{B.2}$$

To employ MCE relaxation, auxiliary variables $\{a, b, c, d\}$ are introduced to to describe the bilinear terms in (B.1i)-(B.1j), and the corresponding linear constraints as described in (B.2). To leverage the MCE relaxation, access to both voltage and current bounds are required (see (B.1e)-(B.1f) and (B.3)).

$$\underline{I}_j^{\phi,R} \leq I_j^{\phi,R} \leq \overline{I}_j^{\phi,R} \tag{B.3a}$$

$$\underline{I}_j^{\phi,I} \leq I_j^{\phi,I} \leq \overline{I}_j^{\phi,I} \tag{B.3b}$$

These are indirectly specified by (B.1e)-(B.1j), $\forall j \in \mathcal{N}, \forall \phi \in \mathcal{P}$. The determination of these upper and lower bounds on nodal current is not trivial. An effective pre-processing technique to define these values is presented in [46].

The convex CI model is then described by the augmented decision vector:

$$\tilde{x} = \left[I^R\ I^I\ V^R\ V^I\ P\ Q\ I^R_{\text{flow}}\ I^I_{\text{flow}}\ a\ b\ c\ d\right] \tag{B.4}$$

where each term in the vector $\tilde{x}$ is of the form $I^R = [I_j^R] \quad \forall j \in \mathcal{N}$, and $I_j^R = \left[I_j^{R,\phi}\right] \quad \forall \phi \in \mathcal{P}$; subject to constraints (B.1b)-(B.1h), (B.3a)-(B.3b), and the linear constraints for auxiliary variables as described in (B.2).

Remark 1: The use of current and voltage to model the power physics permits a simple multi-phase model for the unbalanced distribution grid. In the model presented about the system impedance matrix is used, $Z \in \mathbb{R}^{3N\times 3N}$, which includes cross-phase coupling inherent in multi-phase distribution grids.

Remark 2: The CI model includes hard constraints on the nodal voltages in (B.1e) and (B.1f). These bounds on the real and imaginary components of the voltage phasor at every node are determined by the pre-processing technique, using information of

operating bounds (ex. $\pm 5\%$ in the US) and the forecast for power injections and loads. Thus, any dispatch solution to the CI-OPF model will enforce voltage constraints, as required for grid operation.

Remark 3: Another typical constraint in power systems is a congestion constraint, which limits the line currents. This constraint can be trivially included in our CI-OPF formulation by introducing a variable describing the current flow in the distribution lines, I_{flow}, and rewriting constraint (B.1b) as two constraints:

$$AV = ZI_{\text{flow}} \tag{B.5a}$$

$$I = A^T I_{\text{flow}} \tag{B.5b}$$

where $A \in \mathbb{R}^{3N \times 3N}$ is the 3-phase graph incidence matrix.

Bibliography

[1] Avista chooses smart grid technology for 'smart circuits' project. *T&D World*.

[2] PECO installing smart reclosers in Chester county in major reliability program. *T&D World*.

[3] Smart grid trends in Japan: 7 things to know. *Smart Energy International*.

[4] Smart meters Japan: TEPCO to deploy 27m by 2020 Olympic Games. *Smart Energy International*.

[5] IEEE standard for AC high-voltage circuit breakers rated on a symmetrical current basis–preferred ratings and related required capabilities for voltages above 1000 V. *IEEE Std C37.06-2009*, pages 1–56, 2009.

[6] IEEE/IEC international standard - high-voltage switchgear and controlgear - part 111: Automatic circuit reclosers and fault interrupters for alternating current systems up to 38 kv. *IEEE Std C37.60-2012*, pages 1–134, 2012.

[7] Advanced metering infrastructure and customer systems. Technical report, U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability, 2016.

[8] Visibility of distributed energy resources: Future power system security program. Technical report, Australian Energy Market Operator, 2017.

[9] Ieee standard for smart energy profile application protocol. *IEEE Std 2030.5-2018 (Revision of IEEE Std 2030.5-2013)*, pages 1–361, 2018.

[10] Instituting a proceeding to investigate performance-based regulation, 2018. Docket Number 2018-0088. `https://dms.puc.hawaii.gov/dms/DocumentViewer?pid=A1001001A21E17B53226E00118`.

[11] Making the connection: The importance of DER visibility to grid support and modernization. Technical Report 000000003002013388, Electric Power Research Institute, June 2018.

[12] Massachusetts residential baseline study. Technical report, Guidehouse Inc., March 2020.

[13] Behaviour of distributed resources during power system disturbances. Technical report, Australian Energy Market Operator, May 2021.

[14] Distributed Energy Resources (DER) Project Costs, Jul 2021.

[15] End-use load profiles for the u.s. building stock: Methodology and results of model calibration, validation, and uncertainty quantification. Technical report, National Renewable Energy Laboratory, March 2022.

[16] Abdeen A., Kharvari F., O'Brien W., and Gunay B. The impact of the covid-19 on households' hourly electricity consumption in canada. *Energy Build*, 2021.

[17] Hamed Ahmadi and José R. Martí. Mathematical representation of radiality constraint in distribution system reconfiguration problem. *International Journal of Electrical Power & Energy Systems*, 64:293–299, 2015.

[18] Rob Ardis and Robert Uluski. CVR is here to stay. *T & D World*, 2015.

[19] Leonardo Meeus Athir Nouicer. The EU clean energy package. Technical Report 2, Florence School of Regulation, 2019. doi:10.2870/33236.

[20] M.E. Baran and F.F. Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, 4(2):1401–1407, 1989.

[21] Mohini Bariya, Deepjyoti Deka, and Alexandra von Meier. Guaranteed phase topology identification in three phase distribution grids. *IEEE Transactions on Smart Grid*, 12(4):3605–3612, 2021.

[22] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.

[23] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., USA, 1995.

[24] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[25] Scott Burger, Ian Schneider, Audun Botterud, and Ignacio Pérez-Arriaga. *Fair, Equitable, and Efficient Tariffs in the Presence of Distributed Energy Resources*, pages 155–188. 01 2019.

[26] CAISO. California ISO demand response and energy efficiency roadmap: Maximizing preferred resources, 2013.

[27] Yankai Cao and Victor M. Zavala. A sigmoidal approximation for chance-constrained nonlinear programs, 2020. arXiv:2004.02402 [math.OC].

[28] Audrey Carleton. Renewables met all of California's energy needs for the first time ever. *Vice Media Group*, May 2022.

[29] Yexiang Chen, Subhash Lakshminarayana, Carsten Maple, and H. Vincent Poor. A meta-learning approach to the optimal power flow problem under topology reconfigurations. *IEEE Open Access Journal of Power and Energy*, 9:109–120, 2022.

[30] G. A. Comnes, S. Stoft, N. Greene, and L. J. Hill. Performance-based ratemaking for electric utilities: Review of plans and analysis of economic and resource planning issues. Technical Report LBL-37577, UC-1320, Lawrence Berkeley National Laboratory, November 1995. `https://eta-publications.lbl.gov/sites/default/files/report-lbnl-37577.pdf`.

[31] Emiliano Dall'Anese, Georgios B. Giannakis, and Bruce F. Wollenberg. Optimization of unbalanced power distribution networks via semidefinite relaxation. In *2012 North American Power Symposium (NAPS)*, pages 1–6, 2012.

[32] Emiliano Dall'Anese, Sairaj V. Dhople, and Georgios B. Giannakis. Optimal dispatch of photovoltaic inverters in residential distribution systems. *IEEE Transactions on Sustainable Energy*, 5(2):487–497, 2014.

[33] Federico Delfino, Giulio Ferro, Michela Robba, and Mansueto Rossi. An energy management platform for the optimal control of active and reactive powers in sustainable microgrids. *IEEE Transactions on Industry Applications*, 55(6):7146–7156, 2019.

[34] Roel Dobbe, Oscar Sondermeijer, David Fridovich-Keil, Daniel Arnold, Duncan Callaway, and Claire Tomlin. Toward distributed energy services: Decentralizing optimal power flow with machine learning. *IEEE Transactions on Smart Grid*, 11(2):1296–1306, 2020.

[35] Priya L. Donti, David Rolnick, and J Zico Kolter. DC3: A learning method for optimization with hard constraints. In *International Conference on Learning Representations*, 2021.

[36] EIA. Wholesale power price maps reflect real-time constraints on transmission of electricity. `https://www.eia.gov/todayinenergy/detail.php?id=3150#`, September 2011.

[37] Mark Eisen, Aryan Mokhtari, and Alejandro Ribeiro. Decentralized quasi-newton methods. *IEEE Transactions on Signal Processing*, 65(10):2613–2628, 2017.

[38] Ericsson Technology Review Articles. Critical IoT connectivity: Ideal for time-critical communications, 2020.

[39] F. Evans, J.M. Zolezzi, and H. Rudnick. Cost assignment model for electrical transmission system expansion: an approach through the kernel theory. *IEEE Transactions on Power Systems*, 18(2):625–632, 2003.

[40] Masoud Farivar, Christopher R. Clarke, Steven H. Low, and K. Mani Chandy. Inverter var control for distribution systems with renewables. In *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 457–462, 2011.

[41] Masoud Farivar and Steven Low. Branch flow model: Relaxations and convexification. *2014 IEEE PES T&D Conference and Exposition*, 2014.

[42] Anna Rita Di Fazio, Chiara Risi, Mario Russo, and Michele De Santis. Coordinated optimization for zone-based voltage control in distribution grids. *IEEE Transactions on Industry Applications*, 58(1):173–184, 2022.

[43] Federal Energy Regulatory Commission. Payment for reactive power- commission staff report. `[Online].Available:https://www.ferc.gov/sites/default/files/2020-05/04-11-14-reactive-power.pdf`, April 2014.

[44] Changsen Feng, Zhiyi Li, Mohammad Shahidehpour, Fushuan Wen, Weijia Liu, and Xiaolei Wang. Decentralized short-term voltage control in active power distribution systems. *IEEE Transactions on Smart Grid*, 9(5):4566–4576, 2018.

[45] FERC. Participation of Distributed Energy Resource Aggregations in markets operated by Regional Transmission Organizations and Independent System Operators, September 2020.

[46] G. Ferro. *Competitive and Cooperative Approaches to the Balancing Market in Distribution Grids*. PhD thesis, Università degli studi di Genova, 2020.

[47] Giulio Ferro, Michela Robba, Federico Delfino, Rabab Haider, and Anuradha M. Annaswamy. Distributed operational management of microgrids: a second order dual update approach. *IFAC-PapersOnLine*, 2023. 22nd IFAC World Congress.

[48] Giulio Ferro, Michela Robba, David D'Achiardi, Rabab Haider, and Anuradha M. Annaswamy. A distributed approach to the optimal power flow problem for unbalanced and mesh networks. *IFAC-PapersOnLine*, 53(2):13287–13292, 2020. 21st IFAC World Congress.

[49] Giulio Ferro, Michela Robba, Rabab Haider, and Anuradha M. Annaswamy. A distributed-optimization-based architecture for management of interconnected energy hubs. *IEEE Transactions on Control of Network Systems*, 9(4):1704–1716, 2022.

[50] D. Filev. Applied intelligent control - control of automotive paint process. In *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291)*, volume 1, pages 1–6 vol.1, 2002.

[51] Ferdinando Fioretto, Terrence W.K. Mak, and Pascal van Hentenryck. Predicting ac optimal power flows: Combining deep learning and lagrangian dual methods. In *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, AAAI 2020 - 34th AAAI Conference on Artificial Intelligence, pages 630–637. AAAI Press, 2020.

[52] Ben Foster, Timothy Bialecki, David Burns, David Kathan, Michael P. Lee, and Samin Peirovi. 2019 assessment of demand response and advanced metering. `https://www.ferc.gov/sites/default/files/2020-04/DR-AM-Report2019_2.pdf`, December 2019.

[53] Guilherme Franca, Daniel Robinson, and Rene Vidal. Admm and accelerated admm as continuous dynamical systems. In *International Conference on Machine Learning*, pages 1559–1567. PMLR, 2018.

[54] Guilherme França, Daniel P Robinson, and Rene Vidal. A dynamical systems perspective on nonsmooth constrained optimization. *arXiv preprint arXiv:1808.04048*, 2018.

[55] Guilherme França, Daniel P Robinson, and René Vidal. A nonsmooth dynamical systems perspective on accelerated extensions of admm. *IEEE Transactions on Automatic Control*, 2023.

[56] Euhanna Ghadimi, André Teixeira, Iman Shames, and Mikael Johansson. Optimal parameter selection for the alternating direction method of multipliers (admm): quadratic problems. *IEEE Transactions on Automatic Control*, 60(3):644–658, 2014.

[57] J.D. Glover, T. Overbye, M.S. Sarma, and A.B. Birchfield. *Power System Analysis and Design, SI Edition*. Cengage Learning, 2021.

[58] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, nov 1995.

[59] Tom Goldstein, Brendan O'Donoghue, Simon Setzer, and Richard Baraniuk. Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, 7(3):1588–1623, 2014.

[60] NERC Working Group. System planning impacts from distributed energy resources working group. `https://www.nerc.com/comm/RSTC/SPIDERWG/SPIDERWG%20Scope.pdf`, December 2018. SPIDERWG Scope document.

[61] J. M. Guerrero, M. Chandorkar, T. Lee, and P. C. Loh. Advanced control architectures for intelligent microgrids—part i: Decentralized and hierarchical control. *IEEE Transactions on Industrial Electronics*, 60(4):1254–1262, 2013.

[62] Victor J. Gutierrez-Martinez, Claudio A. Cañizares, Claudio R. Fuerte-Esquivel, Alejandro Pizano-Martinez, and Xueping Gu. Neural-network security-boundary constrained optimal power flow. *IEEE Transactions on Power Systems*, 26(1):63–72, 2011.

[63] R. Haider, S. Baros, Y. Wasa, J. Romvary, K. Uchida, and A.M. Annaswamy. Towards a retail market for distribution grids. *IEEE Transactions on Smart Grids*, 2020. doi: 10.1109/TSG.2020.2996565.

[64] Rabab Haider and Anuradha M. Annaswamy. A hybrid architecture for volt-var control in active distribution grids. *Applied Energy*, 312:118735, 2022.

[65] Rabab Haider, David D'Achiardi, Venkatesh Venkataramanan, Anurag Srivastava, Anjan Bose, and Anuradha M. Annaswamy. Reinventing the utility for distributed energy resources: A proposal for retail electricity markets. *Advances in Applied Energy*, 2:100026, 2021.

[66] Rabab Haider, Giulio Ferro, Michela Robba, and Anuradha M. Annaswamy. Flattening the duck curve: A case for distributed decision making. In *2022 IEEE Power Energy Society General Meeting (PESGM)*, pages 01–05, 2022.

[67] Lee Hall and Tracy Yount. Pacific Northwest Smart Grid Pacific Northwest Smart Grid Demonstration Project Demonstration Project, April 2010.

[68] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, USA, 2nd edition, 1998.

[69] Mhand Hifi and Shohre Sadeghsa. A rounding strategy-based algorithm for the k-clustering minimum biclique completion problem. *Journal of the Operational Research Society*, February 2022.

[70] IPCC. Summary for policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, , and B. Zhou, editors, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021.

[71] IRENA. Innovation landscape for a renewable-powered future: Solutions to integrate variable renewables. Technical report, International Renewable Energy Agency, 2019.

[72] Vineet Jagadeesan Nair, Venkatesh Venkataramanan, Rabab Haider, and Anuradha M. Annaswamy. A hierarchical local electricity market for a der-rich grid edge. *IEEE Transactions on Smart Grid*, 14(2):1353–1366, 2023.

[73] Dušan Jakovetić, Joao Xavier, and José MF Moura. Fast distributed gradient methods. *IEEE Transactions on Automatic Control*, 59(5):1131–1146, 2014.

[74] Chaitanya K. Joshi, Thomas Laurent, and Xavier Bresson. An efficient graph convolutional network technique for the travelling salesman problem, 2019.

[75] Michael Jünger, Thomas M. Liebling, Denis Naddef, George L. Nemhauser, William R. Pulleyblank, Gerhard Reinelt, Giovanni Rinaldi, and Laurence A. Wolsey, editors. pages 619–645. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[76] Siripha Junlakarn and Marija Ilić. Toward implementation of the reconfiguration for providing differentiated reliability options in distribution systems. In *2014 IEEE PES General Meeting | Conference Exposition*, pages 1–5, 2014.

[77] Prashant Kansal and Anjan Bose. Bandwidth and latency requirements for smart transmission grid applications. *IEEE Transactions on Smart Grid*, 3(3):1344–1352, 2012.

[78] W.H. Kersting. *Distribution System Modeling and Analysis*. Taylor & Francis, CRC Press, 2017.

[79] J.L. Kirtley. *Electric Power Principles: Sources, Conversion, Distribution and Use*. Wiley, 2019.

[80] Jonatan Ralf Axel Klemets and Merkebu Zenebe Degefa. A distributed algorithm for controlling continuous and discrete variables in a radial distribution grid. *IEEE Access*, 11:2488–2499, 2023.

[81] Lorenzo Kristov, Paul De Martini, and Jeffrey D. Taft. A tale of two visions: Designing a decentralized transactive electric system. *IEEE Power and Energy Magazine*, 14(3):63–69, May 2016.

[82] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[83] Javad Lavaei and Steven H. Low. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems*, 27(1):92–107, 2012.

[84] Marina Lavorato, John F. Franco, Marcos J. Rider, and Rubén Romero. Imposing radiality constraints in distribution system optimization problems. *IEEE Transactions on Power Systems*, 27(1):172–180, 2012.

[85] Shunbo Lei, Chen Chen, Yue Song, and Yunhe Hou. Radiality constraints for resilient reconfiguration of distribution systems: Formulation and application to microgrid formation. *IEEE Transactions on Smart Grid*, 11(5):3944–3956, 2020.

[86] Wenting Li, Deepjyoti Deka, Michael Chertkov, and Meng Wang. Real-time faulted line localization and pmu placement in power systems through convolutional neural networks. *IEEE Transactions on Power Systems*, 34(6):4640–4651, 2019.

[87] David Littel, Camille Kadoch, Phil Baker, Ranjit Bharvirkar, Max Dupuy, Brenda Hausauer, Carl Linvill, Janine Migden-Ostrander, Jan Rosenow, Wang Xuan, Owen Zinaman, and Jeffrey Logan. Next-generation performance-based regulation. Technical report, NREL, September 2017.

[88] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[89] Steven H. Low. Convex relaxation of optimal power flow—part i: Formulations and equivalence. *IEEE Transactions on Control of Network Systems*, 1(1):15–27, 2014.

[90] Chen Luo, Hongbin Wu, Yiyao Zhou, Yida Qiao, and Mengyi Cai. Network partition-based hierarchical decentralised voltage control for distribution networks with distributed pv systems. *International Journal of Electrical Power & Energy Systems*, 130:106929, 2021.

[91] Ali Makhdoumi and Asuman Ozdaglar. Convergence rate of distributed admm over networks. *IEEE Transactions on Automatic Control*, 62(10):5082–5095, 2017.

[92] Ali Maknouninejad and Zhihua Qu. Realizing unified microgrid voltage profile and loss minimization: A cooperative distributed optimization and control approach. *IEEE Transactions on Smart Grid*, 5(4):1621–1630, 2014.

[93] A. R. Malekpour, A. M. Annaswamy, and J. Shah. Hierarchical hybrid architecture for volt/var control of power distribution grids. *IEEE Transactions on Power Systems*, 35(2):854–863, 2020.

[94] Mohammad MansourLakouraj, Mukesh Gautam, Hanif Livani, Mohammed Benidris, and Poria Fajri. Multi-timescale risk-constrained volt/var control of distribution grids with electric vehicles and solar inverters. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, pages 01–06, 2021.

[95] Hugues Marchand and Laurence A. Wolsey. Aggregation and mixed integer rounding to solve mips. *Operations Research*, 49(3), June 2001.

[96] Paul De Martini and Lorenzo Kristov. Distribution systems in a high distributed energy resources future. Technical report, 10/2015 2015.

[97] Garth P. Mccormick. Computability of global solutions to factorable nonconvex programs: Part i — convex underestimating problems. *Mathematical Programming*, 10(1):147–175, 1976.

[98] John McDonald. Solar power impacts power electronics in the smart grid. *Electronic Design*, Aug 2013.

[99] Ariana Minot, Yue M Lu, and Na Li. A parallel primal-dual interior-point method for dc optimal power flow. In *2016 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE, 2016.

[100] MITEI. Utility of the future. Technical report, MIT Energy Initiative, December 2016.

[101] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei. A survey of distributed optimization and control algorithms for electric power systems. *IEEE Transactions on Smart Grid*, 8(6):2941–2962, 2017.

[102] Fabio Moret and Pierre Pinson. Energy collectives: a community and fairness based approach to future electricity markets. *IEEE Transactions on Power Systems*, 34(5):3994–4004, 2018.

[103] Eric Münsing, Jonathan Mather, and Scott Moura. Blockchains for decentralized optimization of energy resources in microgrid networks. In *2017 IEEE conference on control technology and applications (CCTA)*, pages 2164–2171. IEEE, 2017.

[104] National Academies of Sciences, Engineering, and Medicine. *The Future of Electric Power in the United States*. The National Academies Press, Washington, DC, 2021.

[105] National Renewable Energy Laboratory. Commercial and residential hourly load profiles for all TMY3 locations in the United States [data set], 2014. doi:10.25984/1788456.

[106] National Renewable Energy Laboratory. System advisory model. https://sam.nrel.gov/, 2020.

[107] United Nations. The sustainable development goals report. Technical report, United Nations Publications, 2021. https://unstats.un.org/sdgs/report/2021/The-Sustainable-Development-Goals-Report-2021.pdf.

[108] Yurii E Nesterov. A method for solving the convex programming problem with convergence rate o (1/kˆ 2). In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547, 1983.

[109] NIST. Transactive energy: An overview. https://www.nist.gov/engineering-laboratory/smart-grid/hot-topics/transactive-energy-overview, April 2017.

[110] Thomas R. Nudell, Anuradha M. Annaswamy, Jianming Lian, Karanjit Kalsi, and David D'Achiardi. *Electricity Markets in the United States: A Brief History, Current Operations, and Trends*, pages 3–27. Springer International Publishing, Cham, 2019.

[111] Daniel Olsen, Michael Sohn, Mary Ann Piette, and Sila Kiliccote. Demand response availability profiles for California in the year 2020. *LBNL Report*, 2014.

[112] Xiang Pan, Tianyu Zhao, and Minghua Chen. Deepopf: Deep neural network for dc optimal power flow. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6, 2019.

[113] Xiang Pan, Tianyu Zhao, Minghua Chen, and Shengyu Zhang. Deepopf: A deep neural network approach for security-constrained dc optimal power flow. *IEEE Transactions on Power Systems*, 36(3):1725–1735, 2021.

[114] Zhaoguang Pan, Qinglai Guo, and Hongbin Sun. Feasible region method based integrated heat and electricity dispatch considering building thermal inertia. *Applied Energy*, 192:395–407, 2017.

[115] Michael Perrone and Leon N. Cooper. When networks disagree: Ensemble methods for hybrid neural networks. 1992.

[116] Marouf Pirouti. *Modelling and analysis of a district heating network.* PhD thesis, Cardiff University, 2013.

[117] Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[118] Adam Potter, Rabab Haider, Giulio Ferro, Michela Robba, and Anuradha M. Annaswamy. A reactive power market for the future grid. *Advances in Applied Energy*, 9:100114, 2023.

[119] G. Rancilio, A. Rossi, D. Falabretti, A. Galliani, and M. Merlo. Ancillary services markets in europe: Evolution and regulatory trade-offs. *Renewable and Sustainable Energy Reviews*, 154:111850, 2022.

[120] Brett A. Robbins, Hao Zhu, and Alejandro D. Domínguez-García. Optimal tap setting of voltage regulation transformers in unbalanced distribution systems. *IEEE Transactions on Power Systems*, 31(1):256–267, 2016.

[121] J. Romaine. Biden administration shoots for the sun with goal of 45 percent solar power by 2050, Sept 2021. `https://thehill.com/changing-america/sustainability/energy/571256-biden-administration-shoots-for-the-sun-with-goal-of/`.

[122] J. Romvary, G. Ferro, R. Haider, and A.M. Annaswamy. A distributed proximal atomic coordination algorithm. *IEEE Transactions on Automatic Control*, 2021. doi: 10.1109/TAC.2021.3053907.

[123] Yoav Sharon, Anuradha M. Annaswamy, Alexis L. Motto, and Amit Chakraborty. Topology identification in distribution network with limited measurements. In *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, pages 1–6, 2012.

[124] Tom A. Short. Advanced metering for phase identification, transformer identification, and secondary modeling. *IEEE Transactions on Smart Grid*, 4(2):651–658, 2013.

[125] Priyank Srivastava, Rabab Haider, Vineet J. Nair, Venkatesh Venkataramanan, Anuradha M. Annaswamy, and Anurag K. Srivastava. Voltage regulation in distribution grids: A survey. *Annual Reviews in Control*, 2023.

[126] Ching-Tzong Su, Chung-Fu Chang, and Ji-Pyng Chiou. Distribution network reconfiguration for loss reduction by ant colony search algorithm. *Electric Power Systems Research*, 75(2):190–199, 2005.

[127] Medha Subramanian, Jan Viebahn, Simon H. Tindemans, Benjamin Donnot, and Antoine Marot. Exploring grid topology reconfiguration using a simple deep reinforcement learning approach. In *2021 IEEE Madrid PowerTech*, pages 1–6, 2021.

[128] Xianzhuo Sun and Jing Qiu. Hierarchical voltage control strategy in distribution networks considering customized charging navigation of electric vehicles. *IEEE Transactions on Smart Grid*, 12(6):4752–4764, 2021.

[129] Patricia Mae Tagabe. The power of community batteries: bringing renewables to the masses. *Energy Magazine*, Aug 2020.

[130] Zhiyuan Tang, David J. Hill, and Tao Liu. Distributed coordinated reactive power control for voltage regulation in distribution networks. *IEEE Transactions on Smart Grid*, 12(1):312–323, 2021.

[131] Joshua A. Taylor and Franz S. Hover. Convex models of distribution system reconfiguration. *IEEE Transactions on Power Systems*, 27(3):1407–1413, 2012.

[132] Herman K. Trabish. A user's guide to natural gas power plants. *Utility Dive*, May 2014.

[133] Rasul Tutunov, Haitham Bou-Ammar, and Ali Jadbabaie. Distributed newton method for large-scale consensus optimization. *IEEE Transactions on Automatic Control*, 64(10):3983–3994, 2019.

[134] Vlerick Energy Centre. Outlook on the European DSO landscape 2020. Technical report, Vlerick Business School, 2020. https://home.kpmg/content/dam/kpmg/pdf/2016/05/Energy-Outlook-DSO-2020.pdf.

[135] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106:25–57, 2006.

[136] Shen Wang, Ahmad F Taha, Jianhui Wang, Karla Kvaternik, and Adam Hahn. Energy crowdsourcing and peer-to-peer energy trading in blockchain-enabled smart grids. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(8):1612–1623, 2019.

[137] Ying Wang, Yin Xu, Jiaxu Li, Jinghan He, and Xiaojun Wang. On the radiality constraints for distribution system restoration and reconfiguration problems. *IEEE Transactions on Power Systems*, 35(4):3294–3296, 2020.

[138] Yang Weng, Yizheng Liao, and Ram Rajagopal. Distributed energy resources topology identification via graphical modeling. *IEEE Transactions on Power Systems*, 32(4):2682–2694, 2017.

[139] Joan White. Preparing for FERC Order 2222, a PUC perspective. In *New England Conference of Public Utilities Commissioners Webinar Series*, June 2021. https://energy.sandia.gov/wp-content/uploads/2021/12/2_White_Joan_VTPUC_NECPUC_Session6_6-4-2021.pdf.

[140] F. F. Wu, K. Moslehi, and A. Bose. Power system control centers: Past, present, and future. *Proceedings of the IEEE*, 93(11):1890–1908, 2005.

[141] Mou Wu, Naixue Xiong, Athanasios V Vasilakos, Victor CM Leung, and CL Philip Chen. Rnn-k: A reinforced newton method for consensus-based distributed optimization and control over multiagent systems. *IEEE Transactions on Cybernetics*, 52(5):4012–4026, 2020.

[142] Álinson S. Xavier, Feng Qiu, and Shabbir Ahmed. Learning to solve large-scale security-constrained unit commitment problems. *INFORMS Journal on Computing*, 33(2):739–756, 2020.

[143] Hanchen Xu, Alejandro D. Domínguez-García, and Peter W. Sauer. Optimal tap setting of voltage regulation transformers using batch reinforcement learning. *IEEE Transactions on Power Systems*, 35(3):1990–2001, 2020.

[144] Hanchen Xu, Alejandro D. Domínguez-García, Venugopal V. Veeravalli, and Peter W. Sauer. Data-driven voltage regulation in radial power distribution systems. *IEEE Transactions on Power Systems*, 35(3):2133–2143, 2020.

[145] Xiaoyuan Xu, Yunhong Li, Zheng Yan, Hongyan Ma, and Mohammad Shahidehpour. Hierarchical central-local inverter-based voltage control in distribution networks considering stochastic pv power admissible range. *IEEE Transactions on Smart Grid*, pages 1–1, 2022.

[146] Yan Yang, Zhifang Yang, Juan Yu, Baosen Zhang, Youqiang Zhang, and Hongxin Yu. Fast calculation of probabilistic power flow: A model-based deep learning approach. *IEEE Transactions on Smart Grid*, 11(3):2235–2244, 2020.

[147] Zhu Yang and Shmuel Oren. Line selection and algorithm selection for transmission switching by machine learning methods. In *2019 IEEE Milan PowerTech*, pages 1–6, 2019.

[148] M. Yazdanian and A. Mehrizi-Sani. Distributed control techniques in microgrids. *IEEE Transactions on Smart Grid*, 5(6):2901–2909, 2014.

[149] Haishan Ye, Chaoyang He, and Xiangyu Chang. Accelerated distributed approximate newton method. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[150] Ziyang Yin, Xingquan Ji, Yumin Zhang, Qi Liu, and Xingzhen Bai. Data-driven approach for real-time distribution network reconfiguration. *IET Generation, Transmission & Distribution*, 4(13), March 2020.

[151] Ahmed S. Zamzam and Kyri Baker. Learning optimal solutions for extremely fast ac optimal power flow. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6, 2020.

[152] Michael Zarghamy, Alejandro Ribeiroy, and Ali Jadbabaiey. Accelerated dual descent for constrained convex network flow optimization. In *52nd IEEE Conference on Decision and Control*, pages 1037–1042. IEEE, 2013.

[153] Baosen Zhang, Albert Y.S. Lam, Alejandro D. Domínguez-García, and David Tse. An optimal and distributed method for voltage regulation in power distribution systems. *IEEE Transactions on Power Systems*, 30(4):1714–1726, 2015.

[154] Junbo Zhao, Fei Ding, Hao Zhu, and Shunbo Lei. Behind-the-meter distributed energy resources: Estimation, uncertainty quantification, and control. Technical report, IEEE Power and Energy Society, 2023.

[155] Tianyu Zhao, Xiang Pan, Minghua Chen, Andreas Venzke, and Steven H. Low. Deepopf+: A deep neural network approach for dc optimal power flow for ensuring feasibility. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–6, 2020.

[156] Xinyang Zhou, Zhiyuan Liu, Changhong Zhao, and Lijun Chen. Accelerated voltage regulation in multi-phase distribution networks based on hierarchical distributed algorithm. *IEEE Transactions on Power Systems*, 35(3):2047–2058, 2020.

[157] Yuhao Zhou, Bei Zhang, Chunlei Xu, Tu Lan, Ruisheng Diao, Di Shi, Zhiwei Wang, and Wei-Jen Lee. A data-driven method for fast ac optimal power flow solutions via deep reinforcement learning. *Journal of Modern Power Systems and Clean Energy*, 8(6):1128–1139, 2020.

[158] Hao Zhu and Hao Jan Liu. Fast local voltage control under limited reactive power: Optimality and stability analysis. *IEEE Transactions on Power Systems*, 31(5):3794–3803, 2016.

[159] Owen Zinaman, Mackay Miller, Ali Adil, Doudlas Arent, Sonia Aggarwal, Minnesh Bipath, Carl Linvill, Ari David, Richard Kauffman, Matt Futch, Efrain Villaneuva Arcos, Jose Maria Valenzuela, Eric Martinot, Morgan Bazilian, and Reji Kumar Pillai. Power systems of the future: A 21st century power partnership thought leadership report. Technical report, NREL, February 2015.