

**Coordinated Control of the Modern Grid:
High-bandwidth Optimal Coordination, Inference, and
Verification**

**A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY**

Blake R. Lundstrom

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

Murti V. Salapaka

April, 2020

© Blake R. Lundstrom 2020
ALL RIGHTS RESERVED

Acknowledgements

There are many people that have been a great help to me in completing this doctoral research. I first want to thank those educators that inspired me to begin and continue this academic journey. I've been privileged to learn from many talented teachers that helped give me the knowledge I needed to succeed, but I especially want to thank Mr. Briant McKellips, teacher of Physics at D'Evelyn Jr./Sr. High School, who thoroughly convinced me, and so many others, that "Fysics is Phun!", but also encouraged us to think beyond basic equations, be inquisitive, and see the bigger implications and possibilities of what we were learning. His thoughtful lectures and kind encouragement, including a prescient statement that perhaps I should consider electrical engineering, planted the seeds of my interest in graduate school and ultimately a career in research and development. I also want to thank Professors Ravel Ammerman and P.K. Sen whose inspired teaching, excellent advising, and kind encouragement spurred on my thoughts of graduate studies, helped lead me to a research field I am passionate about, and ultimately advised me on my Master's Thesis.

I offer my sincere thanks to my Ph.D. advisor, Professor Murti Salapaka, who has helped me become a more mature researcher. His rigorous and thoughtful approach and passion for solving meaningful problems has been an inspiration. I also want to thank Professors Andrew Lamperski, Mariko Shirazi, and Peter Seiler for giving their time and offering helpful feedback as members of my thesis committee.

I am grateful to Dr. Ben Kroposki and Dr. Barry Mather at the National Renewable Energy Laboratory (NREL) who have been patient supporters and excellent mentors during my ten year journey of working full time at NREL and being a part-time graduate student. I am thankful for the financial support from NREL during my graduate studies and the project financial support from the U.S. Advanced Research Projects Agency-

Energy that funded some of the work in this thesis. Thanks also to Beth Chacon and team at Xcel Energy who kindly provided real-world power systems modeling data and a very helpful practical advisory perspective that greatly strengthened the validation portion of some of this work.

I want to thank all my colleagues in the Nanodynamics Systems Laboratory at the University of Minnesota—especially Sourav Kumar Patel, Govind Sarawat, Sandeep Attree, Mangal Prakash, Saurav Talukdar, Sivaraman Rajaganapathy, and Shreyas Bhaban—for their helpful collaborations and kind encouragements.

Finally, I thank my wife Rachel, my parents Richard and Jill, and my brother Forrest, for their continual love and support, despite the extra stresses brought on by another ten years of work-school. Without their support, this work would not have been possible. And thanks to many friends not listed here who have been an encouragement along the way. I am grateful to you all.

Abstract

Modernization of electric grids worldwide continues in earnest, with increasing deployments of renewable and distributed energy resources (DERs); flexible loads with evolving usage patterns; and new control solutions in distribution networks. The convergence of these new technologies inherently increases the complexity of the power system; and maintaining safe, reliable, and efficient operation of modern grids relies on the successful integration of these many technologies together. Achieving successful modern grid control solutions requires not only the development of innovative distributed coordination techniques, but also methods to effectively and accurately test such complex solutions prior to field deployment and approaches to achieve cost-effective, high-bandwidth, inference capabilities in buildings and distribution grids.

In the first part of this thesis, we make important contributions to the design of power hardware-in-the-loop (PHIL) simulations, an important tool for derisking modern grid technologies in a controlled laboratory setting. This technique allows actual at-power devices and systems to be interconnected in software simulations of complex, large-scale power system scenarios to evaluate their closed-loop interaction. We develop a novel approach to designing PHIL simulation interfaces that maximizes simulation bandwidth and accuracy by leveraging a modern control framework that explicitly considers objectives on accuracy and can automatically synthesize a single, optimal controller that meets these objectives while stabilizing the closed-loop system. This method improves upon common approaches to PHIL interface design that typically involve multiple steps of manual compensation and stabilization design that can result in interfaces that are stable but have suboptimal bandwidth and accuracy. The approach developed is general and can be applied to most PHIL system configurations. We also present a practical method and metrics for verifying the true accuracy of PHIL simulation results without relying on relative comparisons to potentially inaccurate models or previous simulation results. We demonstrate the accuracy evaluation method and show the improved performance achievable when using the optimal PHIL interface design approach in an experimental case study involving a 100-kVA battery inverter.

The second part of this thesis develops a novel approach for high-frequency, multi-class nonintrusive load monitoring (NILM) that enables effective net-load monitoring capabilities with minimal additional equipment and cost. Relative to existing NILM work, the proposed solution operates at a faster timescale, providing accurate multiclass state predictions for each 60-Hz ac cycle without relying on event-detection techniques. We also introduce an innovative hybrid classification-regression method that allows for the prediction of not only load on/off states via classification but also their individual operating power levels via regression. The overall approach is validated using a test bed with eight residential appliances and shown to have high accuracy, good scaling and generalization properties, and sufficient response time to support building grid-interactive control at fast timescales relevant to the provision of grid frequency support services.

The third part of this thesis develops and experimentally demonstrates a first-of-its-kind hierarchical control solution for optimally dispatching thousands of deferrable loads and distributed energy resources (DERs) across a distribution feeder to provide fast frequency response (FFR) within 500 ms to the bulk power system. This approach rapidly coordinates resources online after a frequency event occurs, allowing fast-changing, behind-the-meter (BTM) resources to be incorporated and aggregate FFR power set points to be achieved more quickly and accurately than existing approaches. We also present a solution for determining the optimal amount of headroom to operate solar inverters with to minimize opportunity cost while ensuring the FFR response viability of a building with the inverter and deferrable loads. We develop practical algorithms for fast, cost-based optimal dispatch at multiple aggregation scales, establish their optimality, and demonstrate via simulation that they are faster than state-of-the-art, coordinated frequency response approaches. The entire control platform developed is implemented and experimentally verified using a unique PHIL demonstration, including more than 100 powered loads and DERs connected to a real-world distribution network model. Experimental results from multiple scenarios confirm that the optimal FFR dispatch approach scales well and can optimally coordinate more than 10,000 net-load resources across a distribution network while achieving hardware response times within 500 ms, which is not possible using existing optimal coordination approaches.

Contents

Acknowledgements	i
Abstract	iii
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Thesis Overview	3
1.2 Thesis Contributions	4
2 Optimal Power Hardware-in-the-Loop Interfacing: A Modern Control Approach For Design and Verification of High-Accuracy Interfaces	7
2.1 Hardware-in-the-Loop Simulation Background	7
2.2 Introduction	10
2.3 Optimal PHIL Interface Design and Analysis	15
2.3.1 Overview	15
2.3.2 System Modeling	16
2.3.3 PHIL Interface Design	18
2.3.3.1 Objectives	18
2.3.3.2 Approach	18
2.3.3.3 H_{fb} Compensator design	19
2.3.3.4 Preparing the generalized model	20
2.3.3.5 K_v compensator design	22

2.3.4	Stability Analysis	23
2.4	Accuracy Evaluation	24
2.5	Experiment Case Study	25
2.5.1	Experiment Configuration	25
2.5.2	System Modeling	26
2.5.3	PHIL Interface Design	26
2.5.3.1	Base case interfaces	26
2.5.3.2	Optimal (“Opt”) Interface	27
2.5.4	Stability Analysis	28
2.5.5	Experiment Results	28
2.5.5.1	Summary Steady-state Results	29
2.5.5.2	Detailed Dynamic Results	32

3	Scalable Hybrid Classification-Regression Solutions for High-Frequency Nonintrusive Load Monitoring in Grid-Interactive Residential Buildings	35
3.1	Introduction	35
3.2	High-Frequency Nonintrusive Load Monitoring Approach	39
3.2.1	Problem Formulation	39
3.2.2	Experiment Configuration	40
3.2.3	Data Sets	41
3.2.4	Approach	41
3.2.4.1	Input Data Processing	42
3.2.4.2	Labeling (Training Phase)	43
3.2.4.3	Feature Engineering	43
3.3	High-frequency, Multiclass Classification of Residential Appliance Loads	47
3.3.1	Classifier Training and Tuning	48
3.3.2	Performance Metrics	48
3.3.3	Results	49
3.3.4	Classifier Generalization	49
3.3.5	Transient Performance	51
3.4	High-frequency, Binary Classification of Residential Appliance Loads . .	55

3.4.1	Classifier Tuning and Training	55
3.4.2	Results	55
3.5	Power Prediction Using Regression	57
3.5.1	Problem Formulation	57
3.5.2	Model Tuning and Training	57
3.5.3	Results	59
4	Distribution Feeder-Scale Fast Frequency Response via Optimal Co-ordination of Net-load Resources	61
4.1	Introduction	61
4.1.1	Motivation and Background	61
4.1.2	Related Work	63
4.1.3	Contribution and Organization	65
4.2	Platform for Fast Frequency Response from Optimally-coordinated Net-load Resources	66
4.2.1	Solution Overview	66
4.3	Net-load Unit Definition	68
4.4	Optimal NLU Inverter Headroom	70
4.4.1	Problem Formulation	71
4.4.2	Solution	72
4.5	Optimal Fast Frequency Response from a Single Net-load Unit	76
4.5.1	Dispatch Problem Formulation	76
4.5.2	Solution	77
4.5.3	Solution Equivalency Proof	78
4.6	Optimal Fast Frequency Response from Multi-building Net-load Groups	81
4.6.1	Problem Formulation	81
4.6.2	Solution	82
4.6.3	Performance Verification	83
4.6.3.1	Single Net-load Group Case	84
4.6.3.2	Full Feeder Case	84
4.7	Optimal FFR Dispatch PHIL	88
4.7.1	Power Hardware Under Test	88

4.7.1.1	Laboratory Homes (HG1–HG2)	88
4.7.1.2	DER Racks (HG3–HG8)	89
4.7.1.3	Utility-Scale PV and Battery Inverters (HG9–HG10)	90
4.7.2	Real-Time Simulation	91
4.7.2.1	Electric Distribution Network Model	92
4.7.2.2	Simulated Devices and NLUs	94
4.7.3	PHIL Interface	94
4.7.4	Hierarchical Control Platform	95
4.8	Single Net-load Group Experiment Results	97
4.8.1	Scenario	97
4.8.2	Results	98
4.8.2.1	Summary	98
4.8.2.2	Single NLU Optimal MPP Tracking	98
4.8.2.3	NLG Optimal FFR Dispatch	100
4.9	Full Distribution Feeder Experiment Results	104
4.9.1	Scenario	104
4.9.2	Results	104
4.9.2.1	Summary	104
4.9.2.2	Distribution Feeder-Scale Optimal FFR Dispatch	106
5	Conclusion and Discussion	111
	References	114

List of Tables

2.1	Instantaneous Waveform Error Metrics	25
2.2	Steady-State Error Metrics	25
2.3	System Modeling Elements	26
2.4	Summary test results	31
3.1	Load conditions for eight-load demonstration	42
3.2	Hyperparameters for different models	48
3.3	Overall Test Results (10-Run Average)	49
4.1	Single NLG case: Summary simulation results	85
4.2	Distribution Network Model Characteristics	93
4.3	Single NLG Experiment Results	98
4.4	NLU6 Dispatch Scenarios	101
4.5	Experiment Configuration Summary	105
4.6	Full Feeder Experiment Results	106

List of Figures

2.1	Block diagram showing the use of software-in-the-loop (SIL), CHIL, PHIL, and co-simulation techniques for integrated smart grid system testing (Graphic from [1])	9
2.2	PHIL simulation configuration for an experiment evaluating a four-quadrant, 100-kVA battery inverter system (the HUT) interconnected at various points (with source impedances $Z_{s,min}$, $Z_{s,mid}$, and $Z_{s,max}$) in a simulated electric distribution network. This configuration uses the voltage ideal transformer model interface algorithm to interconnect a real-time simulated power distribution network (represented by a Thévenin equivalent model) with the HUT through the power interface. For the forward voltage path, the reference voltage, v_o^* , at the point of interconnection in the simulated power system model is used to provide a physical reference signal, $\hat{v}_o^*$, to the power amplifier in order to achieve output voltage, v_o , at the terminals of the HUT. A measurement of this output voltage, $\hat{v}_o$, is used as feedback to the forward voltage path compensator. The HUT responds to the voltage at its terminals and its measured current response, i_o^* , is measured as $\hat{i}_o^*$ and used as a reference for the current injection, i_o , back into the simulated power network model.	12
2.3	Transfer function block diagram of the PHIL simulation	17
2.4	General control configuration showing (a) the generalized plant and controller interconnection and (b) the closed-loop forms	20
2.5	Weighting functions used for $\mathcal{H}_\infty$ synthesis of G_v . W_P is designed to emphasize performance at the nominal frequency and W_u is designed to limit control effort to the bandwidth of the power interface.	28

2.6	Synthesized K_v reduced to 11th order	29
2.7	Phase A instantaneous voltage, voltage error (IVTE), current (from the Optimal “Opt” vs. CVI interface), and current error (ICTE) throughout a 41 kW to 93 kW change in HUT active power	33
2.8	Magnified view of Figure 2.7. Note the significantly better tracking (phase and bandwidth) shown in the Opt Current vs. CVI Current plots.	34
3.1	Comparison of waveform (top plot) and RMS values (bottom plot) from two observations (one 60-Hz ac cycle sampled at 200 kHz) representing two different load configurations in a residential building.	37
3.2	All observations (all training and test data) with their RMS current feature plotted vs. class for a residential building load classification case study.	38
3.3	Experimental configuration with eight residential appliance loads. I_{TOT} measures the total current flowing through the eight loads.	40
3.4	High-frequency NILM approach. Here Model refers to the classification and/or regression model.	46
3.5	Test accuracy (full test data set) of Classifier 1 by class for the four-load experiment configuration when using all training data.	50
3.6	Classifier performance on Class 6 ($Y = [0, 1, 1, 0]$) test data when the middle (by RMS value) 95% of Class 6 instances in the training data set are removed before training. Accuracy is 97%.	50
3.7	Test accuracy by class for the four-load experiment configuration when the middle (by RMS value) 95% of each class instance in the training data set is removed before training. By adding additional features to Classifier1 versus the limited subset used by Classifier B, classification generalization performance can be significantly improved.	51
3.8	Entire test Data Set 5 with predicted versus true labels. It can be seen that the classifier performs well, even with a variety of load transitions.	53
3.9	Magnified view of test Data Set 5 near $t = 133$ s. The classifier is able to predict well even though the turn-on transient of Load 4 exhibits a variety of wave shapes and magnitudes.	54

3.10	Median test accuracy with error bars for Classifier 2, DNN, and the hybrid model for first approach (80/20 split). The hybrid model and DNN curves are shifted slightly left and right, respectively, for visibility.	56
3.11	Median test accuracy with error bars for Classifier 2, DNN, and the hybrid model for second approach (independent data set for testing). The hybrid model and DNN curves are shifted slightly left and right, respectively, for visibility.	56
3.12	Hybrid classification-regression modeling approach. It can be seen that output of the classifiers are used by the regressors to predict individual load RMS current.	58
3.13	True and predicted RMS values for Load 8 for 80/20 training split . . .	59
3.14	True and predicted RMS values for Load 8 with Data Set 3 used for testing	60
4.1	Hierarchical optimal net-load coordination for FFR approach	68
4.2	Hierarchical optimal net-load dispatch approach	86
4.3	Full distribution feeder case: Percent difference in dispatch cost achieved by the hierarchical approach versus the MILP optimal solution for varying participation factor (10%–45%).	87
4.4	Full distribution feeder case: Solution time vs. number of transformers (problem size). (a) Comparison of hierarchical approach (“ha”) and MILP solution (“milp”) average solution time based on 155 runs for each case. (b) Violin plot showing distribution of hierarchical approach’s solution times for varying participation factor (10%-45%) for 155 runs each.	87
4.5	PHIL experiment platform, including control, power hardware under test, PHIL interface, and real-time simulation components (see labels at left)	89
4.6	Photos of the two lab homes (HG1 and HG2) with residential appliance loads controlled denoted. One Type 4 (custom) inverter (pictured at right inset) is used with each HG.	90
4.7	Photos of equipment comprising three of the six DER racks (HG3–HG8).	91
4.8	Photo of equipment comprising two utility-scale PV+battery inverter systems (HG9–HG10).	92

4.9	Trial #2 optMPP results: (top) Power output for inverter in optimal MPP tracking mode (“optMPP”) versus regular mpp tracking mode (“reg”). Measured experimental data (“Meas”) and predicted simulation results (“sim”) are compared. (middle) The real-time calculated α . (bottom) Load profiles for the five NLU loads.	99
4.10	Trial #4 optMPP results	100
4.11	Trial #4 NLG dispatch results: RMS active power for the (a) three dispatched NLUs, (b) six nondispatched NLUs, (c) constituent loads of NLU2, (d) NLU inverters, and (e) the total NLG (all NLUs)	102
4.12	Trial #4 NLG dispatch results: Instantaneous waveforms (top) NLU1 inverter current (“IL1”), NLU2 inverter current (“M1SMA”), grid voltage (“VG”) and approximate grid frequency (“FREQ”). “FTRIG” goes from low to high when the local inverter controller detects the frequency event. “PTRIG” goes from high to low when the local inverter controller receives its dispatch set point from the NLGC. (bottom) Four NLU1 constituent load currents (“FRIDGE”, “WASHER”, “RANGE”, and “AIRCON”), the total load current of NLU1 (“HM3”), and the total load current of NLU2 (“HM2”).	103
4.13	Full feeder results: RMS active power response for the more than 27 different hardware inverters in NLU1-NLU32	108
4.14	Full feeder results: oscilloscope view	109
4.15	Full feeder results: RMS active power response for the appliance loads in NLU2, a selection of the 224 NLGs, all residential and commercial NLUs, and at the distribution feeder head at the substation. Voltages from a selection of points across the full distance from substation are also shown.	110

Chapter 1

Introduction

Modernization of electric grids worldwide continues in earnest, with increasing deployments of renewable and distributed energy resources (DERs); flexible loads with evolving usage patterns; and new control solutions in distribution networks. The convergence of these new technologies inherently increases the complexity of the power system; and maintaining safe, reliable, and efficient operation of modern grids relies on the successful integration of these many technologies together. These modern grid technologies bring both new challenges and new opportunities.

In many grids worldwide, increasing installation of DERs has resulted in the displacement of conventional, high-inertia, synchronous generation by inverter-interfaced low-zero-inertia renewable generation, resulting in a decrease in total system inertia and a decline in frequency response performance from historic values. Thus, regulating the grid’s frequency to maintain reliable operation is becoming more challenging and new solutions that include the participation of demand-side resources, including DERs and flexible loads, are required. Coordinating thousands of distributed resources together to provide grid support is much more challenging than the traditional approach of managing a relatively-small number of large-capacity synchronous generators, but it does offer new opportunities to better optimize system operation and open new value streams for distributed resources.

In this thesis, we develop a novel control solution for optimally coordinating thousands of DER and flexible loads (“net-loads”) across a power distribution network at very fast timescales. We first consider the residential or building scale coordination for

frequency response problem and devise algorithms with provable optimality that we experimentally show to be viable at fast frequency response (FFR) timescales (<500 ms). Note that most existing high-speed frequency response solutions are often autonomous, offering no opportunity for optimal coordination, and existing coordinated response solutions are often an order of magnitude slower than the FFR achieved here. The solution engine is then scaled up in a hierarchical framework, with optimality guaranteed at the subsystem scale, to be applicable at an entire power distribution system scale with tens of thousands of devices being managed. We experimentally validate that the approach yields coordinated response at the distribution grid scale within FFR timescales with minimal loss in optimality. This approach rapidly coordinates resources online after a frequency event occurs, allowing fast-changing, behind-the-meter (BTM) resources to be incorporated and aggregate FFR power set points to be achieved more quickly and accurately than existing approaches. A solution for determining the optimal amount of headroom to operate solar inverters with to minimize opportunity cost while ensuring the FFR response viability of a building with the inverter and deferrable loads is also presented.

We further consider two challenges that arise in achieving large-scale coordination solutions: how to effectively and accurately test such a complex solution prior to field deployment and how to achieve cost-effective, high-speed, inference capabilities in buildings where net-loads are interconnected to enable large-scale coordination solutions.

Power-hardware-in-the-loop (PHIL) simulation is an important tool for derisking modern grid technologies in a controlled laboratory setting in which actual at-power devices and systems are interconnected in software simulations of complex, large-scale power system scenarios. Using PHIL, the closed-loop interaction between hardware under test (HUT) and a simulated power system can be evaluated under a variety of grid configurations and repeatable operational scenarios without the expense or risk of a field deployment; however, for the simulation results to be representative of expected real-world operations, the power interface between the simulated system and the HUT must be carefully implemented to ensure stable and accurate closed-loop simulation. We make important contributions in developing new modern control-based methods to achieve stable and accurate PHIL simulations. This part of the thesis contributes toward the verification methodologies of the modern and future grid by ensuring a stable

simulation framework with optimal performance. The thesis also creates important connections between modern control synthesis methods and the vast area related to PHIL: for the first time the thesis brings robust control tools to bear on PHIL.

Real-time measurement and state awareness capabilities are frequently required for effective system-level control solutions, yet very little of this infrastructure currently exists at the building and distribution system levels. Enabling circuit-level net-load metering at buildings provides better understanding of building performance and customer usage patterns while enabling the critical monitoring of information needed to perform meaningful grid-interactive control. The simplest approach to achieve such metering is to connect a sensor at each load, but this is invasive and very costly. Nonintrusive load monitoring (NILM) can extend many benefits of smart buildings to homes using minimal additional equipment and cost. With this method, a single electrical measurement point at the building’s point of common coupling with the grid is used to infer information (e.g., on/off status, electricity consumption) about the constituent net-load devices in the home. In this thesis, we make novel developments to the field of NILM that enable high-bandwidth, real-time inference of individual on/off status and operating power levels for a group of loads within the same building using only a single aggregate measurement. This machine learning based methodology is a key tool in enabling implementation of our optimal coordinated net-load management scheme in a cost-effective manner in real-world deployments.

1.1 Thesis Overview

Chapter 2 introduces the power hardware-in-the-loop simulation validation approach, including the challenges of achieving stable and accurate simulations. A key challenge is to properly interface the HUT to the simulated model through the power interface. A review of existing approaches to achieve stable simulations and to evaluate the accuracy of simulation results is presented. We develop a novel approach to systematically design PHIL simulation interfaces that achieve stable and accurate simulations and also introduce a method and metrics for verifying their accuracy. PHIL simulation is used to evaluate the optimal DER and flexible load coordination solution developed in Chapter 4. A dedicated experimental validation demonstrating the effectiveness of this PHIL

interface design and verification approach is also presented.

Chapter 3 considers the problem of enabling cost-effective, high-bandwidth inference of multiple loads' on/off status and power consumptions using a single aggregate measurement. We first review existing nonintrusive load monitoring approaches that have largely focused on classification methods to determine on/off status only and are not sufficiently fast for <1 -s grid support response applications. We introduce a novel hybrid classification-regression technique that allows for the prediction of load current consumption levels via regression in addition to the load on/off states via classification and that operates at a high-speed timescale with predictions for each 60-Hz cycle. This solution provides a cost-effective method to monitor multiple net-loads within a building, as is needed for the net-load optimal coordination solution of Chapter 4. Experimental results from an experiment configuration with eight residential appliances are used to validate the approach.

Chapter 4 discusses the challenges of regulating frequency in modernized grids and provides a thorough literature review on existing approaches to regulating frequency, including those that include the participation of demand-side resources. Existing approaches to frequency regulation that incorporate demand-side resources are typically either fast but uncoordinated or coordinated but too slow for fast frequency response (FFR) applications. We develop a system-level solution for optimally coordinating tens of thousands of net-load resources that can achieve very fast (<500 ms) timescale optimal coordination that is suitable for FFR as well as other grid support applications. This solution leverages a hierarchical control structure, which allows for consistently fast response times as the number of managed devices grows. The entire platform is implemented and experimentally verified using a unique PHIL demonstration including more than 100 powered net-loads connected to a real-world distribution network model.

1.2 Thesis Contributions

This thesis makes several contributions towards modern power system and device control, inference, and validation capabilities:

1. **Improved Approach for Attaining Stable and Accurate PHIL Simulation**

- *New PHIL Interface Design Method*: We develop a general, systematic approach to designing PHIL simulation interfaces that explicitly considers objectives on accuracy and enables the synthesis of a single, optimal controller that stabilizes the closed-loop system while maintaining maximum simulation bandwidth and accuracy. This method improves upon common approaches to PHIL interface design that typically involve multiple steps of manual compensation and stabilization design that can result in interfaces that are stable but have suboptimal bandwidth and accuracy.
- *Accuracy Evaluation Approach*: We introduce a practical method and metrics for verifying the true accuracy of PHIL simulation results that is easily applied to actual experiments and can be used to compare results from two different simulations on a common basis. This approach improves upon existing PHIL simulation accuracy evaluation methods that rely on relative comparisons to potentially inaccurate models or previous simulation results.

2. High-bandwidth Nonintrusive Load Monitoring

- *Faster Timescale NILM*: In most existing NILM approaches, system state predictions are made in no faster than 1 s to 15 min and use underlying models that cannot make accurate state predications until after an entire transient event (e.g., an appliance’s typical load profile) has occurred. We develop a high-bandwidth approach that is able to make accurate predictions at a 60-Hz timescale and doesn’t rely on event detection.
- *NILM with Both State and Operating Power Level Predictions*: A hybrid classification-regression approach is developed that can use an aggregate current measurement to predict both the on/off state (via classification) and the operating RMS current (via regression) for multiple loads on a 60-Hz timescale. To the best of our knowledge, there is no method currently capable of load power disaggregation at the timescale of <1-s grid ancillary services.

3. First-of-its-kind Demonstration of Distribution Feeder-Scale Fast Frequency Response via Optimal of Net-load Resources

- *Optimal, Large-scale Coordination Approach:* We develop an approach to optimally coordinate tens of thousands of deferrable loads and DERs across a distribution network to provide bulk power system-scale FFR that is dispatchable and accurate. Compared to existing frequency response solutions, the developed control platform is able to optimally coordinate a much larger number of net-load resources (up to 40,000 is shown), incorporate fast-changing behind-the-meter resources, and accurately meet a specific aggregate power set point while also achieving dispatch on a faster timescale (<500 ms) suitable for FFR.
- *Algorithm with Performance Guarantees:* We provide an analytical proof of the atomic control algorithm's optimality and then verify near-optimal (within 0.1%) performance of the algorithm as it is scaled to configurations with up to 40,000 net-load resources via Monte Carlo-style simulation. In all cases, dispatch solutions are achieved faster than traditional mixed-integer linear program optimization approaches and rapidly enough to support coordinated FFR within 500 ms.
- *Implementation and Large-scale Demonstration:* Many existing optimal DER coordination approaches in the literature are validated using only simulation results that do not account for the practical challenges of interfacing and coordinating real power and control hardware. Here, the entire control platform is implemented and experimentally verified using a unique PHIL demonstration including more than 100 powered loads and DERs connected to a real-world distribution network model from Colorado, USA.

Chapter 2

Optimal Power

Hardware-in-the-Loop

Interfacing: A Modern Control

Approach For Design and

Verification of High-Accuracy

Interfaces

2.1 Hardware-in-the-Loop Simulation Background

Hardware-in-the-loop (HIL) simulation involves closed-loop simulation of one or more physical hardware devices connected to a software-only subsystem executed on a real-time simulator (RTS). Hardware under test (HUT) is interfaced to the RTS using analog or digital communication, which is sampled deterministically by the RTS. Controller hardware-in-the-loop (CHIL) involves an embedded controller or other control device interfaced at signal-level (e.g., <50 V) voltages as the HUT. Power hardware-in-the-loop (PHIL) involves a physical device interfaced at higher voltages (power) as the HUT. Examples of CHIL and PHIL applications are shown in Figure 2.1. The advantage of PHIL

over CHIL is that it allows for closed-loop evaluation of a power device in its native form (e.g., a physical power inverter rather than just the embedded controller of the inverter), which is as close as possible to final implementation. With the requirement of additional power interfacing hardware, which has its own physical hardware limitations, careful attention must be paid to ensure that PHIL simulations are both stable and accurate; additional interfacing algorithms and compensation techniques are often required for this purpose.

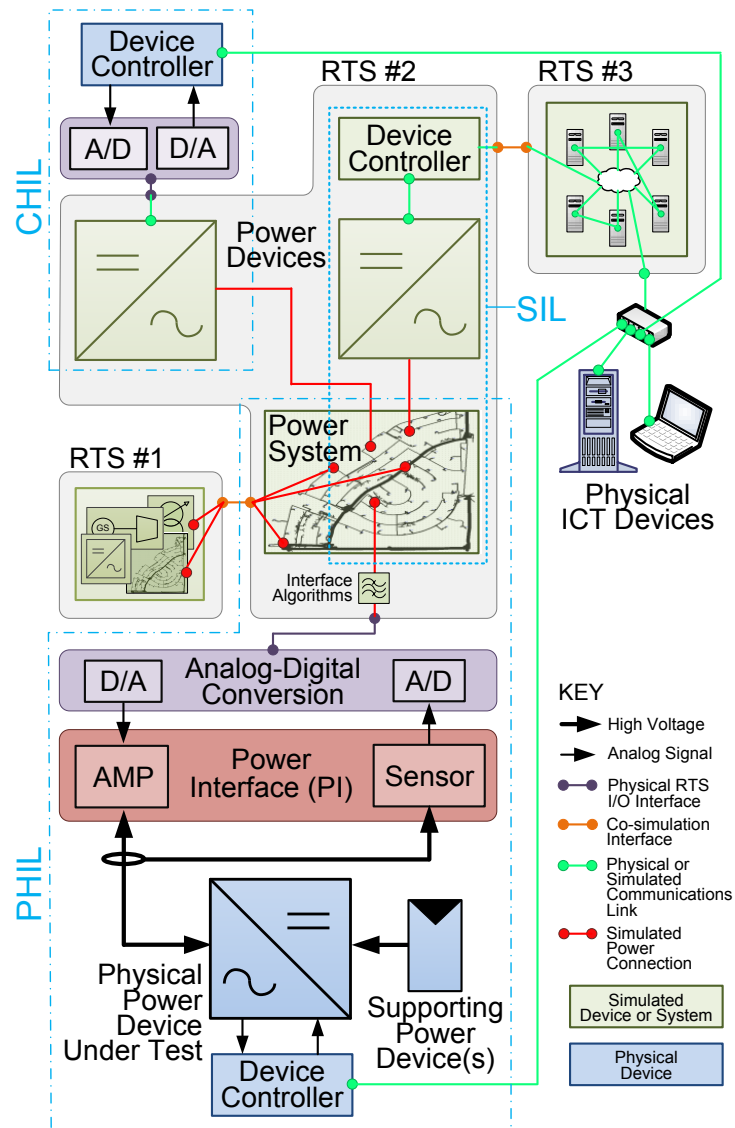


Figure 2.1: Block diagram showing the use of software-in-the-loop (SIL), CHIL, PHIL, and co-simulation techniques for integrated smart grid system testing (Graphic from [1])

2.2 Introduction

PHIL simulation is an important tool for derisking modern grid technologies in a controlled laboratory setting in which actual at-power devices and systems are interconnected in software simulations of complex, large-scale power system scenarios. An example PHIL simulation involving a physical, four-quadrant, 100-kVA battery inverter system interconnected at various points in a simulated electric distribution network (see Figure 2.2) is considered. Using PHIL, the closed-loop interaction between the HUT and the simulated power system can be evaluated under a variety of grid configurations and repeatable operational scenarios without the expense or risk of a field deployment; however, for the simulation results to be representative of expected real-world operations, the power interface between the simulated system and the HUT must be carefully implemented to ensure stable and accurate closed-loop simulation.

In the PHIL literature, significant attention is paid to designing solutions that achieve a stable closed-loop simulation because stability is a necessary condition of any PHIL simulation. Common approaches include augmenting the underlying interface algorithm (i.e., see [2, 3, 4]) used to connect the simulated model to the HUT to enhance stability by adding series inductance to the HUT [5]; moving a portion of the simulated system's impedance to the HUT [6]; adding low-pass filtering of the feedback current [7, 5, 8]; including parallel high-pass filtering [9] or virtual impedance [10] at the power interface; using impedance matching [11, 12]; or employing multirate partitioning [13]. When designing stable PHIL simulation interfaces, however, often the implications of a design on the PHIL simulation's accuracy are not fully considered. Here, we define PHIL simulation accuracy to be how closely the power interface's simulated components, $v_o^*(t)$ and $i_o(t)$, and measured components at the HUT, $v_o(t)$ and $i_o^*(t)$, track one another. For example, early designs of the feedback current filtering (e.g., [7]) technique demonstrated its ability to improve the stability of PHIL simulations using the voltage ideal transformer model (V-ITM) interface algorithm, but results showed significant phase lag and corresponding errors in voltage and current waveforms of approximately 5% and 8%, respectively. Even a seemingly small error in the forward voltage and feedback current can result in a significant error when considering the practical comparison of measured HUT power versus the corresponding power injected in the PHIL simula-

tion. For example, three PHIL simulation studies [14, 15, 16] considering active devices as HUT and applying feedback current filtering reported errors of 10%–40% in active and reactive power before further compensation. Many other stability enhancement approaches similarly add phase lag to the closed-loop simulation, which can result in significant error and diminish the value of such simulations, especially in PHIL simulations where the HUT interacts with other simulated control and power systems. It is critical to stabilize PHIL simulations, but PHIL interfaces must be carefully designed and implemented while explicitly considering accuracy to ensure useful and valid results. In this chapter, we develop a systematic interface design approach that allows for explicit accuracy objectives to be defined while achieving closed-loop stability as a necessary condition.

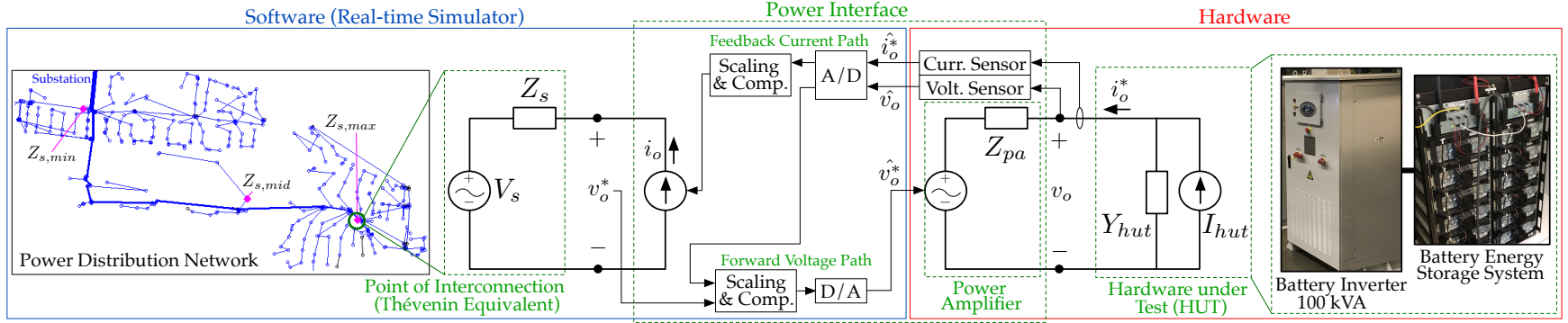


Figure 2.2: PHIL simulation configuration for an experiment evaluating a four-quadrant, 100-kVA battery inverter system (the HUT) interconnected at various points (with source impedances $Z_{s,min}$, $Z_{s,mid}$, and $Z_{s,max}$) in a simulated electric distribution network. This configuration uses the voltage ideal transformer model interface algorithm to interconnect a real-time simulated power distribution network (represented by a Thévenin equivalent model) with the HUT through the power interface. For the forward voltage path, the reference voltage, v_o^* , at the point of interconnection in the simulated power system model is used to provide a physical reference signal, $\hat{v}_o^*$, to the power amplifier in order to achieve output voltage, v_o , at the terminals of the HUT. A measurement of this output voltage, $\hat{v}_o$, is used as feedback to the forward voltage path compensator. The HUT responds to the voltage at its terminals and its measured current response, i_o^* , is measured as $\hat{i}_o^*$ and used as a reference for the current injection, i_o , back into the simulated power network model.

Another potential detriment to achieving accurate PHIL simulations is that many PHIL stability enhancement techniques have been designed using simplified modeling and/or for very specific configurations, and thus they are not simple to generally apply, and they are not optimized for use in simulations evaluating modernized grid system scenarios, which typically involve more complex power systems and devices. Simplified system modeling assumptions often include an idealized power amplifier with no dynamics [17, 18], and they neglect the dynamics or delay of the feedback current sensor (e.g., as in [19, 14, 10]). Ignoring these system modeling assumptions and failing to model their impacts on the closed-loop system’s dynamics can lead to a blunt approach to stable interface design in which stabilizing solutions are overapplied, resulting in lower bandwidth and less accurate simulation results. For example, [15] identified instability in a PHIL experiment that occurred because of resonance between the HUT and the power amplifier’s output filter at a particular frequency. Stability can be achieved by applying a low-pass current feedback filter with a cutoff frequency below the resonant frequency; however, the filter will affect frequencies other than the resonant frequency, thus compromising fidelity of the scenario being simulated. Here the bandwidth of the measured current is needlessly limited and will add phase delay, which will also need to be compensated to achieve useful results (e.g., [14]). Instead, by fully modeling the PHIL interface, more precise stabilization approaches that preserve simulation bandwidth and accuracy can be designed. In [15], the authors employed a manual design to correct this specific resonance and further manual steps to compensate for the phase delay to correct the power mismatch. Instead, we propose using a systematic approach to include all interface component models and then *synthesizing* (instead of manually designing multiple components) a single, optimal controller that stabilizes the closed-loop system while maintaining maximum simulation bandwidth and accuracy.

Many PHIL interface solutions employ the widely used voltage divider circuit consisting of a simulated impedance in series with another impedance as the HUT (e.g., this circuit is used in [20, 21, 18, 22, 23, 19, 24, 25, 26, 7, 27, 12, 9, 3, 28]) as a design case study. This model is easy to use and is appropriate for a simple configuration with two purely passive impedances. However, here, two important aspects of many practical PHIL simulation scenarios are ignored: 1) many HUT are active devices that inject (or sink) current without changing their nominally connected impedance (i.e., the feedback

current is not simply a linear function of the voltage), and 2) scaling factors are often used to amplify the HUT feedback current to represent a larger device in the simulated system. Both aspects can have significant impacts on PHIL simulation stability and accuracy, yet these impacts cannot be considered when evaluating a PHIL interface design using the widely used voltage divider example. In the design approach developed in this chapter, we leverage a modeling framework that explicitly incorporates both important aspects of PHIL simulation scenarios.

Beyond the PHIL interface design stage, rigorous accuracy validation of final PHIL simulation results is often omitted. References [22, 19, 28, 5, 6] make helpful contributions in modeling PHIL simulation accuracy; however, the common approaches they employ including predicting accuracy only for the modeled system, but not for actual results (e.g., [22, 19]); evaluating accuracy only as a comparison between PHIL simulation and modeled results (e.g., [28]); or considering experiment accuracy only on a relative basis among multiple experiments (e.g., [5, 26]) limit their practical use in quantifying how accurate (and therefore valid) an individual PHIL simulation result is. Many of these accuracy evaluation approaches rely on an implicit assumption that models are accurate, but in many cases accurate models of the HUT and possibly other PHIL interface components are not available. Also, most studies only compare between the modeled interconnection point voltage and the voltage at the HUT’s terminals (e.g., [22, 19, 6, 28]). Such evaluations do not provide a complete assessment of closed-loop system accuracy because it considers only the accuracy of the forward path from the simulated model to the power amplifier and ignores common inaccuracies in the feedback path, such as phase shift. For example, it is possible to design a V-ITM interface where the HUT voltage perfectly tracks the model’s reference but still has feedback current with significant phase lag, resulting in the HUT’s actual power not matching the power injected into the PHIL simulation model. In this chapter, we propose a method and metrics to evaluate closed-loop PHIL simulation accuracy that are easily applied to actual experiments, do not rely on accurate models for comparison, and give practical quantification of closed-loop performance.

We now summarize the main contributions of this chapter:

1. We develop a general, systematic approach to designing PHIL simulation interfaces that explicitly considers objectives on accuracy and enables the synthesis of a

single, optimal controller that stabilizes the closed-loop system while maintaining maximum simulation bandwidth and accuracy. The modeling framework used allows the inclusion of practical scaling factors and HUT current injection models common in practical PHIL simulations.

2. We present a practical method and metrics for verifying the accuracy of PHIL simulation results that are easily applied to actual experiments. The method evaluates the performance of the full closed-loop system and uses easy-to-interpret voltage, current, and power-based metrics to quantify PHIL simulation accuracy in terms that can be used to understand the validity of a given simulation's results and benchmark the results against other experiments.
3. We demonstrate the process of applying the PHIL interface design approach and verifying the accuracy of PHIL simulation results on a real-world PHIL simulation case study involving a 100-kVA battery inverter system integrated with an electric distribution system. We examine multiple system interconnection scenarios and HUT operating points to validate the performance of the designed PHIL interface.

The remainder of this chapter is organized as follows: The optimal PHIL interface design approach is presented in Section 2.3, followed by the accuracy evaluation method in Section 2.4. These methods are applied in a practical case study in Section 2.5.

2.3 Optimal PHIL Interface Design and Analysis

2.3.1 Overview

The overall approach employed here is to model the full closed-loop PHIL system with all available detail in a generalized framework and then apply modern control theory—specifically $\mathcal{H}_\infty$ optimal control—to design a single controller that stabilizes and compensates the entire closed-loop system. The advantages of this PHIL interface design approach versus existing approaches are that: 1) a controller is *synthesized* (versus manually designed) using explicit accuracy performance objectives while achieving closed-loop stability as a necessary condition; 2) interface bandwidth and accuracy are optimized by emphasizing and de-emphasizing performance at different frequencies by

designing a single controller based on knowledge of the full closed-loop system instead of multiple stages of manual interface and compensation design; and 3) the approach is general and can be applied easily to most PHIL experiments.

In the remainder of this section, we detail how to apply the proposed interface design approach at key steps in the process of designing and implementing a stable and accurate PHIL simulation. The steps highlighted here provide a thorough overview of the new contributions proposed to the overall PHIL simulation design and verification approach, but they are not comprehensive of the full process. The reader is referred to other key references [2, 29] and the forthcoming IEEE P2004 Recommended Practice [30] for more comprehensive discussions of other important steps for successful PHIL simulation implementation.

2.3.2 System Modeling

For effective interface design and analysis, the PHIL simulation configuration shown in Figure 2.2 must be modeled and should include all possible details for each element in the system’s closed loop. Here, we model the PHIL simulation configuration in the continuous-time domain using the V-ITM interface algorithm and the transfer function block diagram shown in Figure 2.3. The continuous-time domain is most convenient for design, initial stability analysis, and optimal controller synthesis (e.g., most $\mathcal{H}_\infty$ software tools are written for continuous time models), though note that the final PHIL simulation implementation will involve a hybrid model with a discrete-time model coupled to a continuous-time physical system. Note that for three-phase power system configurations, such as that presented in the case study, each phase is modeled separately using the block diagram in Figure 2.3.

The block diagram used here is based on the comprehensive PHIL experiment modeling block diagram presented in our previous work [31] and is similar to that proposed by institutional colleagues in [15]. Although the representation shown includes some control-related additions relative to the most basic V-ITM configuration—notably feed-forward and feedback on the forward voltage path—note that the configuration can be applied to most PHIL experiment configurations, even those without voltage feedback sensing (use $H_{s,v} = 0$) or where the power amplifier does not have an output filter (use $Z_{pa} = 0$). For cases where the HUT seeks to regulate v_o , it may be beneficial to em-

ploy other interface algorithms, such as the current ITM algorithm [2], in place of the V-ITM algorithm used here; however, the optimal interface design approach presented here can still be easily applied. The PHIL system model is unique compared to models used in most existing PHIL literature in that 1) the HUT is represented as a current injection (modeled as a disturbance) behind an impedance rather than only a passive impedance, and 2) voltage and current scaling factors between the simulated model and the physical HUT are explicitly included. These aspects considerably affect the stability and accuracy of the closed-loop simulation, so they must be included for comprehensive design and analysis. In modeling the HUT as a disturbance current injection behind a passive impedance, only the impedance of the HUT's output filter, which can be readily identified using one of the three methods presented in our earlier work [31], must be identified. This is important as more detailed models of the HUT are often not available, a common reason for employing PHIL simulation in the first place.

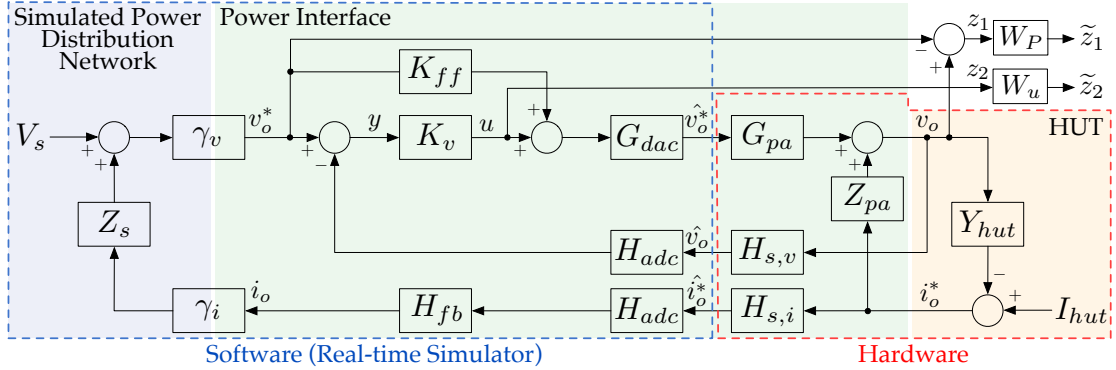


Figure 2.3: Transfer function block diagram of the PHIL simulation

In Figure 2.3, V_s and Z_s are the voltage reference and impedance for a Thévenin equivalent circuit representing the point of interconnection in the PHIL model; γ_v is the voltage scaling factor; K_v is a compensator for the voltage path; K_{ff} is the feed-forward gain for the voltage forward path; G_{dac} is a transfer function representing the digital-to-analog (D/A) conversion delay; $\hat{v}_o^*$ is the output voltage physical reference signal; G_{pa} is the power amplifier's transfer function; Z_{pa} is the output impedance of the power amplifier; Y_{hut} is the output admittance of the HUT; I_{hut} is the current injection from the HUT; $H_{s,v}$ and $H_{s,i}$ are transfer functions representing the voltage and

current sensors, respectively; $\hat{v}_o$ is the measured output voltage; $\hat{i}_o^*$ is the measured HUT current; H_{adc} is a transfer function representing the analog-to-digital (A/D) conversion delay; H_{fb} is a compensator for the current feedback path; and γ_i is the current scaling factor. Our previous work [31] provides a guide on how to identify transfer functions for each component. We summarize the specific models used in the case study section and proceed to interface design with each element represented as labeled in Figure 2.3.

2.3.3 PHIL Interface Design

2.3.3.1 Objectives

For a stable and accurate PHIL simulation of the configuration shown in Figure 2.3, the following objectives should be met:

- **O1. Stability:** The closed-loop PHIL system must be internally stable.
- **O2. Forward path voltage tracking:** $v_o \rightarrow v_o^*$. That is, the voltage at the terminals of the HUT, v_o , should track the voltage at the point of interconnection in the PHIL model (after voltage scaling), v_o^* .
- **O3. Feedback path current tracking:** $i_o \rightarrow i_o^*$. That is, the current injected into the PHIL simulation (before current scaling), i_o , should track the measured current at the terminals of the HUT, i_o^* .

2.3.3.2 Approach

In the block diagram in Figure 2.3, K_v and H_{fb} are the two compensators to be designed to achieve objectives O1–O3. In the modern control approach we employ, it is possible to simultaneously design both by considering K_v and H_{fb} as a single multiple-input, multiple-output compensator; however, for simplicity, and because the optimal form of H_{fb} needed to achieve objective O3 can be easily identified and approximated, we fix the design of H_{fb} and then synthesize a stabilizing, optimal controller for K_v that achieves objectives O1 and O2 using the $\mathcal{H}_\infty$ optimal control framework. Measurement $\hat{v}_o$ is given as feedback to K_v and measurement $\hat{i}_o^*$ is the input to H_{fb} .

2.3.3.3 H_{fb} Compensator design

Compensator H_{fb} should achieve objective O3 by compensating for the A/D conversion delay, H_{adc} , and current sensor response time and dynamics, $H_{s,i}$. It is common for current sensors used in PHIL simulations to have a bandwidth that is significantly greater than other key components in the closed-loop system, such as the power amplifier. Thus, the dynamics of $H_{s,i}$ associated with the current sensor's bandwidth often need not be explicitly compensated for, and instead we focus on the delay portion of $H_{s,i}$ and the A/D conversion delay, H_{adc} . The ideal form of H_{fb} would then be $H_{fb} = e^{sT_{fb}}$, where T_{fb} is the sum of the current sensor and A/D conversion delays; however, this transfer function is noncausal and cannot be realized in practical systems. Instead, we use a simple approach to approximate the same effect; H_{fb} is designed as a second-order, low-pass filter with a relatively high cutoff frequency and an additional zero. The filter is designed to achieve the desired phase lead at the fundamental frequency, ω_0 , (e.g., $\omega_0 = 2\pi 60$ rad/s in the case study presented) while having negligible phase lead and approximately 0 dB gain for all frequencies within the bandwidth of the closed-loop simulation. This is achieved with H_{fb} of the form:

$$H_{fb} = \alpha_{fb} \frac{s + \omega_{fb1}}{(s + \omega_{fb2})^2}, \quad (2.1)$$

where α_{fb} is a gain chosen to achieve 0 dB gain at DC, and $\omega_{fb1} < \omega_{fb2}$. Selecting the best set of parameters might require some iteration because the closed-loop PHIL system model might not be completely accurate. The following approach is suggested:

1. Choose $\omega_{fb2} = 1.5\omega_{bwCL}$, where ω_{bwCL} is the approximate -3 dB bandwidth of the closed-loop system. ω_{bwCL} can be determined using a Bode plot of the closed-loop system model or approximated as the lowest bandwidth of any PHIL interface component in the closed loop (often the power amplifier).
2. Set ω_{fb1} so that the following relation between desired time delay and the phase lead of H_{fb} is satisfied:
$$T_{fb} = \frac{1}{\omega_o} \left[\tan^{-1}\left(\frac{\omega_o}{\omega_{fb1}}\right) - 2 \tan^{-1}\left(\frac{\omega_o}{\omega_{fb2}}\right) \right].$$
3. Set $\alpha_{fb} = \frac{\omega_0^2 + \omega_{fb2}^2}{(\omega_0^2 + \omega_{fb1}^2)^{1/2}}$ to achieve 0 dB gain at DC.
4. (Optional) Experimentally verify how well i_o tracks i_o^* . If additional phase lead

is required (e.g., because the modeled time delay is less than the measured time delay) to achieve phase tracking, increase T_{fb} and repeat steps 1–3. Once the phase tracks well, then ω_{fb2} and α_{fb} can be further adjusted as needed to achieve satisfactory magnitude tracking.

2.3.3.4 Preparing the generalized model

In preparation for G_v compensator design, we simplify the multiloop, closed-loop block diagram in Figure 2.3 into the general control configuration commonly used in modern control theory shown in Figure 2.4(a). Here, P is the generalized plant; K is the controller (e.g., interface compensation design); w is a vector of the exogenous inputs (e.g., commands, disturbances, and noise); z are the exogenous outputs (e.g., signals to be regulated); y are the controller inputs; and u are the control signals. Note that in this continuous-time modeling framework, all variables are functions of the Laplace variable, s , but this is not explicitly shown for notational convenience.

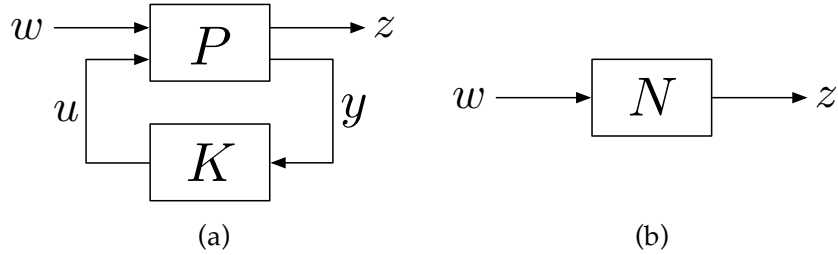


Figure 2.4: General control configuration showing (a) the generalized plant and controller interconnection and (b) the closed-loop forms

For the block diagram in Figure 2.3, the exogenous inputs are $w = \begin{bmatrix} V_s & I_{hut} \end{bmatrix}^\top$, and the exogenous outputs are chosen to be $z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^\top = \begin{bmatrix} (v_o - v_o^*) & u \end{bmatrix}^\top$ because we desire to place a performance target on the error between v_o and v_o^* , and this must be balanced against the control effort, u . The generalized plant maps $\begin{bmatrix} w & u \end{bmatrix}^\top$ to $\begin{bmatrix} z & y \end{bmatrix}^\top$ as follows:

$$\begin{bmatrix} z_1 \\ z_2 \\ y \end{bmatrix} = \begin{bmatrix} v_o - v_o^* \\ u \\ y \end{bmatrix} = \begin{bmatrix} P_{11,1} & P_{11,2} & P_{12} \\ 0 & 0 & 1 \\ P_{21,1} & P_{21,2} & P_{22} \end{bmatrix} \begin{bmatrix} V_s \\ I_{hut} \\ u \end{bmatrix}$$

We then solve for each component of the generalized plant. To begin, consider that v_o can be written as:

$$v_o = G_2 I_{hut} + G_3 u + G_4 V_s \quad (2.2)$$

where:

$$\begin{aligned} G_2 &= G_1^{-1} [Z_{pa} + G_{pa} G_{dac} K_{ff} \gamma_v G_5] \\ G_1 &= 1 + Z_{pa} Y_{hut} + G_{pa} G_{dac} K_{ff} \gamma_v G_5 Y_{hut} \\ G_5 &= Z_s \gamma_i H_{fb} H_{adc} H_{s,i} \\ G_3 &= G_1^{-1} G_{pa} G_{dac} \\ G_4 &= G_3 K_{ff} \gamma_v \end{aligned}$$

Next, see that v_o^* can be expressed as:

$$v_o^* = \gamma_v [V_s + G_5 (I_{hut} - Y_{hut} v_o)] \quad (2.3)$$

and, when substituting (2.2) into (2.3), we obtain:

$$\begin{aligned} v_o - v_o^* &= v_o - \gamma_v [V_s + G_5 (I_{hut} - Y_{hut} v_o)] \\ &= P_{11,1} V_s + P_{11,2} I_{hut} + P_{12} u \end{aligned}$$

where:

$$\begin{aligned} P_{11,1} &= G_4 + \gamma_v (G_5 Y_{hut} G_4 - 1) \\ P_{11,2} &= G_2 - \gamma_v G_5 (1 - Y_{hut} G_2) \\ P_{12} &= (1 + \gamma_v G_5 Y_{hut}) G_3 \end{aligned}$$

Finally, consider that v can be written as:

$$y = (v_o^* - H_{adc}H_{s,v}v_o). \quad (2.4)$$

Then, substituting (2.2) and (2.3) into (2.4), we obtain:

$$y = P_{21,1}V_s + P_{21,2}I_{hut} + P_{22}u$$

where:

$$\begin{aligned} P_{21,1} &= \gamma_v - G_6G_4 \\ P_{21,2} &= \gamma_vG_5 - G_6G_2 \\ P_{22} &= -G_6G_3 \\ G_6 &= \gamma_vG_5Y_{hut} + H_{adc}H_{s,v} \end{aligned}$$

We can close the loop of the generalized control configuration in Figure 2.4(a) to achieve the closed-loop system, N , that maps w to z , as shown in Figure 2.4(b). Here, N is a function of K and can be determined using a lower linear fractional transformation of P with K , $F_l(P, K)$, as follows:

$$N = F_l(P, K) = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}$$

where $P_{11} = \begin{bmatrix} P_{11,1} & P_{11,2} \end{bmatrix}$, $P_{21} = \begin{bmatrix} P_{21,1} & P_{21,2} \end{bmatrix}$, and I is the identity matrix.

2.3.3.5 K_v compensator design

It remains to design compensator H_v to achieve objectives O1 and O2. To do so, we leverage $\mathcal{H}_\infty$ optimal control. This control design approach allows us to shape the magnitude of the system's closed-loop transfer function by synthesizing a controller that minimizes user-defined $\mathcal{H}_\infty$ performance objectives. The user must define reasonable bounds on the designed closed-loop transfer function and then the process of synthesizing a controller can be automated. Here, $\mathcal{H}_\infty$ refers to the $\mathcal{H}_\infty$ norm, which is simply the peak value of a transfer function's magnitude as a function of frequency. That is, if $f(s)$ is a stable scalar transfer function, then its $\mathcal{H}_\infty$ norm is $\|f(s)\|_\infty := \sup_{\omega} |f(j\omega)|$. For the generalized model developed in Section 2.4, maximum bounds on the closed-loop transfer function can be defined by choosing frequency-dependent weights W_P and W_u in the weighted, generalized plant $\tilde{P} = \text{diag}(W_P, W_u, I)P$. This weighted, generalized plant

maps w to $\tilde{z} = \begin{bmatrix} \tilde{z}_1 & \tilde{z}_2 \end{bmatrix}^\top$. Thus, to achieve objective O2, we select a performance weight, W_P , that will be applied to $\tilde{z}_1 = W_P z_1 = W_P(v_o - v_o^*)$, allowing us to place frequency-dependent performance objectives on the closed-loop voltage tracking error transfer function. The best selections for weights W_P and W_u are dependent on the PHIL system configuration (specific examples are given in the case study), but for most PHIL simulations W_P is designed to emphasize performance at the nominal frequency with high gain at that frequency and W_u is designed to limit control effort to the bandwidth of the power interface using a high-pass filter. The user also chooses a scalar gain for K_{ff} , which improves the overall forward voltage path compensator's response time. This gain is generally between 0.5 and 0.95 and can be tuned in the experiment. Once weights are selected, the $\mathcal{H}_\infty$ optimal controller can be obtained by solving:

$$\min_{K \text{ stabilizing}} \|F_l(\tilde{P}, K)\|_\infty \quad (2.5)$$

This problem can be readily solved using the MATLAB Robust Control Toolbox. An algorithm for solving this problem along with the theoretical underpinnings of this optimization problem are given in [32].

Note that this problem finds the *stabilizing* K that minimizes $\|\tilde{N}\|_\infty = \|F_l(\tilde{P}, K)\|_\infty$. That is, if there is a solution, K^* , to (2.5), it will be a controller that stabilizes the closed-loop system. Therefore, if a solution is obtained, then the stability of the closed-loop system is guaranteed (thus meeting objective O1). The optimal controller K^* will have an order greater than or equal to the order of $\tilde{P}$, which could be considerable. Thus, before implementation for real-time simulation, it is recommended to apply model order reduction (e.g., see MATLAB's *balred* command) to obtain a lower order controller prior to discretization. In practice, we have found that a reduction down to the 11th order is sufficient and performs well.

2.3.4 Stability Analysis

When following the design approach outlined in Section 2.3.3 the closed-loop system is guaranteed to be stable as long as a controller can be synthesized successfully; thus, no additional stability verification of the continuous time closed-loop system model is

required. The authors of [23, 17], however, suggest that it is important to also consider the effects of sampling and numerical methods inherent in the discrete time model on the stability of the final implementation. A good way to verify closed-loop stability with the discrete time model is to discretize K_v and H_{fb} (using the Bilinear transform is recommended) and then perform a simulation (e.g., in MATLAB Simulink) of the full hybrid system including discrete time models of all software components and continuous time models of the hardware components. Also, the user should consider the potential impacts of controller model order reduction, if applied, by using the reduced order model in the final stability evaluation.

2.4 Accuracy Evaluation

Here, we introduce metrics for verifying the absolute accuracy of PHIL simulation results without relying on relative comparisons to potentially inaccurate models or previous simulation results. This method provides a valuable means for PHIL practitioners and researchers to easily identify the absolute accuracy of a particular simulation and compare results from two different simulations on a common basis. In short, this absolute method of accuracy evaluation is accomplished by evaluating tracking objectives O2 and O3 as well as how the power injected into the simulation, $S_{inj} = v_o^* i_o$, tracks the measured power, $S_{meas} = v_o i_o^*$.

We introduce two sets of metrics: one set (see Table 2.1) for quantifying accuracy at the instantaneous waveform timescale indicative of the simulation's bandwidth and another set (see Table 2.2) on a steady-state timescale for quantifying average quantities, such as active and reactive power. Note that $\theta_{inj} = \angle v_o^* - \angle i_o$ and $\theta_{meas} = \angle v_o - \angle i_o^*$ are the power factor angles for the injected and measured quantities, respectively; and $S_{hut,nom}$ is the nominal apparent power rating of the HUT. These metrics are demonstrated in the case study.

Table 2.1: Instantaneous Waveform Error Metrics

Metric	Definition
Instantaneous voltage tracking error:	$IVTE(t) = \frac{v_o(t) - v_o^*(t)}{\sqrt{2} \cdot \text{RMS}(v_o^*(t))}$
RMS instantaneous voltage tracking error:	$rIVTE = \text{RMS}(IVTE)$
Instantaneous current tracking error:	$ICTE(t) = \frac{i_o(t) - i_o^*(t)}{\sqrt{2} \cdot \text{RMS}(i_o^*(t))}$
RMS instantaneous current tracking error:	$rICTE = \text{RMS}(ICTE)$
Instantaneous power tracking error:	$IVTE(t) = \frac{S_{inj}(t) - S_{meas}(t)}{2 \cdot \text{RMS}(S_{meas}(t))}$
RMS instantaneous power tracking error:	$rIPTE = \text{RMS}(IPTE)$

Table 2.2: Steady-State Error Metrics

Metric: Definition
Active power tracking error:
$APTE = \frac{\text{RMS}(S_{inj}) \cos \theta_{inj} - \text{RMS}(S_{meas}) \cos \theta_{meas}}{S_{hut, nom}}$
Reactive power tracking error:
$RPTE = \frac{\text{RMS}(S_{inj}) \sin \theta_{inj} - \text{RMS}(S_{meas}) \sin \theta_{meas}}{S_{hut, nom}}$

2.5 Experiment Case Study

2.5.1 Experiment Configuration

Here, the PHIL experiment configuration in Figure 2.2 was used to evaluate a three-phase, 100-kVA, four-quadrant inverter with a 33-kWh Lithium-ion battery energy storage system interconnected at various points in an electric distribution network. The electric distribution feeder model used is from a real-world system in Colorado, USA, and was analyzed to determine the interconnection impedance (determined via short-circuit impedance values) at multiple three-phase interconnection nodes along the length of the approximately 4-mile-long distribution feeder. The power interface for this experiment consisted of a three-phase, 270-kVA Ametek RS-series power amplifier; a set of three LEM DVL1000 voltage transducers for output voltage measurement; and a set of three LEM LTC-600-SF current transducers for current feedback measurement. An Opal RT OP5700-series real-time simulator with a time step of $T_s = 50 \mu s$ was used to execute the real-time distribution system model and discrete-time PHIL interface compensators.

2.5.2 System Modeling

Modeling detail for each element in the closed-loop system are given in Table 2.3. Note that all time delays were approximated by a first-order Padé approximation, $e^{-sT} \approx \frac{-(T/2)s+1}{(T/2)s+1}$, because rational transfer functions are required for controller synthesis. The generalized plant was obtained following the process outlined in Section 2.3.3.4.

Table 2.3: System Modeling Elements

Element	Value	Modeling Basis
T_s	$50 \mu s$	Real-time simulation time step
$Z_{s,min}$	$53.3e^{-3} + s150.3e^{-6} \Omega$	Distribution feeder model, short-circuit impedance near substation
$Z_{s,max}$	$163.4e^{-3} + s565.4e^{-6} \Omega$	Distribution feeder model, short-circuit impedance near end of three-phase trunk
γ_v	$\frac{0.480 \text{ kV}}{13.2 \text{ kV}}$	Scale modeled system nominal voltage to that of physical HUT
G_{dac}	$e^{-sT_s} \approx \frac{-(T_s/2)s+1}{(T_s/2)s+1}$	Single time step delay for D/A conversion
G_{pa}	$\frac{1}{(1/2\pi 1950)s+1} e^{-s50e^{-6}}$	Delay and single pole filter model previously identified in [15]
Z_{pa}	$26e^{-3} + s53.2e^{-6} \Omega$	Physical measurement
Y_{hut}	$s\frac{2}{3}70e^{-6} S$	Vendor-supplied inverter wiring diagram. See [31] for model structure derivation.
$H_{s,v}$	$\frac{1}{(1/2\pi 15000)s+1} e^{-s20e^{-6}}$	Delay and single pole filter model. Delay and bandwidth information provided on datasheet.
$H_{s,i}$	$\frac{1}{(1/2\pi 100000)s+1} e^{-s1e^{-6}}$	Delay and single pole filter model. Delay and bandwidth information provided on datasheet.
H_{adc}	$= G_{dac}$	Single time step delay for A/D conversion
$\gamma_{i,min}$	$\frac{(1 \text{ MVA}/13.2 \text{ kV})}{(0.1 \text{ MVA}/0.480 \text{ kV})}$	Scale 100-kVA physical inverter to represent 1-MVA inverter in simulation model
$\gamma_{i,max}$	$\frac{(1 \text{ MVA}/13.2 \text{ kV})}{(0.1 \text{ MVA}/0.480 \text{ kV})}$	Scale 100-kVA physical inverter to represent 5.5-MVA inverter in simulation model

2.5.3 PHIL Interface Design

2.5.3.1 Base case interfaces

To emphasize the benefits of the optimal interface, we consider two base case interfaces for comparison. The first, denoted “CVI” (compensated voltage and feedback current

filtering), is modeled on a commonly used PHIL interfacing approach: design a simple compensator for the forward voltage path to achieve adequate $v_o \rightarrow v_o^*$ tracking, and then, if the closed-loop system is unstable (as was verified experimentally for this case study's configuration), apply an interface stabilization technique, such as feedback current filtering. Here, we used a proportional-resonant controller— $K_v = k_p + \frac{2k_i\omega_c s}{s^2 + 2\omega_c s + \omega_0^2}$ where $k_p = 0.1$, $k_i = 200$, $\omega_c = 1$, and $\omega_0 = 2\pi 60$ —and parallel feed-forward gain, $K_{ff} = 0.7$, for the forward voltage path compensator and a first-order, low-pass feedback current filter, $H_{fb} = \frac{1}{(1/2\pi 800)s + 1}$.

The second base case interface, “NVC” (no voltage compensation), models the case where no forward voltage path compensation is used ($K_v = 1$ and $K_{ff} = 1$) and the same low-pass feedback current filter was used.

2.5.3.2 Optimal (“Opt”) Interface

With a power amplifier bandwidth of approximately 1,950 Hz and a measured i_o^* to i_o time delay of $T_{fb} = 52.875 \mu s$, compensator H_{fb} was designed according to (2.1) with parameters $\omega_{fb1} = 2\pi \cdot 1000$ rad/s, $\omega_{fb2} = 2\pi \cdot 3000$ rad/s, and $\alpha_{fb} = 56500$. This achieved phase compensation as desired.

For the synthesis of K_v , a performance weight for the closed-loop voltage error transfer function of the form $W_P = k_{DC} \left(k_p + \frac{2k_i\omega_c s}{s^2 + 2\omega_c s + \omega_0^2} \right) \left(\frac{\omega_{lpf}}{s + \omega_{lpf}} \right)^2$ was used with $k_{DC} = 110$, $k_p = 0.6$, $k_i = 5000$, $\omega_c = 0.01$, $\omega_0 = 2\pi 60$, and $\omega_{lpf} = 2\pi 2000$. This performance weight encourages high gain near the fundamental frequency and relatively low gain everywhere else. The input weight was designed to limit control effort at high frequencies and had the form $W_u = \frac{s + c_2\omega_w}{c_1 s + \omega_w}$ with $c_1 = 0.01$, $c_2 = 0.1$, and $\omega_w = 0.1 \cdot 2\pi 60$. The two weighting functions used are shown in Figure 2.5. The feedforward gain $K_{ff} = 0.7$ was found to work best after tuning in the experiment. Using these weights and the process outlined in Section 2.3.3, the K_v with Bode plot shown in Figure 2.6 was synthesized successfully. This controller emphasizes performance near $\omega_0 = 2\pi 60$ as desired, but it also has another benefit: a notch is added to de-emphasize performance near $\omega = 2.01e^4$ rad/s—the resonant frequency of the power amplifier's inductance and the HUT's output filter capacitor ($\omega_{res} = 1/\sqrt{L_{pa}C_{hut}} = 2.01e^4$ rad/s). This notch was added by the optimal controller synthesis process and not because of a specific emphasis in the performance weight, emphasizing the benefit of the optimal

interface design approach. Note that the controller synthesis process outputs a high-order controller, which we then reduced to the 11th order controller shown in Figure 2.6.

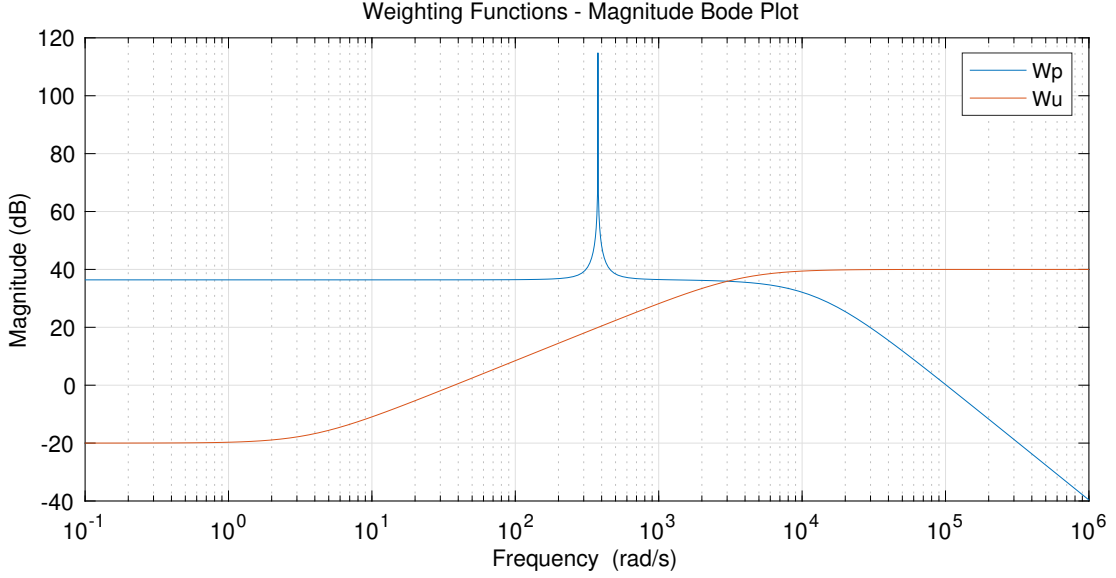


Figure 2.5: Weighting functions used for $\mathcal{H}_\infty$ synthesis of G_v . W_P is designed to emphasize performance at the nominal frequency and W_u is designed to limit control effort to the bandwidth of the power interface.

2.5.4 Stability Analysis

As described in Section 2.3.4, as long as a controller can be synthesized, the closed-loop system is guaranteed to be stable; however, the controller model order reduction process was applied to obtain $K_{v,red}$ from K_v after this controller synthesis. Thus, we additionally verify that the eigenvalues of $F_l(\tilde{P}, K_{v,red})$ are all in the strict left half plane, confirming closed-loop system stability.

2.5.5 Experiment Results

Stable PHIL simulations using the optimal (“Opt”), CVI, and NVC interfaces were achieved, and the 100-kVA inverter HUT was operated at a wide variety of active and reactive power operating points while connected to a simulated grid interconnection

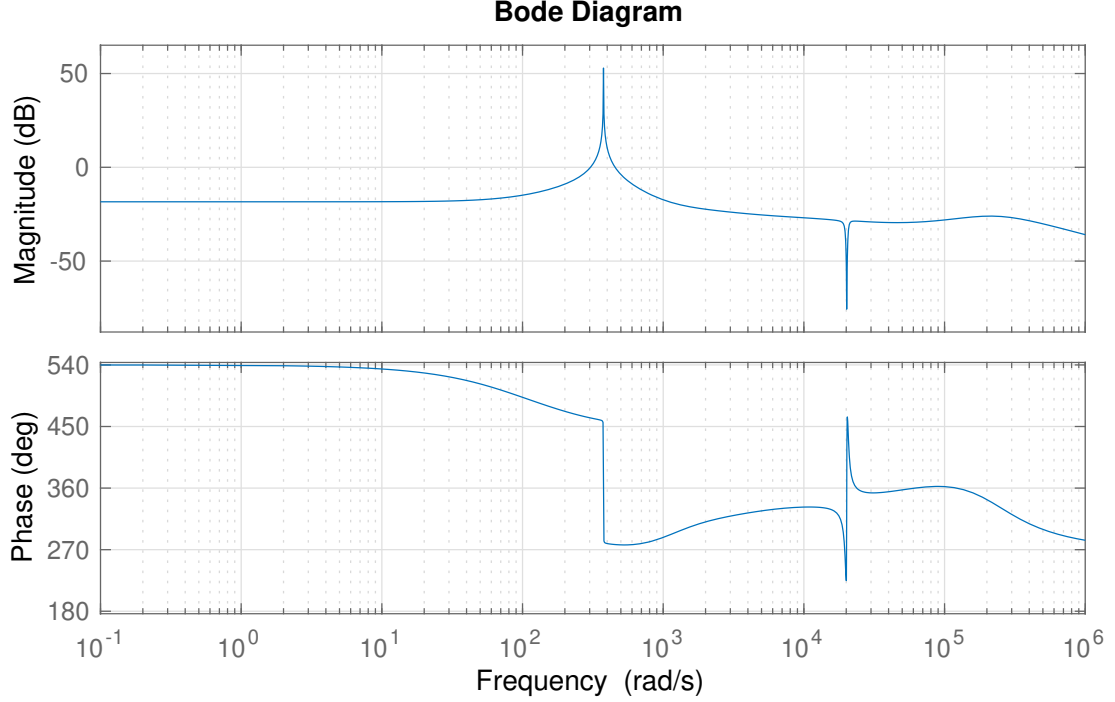


Figure 2.6: Synthesized K_v reduced to 11th order

point with impedance Z_s . The HUT power was also scaled by γ_i to represent a larger capacity inverter (of capacity S_{scl}) in the simulated model. Independently, both Z_s and γ_i were varied between their minimum and maximum values under constant HUT power to evaluate the performance of the interfaces under varying impedance and scaling factors.

2.5.5.1 Summary Steady-state Results

Table 2.4 lists summary results from the many tests examining different combinations of Opt/CVI/NC interface, HUT power, source impedance, and scaling factor that were completed. The following conclusions can be made regarding the summary steady-state results:

1. The optimal (“Opt”) interface offers improved tracking of all quantities (voltage, current, inst. power, active power, reactive power) than the other interfaces. Improvements in current and reactive power tracking are particularly significant.

The CVI and NVC interfaces generally have 3%–10% reactive power tracking error (RPTE), whereas the optimal interface improves this performance two to five times and keeps the RPTE within 1.8% for all cases.

2. When using the optimal interface, the performance is approximately constant (consistently similar tracking errors—e.g., RPTE within $\pm 1.2\%$) when testing across various HUT power factors; however, the CVI and NVC interfaces show greater variation in performance (e.g., RPTE within $\pm 5\%$ and $\pm 7\%$).
3. Interface tracking error is consistent for varying scaling factor and varying source impedance.

Table 2.4: Summary test results

Nominal Test Conditions				Tracking Error Metric (%)														
HUT (kW)		Model		Inst. Power (rIPTE)			Inst. Voltage (rIVTE)			Inst. Current (rICTE)			Active Power (APTE)			Reactive Power (RPTE)		
P_{hut}	Q_{hut}	S_{scl} (MVA)	Z_s	Opt	CVI	NVC	Opt	CVI	NVC	Opt	CVI	NVC	Opt	CVI	NVC	Opt	CVI	NVC
-93	0	2.5	min	1.18	2.91	1.84	0.85	0.91	2.84	2.13	5.74	5.74	0.04	0.52	0.85	1.81	-6.90	-10.36
-62	0	2.5	min	1.38	3.09	2.12	0.83	0.88	2.83	2.57	6.10	6.10	0.02	0.34	0.56	1.20	-4.59	-6.90
-41	0	2.5	min	1.81	3.47	2.65	0.81	0.87	2.83	3.39	6.84	6.84	0.02	0.28	0.44	0.80	-3.06	-4.59
-26	0	2.5	min	2.59	4.30	3.68	0.81	0.85	2.82	4.79	8.41	8.41	0.02	0.17	0.28	0.50	-1.92	-2.88
26	0	2.5	min	2.00	3.78	3.03	0.79	0.83	2.82	3.42	6.88	6.88	-0.03	-0.06	-0.11	-0.52	1.95	2.93
41	0	2.5	min	1.68	3.43	2.60	0.83	0.88	2.83	2.91	6.41	6.41	-0.03	-0.16	-0.27	-0.82	3.09	4.65
62	0	2.5	min	1.42	3.17	2.25	0.85	0.91	2.84	2.39	5.96	5.96	-0.03	-0.30	-0.50	-1.22	4.63	6.95
93	0	2.5	min	1.26	3.00	2.01	0.88	0.93	2.85	2.01	5.67	5.67	-0.06	-0.42	-0.70	-1.82	6.91	10.38
62	30	2.5	min	1.55	3.76	3.68	0.87	0.93	2.85	2.47	6.01	6.01	0.57	-2.56	-3.89	-1.24	4.48	6.71
62	-30	2.5	min	1.32	3.26	2.84	0.82	0.86	2.83	1.98	5.63	5.63	-0.63	2.02	2.99	-1.20	4.77	7.19
-62	-30	2.5	min	1.24	3.48	3.41	0.78	0.83	2.82	2.09	5.75	5.75	-0.58	2.63	4.00	1.22	-4.45	-6.65
-62	30	2.5	min	1.63	3.49	3.07	0.89	0.95	2.85	2.80	6.35	6.35	0.64	-1.94	-2.87	1.18	-4.74	-7.14
-77	0	2.5	min	1.25	2.96	1.92	0.84	0.89	2.83	2.30	5.87	5.87	0.03	0.42	0.69	1.50	-5.76	-8.66
-77	0	3	min	1.21	2.97	1.94	0.84	0.88	2.83	2.24	5.86	5.86	0.04	0.48	0.78	1.51	-5.75	-8.64
-77	0	4.8	min	1.16	2.95	1.91	0.83	0.88	2.83	2.17	5.83	5.83	0.04	0.42	0.69	1.50	-5.75	-8.64
-77	0	5.5	min	1.14	2.95	1.90	0.83	0.89	2.83	2.13	5.82	5.82	0.05	0.39	0.64	1.50	-5.75	-8.63
62	0	1	min	1.40	3.17	2.25	0.86	0.91	2.84	2.35	5.97	5.97	-0.05	-0.22	-0.38	-1.21	4.62	6.95
62	0	1	mid	1.37	3.16	2.24	0.86	0.90	2.84	2.24	5.94	5.94	-0.04	-0.29	-0.48	-1.22	4.63	6.95
62	0	1	max	1.30	3.16	2.22	0.82	0.90	2.84	2.44	5.94	5.94	-0.03	-0.25	-0.42	-1.21	4.63	6.96

2.5.5.2 Detailed Dynamic Results

Detailed performance results from a 41-kW to 93-kW change in HUT active power are shown in Figure 2.7 and Figure 2.8. In particular, Figure 2.7 further demonstrates the significantly improved current tracking performance of the optimal interface. The magnified view of Figure 2.8 emphasizes that the optimal interface both:

1. fully compensates for phase lag (achieves nearly zero phase lag vs. multiple degrees of phase lag for the CVI case), which explains the significantly better reactive power tracking demonstrated in Table 2.4; and
2. results in a significantly higher bandwidth simulation, as evidenced by the fact that injected current includes many of the same higher frequency components present in the measured current, whereas in the CVI case, these higher frequency components are filtered out.

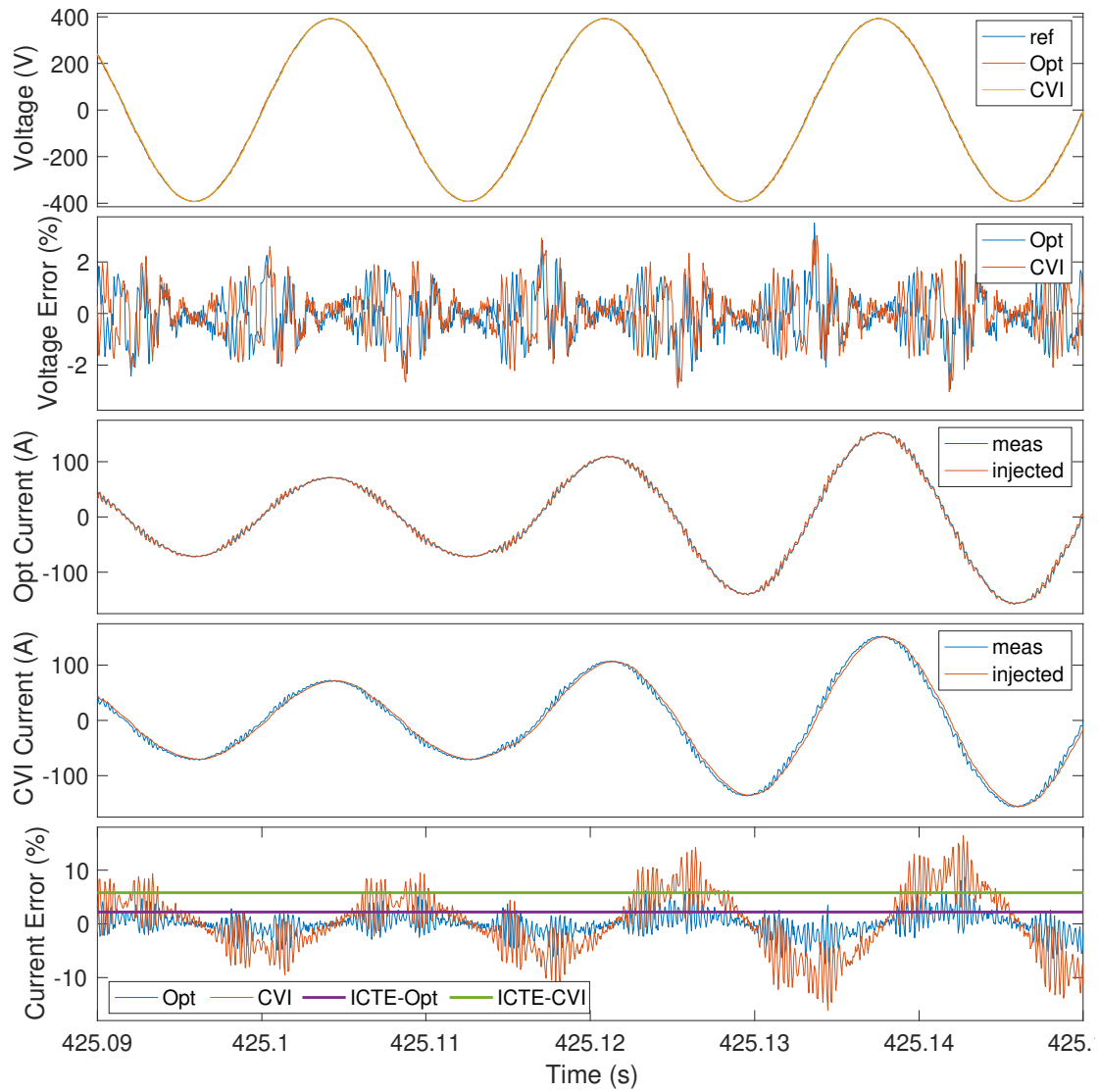


Figure 2.7: Phase A instantaneous voltage, voltage error (IVTE), current (from the Optimal “Opt” vs. CVI interface), and current error (ICTE) throughout a 41 kW to 93 kW change in HUT active power

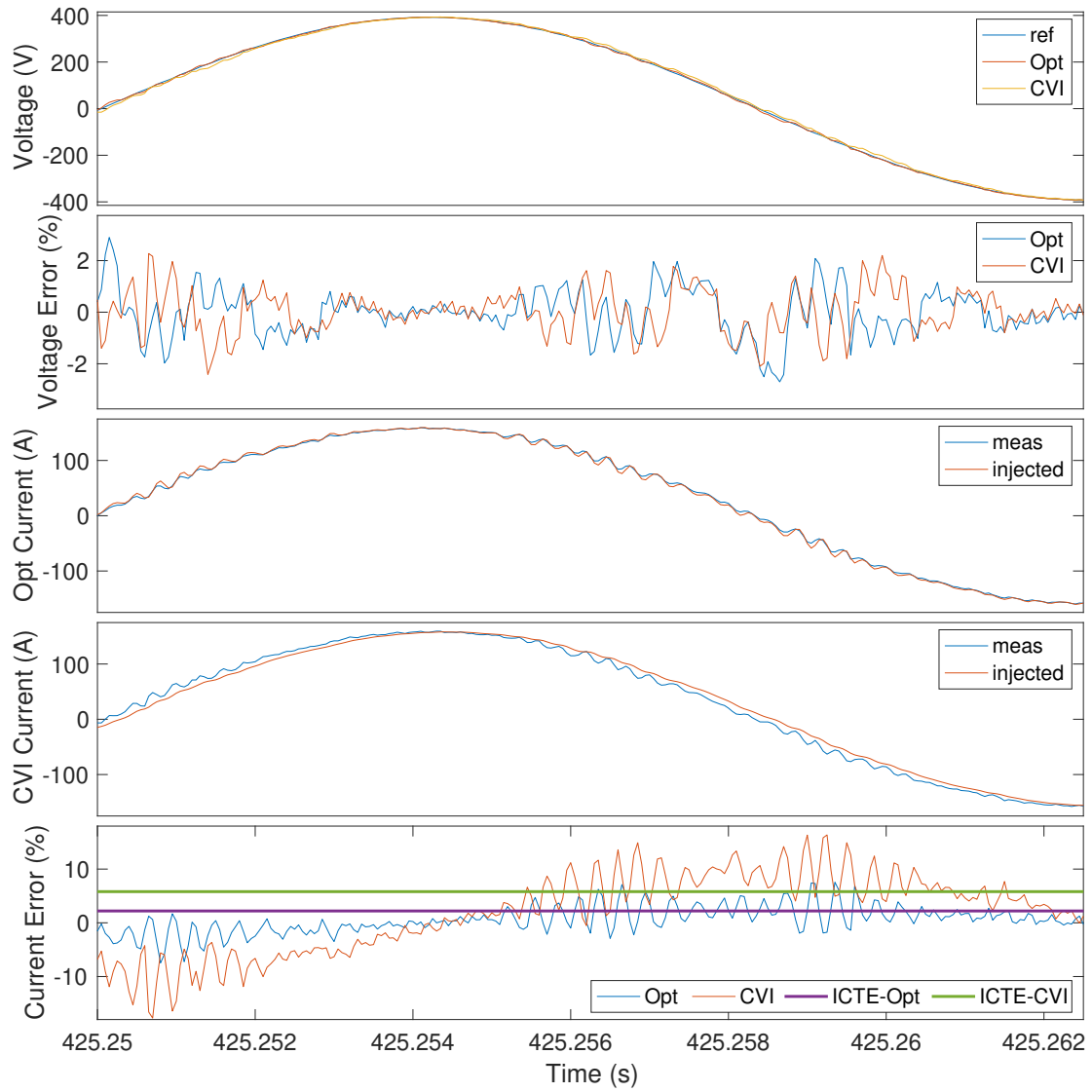


Figure 2.8: Magnified view of Figure 2.7. Note the significantly better tracking (phase and bandwidth) shown in the Opt Current vs. CVI Current plots.

Chapter 3

Scalable Hybrid Classification-Regression Solutions for High-Frequency Nonintrusive Load Monitoring in Grid-Interactive Residential Buildings

3.1 Introduction

Smart buildings that can manage their net loads to provide additional flexibility to grid operators while maximizing customer preferences and energy cost savings offer a valuable tool for a modernized electric grid. Residential buildings in the United States use 38.5% of the total electrical energy produced [33]; thus, if residential buildings can be converted into smart buildings, significant flexibility to support the grid will be unlocked.

It is crucial to enable circuit-level net-load metering at these residential buildings to gain a better understanding of building performance and customer usage patterns

and preferences while enabling the critical monitoring of information needed to perform meaningful grid-interactive control. The simplest approach for such metering is having a sensor at each load, but this is invasive and very costly. Nonintrusive load monitoring (NILM) can extend many benefits of smart buildings to homes using minimal additional equipment and cost. With this method, a single electrical measurement point at the building’s point of common coupling with the grid is used to derive information (e.g., on/off status, electricity consumption) about the constituent net-load devices in the home.

Existing work in NILM can be categorized based on the type of measurement inputs used. Generally, NILM approaches use one or both of: 1) steady-state measurement quantities, such as active and reactive power or root mean square (RMS) current, on a macro timescale (generally 1–60 min [34]); and 2) transient features derived from micro-timescale data (generally > 1 Hz [34]). In the first category, [34, 35, 36, 37, 38, 39] used steady-state power measurements with a variety of contemporary machine learning methods, including multilabel k-nearest neighbor [34, 35], multilabel support vector machines [40], binary relevance [34, 40], hidden Markov models [34, 39], recurrent neural networks [38], and long short-term memory [41, 42] methods. In the second category, micro-timescale data is employed to infer features based on a frequency-domain transformation [43, 44], instantaneous wave shape [43], switching transients [45], or wavelet transformations [45, 46]. References [47, 48] fit into both categories, using micro-timescale data to detect events and macro-timescale data to capture steady-state data throughout the event and build a library of events that can be used later for classification.

Steady-state measurement NILM methods often require a large sample (days to months) of training data and suffer a slow response time following events. Further, the use of steady-state measurements can easily result in misclassification between two different combinations of loads that have nearly identical steady-state measurements but differing instantaneous waveforms. For example, Figure 3.1 shows the instantaneous current waveform (top plot) and corresponding RMS values (bottom plot) associated with two different combinations of home loads. NILM methods based on steady-state measurements, such as the mean RMS current (-rmsMN in Figure 3.1) in this example, will not be able to successfully differentiate between these two combinations; whereas

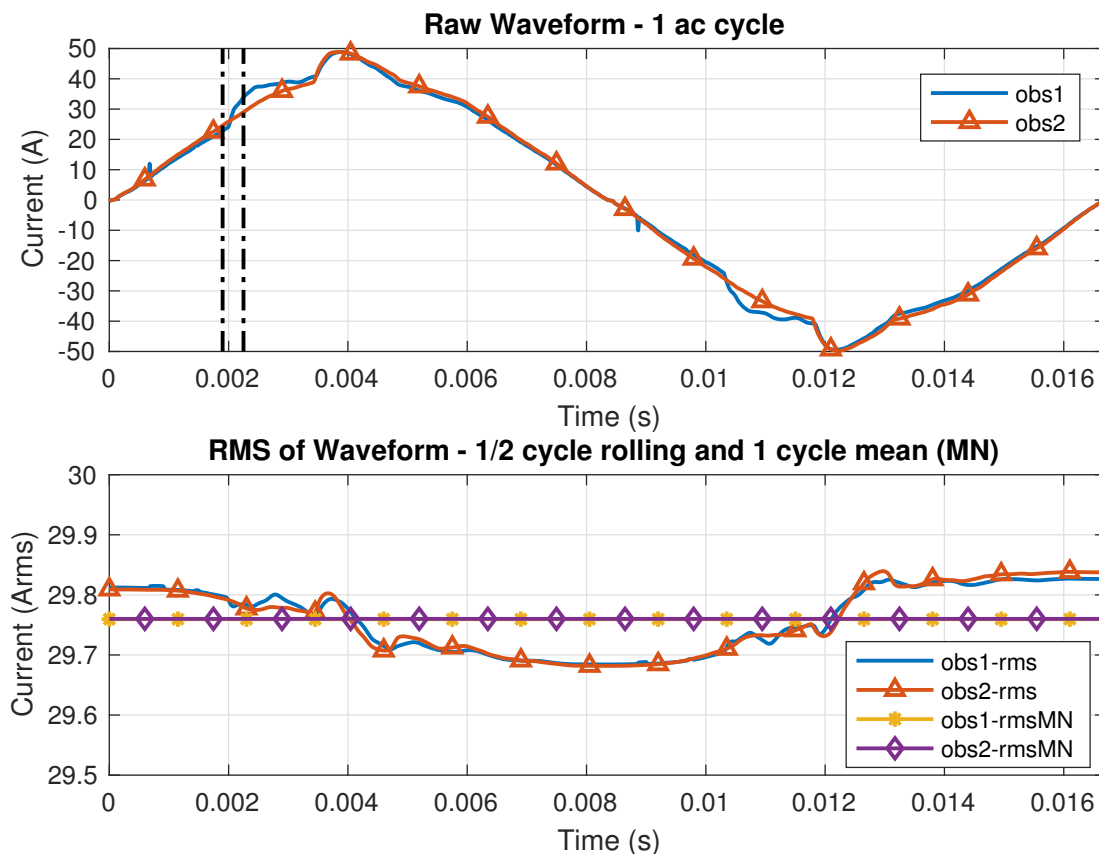


Figure 3.1: Comparison of waveform (top plot) and RMS values (bottom plot) from two observations (one 60-Hz ac cycle sampled at 200 kHz) representing two different load configurations in a residential building.

NILM methods using micro features, such as harmonics or wave shape, can be more successful. This point is further emphasized in Figure 3.2, which shows the distribution of RMS current values (a steady-state quantity) for the four-load case study presented later in this chapter, which demonstrates that the same steady-state value could exist in many different system load configurations.

In both steady-state and transient NILM approaches, it is common to use methods based on event detection; however, such approaches often assume that multiple load transients do not occur at the same time, might suffer confusion between transient events occurring within a single load (e.g., same load switching to a higher power level), and cannot make accurate state predictions until after an entire transient event has

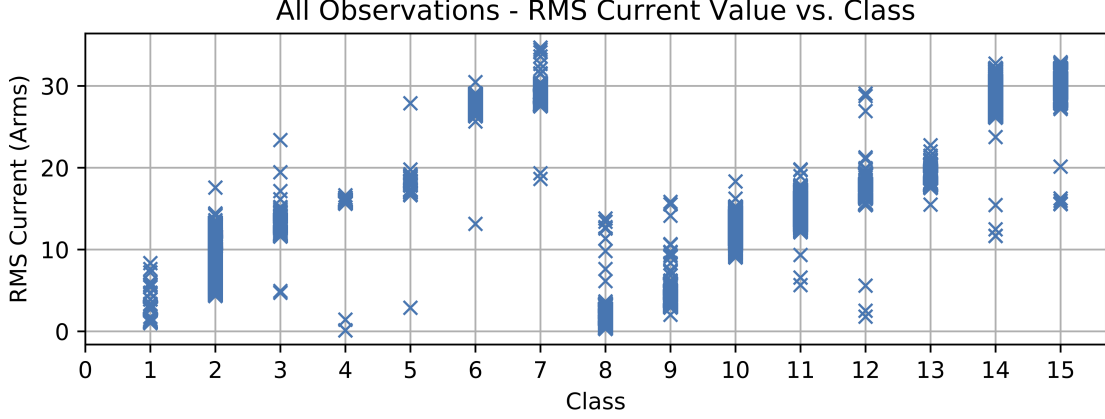


Figure 3.2: All observations (all training and test data) with their RMS current feature plotted vs. class for a residential building load classification case study.

occurred. In most existing NILM approaches, system state predictions are often made in no faster than 1 s to 15 min. This might be sufficiently fast for some functions, such as estimating energy usage or customer load usage patterns, but it is too slow to enable a smart building to provide grid services (e.g., fast frequency response using building net load [49]).

The majority of NILM literature has focused on using an aggregate current or power measurement to predict the on/off state of connected loads via classification. Some existing work (e.g., [41, 42]) also considered regression-based approaches. Reference [41] considered bidirectionality and long-term temporal dependence of power consumption by appliances for load disaggregation. These approaches can disaggregate the power consumption of multiple loads when a long sequence (i.e., multiple hours) of power data is provided, but they cannot accurately make predictions in real time using short intervals of input data. To the best of our knowledge, there is no method currently capable of load power disaggregation at the timescale of <1 -s grid ancillary services. In this chapter, we improve upon these methods by achieving accurate load power disaggregation on a 60-Hz timescale.

In this chapter, we first develop a method for the accurate, high-speed (>60 -Hz) prediction of the multiload on/off state of a four-load residential building experiment configuration. The method employs multiload state classification using a combination of frequency transformation, wavelet transformation, and wave shape features that allow

for good generalization and fast response to transient events without relying on event detection. We then extend this high-speed classification method to improve scalability, and we apply the improved classification method to a larger eight-load residential configuration. In addition, we introduce a novel hybrid classification-regression model that allows for the prediction of load RMS current consumption levels via regression in addition to the load on/off states via classification while still operating at a high-speed timescale with predictions for each 60-Hz cycle. The hybrid classification-regression approach is validated using an experiment configuration including eight residential building appliances, and it is shown to give high-accuracy RMS current and multiload on/off state predictions.

3.2 High-Frequency Nonintrusive Load Monitoring Approach

3.2.1 Problem Formulation

A total of N_L loads are connected to the main load center panel of a residential building, and each consumes an instantaneous current $I_i(t) = q_i(t)\bar{I}_i(t)$, where $\bar{I}_i(t)$ is the load signature, $q_i = \{0, 1\}$ is the on/off status (where $q_i = 1$ denotes that the load is on), and I_i is the resulting current consumption of the i^{th} load. The total current consumption of the residential building is $I_{TOT}(t) = \sum_{i=1}^{N_L} q_i(t)\bar{I}_i(t)$. With N_L loads, there are 2^{N_L} possible system load on/off states, $Y_C = [q_1, q_2, \dots, q_i, \dots, q_{N_L}]$, for the building. The first objective is to train a classifier that uses an input vector, X , of features calculated for each 60-Hz ac cycle observation of $I_{TOT}(t)$ to predict the correct class label, Y_C , representing the building's complete load on/off state. The classifier should provide a state label prediction for every 60-Hz ac cycle of $I_{TOT}(t)$. The second objective is to build a regression model that takes an input feature vector X (including the result from the classifier) and predicts the RMS current consumption of each load, $Y_{reg} = [\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_i, \dots, \tilde{I}_{N_L}]$. Similar to the classifier, the regression model should provide a prediction for every 60-Hz ac cycle of $I_{TOT}(t)$. For the application of high-frequency smart building control, the entire process of splitting the instantaneous $I_{TOT}(t)$ into single-cycle observations, calculating features, and using the classifier and regression

model to predict an output label should occur within ten AC cycles or $\tau = 10/60 = 0.167$ s. This response time will ensure state awareness commensurate with the needs of fast frequency response (generally $< 0.33 - 0.5$ s total response is time required). To be of practical value, the classifier and regression model should be able to generalize well to any $I_{TOT}(t)$ observations from the same group of appliances they are trained on, regardless of whether a similar magnitude $I_{TOT}(t)$ observation is seen during training.

3.2.2 Experiment Configuration

Two residential-scale demonstrations using four and eight household appliances in the Energy Systems Integration Facility at the National Renewable Energy Laboratory were completed. The full experimental configuration, shown in Figure 3.3, includes eight residential appliances, including a combination refrigerator/freezer (General Electric Profile PSQS6YGY), combination oven/range (Maytag MER8674), space heater (Bovado 0CZ449), a bank of (10) compact fluorescent light bulbs, a water heater (AO Smith HP10), clothes dryer (Whirlpool WEL98HEBU), window air conditioner unit (Frigidaire FFRE0633S1), and the condenser unit of a central air conditioner unit. Nine current probes and one voltage probe at the locations indicated were used to gather instantaneous measurements sampled at 200 kHz over each 200-s data set period using a digital oscilloscope. The first experiment configuration consisted of only appliances #1–4 (the other appliances were disconnected), and the second experiment used all eight appliances.

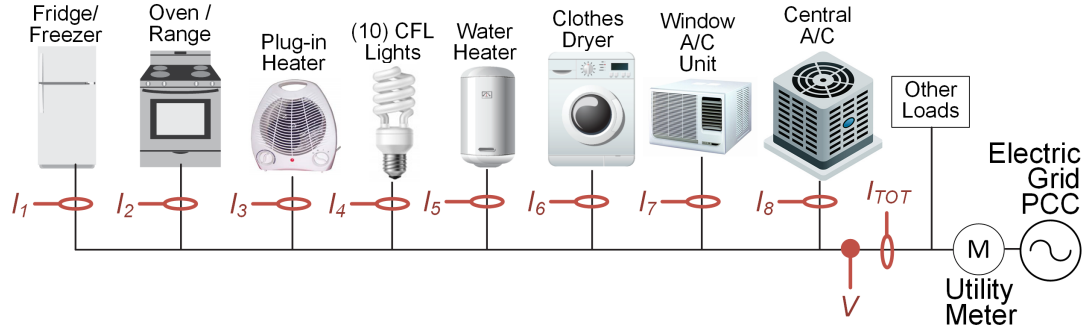


Figure 3.3: Experimental configuration with eight residential appliance loads. I_{TOT} measures the total current flowing through the eight loads.

3.2.3 Data Sets

For the first demonstrative case with a four-load configuration, seven independent 200-s data sets—each including measurements of V , I_1 , I_2 , I_3 , and I_4 (but not I_{TOT})—were collected. I_{TOT} was obtained as $I_{TOT} = I_1 + I_2 + I_3 + I_4$. For each data set, the appliance loads were put into a different starting condition, and then throughout the data set period, loads were turned on and off to obtain many instances of the 15 different classes. Throughout all data sets, the fridge, oven, and space heater were manually perturbed (e.g., fridge door opened, oven opened, space heater fan speed changed) to capture the variety of operating conditions of the loads. The lights were considered to be either on or off and were not manually perturbed. These data were used to train the model.

For testing, six additional independent 200-s data sets—each including V , I_1 , I_2 , I_3 , I_4 , and I_{TOT} —were collected. For testing, only I_{TOT} and V were used as input data to the presented NILM approach, and I_1 , I_2 , I_3 , and I_4 were used to assign the true label to each observation. As with the training, appliances were cycled on and off and manually perturbed to obtain instances of all 15 classes as well as significant steady-state load variations across their ranges of operation. An example of one testing data set is shown in Figure 3.8.

For the second demonstration, which used the eight-load configuration, seven independent 200-s data sets—each including measurements V , I_1 , I_2 , I_3 , I_4 , I_5 , I_6 , I_7 , I_8 , and I_{TOT} —were collected. Similar to first demonstration, the loads were perturbed in a random manner. They are described in Table 3.1.

3.2.4 Approach

A summary of the high-frequency NILM approach implemented for classification is shown in Figure 3.4. This approach extends previous NILM work by using a wider combination of features, being able to predict both on/off state and operating current level, and by achieving much higher frequency output while maintaining exceptional accuracy.

The method presented here is applicable to a wide variety of load configurations and types. We demonstrate scalability of the approach from a four- to an eight-load

Table 3.1: Load conditions for eight-load demonstration

Data set	Conditions
1	All loads are on except water heater is cycling on/off
2	All loads are on. Perturbations including opening the fridge/freezer doors and changing fan speed on window air conditioner are applied.
3	Loads 5–8 always on (except water heater cycling), then ran through random combinations for loads 1–4
4	Loads 1–4 always on, then ran through random combinations for loads 5,7, and 8. Dryer finished cycle in middle of test.
5	All loads random on/off
6	All loads random on/off
7	All loads random on/off

configuration and we note that the approach can also be applied in a modular fashion for multiple load groups on a single grid PCC as needed by configuring and measuring an I_{TOT} for each load group, as depicted in Figure 3.3.

3.2.4.1 Input Data Processing

As shown in Figure 3.4, high-speed instantaneous voltage and current waveforms are the inputs to this approach. During the training phase, current measurements from each circuit are used. Voltage measurement data are not strictly required, but if available, they can be used to provide more accurate detection of zero crossings. The current data are divided at each 60-Hz ac cycle using zero-crossing detection, then each cycle of the aggregate current measurement becomes an observation for which a label and features are derived.

For the eight-load configuration, we employ a novel strategy for creating a synthetic data set from the original data that expands each observation into multiple observations by using the superposition principle of electric current. We then employ a reduction strategy to remove samples that are very similar to each other. These two strategies are described in detail as follows:

Synthetic data set: Each observation, i , in the original data set has a total current measurement, I_{TOT} , based on $1 \leq N_{L,i} \leq 8$ superimposed currents. $N_{L,i}$ is

determined by thresholding all I_i and counting the number of nonzero I_i . Then, the $N_{L,i}$ nonzero currents can be used to generate $2^{N_{L,i}}$ unique observations by superimposing the corresponding I_i for each combination together and calculating the features (see Section 3.2.4.3). All $2^{N_{L,i}}$ unique observations are added to the synthetic data set, and the index of the original, nonsynthetic observation is preserved and added as a column to all associated observations in the synthetic data set.

Data set reduction: Within each data set, all the observations are first sorted, the distance (in terms of the L_1 -norm) between successive observations is determined, and new observations are discarded if they are not significantly different from the previous observation. We tried multiple thresholds and finally chose $\pm 5\%$ as the threshold for discarding the new observation. The resulting reduced data set is almost 1/5th the size of the original synthetic data set. With this reduced data set, computational complexity and memory requirements for training the models are significantly reduced. This reduced data set is used for training.

3.2.4.2 Labeling (Training Phase)

For classification, observations are generally categorized into multiple classes. When classifying between more than two classes, multiclass classification is used. Each aggregate current cycle observation is labeled by examining the instantaneous current data at each circuit, using level detection to derive the on/off state of that circuit, and finally building the multiclass label. This multiclass label could be a multilabel vector or a single integer label with the class encoded using a binary encoding scheme (e.g., for the four-load case, Label 11 corresponds to $[1,0,1,1]$), referred to as the label power-set method. For the four-load case, a single integer label is used, whereas for the eight-load case, each load's on/off state is used as a multilabel vector for the corresponding model. For regression, the RMS current of each individual load is used as a label for the corresponding regression model.

3.2.4.3 Feature Engineering

The individual data points of the instantaneous aggregate current cycle observation are not used directly as the input to the classifier (this approach does not result in good prediction). Instead, the instantaneous data are processed into features to reduce

the dimensionality of the problem while retaining the important characteristics of the waveform. Four key categories of features are used:

1. Harmonic features: A 1-D discrete Fourier transform is used to compute the magnitudes of the Fourier series coefficients at the fundamental and 3rd, 5th, 7th, 9th, 11th, and 13th harmonics of the 60-Hz nominal aggregate current waveform. These coefficients are used in both their unscaled and normalized (using the L_2 -norm) form, giving 14 possible features.
2. Steady-state RMS feature: A single feature based on calculating the RMS of all data points in the single-cycle aggregate current observation is used. For training the regression model, the RMS values of each single-cycle individual current observation for all eight loads are retained.
3. Wavelet features: An eight-level (the maximum decomposition level possible for the wavelet and sampling rate used) 1-D discrete wavelet transform using the Daubechies 7 (db7) wavelet [50] is used to derive the discrete wavelet transform coefficients at multiple levels of decomposition. The db7 wavelet is used because it has been shown to have good performance for power system applications [51]. The Level 8 detail coefficients were studied as potential features, and coefficients # 12–23 for a total of 12 out of the 25 total coefficients are selected as the final features.
4. Wave shape features: Here, wave shape features refer to the features derived using individual data points in the raw instantaneous current waveform that represent unique characteristics of the waveform’s shape. Three such features are used. The first is the maximum value in the middle third of the waveform’s positive half cycle. The last two are ratios of two specific points in the ac waveform that capture local minima present in a waveform when particular loads are present. For example, in Figure 3.1 (top plot), Obs1 and Obs2 contain the same three loads, but Obs1 also includes a fourth, relatively low-magnitude load. Overall, the two observations’ waveforms are very similar, making it difficult to differentiate between these two classes using traditional features. In the region between the plot’s two vertical black dotted lines, however, a difference in wave shape can be observed. A feature

based on the ratio of the waveform’s values at the first and second lines allows for reliable separability of these two very similar classes and provides a valuable measure of whether this additional load is present in any class. Overall, two such ratios are used for the four-load case, and five such ratios are used for the eight-load case. Ratios are derived based on local maxima present in the different loads’ characteristic waveforms. This approach can be applied generally by deriving a ratio feature for any constituent load that contains local maxima outside of the normal 60-Hz peak in its waveform.

We initially used only unscaled harmonic and steady-state RMS features [52]. However, we found that this set of features was not sufficient to achieve classifiers with both high prediction accuracy and good generalization properties [52]. Adding normalized harmonic features, wavelet features, and the wave shape features corrected this problem, as shown in Section 3.3.4. We expect that these categories of features will generalize well to a variety of possible load types, though the designer may wish to fine tune which particular subsets of these feature categories (e.g., particular Fourier or wavelet transform coefficients) are used for a given load configuration at training time to obtain the best performance.

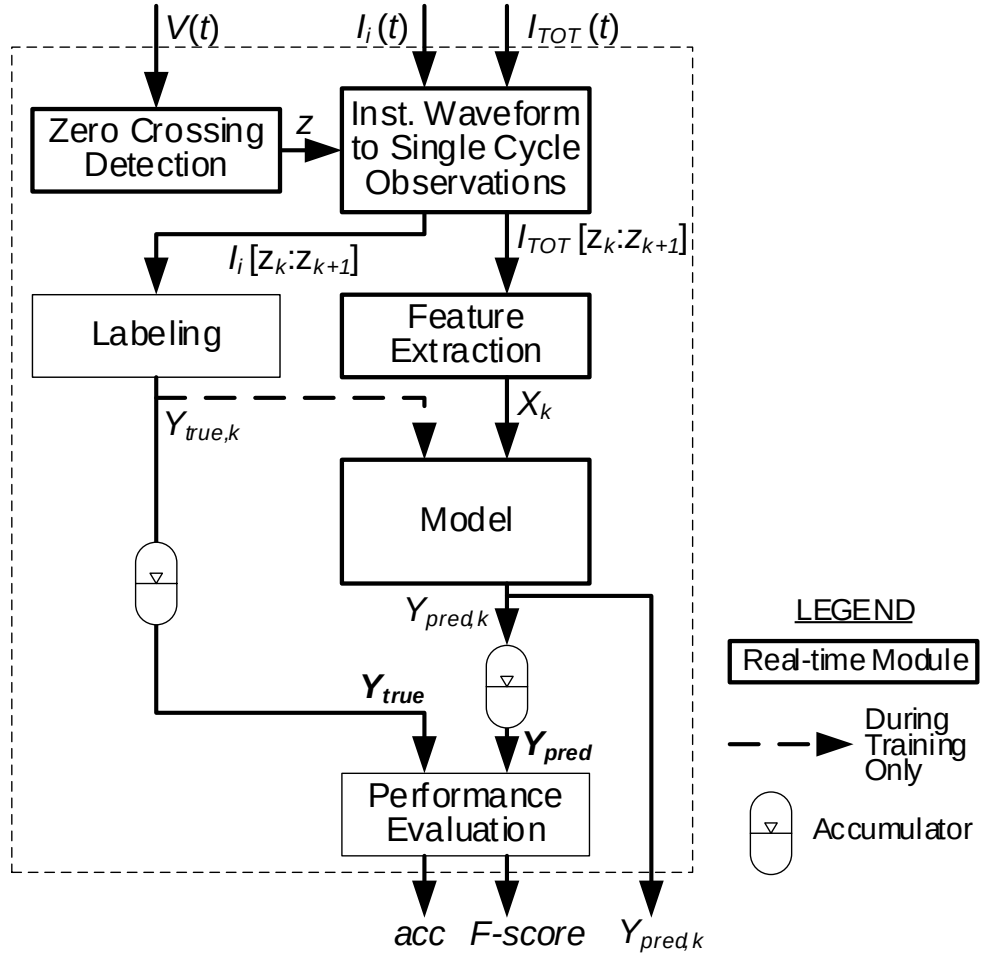


Figure 3.4: High-frequency NILM approach. Here Model refers to the classification and/or regression model.

3.3 High-frequency, Multiclass Classification of Residential Appliance Loads

A random forest classifier (RFC) [53], which is an ensemble method, is used in this work. Ensemble methods use multiple uncorrelated base classifiers for better predictive performance. Uncorrelated models usually produce ensemble predictions that are more accurate than individual predictions [54]. RFCs use decision trees as their base classifier. Each tree provides a class prediction, and the class with the most votes becomes the RFC’s predicted class. Decision trees separate a data set so that the resulting subgroups are as different from each other as possible at every split and the observations in the same subgroup are as similar to each other as possible. Decision trees depend heavily on data they are trained on and can be very different if those training data change. RFCs consist of multiple decision trees where each tree is formed by randomly sampling data from the data set (with replacement). Further, each tree in a random forest can use only a random subset of the feature set. This leads to the individual trees being highly uncorrelated, leading to the ensemble’s prediction being more accurate than that of any individual tree. Because multiple loads and any number of loads can be turned on, multiclass classification is required for this problem. The “label power-set method,” which takes each combination of possible load states as one class, was initially used when considering the four-load test configuration (Classifier 1). This method considers the correlation of different labels and can provide accurate predictions when the number of loads is small; however, the total number of classes in this method increases exponentially with the number of loads. Thus, when applying the approach to the larger eight-load configuration, the label power-set method does not scale well because the class size explodes to 255, and accuracy is significantly reduced. To address this, we use a multi-model approach (Classifier 2) wherein we train simple binary classifiers to detect the presence of a single load, as discussed in detail in Section 3.4. This leads to much better accuracy and easier implementation for parallel processing of multiple load classifications.

3.3.1 Classifier Training and Tuning

Multiple RFCs using different subsets of the features described in Section 3.2 are considered in the search for a classifier that is accurate in differentiating the 15 different classes pertaining to the four-load configuration (as described in Section 3.2.3) but that also generalizes well to make correct predictions on observations not previously seen before in training data.

The final classifier, termed Classifier 1, is an RFC ensemble using bootstrapping and including 200 decision tree classifiers, each with a max depth of 35 and using entropy (information gain) as the split criterion. These optimal hyperparameters were determined using a grid search wherein for each hyperparameter combination, seven-fold cross-validation was performed to obtain the classifier’s accuracy as the scoring metric. Seven-fold cross-validation was implemented such that each combination of using 1/7 training data sets for testing and the remaining 6/7 training data sets for training were considered. The feature set used for this classifier is summarized in Table 3.2. Further details on RFCs hyperparameters and testing methodologies can be found in [53]. Here, the RFC was implemented in Python using the scikit-learn package [55].

Table 3.2: Hyperparameters for different models

Model	Features Included			
	Harmonic Co-efficients	Normalized Harm. Coeff.	Wavelet	Wave Shape
Classifier 1	All	5,7	All	1,2
Classifier 2	All	5,7	All	1,2,3,4,5
DNN	All	All	All	None

3.3.2 Performance Metrics

Both accuracy and weighted F1-score are used as measures of classifier performance. Accuracy provides an intuitive evaluation of the ratio of correctly labeled observations to the total number of observations. The F1 score, a common metric for NILM implementations, is the harmonic mean of the precision and recall [53] (i.e., $F_1 = \frac{2}{(R^{-1} + P^{-1})}$, where R is recall, and P is precision). Although an effort was made to get a sizable number of instances of each class, the distribution of instances across classes is not

equal, and hence an F1 score weighted by the number of true instances in a class is used to account for class imbalance (it is essentially a “macro” F1 score that accounts for class imbalance).

3.3.3 Results

Table 3.3 presents accuracy and F1 score results as averaged over 10 repeated (there is a small variation in results from one run to another because of the randomization present in the RFC) test runs evaluating the trained classifier using the six independent test data sets.

Table 3.3: Overall Test Results (10-Run Average)

Test Data Set #	Test Accuracy [%]	F1 Score (Weighted) [%]
All	99.44	99.45
1	99.02	99.44
2	99.07	99.06
3	99.57	99.68
4	99.99	99.99
5	99.97	99.97
6	98.87	98.90

Figure 3.5 presents the classifier’s accuracy in predicting each class in the full test data set (all six data sets). These results show that the classifier performs well in terms of accuracy and F1 score on all data sets and among all classes.

3.3.4 Classifier Generalization

Generalization is important because practical NILM implementations should have good multiclass prediction performance across a variety of load operating conditions and it is not practical to train the classifier with every possible condition. To evaluate generalization performance, the training data set is first segmented by class, then each segment is sorted by cycle observation RMS value, and the middle 95% of that class segment’s training data is discarded before proceeding to training and testing. An example of this for Class 6 is shown in Figure 3.6, where the difference between the green triangular and red diamond line plots of the training data show that all training

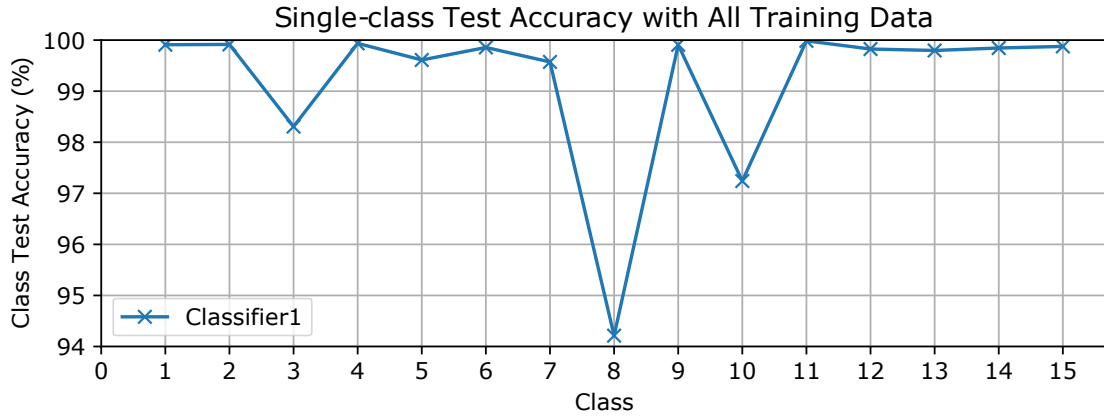


Figure 3.5: Test accuracy (full test data set) of Classifier 1 by class for the four-load experiment configuration when using all training data.

instances with RMS current values between 26.5 and 28.75 have been removed in the 5% training data set.

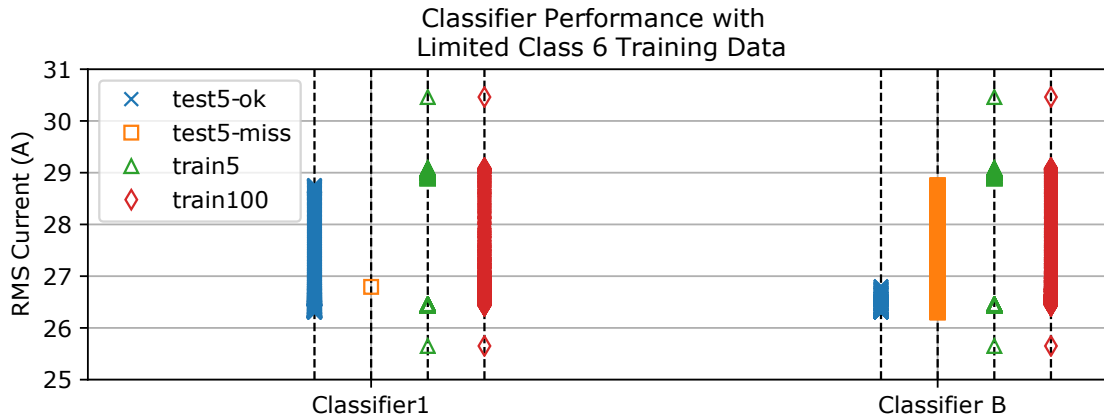


Figure 3.6: Classifier performance on Class 6 ($Y = [0, 1, 1, 0]$) test data when the middle (by RMS value) 95% of Class 6 instances in the training data set are removed before training. Accuracy is 97%.

The blue cross and orange square line plots in Figure 3.6 show how the range of instances in the test data set—which includes many values of RMS current in the range from 26.5–28.75 where no similar training instances were used—were predicted with an accuracy of 97%. To emphasize the generalization capability of this classifier, we trained another classifier, Classifier B, which did not have normalized harmonics and

wavelet features (further detail on the training of this classifier is given in [52]). This classifier had very good accuracy (similar to Classifier 1) when trained with the entire data set, but the accuracy significantly reduced when trained with the reduced training data; results show that Classifier 1 generalized significantly better than Classifier B when testing across all four-load configuration classes (see Figure 3.7) and especially for Class 6 (see Figure 3.6). For Class 6, Classifier B misclassified most samples whose RMS values were not in the reduced training data set.

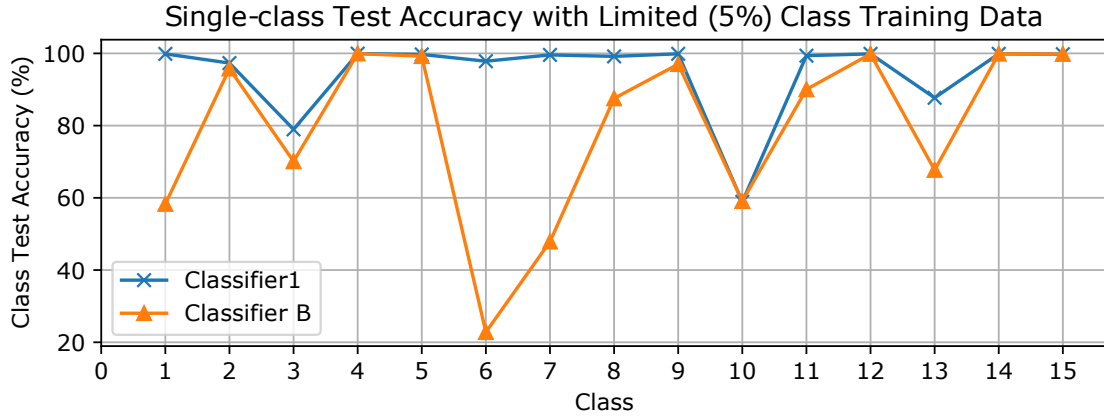


Figure 3.7: Test accuracy by class for the four-load experiment configuration when the middle (by RMS value) 95% of each class instance in the training data set is removed before training. By adding additional features to Classifier1 versus the limited sub-set used by Classifier B, classification generalization performance can be significantly improved.

3.3.5 Transient Performance

Figure 3.8 shows that label predictions match the true values throughout the entire data set (as expected from the 99.97% test accuracy in Table 3.3), even including most transient periods. Figure 3.9 shows an example of one such transient period (near $t=133$ s) where the Load 4 (compact fluorescent lights) turn-on event transient lasts nearly 14 ac cycles. During this event, the classifier correctly predicts each cycle, despite the cycles having a variety of wave shapes and magnitudes. This demonstrates the high-frequency and high-accuracy performance of the approach and its advantage over many existing NILM approaches, which would not predict a state change until after the entire

14 ac cycle event signature was detected.

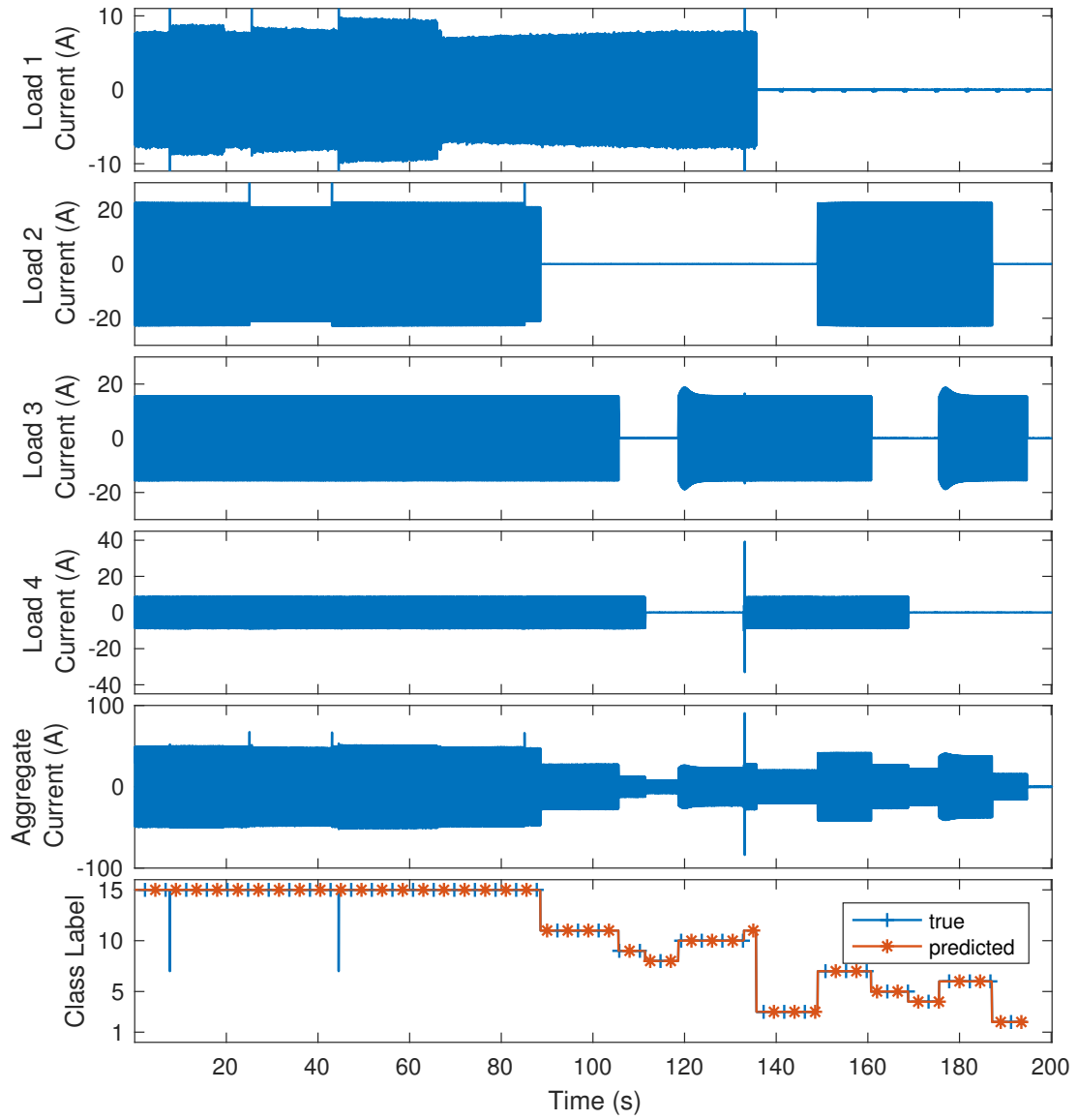


Figure 3.8: Entire test Data Set 5 with predicted versus true labels. It can be seen that the classifier performs well, even with a variety of load transitions.

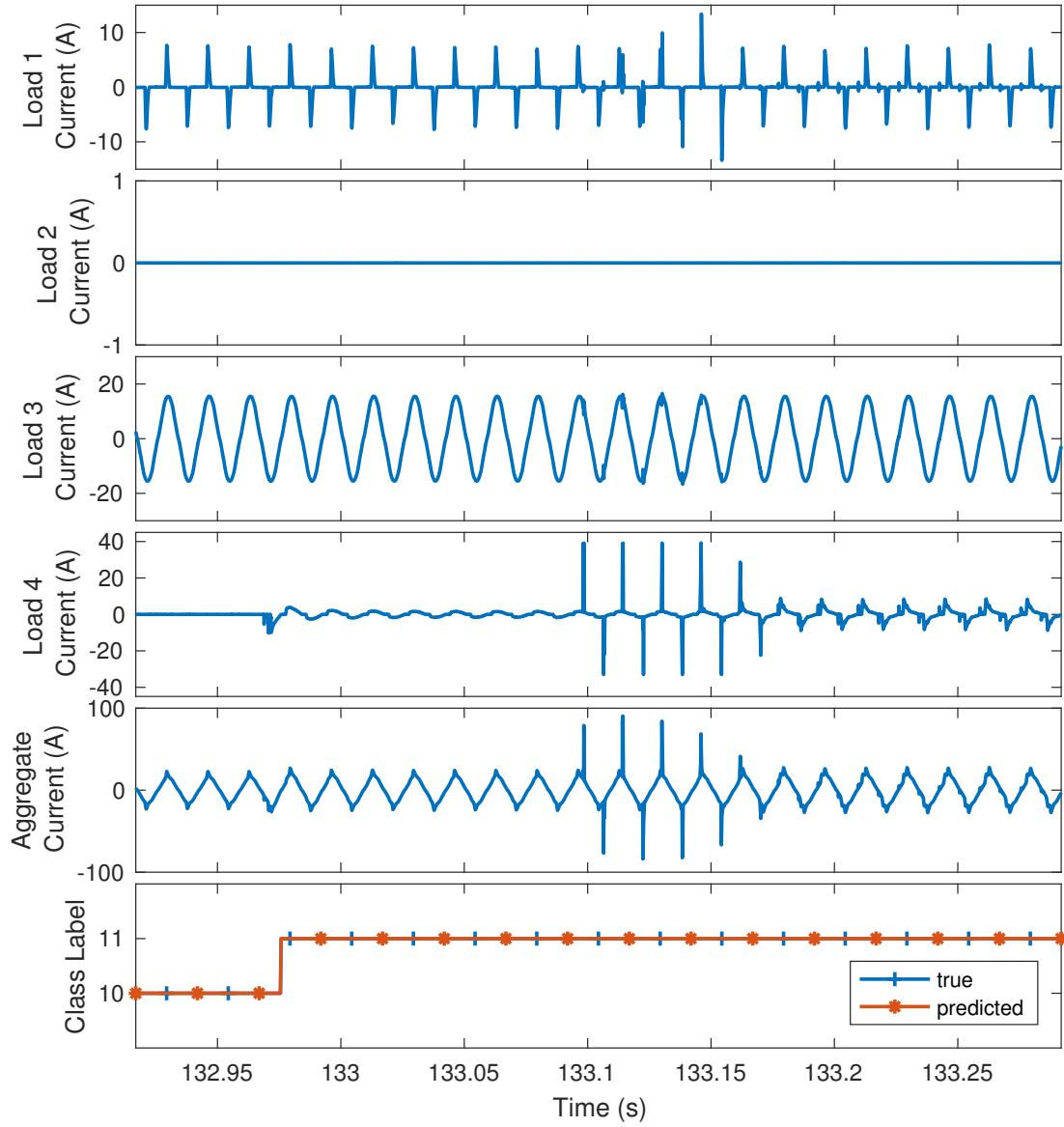


Figure 3.9: Magnified view of test Data Set 5 near $t = 133$ s. The classifier is able to predict well even though the turn-on transient of Load 4 exhibits a variety of wave shapes and magnitudes.

3.4 High-frequency, Binary Classification of Residential Appliance Loads

As mentioned in the previous section, when we consider the second configuration with eight loads, Classifier 1 does not scale well as the class size explodes to 255, and accuracy is significantly reduced. In this section, we present a multi-model approach where we train simple binary classifiers to detect the on/off state of a single load. This approach trains N_L RFC models for N_L loads (see left side of Figure 3.12). Here, model $CL2_i$ corresponds to load i . The set of all N_L models is termed as Classifier 2. The output of the classifier is a vector comprising individual predictions of each model. Let y_i be the prediction of model i . Then the output of Classifier 2 is $Y_c = [y_{c1}, \dots, y_{ci}, \dots, y_{cN_L}]$.

3.4.1 Classifier Tuning and Training

Here, we use the reduced synthetic data from the second eight-load experiment configuration for training, as described in Section 3.2. Two training methodologies are employed. The first training strategy follows the traditional approach of randomly shuffling the full data set (all seven independent data sets are used) and then splitting the data into a training set (80%) and test set (20%). The second approach is to use six of the seven independent data sets for training and testing on the remaining data set. This leads to seven distinct runs. The second approach emphasizes the generalization capabilities of the classifier. As detailed in Table 3.1, many data sets are quite different because they capture different running conditions of each load. These two methodologies are employed for the regression model as well. The hyperparameters used here are the same as with Classifier 1. Three additional wave shape features (described in Table. 3.2) were added to help differentiate the more complex eight-load configuration (see Section 3.2.4.1 for more details).

3.4.2 Results

When testing with 10 independent runs using the 80/20 split approach, the classifier's overall test accuracy is 99.2%, as averaged over 10 runs across all loads. The median accuracy for each load is shown in Figure 3.10 (Classifier), with the first and third quantiles plotted as error bars. As expected, the classifier does extremely well when

tested on similar data. With the second validation approach, when a completely new data set is used for testing, overall test accuracy is 95.0%, as averaged over seven runs across all loads. The median accuracy for each load is shown in Figure 3.11 (Classifier), with the first and third quantiles plotted as error bars. Clearly, the classification model generalizes reasonably well when tested on completely independent data sets.

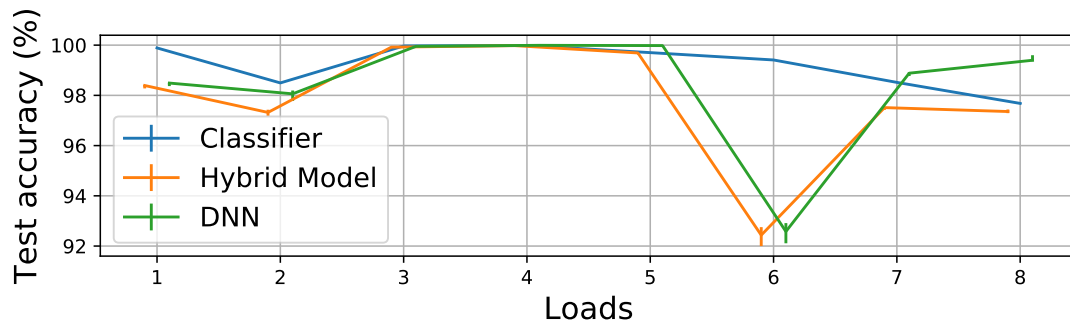


Figure 3.10: Median test accuracy with error bars for Classifier 2, DNN, and the hybrid model for first approach (80/20 split). The hybrid model and DNN curves are shifted slightly left and right, respectively, for visibility.

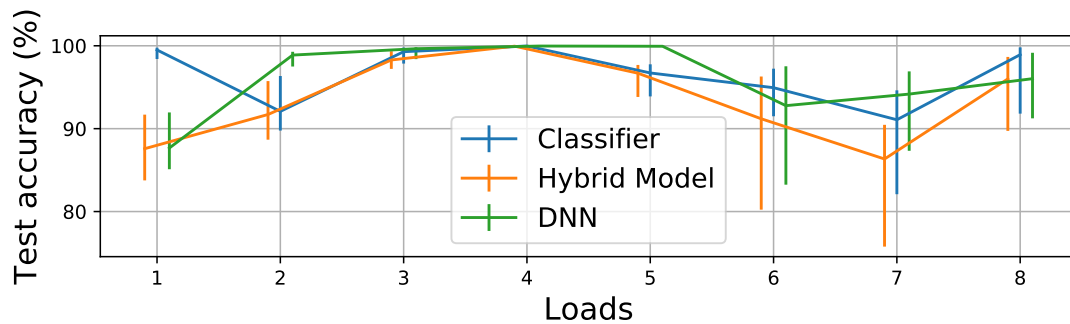


Figure 3.11: Median test accuracy with error bars for Classifier 2, DNN, and the hybrid model for second approach (independent data set for testing). The hybrid model and DNN curves are shifted slightly left and right, respectively, for visibility.

3.5 Power Prediction Using Regression

Deep neural networks (DNNs) are used in this work to predict the RMS currents of each load. Neural networks are biologically inspired mathematical models which are used to enable computers to learn from observational data. Neural networks are a network of "neurons," where each neuron is defined by a multi-input, single-output "activating function." For example, this function could be a weighted sum of all its inputs. Each layer of the neural network consists of multiple neurons, where each neuron takes outputs of all the neurons of the previous layer as its input. For the first layer, the number of inputs for each neuron is equal to the number of features. The number of neurons in the output layer is fixed and equal to the number of outputs of the model. In this work, we train a single neural network for each load; thus, the number of neurons in the outer layer is one. Intermediate layers are called "hidden" layers. Usually, a neural network is called a DNN if it has more than one hidden layer. DNNs are usually able to form complex non-linear links between features and output and are well suited to model complex NILM applications [42]. A detailed description of DNNs can be found in [53].

3.5.1 Problem Formulation

Here, we have a formulation similar to those in the previous sections with N_L loads, but the objective is to predict $\tilde{I}_i$, the RMS current consumption of the i^{th} load. To this end, we adopt a two-step hybrid approach (see Figure 3.12). The first step uses Classifier 2 of the previous section to predict the on/off state of each load. As mentioned earlier, classifier $CL2_i$ is the classifier for the i^{th} load. The second step applies a DNN model, DNN_i , to predict the current consumption of load i using the output of classifier $CL2_i$ as well as the same feature inputs given to Classifier 2.

3.5.2 Model Tuning and Training

The same reduced synthetic data sets and two training approaches that were used for Classifier 2 are used for regression. Similar to the tuning of RFCs, multiple DNNs are considered in a search for a model that can accurately predict the RMS current. The final classifier is a DNN with three hidden layers, each having 256 neurons. The input

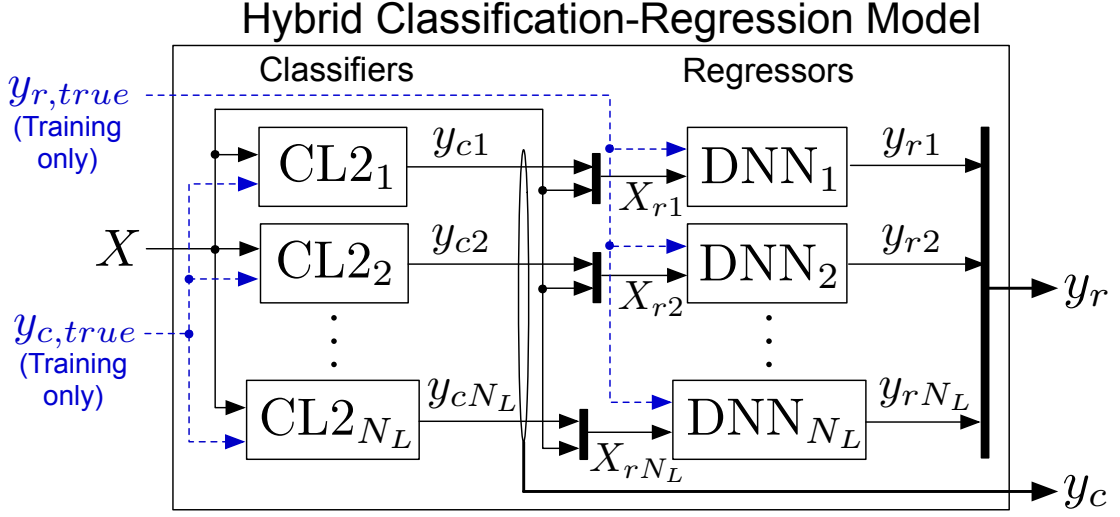


Figure 3.12: Hybrid classification-regression modeling approach. It can be seen that output of the classifiers are used by the regressors to predict individual load RMS current.

layer is chosen to have 128 neurons. The hyperparameters considered for tuning are the dropout rate for the hidden layers and the activation functions (including its internal parameters). We also consider removing or keeping the outliers (.001 percentile) as part of the hyperparameters. The final hyperparameters chosen depend on the training data and load being predicted. For example, one model has a dropout rate of 0.05, activation function as "LeakyRelu" with $\alpha = 0.05$, loss function as "mean_squared_error," optimizer as "adam," and performance metric as "mean_absolute_error." The output layer has a "linear" activation function, which is usual when using DNN for regression. Features used for regression are described in Table 3.2.

In contrast with RFCs, neural networks are very susceptible to the scaling of inputs and outputs. Thus, we apply a "power transformation" to the feature matrix X to make the input data more Gaussian. Specifically, we use a "Yeo-Johnson" transform [56], which supports both positive or negative data. For scaling the output y , we employ standard scaling, which removes the mean and scales the data to unit variance. The above methodology is implemented in Python using the keras package [57].

3.5.3 Results

To evaluate the regression model, we present two sets of results. The first set is the test accuracy of just the DNN model, assuming X has an accurate load on/off state input, whereas the second set of results is for the hybrid model. where X contains predicted state values from the classifier. To quantify the accuracy of the regression model, a test sample is deemed predicted true if the predicted RMS value from the DNN model is within $\pm 10\%$ of the true RMS value. In the case when all load is OFF, a test sample is deemed predicted true if the predicted RMS value is within $\pm 0.2A$. Figure 3.13 shows the true and predicted RMS currents from the hybrid model for Load #8 when using the 80/20 split training. These are 100 test samples (sorted by RMS) from the test data set. Here, we can see the model predicting the RMS current very accurately across different operating regimes (different power consumption) of this load. The overall test accuracy of the DNN and hybrid models for this approach is 98.35% and 97.77%, respectively, when taken as an average over 10 runs across all loads. The median accuracy for each load is shown in Figure 3.10 (DNN, hybrid model), with the first and third quantiles plotted as error bars. Similar to Classifier 2, both the DNN and hybrid models do very well when tested on data for which similar samples have been seen before.

With the second approach, when a completely new data set is used for testing, the overall test accuracy of the DNN and hybrid model is 94.77% and 91.77%, respectively, when averaged over 7 runs across all loads. The median accuracy for each load is shown in Figure 3.11 (DNN, hybrid model), with the first and third quantiles plotted as error bars. Figure 3.14 shows the true and predicted RMS current (by the hybrid model) of

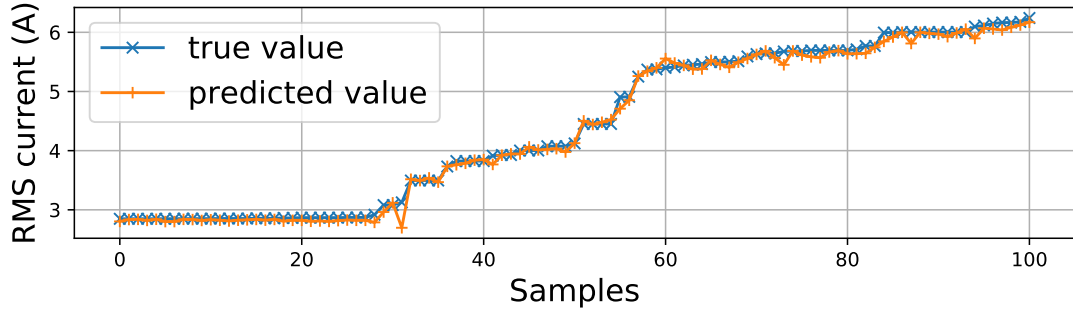


Figure 3.13: True and predicted RMS values for Load 8 for 80/20 training split

Load 8 using Data Set #3 for testing. These are 100 test samples (sorted by RMS) from the test data set. When compared to Figure 3.13, it is clear that for this testing data set, the model performs very well when predicting the RMS current for the main operating conditions, but it is not very accurate for the intermediate operating conditions. Similar to Classifier 2, the DNN model generalizes reasonably well when tested on completely independent data sets.

For this proof-of-concept study, the NILM approach is not implemented in real time, but the total processing time—including zero-crossing detection, feature extraction, and classifier and regression prediction—is measured (averaged over 80,000+ cycles) to be 10.42 ms, which is fast enough to support high-frequency load disaggregation and certainly well within the desired $\tau = 0.167$ s response time. This time includes 10.37 ms for feature extraction, 0.016 ms for classification, and 0.036 ms for DNN model prediction. This is measured on a laptop with an Intel i7 processor and Nvidia GeForce GTX 1050 Ti GPU and assuming no data acquisition time. With a response time 10 times faster than the desired response time, however, there is sufficient room for data acquisition and decreased performance if implemented on a slower computational platform while still meeting the response time requirement.

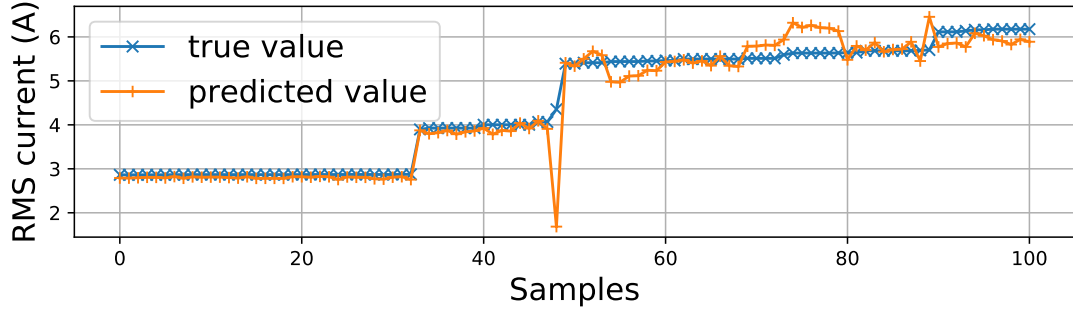


Figure 3.14: True and predicted RMS values for Load 8 with Data Set 3 used for testing

Chapter 4

Distribution Feeder-Scale Fast Frequency Response via Optimal Coordination of Net-load Resources

4.1 Introduction

4.1.1 Motivation and Background

Reliable operation of a power system depends on maintaining its frequency close to its nominal value at all times. In many grids worldwide, the displacement of conventional, high-inertia, synchronous generation by inverter-interfaced low-zero-inertia renewable generation has resulted in a decrease in total system inertia and a decline in frequency response performance from historic values. These trends are especially prevalent in isolated grids with high penetrations of renewable generation, such as Ireland/Northern Ireland and South Australia [58], but also evident in much larger interconnections in Europe [59] and the United States [60]. The outcome is that sudden imbalances in grid load and generation result in faster rates of change of frequency and larger frequency nadirs, making frequency regulation more challenging.

With continued grid modernization, new approaches to ensure sufficient frequency

regulation are required. Traditionally, mitigation of frequency deviations after a sudden loss of generation or load is driven by three primary factors: inertial response, primary frequency response (PFR), and secondary frequency response (SFR). System operators, such as ERCOT in Texas, USA and ENTSO-E in Europe, have conducted recent studies (e.g., [61, 59]) to examine the decreasing inertia in their systems and propose mitigation measures. ERCOT’s study highlights faster responding reserves as a key mitigation measure. Frameworks in different regions vary but generally include Fast Frequency Response (FFR), a response of the resource’s full allocated capacity within 1 s or less after the frequency crosses a predefined threshold or a dispatch signal is received. FFR is particularly effective because it can increase the time to system frequency nadir, allow additional time for PFR to deploy, and contribute to reducing the nadir’s magnitude [62, 63]. In [62], a study on the ERCOT system found that FFR can be up to 2.2 times more effective than PFR, depending on the system inertia. As of 2019, many system operators have implemented or are finalizing the implementation of new FFR ancillary service market products that are distinct from PFR (e.g., FFR under ERCOT’s NPRR863 or the U.K. National Grid’s Enhanced Frequency Response Service).

Demand-side resources, including DER and deferrable load (“net-load” resources), offer a largely untapped opportunity to provide valuable FFR services. DERs, such as solar inverters, have traditionally not contributed to frequency response [60], but recent work demonstrated new control methods to allow BTM solar inverters [64] and battery inverters [65] to rapidly respond (within 75 ms) to changes in grid frequency. At the solar power plant scale, the provision of FFR from multiple +20-MW solar power plants was demonstrated in [63]. Note that providing FFR from a solar inverter represents an opportunity cost; to provide increased active power output in response to a system loss of generation, the inverter must nominally operate off of its maximum power point (leaving “headroom” to increase the power output), forgoing some energy production and the associated revenue. Determining the optimal amount of headroom to leave for a given solar inverter’s operating conditions is not studied in the literature. Here, we propose an approach to determine the optimal headroom. Deferrable load is particularly well suited to provide FFR services because it does not incur an opportunity cost in preparing for FFR provision, and it can respond with its full capacity as quickly as a relay can be operated or a modern appliance’s power electronic front-end can be

controlled. In particular, thermostatically controlled loads (TCLs) and electric vehicles (EVs) are well-suited because of their inherent energy storage properties, which allow them to be switched off for a short period (typically at least a few minutes) without significantly compromising user comfort [66, 67]. TCLs represent a significant portion of total system load: more than 48% of residential electricity consumption in the U.S. in 2018 [68].

A key challenge to leveraging demand-side resources to provide FFR in associated ancillary service markets is the ability to provide sufficient capacity (historically, ≥ 1 -MW capacity is required) and reliable dispatchability to a particular set point. Recently, regulators are considering actions to remove barriers to the participation of DER aggregations in ancillary service markets, including reducing minimum capacity requirements [69]. Coordinating deferrable load and DERs as a virtual power plant (VPP) capable of FFR is particularly attractive because it not only allows aggregations of resources to meet market capacity participation limits but also leverages diversity, manages the variability of load and generation to enable dispatchability, and provides a framework to incentivize BTM participation and achieve the lowest-cost global dispatch solution. Also, by coordinating variable DERs and load in the same framework, a better determination of optimal DER operating headroom can be made.

4.1.2 Related Work

Few existing works address the coordination of net-load resources to provide FFR given the relatively new definition of FFR as a grid ancillary service product. For context, we examine related work, including existing PFR response approaches, which are often uncoordinated, as well as more general frequency response approaches that are coordinated.

Using deferrable load to contribute to grid services such as peak demand reduction and load following was explored as early as the late 1970s, though work on controlling loads fast enough to provide PFR has occurred only more recently. Earlier work includes a simulation study [70] that implemented autonomous PFR via temperature-based frequency trip points at individual refrigerators; simulation results demonstrated how autonomous PFR from 1,000 refrigerators could, in aggregate, provide PFR similar to that of a synchronous generator to the bulk power system. In [71], more than

50 electric water heaters and 150 clothes dryers were controlled in one of the earlier field demonstrations of fast (~ 400 -ms response times), autonomous PFR from loads. Later simulation studies identified overresponse, frequency oscillation, and loss of load diversity as potential issues with uncoordinated load frequency response [72, 73, 74, 75]. Strategies to mitigate potential issues in autonomous approaches have been proposed, including using uniformly distributed response time delays [67] and implementing random activation and re-activation frequency thresholds [76]. These *uncoordinated* PFR approaches demonstrate meaningful contribution to system PFR, but they are unable to provide dispatchability to the operator and will result in uncertainty of response and suboptimal dispatch.

Semicoordinated load PFR approaches (e.g., [77, 78, 79, 73, 80]), in which individual resources respond rapidly via autonomous control while having their control parameters periodically updated for coordination purposes, provide some improvement. Reference [73] proposed a supervised threshold determination approach in which electric water heaters are given updated frequency thresholds every 15 minutes and a given aggregate droop characteristic is achieved. Reference [80] demonstrated a similar approach applied to HVAC equipment with room temperature as feedback and updates made every minute. Reference [79] developed an hourly optimization-based approach that accounts for individual resource costs, comfort constraints, and global cost. Reference [77] extended the approach in [79] by adding load priorities and appliance off-time constraints to the optimization problem. Reference [78] solved a minimum cost commitment problem to compute the lowest cost interruptible load action to take for every given frequency deviation with hourly updates. These semicoordinated PRF approaches add increased capabilities to manage the penetration level and specific settings of resources participating autonomously to achieve a more predictable overall PFR. However, they all assume simplistic, slow-changing, aggregated load models and perform coordination updates on a minute to hour basis, resulting in significant uncertainty in their dispatch when applied to controlling real-world net-load resources that change their active power significantly faster. Given this uncertainty, it is unlikely that such approaches will be able to accurately achieve an aggregate power set point and are much more likely to make sub-optimal dispatch decisions.

Online coordination approaches, in which individual resources are coordinated after

a frequency event occurs, often have the advantage of optimal dispatch and the ability to integrate with automatic generation control (AGC), but they are generally too slow to meet PFR or FFR timescales. For example, in [81], eight optimization-based coordination frameworks in the literature are compared, and all have computation times of 10 s or greater. Distributed approaches are used to solve an optimal load-shedding problem [82] and an optimal load-control problem [83] on the +5-s timescale. Reference [84] developed a distributed approach for coordinating load aggregators that integrates with AGC for SFR. Reference [85] developed a centrally-managed solution where aggregated loads are controlled autonomously using control parameters precoordinated based on day-ahead estimates to provide PFR and a SFR supervisory control can be integrated with AGC. Reference [86] demonstrated similar combined SFR and PFR capabilities wherein resources with static droop characteristics provide PFR but then use a direct load frequency control approach for secondary control. These combined approaches have the advantage of fast control followed by achieving a dispatched set point on a slower timescale, but they still have significant uncertainty and nonoptimal dispatch in their PFR and are not fast enough for FFR. In our previous work [49], we experimentally demonstrated optimal coordination of four loads and an inverter within a single building for FFR within 142 ms.

4.1.3 Contribution and Organization

This chapter makes the following contributions:

- 1) We develop a first-of-its-kind control solution that optimally coordinates tens of thousands of deferrable loads and DERs across a distribution network to provide bulk power system-scale FFR that is dispatchable and accurate. Compared to existing frequency response solutions, the developed control platform is able to optimally coordinate a much larger number of net-load resources (up to 40,000 is shown), incorporate fast-changing BTM resources, and accurately meet a specific aggregate power set point while also achieving dispatch on a faster timescale (<500 ms) suitable for FFR.
- 2) We provide an analytical proof of the atomic control algorithm's optimality and then verify near-optimal (within 0.1%) performance of the algorithm as it is scaled to configurations with up to 40,000 net-load resources via Monte Carlo-style simulation. In all cases, dispatch solutions are achieved faster than traditional mixed-integer linear

program (MILP) optimization approaches and sufficiently fast to support coordinated FFR within 500 ms.

3) We present a solution for determining the optimal amount of headroom to operate solar inverters with to minimize opportunity cost while ensuring the FFR response viability of a building with the inverter and deferrable loads. A practical optimal headroom algorithm is given and analytically proven to match a bi-level optimization-based approach.

4) The entire platform is implemented and experimentally verified using a unique power hardware-in-the-loop (PHIL) demonstration including more than 100 powered loads and DERs connected to a real-world distribution network model from Colorado, USA. Test cases demonstrate the approach at multiple scales, including a full-scale demonstration with more than 10,000 deferrable loads and DERs.

The remainder of this chapter is organized as follows: An overview of the control platform is presented in Section 4.2, a definition of the net-load unit (NLU) building block in Section 4.3, an algorithm for determining optimal NLU inverter headroom that is developed with analytical guarantees in Section 4.4, algorithms for optimal FFR dispatch from a single NLU in Section 4.5, and algorithms for optimal FFR dispatch from multiple NLU and net-load groups (NLGs) that are developed and validated via simulation results in Section 4.6. An overview of the PHIL experiment design is presented in Section 4.2, results from a PHIL experiment representing an NLG with nine residential buildings are shown in Section 4.8, and results from the full-scale electric distribution system-scale PHIL experiment with coordinated FFR are given in Section 4.9.

4.2 Platform for Fast Frequency Response from Optimally-coordinated Net-load Resources

4.2.1 Solution Overview

The objective of this work is to achieve a platform that can optimally coordinate tens of thousands of net-load resources connected in an electric distribution network as a VPP to provide FFR to the bulk power system that is dispatchable and meets general FFR service performance requirements (i.e., ≥ 1 -MW capacity, response within 500 ms, and

achieving set point within 1%). As shown in Figure 4.1, the developed platform leverages a hierarchical structure that is scalable, preserves user privacy, enhances reliability, and provides optimal coordination at selectable levels. At the lowest level, individual DERs and flexible loads are organized into NLUs wherein a net-load unit controller (NLUC) manages participating net-load resources. Loads are controlled through plug-or circuit-level load sensing and control interfaces (LSCIs), which provide high-speed (half ac cycle) power measurement and on/off actuation capabilities (see [49] for more details). NLUs, typically representing individual buildings, are organized into NLGs, which generally represent all the NLUs connected to a single distribution transformer and are managed by a net-load group controller (NLGC). For global coordination, a distribution system master controller (DSMC) can optimally dispatch all NLUs across a distribution network. In this configuration, the computational burden of optimal dispatch calculation is distributed across the NLGCs, allowing nearly constant dispatch calculation times regardless of how many NLGs are participating.

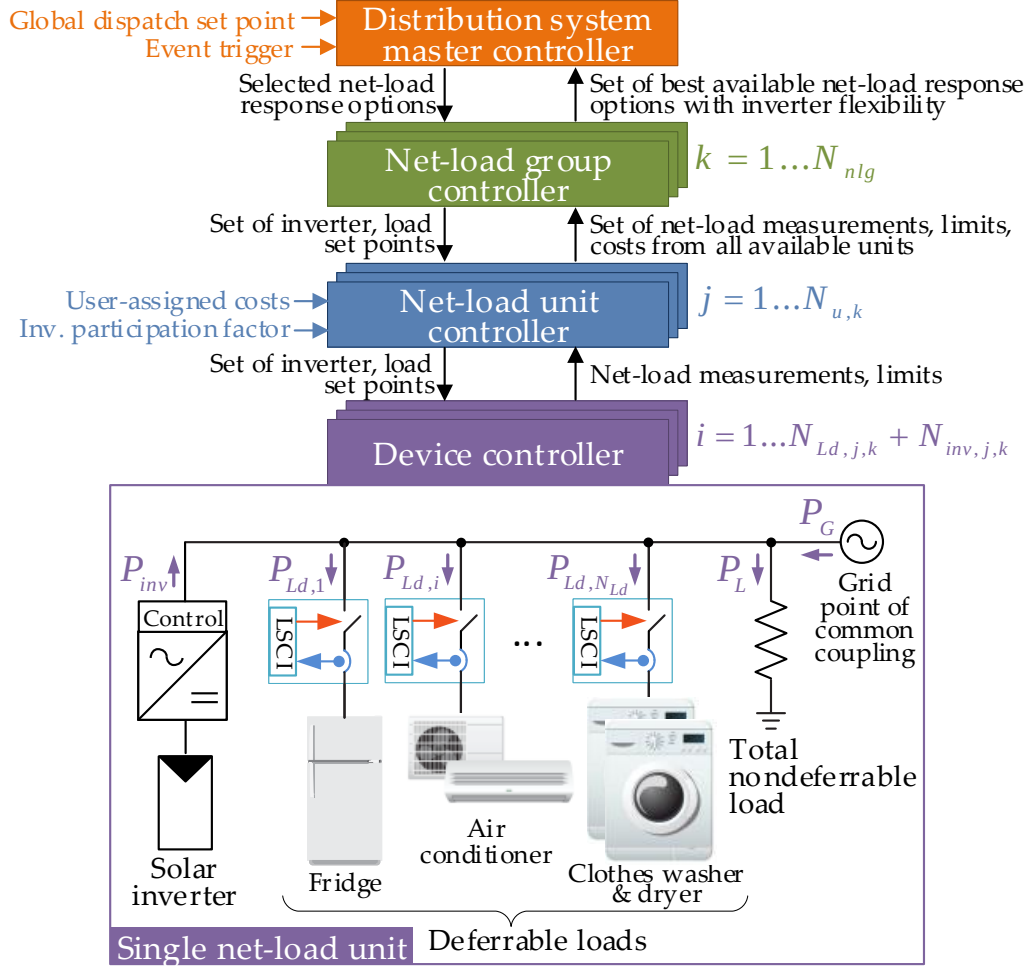


Figure 4.1: Hierarchical optimal net-load coordination for FFR approach

4.3 Net-load Unit Definition

At its most basic level, an NLU (e.g., see the bottom of Figure 4.1) contains nondeferrable load with aggregate active power consumption, P_L ; N_{Ld} deferrable loads with active power consumptions, $\{P_{Ld,i}\}_{i=1}^{N_{Ld}}$; an inverter that outputs active power, P_{inv} ; and active power consumption, P_G (i.e., $P_G > 0$ represents power consumed by the NLU), from the NLU's grid connection. Note that power consumed or supplied at each point throughout this problem formulation are all time-varying quantities; however, in most cases, the time dependence (t) is dropped for ease of notation. All $P_{Ld,i}$ are deferrable

such that their final power consumption is $\hat{P}_{Ld,i} = \bar{x}_i P_{Ld,i} = (1 - x_i) P_{Ld,i}$ after being issued an on/off command, $x_i \in \{0, 1\}$. That is, $x_i = 1$ denotes that the load has been deferred or switched off. The inverter is initially operating at power $P_{inv}(t_0) = P_{inv,base}$ and has operating limits $0 \leq P_{inv} \leq P_{inv}^{max}$. Upon command, the inverter can change its output power to $P_{inv} = P_{inv,base} + P_{inv\Delta}$, where $P_{inv\Delta}$ is the inverter's change in power from $P_{inv}(t_0)$. The inverter's change in power is limited: $P_{inv\Delta,dn} \leq P_{inv\Delta} \leq P_{inv\Delta,up}$, where $-P_{inv,base} \leq P_{inv\Delta,dn} \leq 0$ limits how much the inverter can decrease its power output, and $0 \leq P_{inv\Delta,up} \leq (P_{inv}^{max} - P_{inv,base})$ limits how much the inverter can increase its power output. Throughout this work, we assume $P_{inv\Delta,dn} = -P_{inv,base}$ for simplicity because curtailment of inverter output is generally not a significant constraint. $P_{inv\Delta,up}$ is known as the inverter's headroom and will be determined in real time using the algorithm developed in Section 4.4. By Kirchoff's laws, the power balance on the net-load system is given by: $\sum_{i=1}^{N_{Ld}} \hat{P}_{Ld,i} + P_L - P_{inv} = P_G$.

The NLU owner determines in advance of the frequency contingency event which loads are deferrable vs. nondeferrable and also assigns a participation unit cost, c_i , for each deferrable load as well as the inverter (c_{inv}) such that the owner will be compensated a total of $c_{tot} = \sum_{i=1}^{N_{Ld}} x_i(t) \cdot c_i \cdot P_{Ld,i}(t_0) + c_{inv} |P_{inv}(t) - P_{inv}(t_0)|$ after providing FFR. These determinations are updated by the building manager in real time so that any frequency event response considers the owner's latest preferences. This simplistic compensation scheme is meant to replicate key aspects of market participation, including users determining if they want to participate and at what price, determining total cost for an aggregated response action, and compensating users. Each NLU periodically (and again immediately after receiving an event response trigger) sends its measured P_G , $\{\hat{P}_{Ld,i}\}_{i=1}^{N_{Ld}}$, P_{inv} , $P_{inv\Delta,dn}$, and $P_{inv\Delta,up}$ to the NLGC. The NLGC uses this information to determine an optimal $\Delta P_G^* = P_G(t) - P_G(t_0)$ for each NLU shortly after an FFR event is triggered and subsequently verify the response performance of each NLU. The NLGC will ensure that NLU's are provided with ΔP_G^* set points within their operating capability; that is, $-(P_{inv\Delta,up} + \sum_{i=1}^{N_{Ld}} \hat{P}_{Ld,i}) \leq \Delta P_G^* \leq P_{inv\Delta,dn}$.

4.4 Optimal NLU Inverter Headroom

Here, we consider the problem of determining the optimal amount of headroom, $P_{inv\Delta,up}$, that a solar inverter within an NLU should maintain prior to an FFR underfrequency response dispatch event. This issue is important in the context of the overall NLU optimal dispatch problem (see Section 4.5.1) because the inverter headroom is set in advance of a dispatch event and determines what values of dispatch command ΔP_G^* can be achieved by the NLU. Insufficient inverter headroom can make the single NLU optimal dispatch problem infeasible or result in a sub-optimal dispatch solution.

To illustrate the need for headroom, consider the simple case where an NLU has $N_{Ld} = 2$ deferrable loads, $P_{Ld}(t_0) = [100, 300]W$, which are both initially on ($x(t_0) = [0, 0]$), and one solar inverter. A grid underfrequency event occurs and shortly later the NLU is given a dispatch command, $\Delta P_G^* = P_G(t) - P_G(t_0)$, to consume less grid power ($\Delta P_G^* < 0$). To meet this dispatch command, the NLU must curtail load(s) and/or modify inverter output. Controlling the on/off state of the N_{Ld} deferrable loads in the NLU achieves $2^{N_{Ld}}$ possible NLU power responses ($P_{LdT_p} = \{0, 100, 300, 400\}W$ for this case). Unless $|\Delta P_G^*| \in P_{LdT_p}$, a change in inverter output will be required for the NLU's total grid power response, $\Delta P_G = -\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} - P_{inv\Delta}$, to meet the dispatch command, ΔP_G^* , exactly, which is a condition of NLU dispatch problem feasibility (see Section 4.5.1). For example, if a command of $\Delta P_G^* = -125W$ was given, there are two possible solutions: (i) loads are curtailed to $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = 100W$, requiring $P_{inv\Delta} = 25W$; or (ii) loads are curtailed to $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = 300W$, requiring $P_{inv\Delta} = -175W$. Solution (i) requires both less curtailment of load and less total change in inverter power, resulting in a lower cost dispatch solution, but this requires the solar inverter to increase its power output. However, this solution would not be feasible if the solar inverter is already operating at its maximum power ($P_{inv} = P_{inv}^{max}$, which is a typical goal of solar inverters) with no headroom available. Solution (ii) is a higher cost solution, but it does not require any upward headroom. For this example ($\Delta P_G^* = -125W$), having ensured inverter headroom of $P_{inv\Delta,up} \geq 25W$ in advance of the dispatch event would have allowed solution (i) to be feasible and the NLU dispatch problem to achieve the global optimal result. If insufficient inverter headroom ($P_{inv\Delta,up} < 25W$) is available in advance, then the NLU dispatch problem would be forced to use solution (ii), a

less optimal result. This illustrates the needed headroom for a particular ΔP_G^* . The value of ΔP_G^* is not known in advance, however, and therefore to ensure feasibility for any possible $-(P_{inv\Delta,up} + \sum_{i=1}^{N_{Ld}} \hat{P}_{Ld,i}) \leq \Delta P_G^* \leq P_{inv\Delta,dn}$ given to the NLU, the optimal inverter headroom problem must consider all possible ΔP_G^* . In the simple case illustrated, the global maximum headroom required occurs for $\Delta P_G^* = -200 W$, where $P_{inv\Delta,up} = 100 W$ is required.

4.4.1 Problem Formulation

The optimal NLU inverter headroom problem objective is to ensure sufficient headroom, $P_{inv\Delta,up}$, such that the optimal result of the subsequent NLU optimal dispatch problem at any $-(P_{inv\Delta,up} + \sum_{i=1}^{N_{Ld}} \hat{P}_{Ld,i}) \leq \Delta P_G^* \leq 0$ is feasible while simultaneously minimizing $P_{inv\Delta,up}$ to minimize the opportunity cost of operating the solar inverter off of its maximum power point. This can be accomplished by solving the following bi-level optimization problem:

$$\alpha = \max_{\Delta P_G^*} f_2 \quad (4.1)$$

$$\text{s.t.} \quad -(P_{inv}^{max} + \sum_{i=1}^{N_{Ld}} P_{Ld,i}) \leq \Delta P_G^* \leq 0 \quad (4.2)$$

$$\text{where } f_2 = \min_{\tilde{x}_i} \left| \Delta P_G^* + \sum_{i=1}^{N_{Ld}} \tilde{x}_i P_{Ld,i} \right| \quad (4.3)$$

$$\text{s.t.} \quad \tilde{x}_i \in \{0, 1\} \forall i \quad (4.4)$$

and then setting $P_{inv\Delta,up} = \min(\alpha, P_{inv}^{max})$. Note that if $\alpha > P_{inv}^{max}$, then the NLU will be unable to meet every possible given ΔP_G^* , though setting $P_{inv\Delta,up} = P_{inv}^{max}$ will give the NLU the greatest probability of meeting all possible ΔP_G^* . In this bi-level optimization problem, the inner problem (4.3)–(4.4) determines the inverter headroom required at the NLU dispatch solution for a given ΔP_G^* when utilizing the closest possible load combination, and the outer problem (4.1)–(4.2) searches for the maximum inverter headroom result across all ΔP_G^* possible for the NLU's capacities. Here, (4.2) considers an upper limit of $\Delta P_G^* \leq 0$ because we focus on underfrequency events where inverter upward headroom is utilized.

4.4.2 Solution

Here, we develop a fast algorithm that solves the optimal inverter headroom problem (4.1)–(4.4) and without requiring a bi-level optimization solver. A simple recurrence-based implementation is given in Algorithm 1. A proof that this algorithm achieves the optimal result is given in Theorem 1.

Theorem 1. *Let $P_{Lds} = \langle P_{Ld}, \leq \rangle$, where P_{Lds} is an ordered version of P_{Ld} with order given by $\leq$. Then, the result, α , of the optimal inverter headroom problem (4.1)–(4.4) can be obtained as $\alpha = \max_{1 \leq i \leq N_{Ld}} (P_{Lds,i} - \sum_{n=1}^{i-1} P_{Lds,n})/2$.*

Proof. The net-load system's actual change in grid power consumption is $\Delta P_G = -\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} - P_{inv\Delta}$. To ensure problem feasibility, the following must be achieved: $\Delta P_G = \Delta P_G^*$ with $P_{inv\Delta,dn} \leq P_{inv\Delta} \leq P_{inv\Delta,up}$ for $0 \leq P_{inv\Delta,up} \leq (P_{inv}^{max} - P_{inv,base})$ and $0 \leq P_{inv,base} \leq P_{inv}^{max}$.

Consider all possible $-(P_{inv\Delta,up} + \sum_{i=1}^{N_{Ld}} \hat{P}_{Ld,i}) \leq \Delta P_G^* \leq P_{inv\Delta,dn}$ for this problem. If $\Delta P_G^* \leq -\sum_{i=1}^{N_{Ld}} P_{Ld,i}$, a trivial solution of $x_i = 1 \forall i$ and $P_{inv\Delta} = -\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} - \Delta P_G^*$ is used. Similarly, if $\Delta P_G^* \geq 0$, $x_i = 0 \forall i$ and $P_{inv\Delta} = \Delta P_G^*$. We then consider the remaining cases where $-\sum_{i=1}^{N_{Ld}} P_{Ld,i} < \Delta P_G^* < 0$. For convenience, let $\Delta P^* = -\Delta P_G^*$ and consider the remaining cases where $0 < \Delta P^* < \sum_{i=1}^{N_{Ld}} P_{Ld,i}$:

Let $P_{Ldc} = (P_{Ldc,i})_{i=1}^{N_{Ldc}}$ be the ordered set of all $N_{Ldc} = 2^{N_{Ld}}$ possible load combinations from the elements of P_{Lds} in sequential binary order with $P_{Lds,1}$, the smallest load, being associated with the least significant bit. That is, $P_{Ldc,i} = \sum_{j=1}^{N_{Ld}} y_{ij} P_{Lds,j}$ and

$$y_{ij} = \begin{cases} 1 & \text{if } (i-1) \bmod 2^j \geq 2^{j-1} \\ 0 & \text{otherwise} \end{cases}.$$

Claim 1: For any given $0 < \Delta P^* < \sum_{i=1}^{N_{Ld}} P_{Ld,i}$, there exists at least one $1 \leq j \leq (N_{Ldc} - 1)$, such that $P_{Ldc,j} \leq \Delta P^* \leq P_{Ldc,j+1}$.

Thus, we can consider all cases $0 < \Delta P^* < \sum_{i=1}^{N_{Ld}} P_{Ld,i}$ by examining every discrete interval $P_{Ldc,j} \leq \Delta P^* \leq P_{Ldc,j+1}$ for $1 \leq j \leq (N_{Ldc} - 1)$ and considering problem feasibility and inverter headroom requirements for any ΔP^* in that interval.

Claim 2: Within every discrete interval $P_{Ldc,j} \leq \Delta P^* \leq P_{Ldc,j+1}$, the net-load dispatch problem remains feasible as long as $P_{inv\Delta,up} \geq \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2}$.

Proof: Let $\Delta P^* = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2}$. Within this interval, there are two possible resulting load curtailment outcomes:

Case 1: $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j}$ and so

$$P_{inv\Delta} = \Delta P^* - \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} - P_{Ldc,j} = \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2}$$

Case 2: $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j+1}$ and so

$$P_{inv\Delta} = \Delta P^* - \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} - P_{Ldc,j+1} = -\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2}$$

Therefore, $-\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2} \leq P_{inv\Delta} \leq \frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2}$ and $P_{inv\Delta,up} \geq \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2}$.

We then show that increasing or decreasing $\Delta P^* = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2}$ will not extend these bounds:

Case 1 (decrease): $\Delta P^* = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} - \epsilon$. Here, $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j}$ because this will minimize $P_{inv\Delta,up}$.

$$\text{Therefore, } P_{inv\Delta} = \Delta P^* - \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} - \epsilon - P_{Ldc,j} = \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} - \epsilon$$

Clearly, $-\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2} \leq \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} - \epsilon \leq \frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2}$, so the bounds on $P_{inv\Delta}$ are not extended for this case.

Case 2 (increase): $\Delta P^* = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} + \epsilon$. Here, $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j+1}$

because this will minimize $P_{inv\Delta,up}$.

$$\text{Therefore, } P_{inv\Delta} = \Delta P^* - \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j} + \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} + \epsilon - P_{Ldc,j+1} = -\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2} + \epsilon$$

Clearly, $-\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2} \leq -\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2} + \epsilon \leq \frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2}$, so the bounds on $P_{inv\Delta}$ are not extended for this case.

Therefore, increasing or decreasing ΔP^* does not extend the bounds on $P_{inv\Delta}$, so $P_{inv\Delta,up} \geq \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2}$ remains.

Next, we show that the bounds $-\frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2} \leq P_{inv\Delta} \leq \frac{(P_{Ldc,j+1} - P_{Ldc,j})}{2}$ on $P_{inv\Delta}$

cannot be made narrower (e.g., $\frac{-(P_{Ldc,j+1}-P_{Ldc,j})}{2} + \epsilon \leq P_{inv\Delta} \leq \frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} - \epsilon$) without causing problem infeasibility. To show this, choose $\Delta P^* = P_{Ldc,j} + \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2}$.

As before, there are then two possible cases:

Case 1: $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j}$ and so

$$P_{inv\Delta} = \Delta P^* - \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j} + \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2} - P_{Ldc,j} = \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2}$$

However, $\frac{P_{Ldc,j+1}-P_{Ldc,j}}{2} \not\leq \frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} - \epsilon$ and therefore the problem is infeasible with bounds $\frac{-(P_{Ldc,j+1}-P_{Ldc,j})}{2} + \epsilon \leq P_{inv\Delta} \leq \frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} - \epsilon$.

Case 2: $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j+1}$, so:

$$P_{inv\Delta} = \Delta P^* - \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} = P_{Ldc,j} + \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2} - P_{Ldc,j+1} = -\frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2}$$

However, $-\frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} \not\geq -\frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} + \epsilon$ and therefore the problem is infeasible with bounds $\frac{-(P_{Ldc,j+1}-P_{Ldc,j})}{2} + \epsilon \leq P_{inv\Delta} \leq \frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} - \epsilon$.

Thus, we have shown that the bounds of at least $-\frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2} \leq P_{inv\Delta} \leq \frac{(P_{Ldc,j+1}-P_{Ldc,j})}{2}$ on $P_{inv\Delta}$ are required to maintain problem feasibility, and therefore $P_{inv\Delta,up} \geq \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2}$.

Claim 3: The maximum $P_{inv\Delta,up,j} \geq \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2}$ across all discrete intervals $P_{Ldc,j} \leq \Delta P^* \leq P_{Ldc,j+1}$ for all $1 \leq j \leq (N_{Ldc}-1)$, and therefore for any $0 < \Delta P^* < \sum_{i=1}^{N_{Ld}} P_{Ld,i}$, can be obtained using the simplified expression:

$$P_{inv\Delta,up}^{max} = \alpha = \max_{1 \leq i \leq N_{Ld}} \frac{P_{Lds,i} - \sum_{n=1}^{i-1} P_{Lds,n}}{2} \quad (4.5)$$

Proof: $P_{inv\Delta,up}^{max}$ can be obtained by simply examining the $P_{inv\Delta,up,j} \geq \frac{P_{Ldc,j+1}-P_{Ldc,j}}{2}$ at each discrete interval $j = 1 \dots N_{Ldc} - 1$ and taking the maximum. That is:

$$P_{inv\Delta,up}^{max} = \max_{1 \leq j \leq N_{Ldc} - 1} \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} \quad (4.6)$$

Note that each $\frac{P_{Ldc,j+1}-P_{Ldc,j}}{2}$ can be evaluated by considering the constituent loads $P_{Lds,i}$ for all $i = 1 \dots N_{Ld}$ that change state (i.e., $y_{ij} \neq y_{i(j+1)}$) between $P_{Ldc,j}$ and $P_{Ldc,j+1}$. Because P_{Ldc} is in sequential binary order and P_{Lds} is in sequential order, the

sequence of loads changing as j increases $j = 1 \dots N_{Ldc} - 1$ follows a binary pattern wherein:

$$\begin{aligned} & \text{for each } i \quad \text{in} \quad i = 1 \dots N_{Ld} : \\ & (P_{Ldc,k+1} - P_{Ldc,k}) = P_{Lds,i} \\ & \quad \text{where} \quad k = 2^i \cdot m - 2^{i-1} \\ & \quad \text{for all} \quad m = 1 \dots 2^{N_{Ld}-1} \end{aligned}$$

This repeating pattern has the consequence that the only unique $(P_{Ldc,j+1} - P_{Ldc,j})$ differences not already considered in earlier intervals $(P_{Ldc,j+1-\epsilon} - P_{Ldc,j-\epsilon})$ (where $\epsilon \in \mathbb{Z}$, $\epsilon = 1 \dots (j-1)$) occur at the intervals $(P_{Ldc,(2^i-2^{i-1}+1)} - P_{Ldc,(2^i-2^{i-1})}) \forall i = 1 \dots N_{Ld}$ where:

$$\begin{aligned} (P_{Ldc,(2^i-2^{i-1}+1)} - P_{Ldc,(2^i-2^{i-1})}) &= P_{Lds,i} - \sum_{n=1}^{i-1} P_{Lds,n} \\ &\forall i = 1 \dots N_{Ld} \end{aligned}$$

Therefore:

$$\begin{aligned} P_{inv\Delta,up}^{max} &= \max \frac{P_{Ldc,j+1} - P_{Ldc,j}}{2} \\ &\quad \text{across all } j = 1 \dots (N_{Ldc} - 1) \\ \Rightarrow P_{inv\Delta,up}^{max} &= \max \frac{(P_{Ldc,(2^i-2^{i-1}+1)} - P_{Ldc,(2^i-2^{i-1})})}{2} \\ &\quad \text{across all } i = 1 \dots N_{Ld} \\ \Rightarrow \alpha = P_{inv\Delta,up}^{max} &= \max \frac{P_{Lds,i} - \sum_{n=1}^{i-1} P_{Lds,n}}{2} \\ &\quad \text{across all } i = 1 \dots N_{Ld} \end{aligned}$$

□

```

1: Require  $P_{Lds}, P_{inv}^{max}$ 
2: function maxLD( $lds, ldSum, n$ )
3:   if  $n == (\text{length}(lds) - 1)$  then
4:     return  $lds[n] - ldSum$ 
5:   else
6:     return  $\max((lds[n] - ldSum), \text{maxLD}(lds, ldSum + lds[n], n + 1))$ 
7:   end if
8: end function
9:
10: Begin
11:  $P_{inv\Delta,up} = \min(\text{maxLD}(\text{sorted}(P_{Lds}), 0, 0)/2, P_{inv}^{max})$ 
12: return  $P_{inv\Delta,up}$ 

```

Algorithm 1: Optimal NLU inverter headroom recurrence

4.5 Optimal Fast Frequency Response from a Single Net-load Unit

4.5.1 Dispatch Problem Formulation

Prior to an FFR dispatch event occurring, a single NLU has initial conditions of $x_i = 0$ for all i (all participating deferrable loads are on) and $P_{inv} = P_{inv,base}$ with headroom $P_{inv\Delta,up}$. For simplicity in this single-unit problem, resources are all considered to have unit cost (more complex cases with varying cost are considered in Section 4.6). The NLU is given a commanded change in grid active power consumption, ΔP_G , to follow to provide the requested FFR. The objective of the NLU is to track an active power consumption at its grid connection of $P_G^* = P_G(t_0) + \Delta P_G$, while minimizing the amount of deferrable load that must be curtailed and the change in inverter power required. This objective is given by:

$$\mu = \min_{x, P_{inv\Delta}} \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} \quad (4.7)$$

$$\text{subject to } \sum_{i=1}^{N_{Ld}} (1 - x_i) P_{Ld,i} + P_L - P_{inv,base} \quad (4.8)$$

$$- P_{inv\Delta} = P_G^* \quad (4.9)$$

$$P_{inv\Delta,dn} \leq P_{inv\Delta} \leq P_{inv\Delta,up} \quad (4.10)$$

$$x_i \in \{0, 1\} \forall i = 1 \dots N_{Ld}$$

Solving this optimization problem yields the solution $(x^\mu, P_{inv\Delta}^\mu)$. Let the feasible set for this optimization problem be $\Omega^\mu := \{(x, P_{inv\Delta}) \mid (4.8)\text{--}(4.10) \text{ are met}\}$.

4.5.2 Solution

The dispatch problem (4.7)–(4.10), with its mix of integer (e.g., x_i) and real-valued (e.g., $P_{inv\Delta}$) decision variables, can be solved using an MILP. This approach is possibly tractable for the single NLU dispatch problem with a modest number of decision variables; however, it is too slow to support dispatch on FFR timescales for larger problems (e.g., dispatching 1000s of net-load resources), as shown in Section 4.6.3.2. Instead, we develop a simpler approach that obtains the same optimization result for a single NLU (and nearly optimal for the aggregated case) but with faster performance and better scalability.

The dispatch problem (4.7)–(4.10) can be simplified to a binary programming problem with relaxed constraint (4.8), followed by a single variable update, as shown in (4.15). The advantage is that it can be solved using a dynamic program, and we show that it can be solved faster using this method than with a standard MILP. The simplified approach is formulated as:

$$\nu = \min_x \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} \quad (4.11)$$

$$\text{subject to } \sum_{i=1}^{N_{Ld}} (1 - x_i) P_{Ld,i} + P_L - P_{inv,base} \quad (4.12)$$

$$- P_{inv\Delta,up} \leq P_G^*$$

$$\sum_{i=1}^{N_{Ld}} (1 - x_i) P_{Ld,i} + P_L - P_{inv,base} \quad (4.13)$$

$$- P_G^* \geq P_{inv\Delta,dn}$$

$$x_i \in \{0, 1\} \forall i = 1 \dots N_{Ld} \quad (4.14)$$

Let x^ν be the solution to this optimization problem. Finally, to complete the simplified net-load dispatch, $P_{inv\Delta}$ is set so that P_G^* is met exactly:

$$P_{inv\Delta}^\nu := \sum_{i=1}^{N_{Ld}} (1 - x_i^\nu) P_{Ld,i} - P_{inv,base} - P_G^*. \quad (4.15)$$

Let the feasible set for this optimization problem be $\Omega^\nu := \{(x, P_{inv\Delta}) \mid (4.12)-(4.14) \text{ are met}\}$. We prove that this simplified approach yields the same optimal solution as the optimal MILP approach in Section 4.5.3.

4.5.3 Solution Equivalency Proof

Theorem 2. $\nu = \mu$. *That is, the simplified net-load dispatch problem (4.11)-(4.15) yields the same solution as the net-load dispatch optimization problem (4.7)-(4.10).*

Proof. When solving the optimization problem (4.7)–(4.10), a final optimal solution $(x^\mu, P_{inv\Delta}^\mu)$ is obtained. Note that $(x^\mu, P_{inv\Delta}^\mu) \in \Omega^\mu$. When solving the optimization problem (4.11)–(4.14), x^ν is obtained, and then $P_{inv\Delta}^\nu$ is set as given in (4.15). This yields a final solution $(x^\nu, P_{inv\Delta}^\nu)$. Note that $(x^\nu, P_{inv\Delta}^\nu) \in \Omega^\nu$.

First, we show $\mu \leq \nu$:

We will compare $(x^\nu, P_{inv\Delta}^\nu) \in \Omega^\nu$ of the ν problem and the feasible set Ω^μ of the μ problem. To begin, consider the result of (4.15):

$$P_{inv\Delta}^\nu = \sum_{i=1}^{N_{Ld}} (1 - x_i^\nu) P_{Ld,i} - P_{inv,base} - P_G^*$$

By (4.12), $\sum_{i=1}^{N_{Ld}} (1 - x_i) P_{Ld,i} - P_{inv,base} - P_G^* \leq P_{inv\Delta,up}$ so:

$$P_{inv\Delta}^\nu \leq P_{inv\Delta,up}$$

Equation (4.13) guarantees $P_{inv\Delta}^\nu \geq P_{inv\Delta,dn}$, and therefore:

$$P_{inv\Delta,dn} \leq P_{inv\Delta}^\nu \leq P_{inv\Delta,up}$$

This and (4.15) satisfy (4.8)–(4.9), therefore $P_{inv\Delta}^\nu$ meets all constraints of Ω^μ placed on $P_{inv\Delta}$.

By (4.14):

$$x^\nu \in \{0, 1\}$$

which satisfies (4.10), so x^ν meets all constraints of Ω^μ placed on x .

Therefore:

$$(x^\nu, P_{inv\Delta}^\nu) \in \Omega^\mu \quad (4.16)$$

As:

$$\begin{aligned} \mu &= \min_x \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} \\ \text{subject to } & (x^\mu, P_{inv\Delta}^\mu) \in \Omega^\mu \end{aligned}$$

Because $(x^\nu, P_{inv\Delta}^\nu) \in \Omega^\mu$, there exists:

$$\nu = \sum_{i=1}^{N_{Ld}} x_i^\nu P_{Ld,i}$$

and because μ is the minimum $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i}$ for any $x \in \Omega^\mu$, therefore:

$$\mu \leq \nu \quad (4.17)$$

We now show $\nu \leq \mu$:

Examining $(x^\mu, P_{inv\Delta}^\mu) \in \Omega^\mu$ and comparing Ω^μ to the constraint set Ω^ν defined by (4.12)–(4.14):

Combining (4.8) with $P_{inv\Delta} \leq P_{inv\Delta,up}$ yields:

$$\sum_{i=1}^{N_{Ld}} (1 - x_i) P_{Ld,i} - P_{inv,base} - P_{inv\Delta,up} \leq P_G^*$$

which matches (4.12). Rearranging (4.8) yields:

$$\sum_{i=1}^{N_{Ld}} (1 - x_i) P_{Ld,i} - P_{inv,base} - P_{inv\Delta} = P_G^*$$

which satisfies (4.13). Finally, (4.10) satisfies (4.14) directly. All constraints of Ω^ν are satisfied, therefore:

$$(x_i^\mu, P_{inv\Delta}^\mu) \in \Omega^\nu$$

Then, from (4.11)–(4.15):

$$\begin{aligned} \nu &= \min_x \sum_{i=1}^{N_{Ld}} x_i P_{Ld,i} \\ \text{subject to } & (x^\nu, P_{inv\Delta}^\nu) \in \Omega^\nu \end{aligned}$$

Because $(x^\mu, P_{inv\Delta}^\mu) \in \Omega^\nu$, there exists:

$$\mu = \sum_{i=1}^{N_{Ld}} x_i^\mu P_{Ld,i}$$

and because ν is the minimum $\sum_{i=1}^{N_{Ld}} x_i P_{Ld,i}$ for any $x \in \Omega^\nu$:

$$\nu \leq \mu \tag{4.18}$$

Finally, by (4.17) and (4.18):

$$\mu = \nu$$

□

4.6 Optimal Fast Frequency Response from Multi-building Net-load Groups

4.6.1 Problem Formulation

Optimal coordination of net-load resources for FFR can occur for multiple buildings within a single NLGC or, for globally optimal results, via coordination of all NLGCs across the distribution system managed by a single DSMC. Here, we consider the more general problem formulation: a global, distribution feeder-wide dispatch of N_{nlg} NLGCs for FFR. The k^{th} NLGC coordinates $N_{u,k}$ NLUs and the j^{th} NLU of the k^{th} NLGC has $N_{Ld,k,j}$ deferrable loads, which are all active at t_0 . The i^{th} deferrable load and the inverter of the j^{th} NLU under the k^{th} NLGC have cost $c_{kji} > 0$ and $c_{inv,kj} > 0$, respectively. The i^{th} deferrable load consumes power $P_{Ld,kji}$, and the inverter produces power $P_{inv,k,j} = P_{inv,k,j,base} + P_{inv\Delta,k,j}$. The bounds on the change in the inverter power allowable are constrained by the headroom given by $P_{inv\Delta,up,k,j}$ as an upper bound and $P_{inv\Delta,dn,k,j}$ as a lower bound. The group of NLGCs, managed by a single DSMC, is given a commanded change in total active power, ΔP_{Gg} , that the aggregated response of the NLUs should collectively follow to provide the requested FFR. The objective of the coordinated NLGs is to track a total grid active power consumption $P_{Gg}^* = P_{Gg}(t_{dg}) + \Delta P_{Gg}$ (where $P_{Gg} = \sum_{k=1}^{N_{nlg}} \sum_{j=1}^{N_{u,k}} P_{G,k,j}$ and $P_{G,k,j}$ is the active power consumed at the grid connection of the j^{th} NLU in the k^{th} NLG) while minimizing the total cost. This objective can be formalized as:

$$\min_{x_{k,j}, P_{inv\Delta,k,j}} \sum_{k=1}^{N_{nlg}} \left(\sum_{j=1}^{N_{u,k}} \left(\sum_{i=1}^{N_{Ld,k,j}} x_{kji} c_{kji} P_{Ld,kji} + c_{inv,k,j} |P_{inv,k,j}(t) - P_{inv,k,j}(t_0)| \right) \right) \quad (4.19)$$

$$\text{subject to } \sum_{k=1}^{N_{nlg}} \left(\sum_{j=1}^{N_{u,k}} \left(\sum_{i=1}^{N_{Ld,k,j}} (1 - x_{kji}) P_{Ld,kji} + P_{L,k,j} - P_{inv,k,j,base} - P_{inv\Delta,k,j} \right) \right) = P_{Gg}^* \quad (4.20)$$

$$P_{inv\Delta,dn,k,j} \leq P_{inv\Delta,k,j} \leq P_{inv\Delta,up,k,j} \quad (4.21)$$

$$\forall j = 1 \dots N_{u,k}, k = 1 \dots N_{nlg}$$

$$P_{inv\Delta,k,j} = P_{inv,k,j}(t) - P_{inv,k,j}(t_0) \quad (4.22)$$

$$\forall j = 1 \dots N_{u,k}, k = 1 \dots N_{nlg}$$

$$x_{kji} \in \{0, 1\}$$

$$\forall i = 1 \dots N_{Ld,k,j}, j = 1 \dots N_{u,k}, \quad (4.23)$$

$$k = 1 \dots N_{nlg}$$

Here, $P_{inv\Delta,j}$ is the determined change in inverter power for the j^{th} NLU; $P_{L,j}$ is the aggregate nondeferrable load active power for the j^{th} NLU; and $P_{Ld,j,i}$ is the active power of the i^{th} deferrable load in the j^{th} NLU. Solving this optimization problem yields the solution $(x^\rho, P_{inv\Delta}^\rho)$.

4.6.2 Solution

Similar to the single-NLU dispatch problem, the optimal dispatch problem (4.19)–(4.23) can be solved directly using an MILP; however, even when using state-of-the-art optimization solvers, solutions are not always possible for large problems because of limits on the number of variables or max recursion depth that can be achieved. When solutions are possible, solution times are often too slow to support dispatch on FFR timescales, as shown in Section 4.6.3.2. Thus, we develop a hierarchical approach that builds on the simplified NLU dispatch approach of Section 4.5.2 and distributes computation across NLG controllers to calculate dispatch results on a timescale fast enough for FFR while maintaining near-optimal dispatch results. This control approach can be applied within a single NLG or for all NLGs across a distribution feeder. Methods for both scales

are summarized in Figure 4.2. These methods extend the underlying optimal dispatch approach developed and proven in Section 4.5.2 and evaluate all net-load combinations as binary variables and then use aggregate inverter flexibility (e.g., many distributed inverters with individual costs are the same as a single large inverter with piecewise-linear cost for dispatch purposes) to achieve the constraint of meeting the global power set point exactly while achieving the minimum cost. The methods developed trade off complete optimality with solution time and are cognizant of practical implementation limitations, such as (i) only being able to compute and consider a limited number (i.e., 2^{N_o}) of net-load combinations in the short dispatch solution time allowed and (ii) only communicating a smaller subset of net-load options, representing the "best", lowest-cost options across a range of possible operating points, to the DSMC for final dispatch decision to minimize network bandwidth requirements and encourage overall solution speed.

Here, A_{sel} is a matrix of length $|A_{sel}| = 2^{N_o}$ ($N_o = 20$ was used herein), where each row is the N_o -bit binary representation of that row index, $interp1d(y, f(x), x)$ represents a 1-D linear interpolation of function $f(x)$ at point y , $\dagger$ indicates a wild card for the two inputs to that block (i.e., the function with $\dagger$ should be applied to both inputs independently), $\cdot^*$ indicates a set point or optimal point, and ΔP^* is the desired change in global net-load power consumption for the FFR event. Optimal dispatch within a single NLGC does not involve the DSMC. At the conclusion of a global dispatch, which involves NLGCs and the DSMC, each NLGC with id, z , is sent its $N_{r,z}$ corresponding dispatch rows $\{[\Delta P_{L,m}, \Delta P_{inv,m}, \Delta P_{Tnl,m}, i_m, z]_{m=1}^{N_{r,z}}\}_z$. Each NLGC then compares these dispatch rows with the array of net-load options it stored at the event's initiation and executes the appropriate load curtailments and inverter set point changes.

4.6.3 Performance Verification

The hierarchical approach developed employs a heuristic method that can be executed significantly faster than optimization-based approaches, trading complete optimality for speed. Thus, it isn't tractable to analytically prove that the hierarchical approach will achieve the optimal dispatch result 100% of the time. Instead, we evaluate performance using Monte Carlo-style simulation and determine distributions of probability vs. per-

cent optimality (measured via solution cost) and probability vs. solution time. To do so, all simulation runs solve a multi-NLU dispatch problem twice—once using the hierarchical approach implemented in Python and again using an MILP solution deployed on a state-of-the-art, commercial (Gurobi) mixed integer program solver. The MILP solution provides the reference global optimal solution cost and the solution times of both approaches are compared. Both simulation types use the same scenario inputs and are executed on the same computer with a six-core 2.9 GHz Intel Core i9 processor. The hierarchical approach is designed to use only one processor core (to allow for greater portability to a variety of embedded platforms) and is implemented as such, whereas the MILP solution is allowed to use all six processor cores.

4.6.3.1 Single Net-load Group Case

Simulations were performed for optimal FFR dispatch of a single NLG with $N_u = 10$ NLUs; each had 4 loads and 1 inverter, for a total of 50 time-varying net-load resources to be dispatched. More than 2 million simulation runs with deferrable load capacities uniformly distributed in the interval $P_{Ld,j,i} \in [100, 6000] W$; inverters with uniformly distributed capacity $P_{inv,j}^{max} \in [2000, 5000] W$ and headroom $P_{inv\Delta,j,up} \in [0, 2000] W$; uniformly distributed NLU costs $c_{ji} \in [2, 10]$; and an assigned NLG participation factor (the percentage of the NLG’s total net grid power consumption that is dispatched) $\eta \in (0, 45]\%$ were run. Solution optimality results in Table 4.1 confirm that, on average, the hierarchical approach achieves an optimal dispatch decision with a solution cost very close (within 0.00034%) to the optimal, that solution cost is optimal 99.761% of the time, that the hierarchical approach has a mean solution time of 52.2 ms, and that there is a 99.992% probability of achieving a solution within 120 ms.

4.6.3.2 Full Feeder Case

The optimal FFR dispatch problem for a full-scale electric distribution system feeder with a varying number of participating NLGs was simulated in 9,920 simulation runs to evaluate the performance of the proposed hierarchical approach versus the MILP solution and determine how their performances scaled as the number of participating net-load resources and the total feeder participation factor varied. In the simulation, each NLG included $N_u = 10$ NLUs, each with 4 loads and 1 inverter participating,

Table 4.1: Single NLG case: Summary simulation results

Number of simulation runs	2,007,890
Mean % difference in solution cost, HA vs. MILP	0.00034%
Probability of HA solution cost within 1% of optimal	99.997%
Probability of HA solution cost within 0.1% of optimal	99.912%
Probability of HA solution cost within 0.01% of optimal	99.842%
Mean solution time, HA	52.2 ms
Mean solution time, MILP	107.1 ms
Probability of HA solution time < 120 ms	99.992%
Probability of HA solution time < 100 ms	99.962%
Probability of HA solution time < 80 ms	98.935%

connected to a single distribution transformer and coordinated by an NLGC. The simulations included 155 runs at each of 64 possible combinations of the number of NLGs participating, $N_{NLG} = [100, 200, 300, 400, 500, 600, 700, 800]$, and the total feeder dispatch participation factor, $\eta = [10, 15, 20, 25, 30, 35, 40, 45]\%$. Each simulation run utilized deferrable load capacities uniformly distributed in the interval $P_{Ld,k,j,i} \in [100, 6000] W$; inverters with uniformly distributed capacity $P_{inv,k,j}^{max} \in [2000, 5000] W$ and headroom $P_{inv\Delta,k,j,up} \in [0, 2000] W$; and uniformly distributed NLU costs $c_{kji} \in [2, 10]$.

Results shown in Figure 4.3 demonstrate that the hierarchical approach for optimal FFR dispatch can achieve the same optimal dispatch result with very similar ($<0.5\%$ difference for $\eta = [15, 20, 25, 30, 35, 40, 45]\%$ and $<1\%$ for all cases) dispatch cost as the optimal MILP solution for each run. Critically, the hierarchical approach can achieve these very nearly optimal dispatch results at a much faster timescale: Figure 4.4 shows that the hierarchical approach can be consistently solved within ~ 150 ms with low variance regardless of the problem size, compared to the MILP solution's up to 4-s solution time for the largest problem size (800 transformers or 40,000 total net-load resources). Clearly, the MILP solution's 550–4000+ ms solution time is not fast enough to support FFR dispatch within 500 ms, whereas the designed hierarchical approach provides consistently fast performance to support such FFR dispatch.

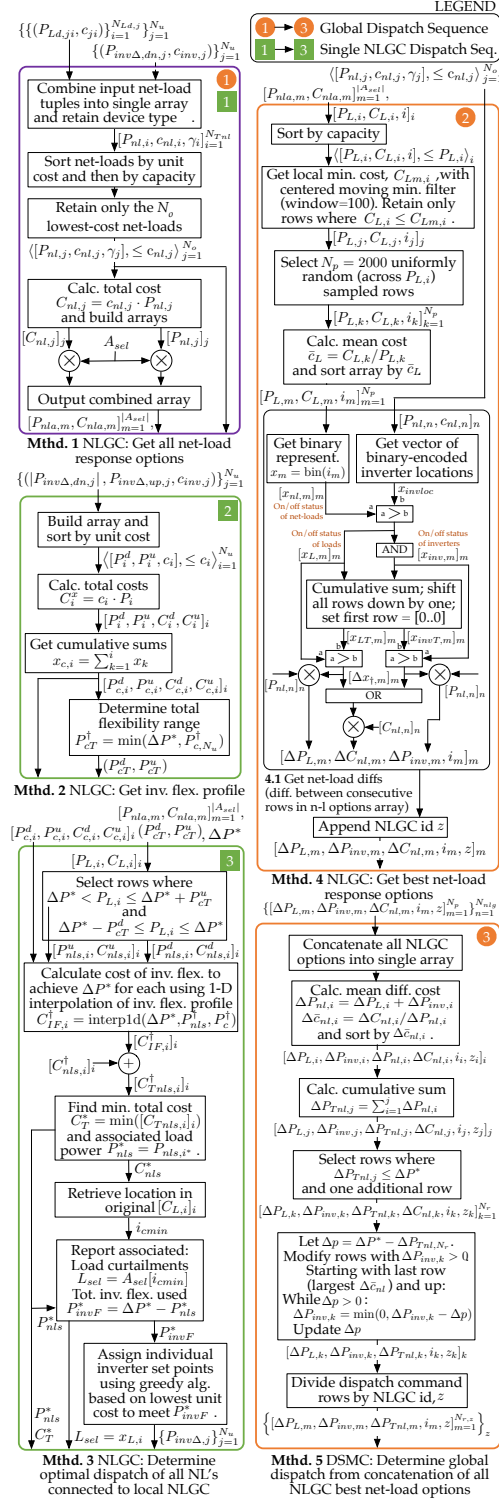


Figure 4.2: Hierarchical optimal net-load dispatch approach

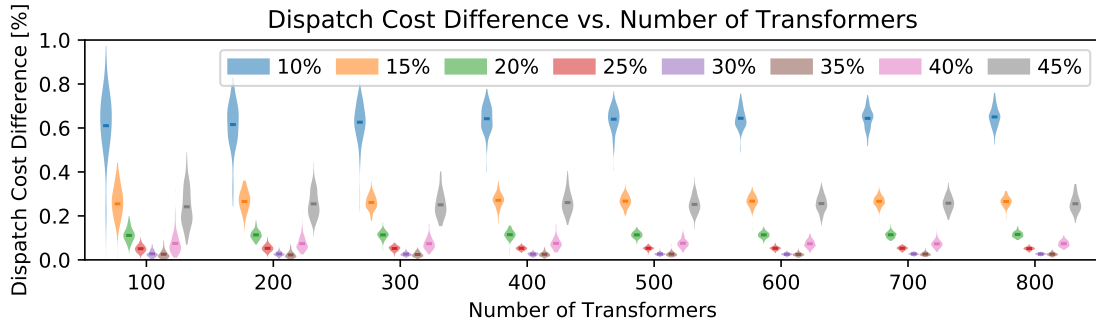


Figure 4.3: Full distribution feeder case: Percent difference in dispatch cost achieved by the hierarchical approach versus the MILP optimal solution for varying participation factor (10%–45%).

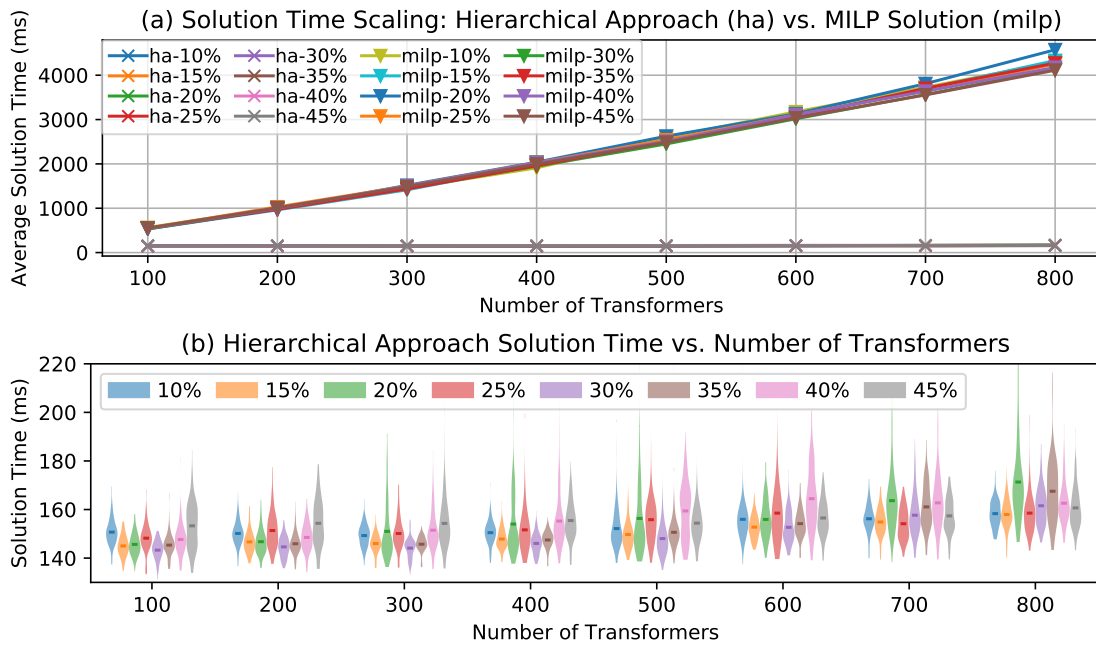


Figure 4.4: Full distribution feeder case: Solution time vs. number of transformers (problem size). (a) Comparison of hierarchical approach (“ha”) and MILP solution (“milp”) average solution time based on 155 runs for each case. (b) Violin plot showing distribution of hierarchical approach’s solution times for varying participation factor (10%–45%) for 155 runs each.

4.7 Optimal FFR Dispatch PHIL

A novel PHIL experiment platform (see Figure 4.5) incorporating more than 100 powered DER and loads, a real-time simulated electric distribution system model, more than 10,000 cosimulated net-load resources, and a hierarchical control solution deployed on embedded hardware platforms and real communications infrastructure is developed. This allows for the optimal FFR dispatch approach developed in Sections 4.5 and 4.6 to be scaled and tested in a realistic electric distribution system simulation. This platform is unique versus state-of-the-art PHIL experiments because it includes a large number of distributed at-power devices under test (more than 100 individually controllable, powered devices) connected at nine different interconnection points in the simulated distribution system model; we are unaware of any other PHIL demonstration to date including this many powered devices and individual PHIL interconnection points. Further detail on the components of the platform are presented in the following subsections.

4.7.1 Power Hardware Under Test

4.7.1.1 Laboratory Homes (HG1–HG2)

Each lab home contains four residential appliance loads that are controlled, additional appliance loads that are uncontrolled, and one custom (“Type 4 inverter”) 2-kVA inverter system (see Figure 4.6). The four controlled appliance loads are connected through load sensing and control interfaces (LSCIs), which provide high-speed measurement feedback and on/off load controllability to the endDevice controller, which is a controller with a common communications interface and generalized structure with specific logic for the various devices controlled (i.e., load, multiload, PV inverter, battery inverter, and PV+batt. inverter). In HG1, a refrigerator (Whirlpool WRL767SIAM00), clothes washer (Whirlpool WFL98HEBU0), window air conditioner (Frigidaire FFRE0633S1), and oven (Bosch HBLP651UC) are controlled; and in HG2, the controlled loads are a refrigerator (GE Profile PSQS6YGY), plug-in heater (Bovado 0CZ449), set of light bulbs (bank of 5 incandescent, 10 compact fluorescent, and 10 LED light bulbs), and oven (Maytag MER8674). A custom 2-kVA inverter system is developed based on a simple H-bridge topology to achieve faster response (within 1–2 AC cycles) to AC power commands than was available via commercially-available residential invert-

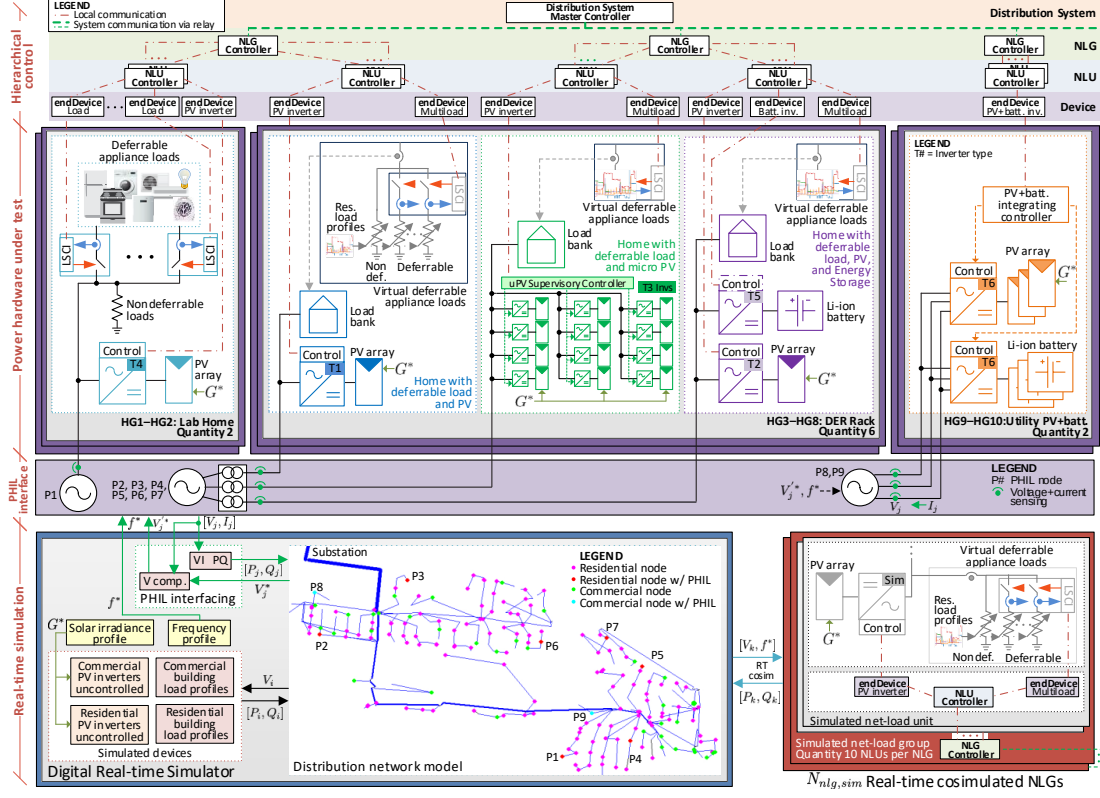


Figure 4.5: PHIL experiment platform, including control, power hardware under test, PHIL interface, and real-time simulation components (see labels at left)

ers. The inverter DC inputs are powered by commercial photovoltaic (PV) simulators. More detail on our LSCI and custom inverter implementations can be found in [49]. HG1 and HG2 are connected to the same 240-V point of common coupling (PCC) in the simulated distribution network.

4.7.1.2 DER Racks (HG3–HG8)

Each DER rack (e.g., see Figure 4.7) contains hardware representing three different residential buildings: (i) a home with deferrable load and a PV string inverter (Type 1); (ii) a home with deferrable load and 12 PV microinverters (collectively referred to as a Type 3 inverter); and (iii) a home with deferrable load, a PV string inverter (Type 2), and a battery energy storage system with inverter (Type 5). All inverters are commercial off-the-shelf (COTS) residential-scale inverters with AC output ratings of 240 V, 3.8–5



Figure 4.6: Photos of the two lab homes (HG1 and HG2) with residential appliance loads controlled denoted. One Type 4 (custom) inverter (pictured at right inset) is used with each HG.

kVA and are managed via Modbus TCP from their local endDevice controllers. Battery energy storage systems are also COTS with ratings of 3.3 kW/6.4 kWh (HG3) and 5 kW/9.8 kWh (HG4–HG8). The load in each home is represented by a set of virtual loads (deferrable and nondeferrable) driven by real-world residential load profiles (see Section 4.7.2.2 for details). In HG3 and HG4, the output of each home’s total virtual load is given as the command to one of six resistive load banks, resulting in physical power consumption for each home’s load. In HG5–8, the total virtual load is given as a command to a real-time simulated load connected at the same PCC of the simulated distribution network as the associated physical DER rack. Each DER rack is connected at an independent 240-V PCC (interface points P2–P7 shown in PHIL interface and real-time simulation panels of Figure 4.5) in the simulated distribution network model. PV inverter DC inputs are powered by commercial PV simulators.

4.7.1.3 Utility-Scale PV and Battery Inverters (HG9–HG10)

HG9 and HG10 each include one utility-scale (100-kW, three-phase) battery inverter and one PV inverter (250-kW, three-phase) that are coordinated to track a commanded AC power output while simultaneously maximizing the use of available input solar energy. This coordinated action is accomplished via an integrating controller (see Figure 4.8). Each battery inverter is powered by a 100-kW/33-kWh Lithium-ion battery rack, and the PV inverters are powered by commercial PV simulators. Each HG is interfaced at

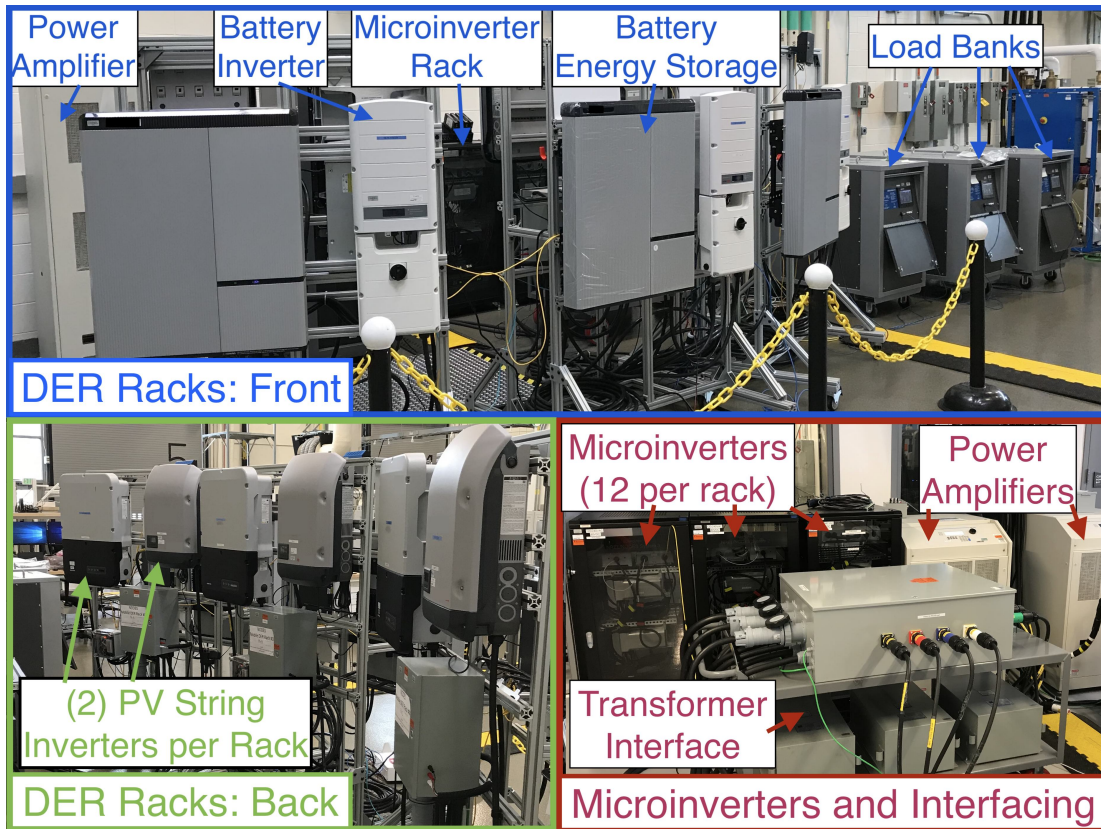


Figure 4.7: Photos of equipment comprising three of the six DER racks (HG3–HG8).

an independent, three-phase PCC (P8,P9 in Figure 4.5) in the simulated distribution network model.

4.7.2 Real-Time Simulation

As shown in Figure 4.5, the real-time simulation includes an electric distribution network model with additional building and solar device inputs implemented on a digital real-time simulator (DRTS) as well as additional NLG models that are cosimulated in real time on a separate computational platform. For this experiment, a 16-core Opal-RT OP5700 is used as the DRTS and up to (20) 36-core servers are used to cosimulate the NLGs—including their constituent load profiles, devices, and controllers—in real time. UDP/IP communications are used to interface the two platforms for real-time cosimulation.

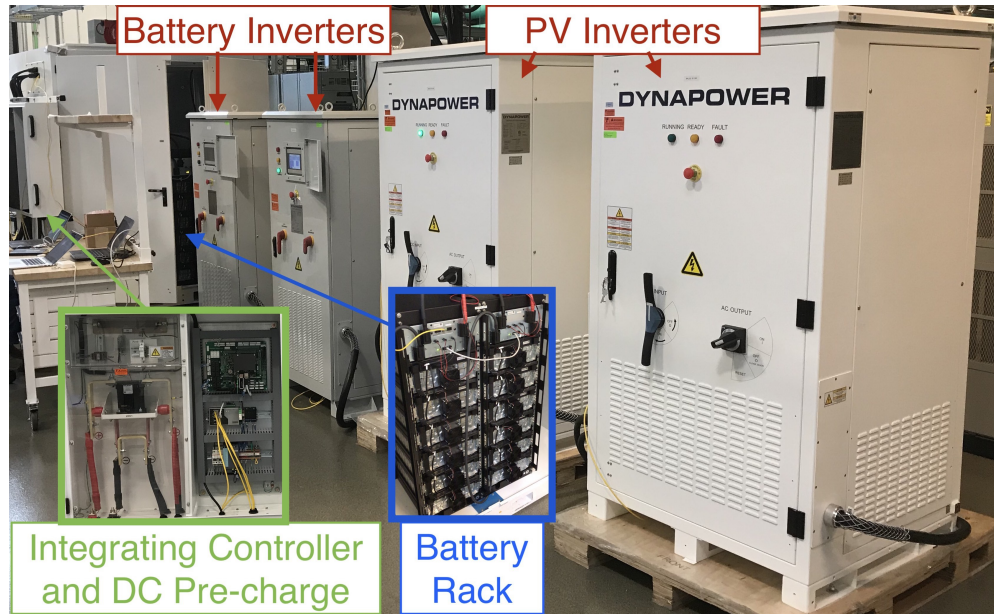


Figure 4.8: Photo of equipment comprising two utility-scale PV+battery inverter systems (HG9–HG10).

4.7.2.1 Electric Distribution Network Model

An electric distribution network model based on a real-world system with more than 2,000 customers in Colorado, USA is used as the basis for this study. The original distribution network model is augmented with additional single-phase residential and three-phase commercial-scale PV generation to achieve a high-penetration solar case study, as detailed in Table 4.2. Distributed PV systems are added at each residential transformer secondary with individual capacities allocated to follow the statistical distribution of PV system capacities currently in the field while respecting existing transformer capacities. The network model was converted from (and validated against) the native format provided by the electric utility to Opal RT’s ePHASORSIM [87] format to allow for real-time phasor domain power system simulation at a time step of 10 ms. The seven residential (P1–P7 in Figure 4.5) and two commercial (P8,P9 in Figure 4.5) PHIL PCCs are chosen as shown in Figure 4.5 to include nodes at a variety of locations. Details on the specific device capacities and mappings are given in Table 4.5.

The network model’s feeder head is connected to a stiff voltage source with source

Table 4.2: Distribution Network Model Characteristics

Location	Colorado, USA
# Customers	Approximately 2,000
% Residential / % commercial load	52% / 48%
# Nodes	1,122
Peak load	7 MVA
Annual energy consumption	25,346 MWh
PV inverter capacity	8.06 MVA
Residential	4.96 MVA
Commercial/utility	3.1 MVA
Annual PV energy delivered	12,714 MWh
PV penetration (by annual energy)	50.2%
Battery inverter capacity	1.01 MW
Residential	0.02 MW
Commercial	0.99 MW

impedance matching that of the substation transformer. Further bulk power system modeling—including the closed-loop impact of feeder power flow on bulk power system frequency—is beyond the scope of this work. The focus of this work is to demonstrate that tens of thousands of distributed net-load resources can be optimally coordinated to provide an aggregate feeder head active power response of desired magnitude and response time commensurate for FFR. The network model provides a realistic power network configuration for interconnection of net-load resources and is used to identify any potential distribution-level impacts of providing FFR (e.g., voltage or power flow beyond desired operating limits). System frequency is determined using an open-loop frequency profile. A frequency profile with an underfrequency step from 60 to 59.9 Hz is used to more easily quantify FFR dispatched device response times. The framework and methodology are also tested using frequency profiles based on real-world measured data, which further corroborate their effectiveness, though these results are not reported here. All NLU controllers, controlled net-load resources, and power hardware under test (HUT) are frequency dependent in this work, whereas the line, load, and transformer models within the distribution network model are frequency independent (a common assumption that doesn’t adversely affect the results shown here).

4.7.2.2 Simulated Devices and NLUs

The simulated PV inverter, battery inverter, and load devices in the real-time simulation (both in the DRTS and cosimulated models) use constant power injection models fed by solar irradiance and load profile data. All load and uncontrolled inverter device models are frequency independent in this work. Commercial and nondeferrable residential load profiles are determined using load allocation techniques based on transformer capacities and time-series feeder head power data provided by the electric utility. To model realistic deferrable residential loads, 1-Hz power readings from real homes in the Reference Energy Disaggregation Data set (REDD) [88] are used. This data set provides load data at the load circuit level that is time-correlated, driven by realistic user interaction, and at a high enough resolution (1 Hz) to represent fast load transitions as well as longer timescale characteristic profiles. From the REDD, we sample sets of four appliance (e.g., electric heat, dishwasher, water heater, and fridge in one home) loads in five different homes during the same 1-hour period (18:00–19:00) from each day in spring to obtain more than 1,000 sets of (4) time-correlated plug loads. The specific load profiles for the four deferrable loads within a given NLU are selected at random from the 1,000 sets of sampled plug load data, and then a uniformly distributed time offset $[-30 \text{ min}, 30 \text{ min}]$ is applied to each profile to maintain appropriate load diversity. Solar irradiance data from the same geographic location as the real-world distribution network and at the same time of day and season as the load data are used.

4.7.3 PHIL Interface

The PHIL interface connects the power HUT in HG1–HG10 (see Figure 4.5) in a closed loop to the real-time simulated distribution network at $N_j = 9$ different PCCs (P1–P9). For P1, two phases of a three-phase, 45-kVA power amplifier (Ametek MX45) are used to obtain a split-phase 240-V output. For P2–P7, two three-phase power amplifiers are used (one MX45 and one 90-kVA Ametek RS90); each phase of these three-phase power amplifiers is controlled independently and interconnected through a transformer to derive an independent, split-phase 240-V output. Finally, P8–P9 are three-phase PCCs driven by separate 270-kVA three-phase power amplifiers (Ametek RS270) with independent phase control. All five power amplifiers have a -3dB bandwidth of approx-

imately 1,950 Hz. Thirteen sets of voltage (LEM DVL series, 15-kHz bandwidth) and current (LEM LF series, 100-kHz bandwidth) transducers at each PCC phase output (i.e., after power amplifier and, if applicable, the transformer interface) are used to measure instantaneous voltage and current for feedback.

In the PHIL interconnection forward path, root mean square (RMS) voltage, V_j^* , measured at each PCC in the simulated network model and the modeled system frequency, f^* , are used as references to the power amplifiers. A simple proportional-integral control-based voltage compensator drives the measured voltage, V_j , at each physical PCC's output terminals to follow V_j^* , ensuring that the HUT experiences the same voltage as in the network model at that point. The final voltage reference, $V_j'^*$, and frequency reference, f^* , are communicated digitally to each power amplifier. In the PHIL interconnection return path, instantaneous voltage and current measurements at each physical PCC's output terminals are connected to the DRTS's analog-to-digital converter, scaled, and then used to calculate the corresponding active and reactive power for injection into the distribution network model. The PHIL interfacing model is executed at a time step of $200 \mu s$ with communication to the power amplifiers and distribution network model every 10 ms (the time step of the phasor-domain network simulation).

4.7.4 Hierarchical Control Platform

As shown in Figure 4.5, four levels of control are employed in the hierarchical control platform. The software for this control platform is developed in Python and designed to be deployable in a variety of embedded platforms, from low-cost single-board computers (SBCs) to high-performance servers. At the device layer, an endDevice controller interfaces one or more net-load resources within an NLU to the NLU controller (NLUC). The endDevice controller relays device RMS power measurements and operating limits to the NLUC and translates NLUC dispatch signals into device commands. In the experiment configuration of Figure 4.5, the four deferrable loads within HG1 have individual endDevice controllers implemented on Raspberry Pi 3B+ SBCs [89] that also function as the LSCI controller and provide measurements to the NLUC every 50 ms. In HG2, a multiload endDevice controller on a single Raspberry Pi 3B+ manages the four deferrable loads through a circuit-level LSCI. For HG3–HG10, the 50 endDevice

controllers, as well as the associated 25 NLUCs and 3 NLG controllers (NLGCs), are deployed on a single 12-core Linux server. The inverter endDevice controllers interface with their physical inverters via Modbus TCP.

At the second layer of control, NLUCs gather measurements from each connected endDevice and maintain a table of assigned device IDs, measurements, capacities, and user-assigned costs that is updated each time a new measurement is received. The NLU periodically (every 100 ms) publishes this table in a standard format to the NLGC, and also translates NLGC dispatch commands and routes them to the appropriate endDevice. Each NLUC and NLGC hosts a local publish/subscribe communication bus (implemented using Zero MQ [90] over TCP/IP) to communicate with their endDevices and child NLUs, respectively. This communication architecture—in which controllers can communicate only one layer up or down—preserves privacy (e.g., NLGs cannot directly address endDevices and do not have knowledge of specific devices) and enables specific communications protocols or security measures to be implemented at each layer as needed. Building owners interact with the NLUC by setting unit costs and enabling/disabling dispatch for each NLU device.

At the third layer of control, NLGCs gather device status tables from each connected NLU and participate in optimal dispatch calculation. If operating in single NLG dispatch mode, NLGCs calculate an optimal dispatch for only the devices within their NLG using Mthd. 1–3 described in Section 4.5. If operating in global dispatch mode, NLGs calculate the best net-load response options for their NLG using Mthd. 1 and Mthd. 4 (described in Section 4.5), and this information is sent to the Distribution System Master Controller (DSMC) to support calculation of a globally-optimal dispatch solution. After optimal dispatch decisions have been made, the NLGC sends these commands to each NLUC for execution within each NLU. The DSMC and all NLGCs communicate through a communications relay that hosts a publish/subscribe communications bus. This architecture removes the need for the DSMC or NLGCs to keep track of device addresses. Instead, each controller simply addresses the relay, which routes all communications between participating controllers.

At the fourth layer of control, the DSMC receives measurements of NLG net-load resources periodically (every 200 ms), and, on command, receives each NLGC’s best net-load response options. The DSMC implements Mthd. 5 described in Section 4.5 to

rapidly calculate the globally-optimal dispatch decision and then distributes dispatch commands to each NLGC. More specifics on the FFR optimal dispatch algorithm and its performance are given in Section 4.5. The controllers for all four layers of the cosimulated NLGs are deployed on up to (20) 36-core linux servers.

4.8 Single Net-load Group Experiment Results

4.8.1 Scenario

The aim of this scenario is to experimentally validate the response time of the FFR optimal dispatch solution when operating in single NLG dispatch mode and to highlight the performance and benefit of the NLU inverter optimal power point approach developed in Section 4.4. A single NLG with nine NLUs (NLU1–9 as described in Table 4.5), representing nine residential buildings with a total of 45 controlled net-load resources, is used as the basis for this scenario. Scenario data inputs include real-world appliance load profiles (as described in Section 4.7.2.2) as well as a variable solar irradiance profile (see Figure 4.9). The physical appliance loads in NLU1–NLU2 are perturbed with user action throughout each trial. Multiple trials of the same scenario, but with different frequency event times and NLG participation factors, are conducted to demonstrate the performance under a variety of starting conditions. For all trials, the FFR dispatch event is initiated using an underfrequency event profile with a 60- to 59.9-Hz step. The NLU1 inverter is configured to monitor system frequency and to send a trigger to the NLGC to start its dispatch sequence whenever frequency declines to less than 59.92 Hz. NLU6 is configured to operate its inverter in optimal MPP tracking mode. That is, the inverter is operated with sufficient headroom to ensure the NLU is capable of providing an FFR event dispatch of any percentage of its controlled net-load capacity at all times, while minimizing the amount of headroom required (see Section 4.4 for further details on the algorithm). To do so, NLGC #6 implements Algorithm 1 from Section 4.5.2, uses the latest load measurements available, and sends power commands to the inverter every 1 s. The inverters of NLU1–5 and NLU7–8 are operated with a fixed headroom of 20%.

4.8.2 Results

4.8.2.1 Summary

Summary results are shown in Table 4.3. It can be seen that the hierarchical approach was able to achieve the same global optimal dispatch solution as the mixed-integer linear program (MILP) solution in 23–56% less time. Global optimum dispatch results for comparison were obtained by running the MILP solution for the same set of net-load measurements and costs as were recorded by the hierarchical approach during its dispatch. Complete closed-loop responses, from system frequency change to all loads and inverters actuated, were achieved within 247 ms for all trials, well within the targeted 500 ms for FFR.

Table 4.3: Single NLG Experiment Results

Trial #		1	2	3	4
η		0.4	0.37	0.45	0.4
ΔP^* (W)	Dispatched	11,298	5,724	13,486	9,553
	Achieved (measured)	11,257	5,762	13,539	9,510
	% diff.	-0.36%	0.67%	0.39%	-0.45%
Cost	Hierarchical Approach (HA)	24,108	10,710	24,455	16,517
	MILP solution	24,108	10,710	24,455	16,517
	% diff.	0%	0%	0%	0%
Solution Time (ms)	HA	88	90	79	85
	MILP	202	117	146	116
	% diff.	-56.4%	-23.2%	-45.8%	-26.9%
HA Solution Time Breakdown (ms)	Detect Δf	55	57	64	63
	Dispatch alg.	88	90	79	85
	Misc. ¹	6	19	13	37
	Initial response	149	166	156	185
	Complete response	225	208	231	247

4.8.2.2 Single NLU Optimal MPP Tracking

Here, detailed results are presented from Trial #2 and Trial #4 to highlight NLU optimal MPP tracking functionality. During both trials, two NLUs—NLU6 using optimal

¹Time for NLGC to send dispatch commands, communications delays, local controller processing time, and local device actuation.

MPP tracking mode and NLU9 using regular MPP tracking mode—with identical equipment ratings, load profiles, and irradiance profile were run simultaneously. Figure 4.9 shows simulation (expected operation, as predicted by a simulation using the same solar irradiance and load profile inputs) and experimental results for the two NLUs during Trial #2. It was verified that the NLU inverter operating in optimal MPP tracking mode operated off of the maximum power profile in proportion to the real-time calculated headroom, α (i.e., the “optMPP-meas” trace in Figure 4.9 follows the simulated “optMPP-Sim” trace, which is the difference between “reg-Meas” and α). Measured operation matches the expected operation except at low power values, where the measured inverter maintains a minimum power of 300 W to avoid going into standby mode. Trial #2 did involve an FFR dispatch event, though neither NLU was dispatched.

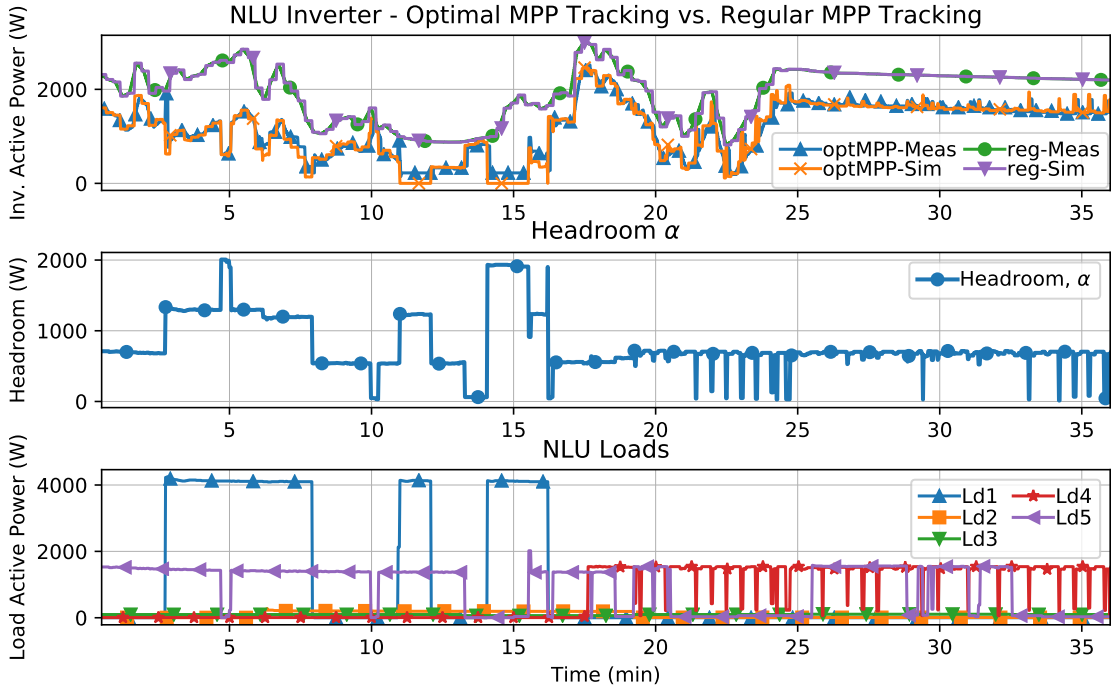


Figure 4.9: Trial #2 optMPP results: (top) Power output for inverter in optimal MPP tracking mode (“optMPP”) versus regular mpp tracking mode (“reg”). Measured experimental data (“Meas”) and predicted simulation results (“sim”) are compared. (middle) The real-time calculated α . (bottom) Load profiles for the five NLU loads.

In Trial #4, NLU6 is dispatched near $t = 6.1$ min by -1313 W, or 25.9% of its pre-

event net-load consumption, as shown in Table 4.4. The NLU6 dispatch set point was achieved entirely by the inverter. Had the NLU6 inverter been operating in regular MPP tracking mode, the NLU would be limited to responding with a few discrete options. E.g., for $\eta = 25.9\%$, the two nearest discrete options would have been either 2.7% or 37.8% of the NLU's net-load power. This demonstrates that the optimal MPP tracking strategy was effective for ensuring the NLU could be dispatched at any percentage of its net-load consumption. Detailed results for NLU6 and NLU9 operation during Trial #4 are shown in Figure 4.10. Note that after the inverter is dispatched near $t = 6.1$ min, its output remains fixed at the minimum of the dispatched set point or the maximum available solar power.

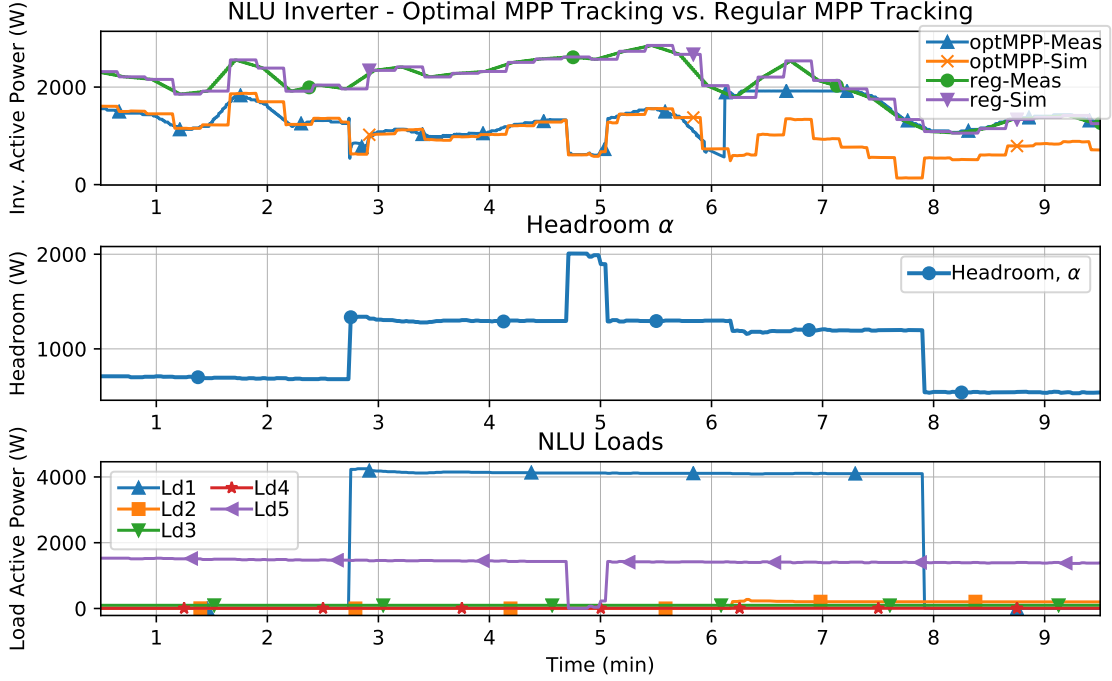


Figure 4.10: Trial #4 optMPP results

4.8.2.3 NLG Optimal FFR Dispatch

Measured RMS power results showing the operation of all NLUs and their constituent inverters and loads throughout Trial #4 are shown in Figure 4.11, and summary results are given in Table 4.3. A dispatch event with participation factor $\eta = 0.4$ is completed,

Table 4.4: NLU6 Dispatch Scenarios

		Load Capacity (W) Under Mode				
		Opt. MPP		Reg. MPP		
	Cost	Before	After	Before	After #1	After #2
Inverter	4	-566	-1879	-1879	-1879	-1879
Load1	3	4110	4110	4110	4110	4110
Load2	4	0	0	0	0	0
Load3	5	102	102	102	0	102
Load4	6	0	0	0	0	0
Load5	7	1420	1420	1420	1420	0
Net		5066	3753	3753	3651	2333
α^*		1294				
α		1313		0		
ΔP			-1313		-102	-1420
η			0.259		0.027	0.378
Total cost			5252		510	9940

and assets in NLU1 (three loads and inverter), NLU2 (three loads and inverter), and NLU6 (inverter only) are dispatched to achieve a total NLG dispatch of $\Delta P_{meas}^* = 9510$ W. Instantaneous current and voltage waveforms of NLU1 and NLU2 assets close to the instant of the dispatch event are shown in Figure 4.12. This plot confirms that initial response is seen approximately 185 ms after a step change in the system frequency. These results experimentally validate that an NLG of nine NLUs can be optimally dispatched to provide FFR within 247 ms and while matching the global optimum dispatch result.

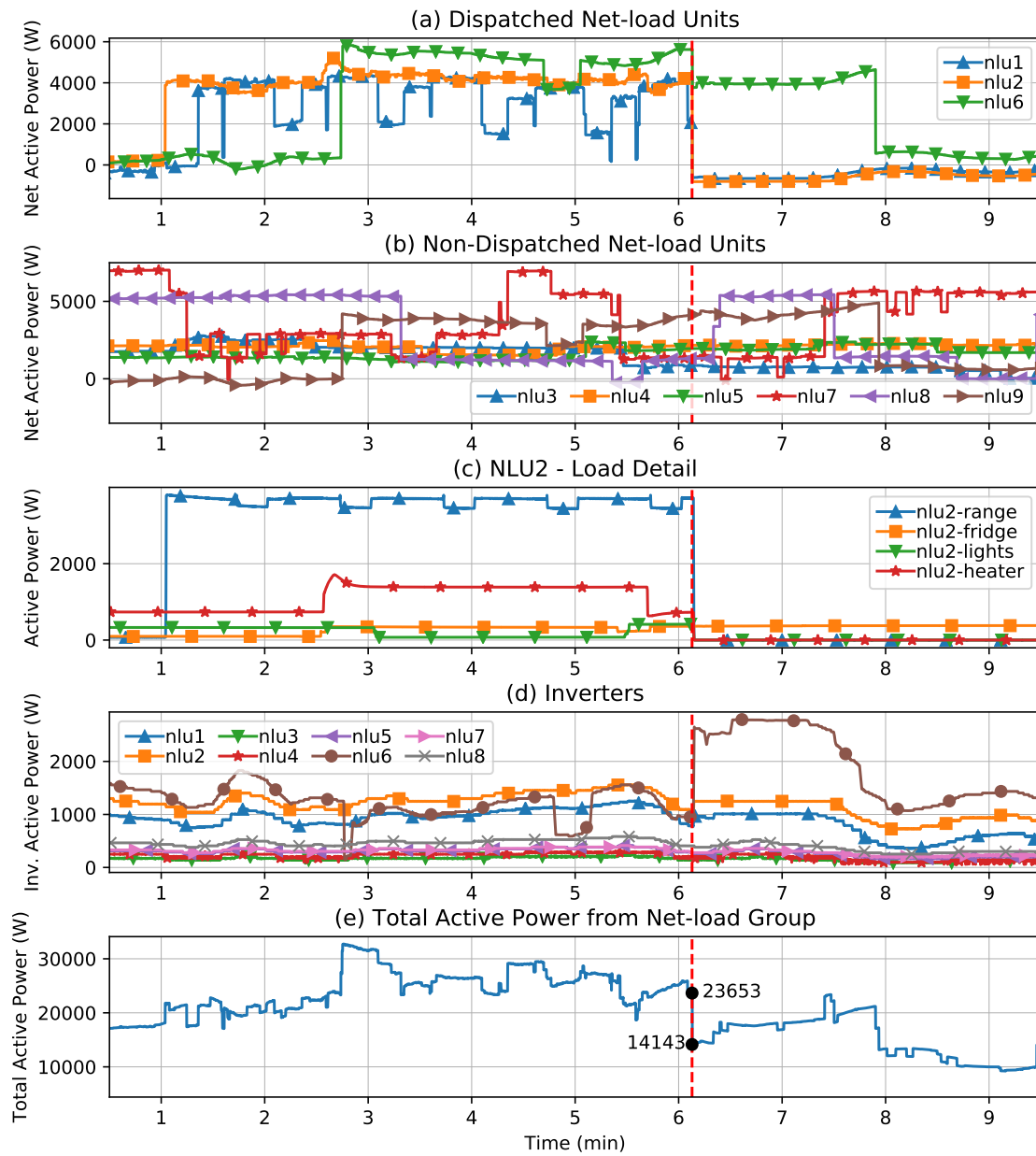


Figure 4.11: Trial #4 NLG dispatch results: RMS active power for the (a) three dispatched NLUs, (b) six nondispatched NLUs, (c) constituent loads of NLU2, (d) NLU inverters, and (e) the total NLG (all NLUs)



Figure 4.12: Trial #4 NLG dispatch results: Instantaneous waveforms (top) NLU1 inverter current (“IL1”), NLU2 inverter current (“M1SMA”), grid voltage (“VG”) and approximate grid frequency (“FREQ”). “FTRIG” goes from low to high when the local inverter controller detects the frequency event. “PTRIG” goes from high to low when the local inverter controller receives its dispatch set point from the NLGC. (bottom) Four NLU1 constituent load currents (“FRIDGE”, “WASHER”, “RANGE”, and “AIRCON”), the total load current of NLU1 (“HM3”), and the total load current of NLU2 (“HM2”).

4.9 Full Distribution Feeder Experiment Results

4.9.1 Scenario

This scenario experimentally validates the response time of the FFR optimal dispatch solution when operating in global dispatch mode with increasing numbers of NLGs controlled. The full distribution network configuration described in Section 4.7.2.1 and detailed in Table 4.5 is used for this scenario. As before, scenario data inputs include real-world appliance load profiles as well as a variable solar irradiance profile, and the physical appliance loads in NLU1–NLU2 are perturbed with user action throughout each trial. Multiple trials of the same scenario, but with different frequency event times, numbers of NLGs being controlled, and total feeder response magnitude are conducted. For all trials, the FFR dispatch event is initiated using an underfrequency event profile with a 60- to 59.9-Hz step. The NLU1 inverter is configured to monitor system frequency and to send a trigger to the DSMC to start its dispatch sequence whenever the frequency declines to less than 59.92 Hz. All NLU inverters are operated with a fixed headroom of 20% for this scenario.

4.9.2 Results

4.9.2.1 Summary

Summary results for the full-scale experiments are shown in Table 4.6. The hierarchical approach was able to achieve nearly the same (within 0.2%) global optimal dispatch solution as the MILP solution and does so in 41–83% less time. Total response time is crucial as the problem size (i.e., N_{nlg}) grows and should remain less than 500 ms to show viability for providing FFR. The results in Table 4.6 demonstrate that the hierarchical approach developed here has a nearly constant dispatch solution time, whereas the MILP approach’s solution time grows significantly with problem size. In particular, the MILP approach solution time for Trial #3 and Trial #4 is too long to be able to support FFR requirements, whereas the hierarchical approach is suitably fast. These

²Appliance (App) , virtual loads with load bank output (VLB), and virtual loads with simulated output (V) are used.

³Groups of 12 microinverters are individually controlled, but they have a combined capacity as listed.

⁴ P_R refers to an inverter capacity chosen at random to follow the statistical distribution of PV system capacities already in the field.

Table 4.5: Experiment Configuration Summary

PHIL				Inverters			Load Type ²	Total Controllable Devices					
				Type	Capacity (kW)			Physical				Virtual	
NLG	NLU	Node	HG		Phys.	Sim.		PV	Batt.	Load	LB	Inv.	Load
1	1	1	1	T4	2	2	App	1	-	4	-	-	-
1	2	1	2	T1	2	2	App	1	-	4	-	-	-
1	3-8	2-7	3-8	T1	3	3	VLB	6	-	-	6	-	24
1	9	-	-	V	-	3	V	-	-	-	-	1	4
1	10	3	4	T2	3.6	3.6	V	1	-	-	-	-	4
2	11-14	4-7	5-8	T2	3.6	3.6	V	4	-	-	-	-	16
2	15-20	2-7	3-8	T5	3	3	V	-	6	-	-	-	24
3	21-26	2-7	3-8	T3	3.8	3.8	V	72 ³	-	-	-	-	24
3	27-30	-	-	V	-	P_R^4	V	-	-	-	-	4	16
4	31	8	9	T6	90	900	-	1	1	-	-	-	-
4	32	9	10	T6	90	90	-	1	1	-	-	-	-
5-26	10 ea.	-	-	V	-	P_R	V	-	-	-	-	220	880
27-114	10 ea.	-	-	V	-	P_R	V	-	-	-	-	880	3,520
115-224	10 ea.	-	-	V	-	P_R	V	-	-	-	-	1,100	4,400
Total Device Counts								87	8	8	6	2,205	8,912

dispatch solution time results confirm the conclusions of the simulation results comparing the two approaches in Section 4.6.3. The hierarchical approach’s dispatch solution times are nearly independent of problem size; however, note that the “Total gather all NL response options”⁵ and “Misc.”⁶ times do increase some with problem size. This is a consequence of the DSMC needing to receive, process, and send out results from/to more NLGCs. The implementation of the DSMC used here completes the tasks of receiving data and sending out dispatch commands in a single thread (in series); however, the implementation can be upgraded to parallelize these tasks, which will allow the implementation to continue scaling as expected to upwards of $N_{nlg} = 800$. The important result here is that the key parts of the dispatch solution—“Dispatch solution time” and “Get local NL response options (max)” —remain nearly constant despite greatly increasing problem size, demonstrating that the underlying algorithmic implementation is sound. For the largest problem tested, a case with $N_{nlg} = 224$ that represents more than 10,000 net-load resources to be dispatched, the hierarchical approach is able to compute the entire dispatch and achieve output response from hardware within 500 ms.

Table 4.6: Full Feeder Experiment Results

Trial #		1	2	3	4
η		0.05	0.13	0.20	0.25
N_{nlg}		26	26	114	224
ΔP^* (W)	Dispatched - Total	172,359	482,646	854,693	1,034,529
	Dispatched - Com.	-	383,130	402,270	391,231
	Achieved (measured)	172,090	478,850	853,922	1,030,202
	% diff.	-0.26%	-0.79%	-0.09%	-0.42%
Cost	Hierarchical Approach (HA)	597,417	1,053,503	1,742,412	2,858,176
	MILP solution	597,218	1,051,374	1,740,908	2,857,474
	% diff.	0.03%	0.20%	0.09%	0.02%
Dispatch Solution Time (ms)	HA	187	197	195	205
	MILP	319	355	543	1,206
	% diff.	-41%	-45%	-64%	-83%
HA Solution Time Break- down (ms)	Detect Δf	39	41	24	21
	Get local NL options (avg)	132	113	125	118
	Get local NL options (max)	149	147	151	163
	Total gather all NL response options ⁵	205	214	252	283
	Central dispatch alg.	38	50	44	42
	Misc. ⁶	52	46	61	79
	Total response - fastest hardware	334	351	381	425
	Total response - fastest COTS hardware	382	404	429	463

4.9.2.2 Distribution Feeder-Scale Optimal FFR Dispatch

Detailed results for the FFR dispatch of Trial #4 are presented here. Figure 4.13 shows that the majority of inverters in the experiment respond within 500 ms of the frequency event occurring. A key exception are the Type 3 inverters, which all respond slower than 500 ms and with some response time variation. This demonstrates the reality of managing a variety of device types in the field; some existing devices might not be fast enough to participate in FFR without further firmware upgrades. Note that the active power responses shown in Figure 4.13 are based on RMS values that

⁵Total time required to send out request for NL options, wait for responses, receive all responses, and concatenate them together.

⁶Time for send/receive dispatch commands and communications delays at central+local controllers, local controller processing, and local device actuation.

are calculated using a rolling window of multiple cycles. This means that the initial response times (i.e., when the slope of the trace changes) indicated are accurate, but the time to ramp up might be delayed versus reality. Thus, instantaneous current waveform measurements of selected inverters (one of each type) and the appliance loads are presented in Figure 4.14 to confirm total response times. As shown, the appliance loads are all deferred quickly and the custom-designed Type 4 inverter also responds rapidly. All other inverters, with the exception of the Type 3 inverters, begin responding within 500 ms and generally ramp quickly. Figure 4.15 shows additional results from a sampling of other NLGs which confirms that substation power ramps quickly and that the desired dispatch target is met within ~500 ms. The bottom graph of Figure 4.15 shows that this rapid provision of FFR does have a small effect on voltages across the distribution feeder, though not enough to cause any deviations anywhere close to ANSI C84.1 [91] $\pm 5\%$ limits, especially for this relatively-stiff feeder.

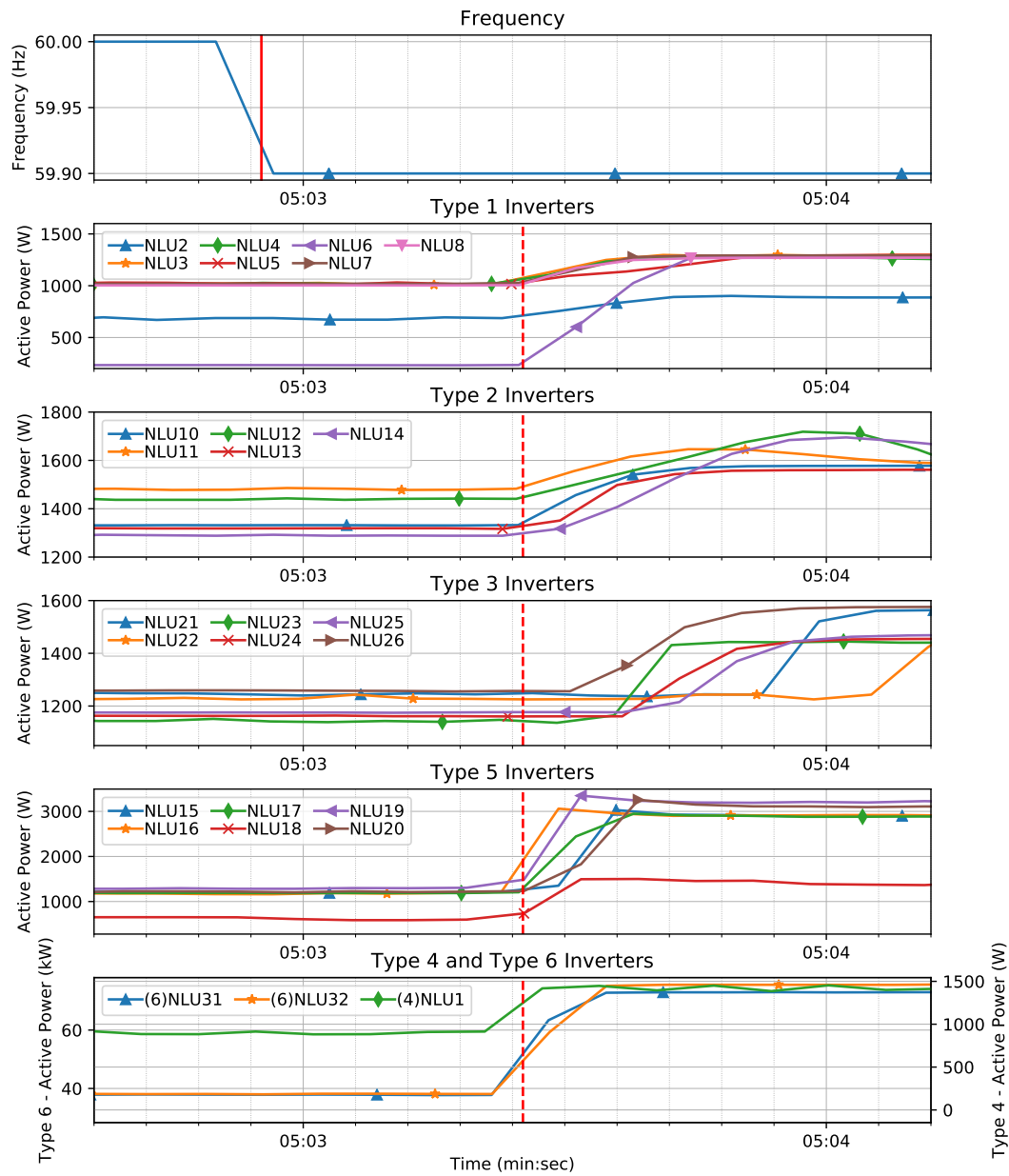


Figure 4.13: Full feeder results: RMS active power response for the more than 27 different hardware inverters in NLU1-NLU32

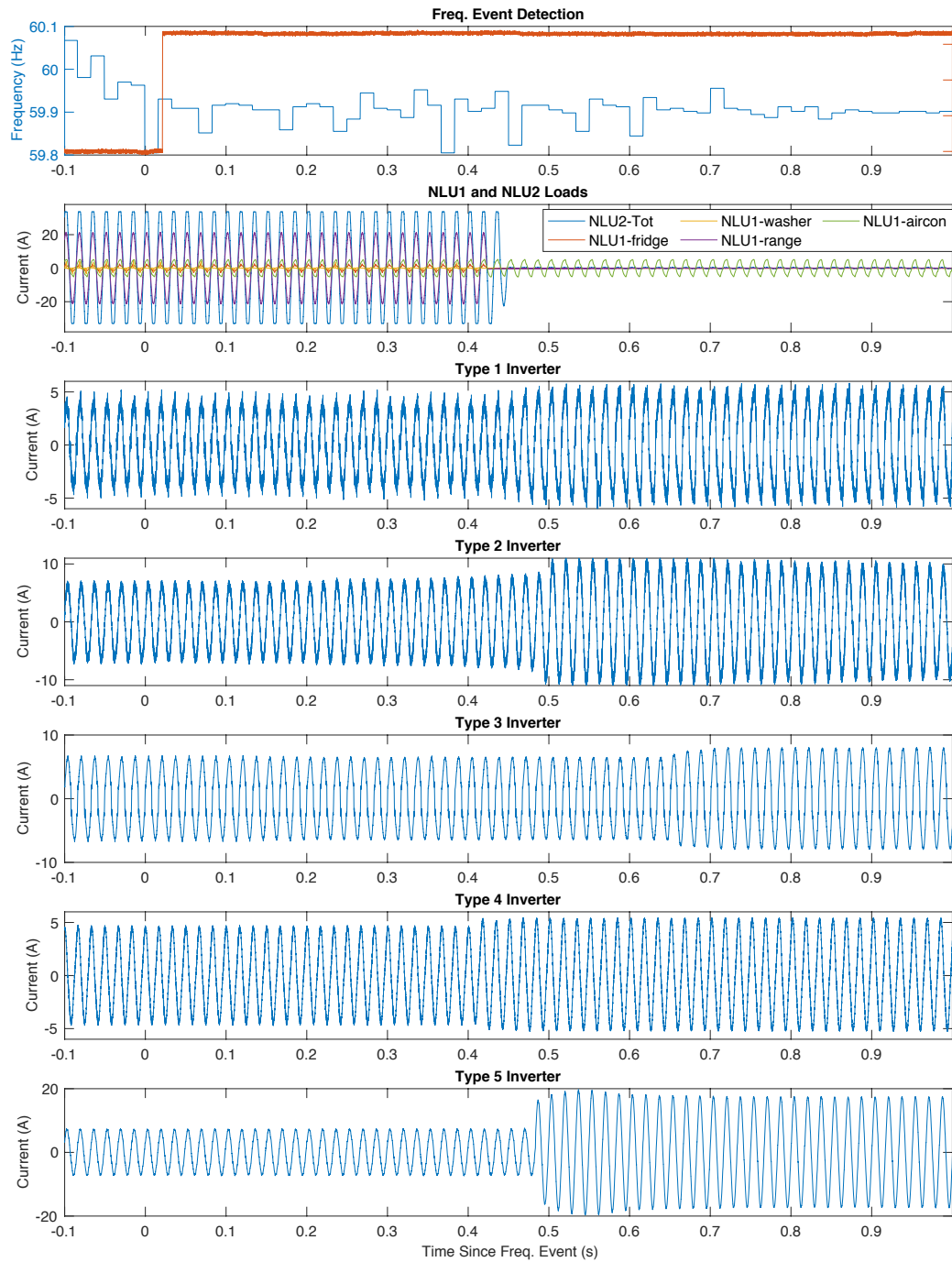


Figure 4.14: Full feeder results: oscilloscope view

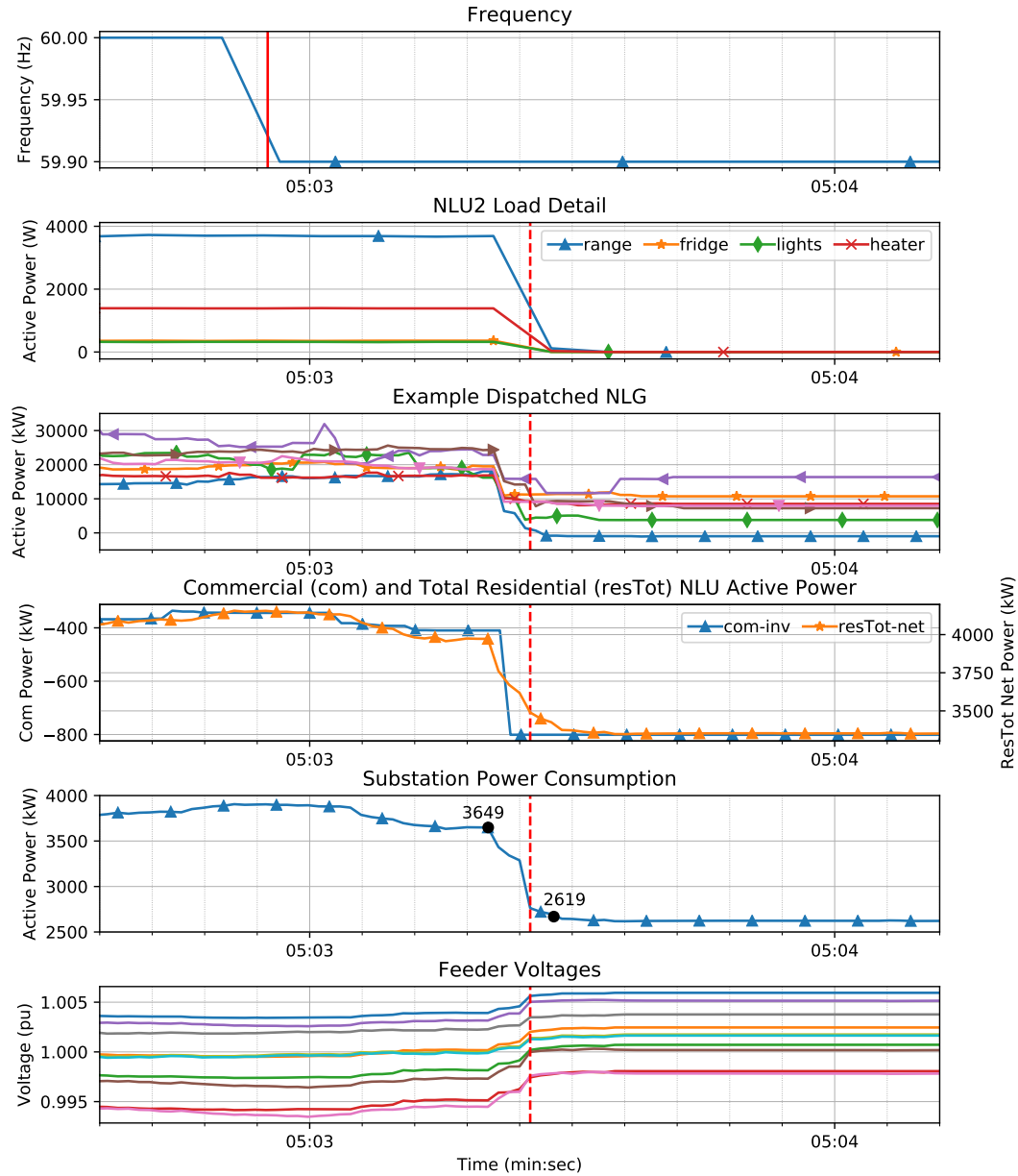


Figure 4.15: Full feeder results: RMS active power response for the appliance loads in NLU2, a selection of the 224 NLGs, all residential and commercial NLUs, and at the distribution feeder head at the substation. Voltages from a selection of points across the full distance from substation are also shown.

Chapter 5

Conclusion and Discussion

This thesis makes several contributions towards implementing innovative solutions for integrating and controlling net-load resources in modern grids. A novel control solution for optimally coordinating thousands of net-load resources across a power distribution network at very fast timescales is developed and experimentally verified. We further the practical value of this coordination solution by innovating solutions to prove such complex coordination solutions prior to field deployment and develop methods to achieve cost-effective, high-speed, inference capabilities in buildings to support this solution in practical implementations.

In Chapter 2, we developed a novel approach for the optimal design of PHIL interfaces and showed how to apply this generalized method to a PHIL simulation case study involving a 100-kVA battery inverter. We demonstrated that this optimal design approach can simplify the process of developing high-accuracy PHIL interfaces—whereas common approaches involve multiple steps of manual compensation and stabilization design—by using a modern control design framework, specifying performance bounds, and then automatically synthesizing an optimal controller. This considerably simplifies the design of a stabilizing PHIL interface compensator by removing the need to iterate over controllers and allowing explicit definition of accuracy objectives, rather than leaving performance to subjective measures. In the case study, the bandwidth and accuracy of the PHIL simulation results were compared when using the optimal PHIL interface compensation design versus a commonly used interface design approach, and significant performance improvements were shown for the optimal approach. We

also introduced a practical method and metrics for verifying the accuracy of the PHIL simulation results that provides a quantitative measure of how accurate a particular simulation result is without relying on relative comparisons to potentially inaccurate models or previous simulation results. This accuracy evaluation method provides a valuable means for PHIL practitioners and researchers to easily identify the absolute accuracy of a particular simulation and compare results from two different simulations on a common basis.

Chapter 3 developed a novel hybrid classification-regression approach for high-frequency, real-time, multiclass NILM. This method was validated using a test bed with eight residential appliances, and it was shown to have high accuracy, good scaling and generalization properties, and sufficient response time to support building grid-interactive control at fast timescales (e.g., within 10 ac cycles) relevant to the provision of grid frequency support services. With minimal additional equipment and cost, the solution developed provides high-speed monitoring capabilities to better measure building performance and understand load usage patterns, including on/off state along with operating power levels, while further enabling smart buildings to support a modernized grid.

In Chapter 4, we developed a first-of-its-kind hierarchical control solution that can optimally coordinate tens of thousands of heterogeneous, fast-changing DERs and flexible loads across an electric distribution network to provide FFR to the bulk power system that meets the general FFR service performance requirements. A method for operating solar inverters with the optimal amount of headroom to minimize opportunity cost while ensuring FFR viability of an NLU was also presented. Practical algorithms for achieving optimal FFR dispatch at multiple scales were developed, their optimality proven, and their speed shown to be faster than state-of-the-art approaches. These simulation results provide validation that coordination of BTM resources in distribution networks can be achieved rapidly enough and on a sufficiently large scale to provide valuable FFR services to the bulk power system. The entire platform is implemented and experimentally verified on a large-scale PHIL platform with a real-world distribution system model and more than 100 at-power DERs. Although additional implementation work to achieve robust, large-scale field deployments and further development of regulatory and market structures will be needed to allow full adoption of this scheme,

this work provides a first-of-its-kind experimental demonstration that this approach is viable.

References

- [1] B. Lundstrom, S. Chakraborty, G. Lauss, R. Bründlinger, and R. Conklin. Evaluation of system-integrated smart grid devices using software- and hardware-in-the-loop. In *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2016.
- [2] G. Lauss, M. O. Faruque, K. Schoder, C. Dufour, A. Viehweider, and J. Langston. Characteristics and Design of Power Hardware-in-the-Loop Simulations for Electrical Power Systems. *IEEE Transactions on Industrial Electronics*, 63(1):406–417, January 2016.
- [3] W. Ren, M. Sloderbeck, M. Steurer, V. Dinavahi, T. Noda, S. Filizadeh, A. R. Chevretils, M. Matar, R. Iravani, C. Dufour, J. Belanger, M.O. Faruque, K. Strunz, and J. A. Martinez. Interfacing Issues in Real-Time Digital Simulators. *IEEE Transactions on Power Delivery*, 26(2):1221–1230, April 2011.
- [4] Wei Ren, Michael Steurer, and Thomas L. Baldwin. Improve the Stability and the Accuracy of Power Hardware-in-the-Loop Simulation by Selecting Appropriate Interface Algorithms. *IEEE Transactions on Industry Applications*, 44(4):1286–1294, July 2008.
- [5] Alexander Viehweider, Georg Lauss, and Felix Lehfuss. Stabilization of Power Hardware-in-the-Loop simulations of electric energy systems. *Simulation Modelling Practice and Theory*, page 10, 2011.
- [6] P. C. Kotsampopoulos, F. Lehfuss, G. F. Lauss, B. Bletterie, and N. D. Hatziargyriou. The Limitations of Digital Simulation and the Advantages of PHIL Testing

- in Studying Distributed Generation Provision of Ancillary Services. *IEEE Transactions on Industrial Electronics*, 62(9):5502–5515, September 2015.
- [7] Mahdi Dargahi, Arindam Ghosh, and Gerard Ledwich. Stability synthesis of power hardware-in-the-loop (PHIL) simulation. In *2014 IEEE PES General Meeting | Conference Exposition*, pages 1–5, July 2014.
 - [8] G. Lauss, F. Lehfuß, A. Viehweider, and T. Strasser. Power hardware in the loop simulation with feedback current filtering for electric systems. In *IECON*, pages 3725–3730, 2011.
 - [9] Il Do Yoo and A. M. Gole. Compensating for Interface Equipment Limitations to Improve Simulation Accuracy of Real-Time Power Hardware In Loop Simulation. *IEEE Transactions on Power Delivery*, 27(3):1284–1291, July 2012.
 - [10] Xiaoming Zha, Chenxu Yin, Jianjun Sun, Meng Huang, and Qionglin Li. Improving the Stability and Accuracy of Power Hardware-in-the-Loop Simulation Using Virtual Impedance Method. *Energies*, 9, 2016.
 - [11] S. Jiang, G. Li, Y. Xin, L. Wang, and W. Wang. Interface algorithm development for PHIL simulations of MMC-HVDC devices via real-time impedance matching. *The Journal of Engineering*, 2019(16):3234–3239, 2019.
 - [12] Sanaz Paran and Chris S. Edrington. Improved power hardware in the loop interface methods via impedance matching. In *2013 IEEE Electric Ship Technologies Symposium (ESTS)*, pages 342–346, April 2013.
 - [13] Georg Lauss and Kai Strunz. Multirate Partitioning Interface for Enhanced Stability of Power Hardware-in-the-Loop Real-Time Simulation. *IEEE Transactions on Industrial Electronics*, 66(1):595–605, January 2019.
 - [14] Kui Luo, Wenhui Shi, Yongning Chi, Qiuwei Wu, and Weisheng Wang. Stability and accuracy considerations in the design and implementation of wind turbine power hardware in the loop platform. *CSEE Journal of Power and Energy Systems*, 3(2):167–175, June 2017.

- [15] Nathan Ainsworth, Ali Hariri, Kumaraguru Prabakar, Annabelle Pratt, and Murali Baggu. Modeling and compensation design for a power hardware-in-the-loop simulation of an AC distribution system. In *2016 North American Power Symposium (NAPS)*, pages 1–6, September 2016.
- [16] W. Ren, M. Steurer, and S. Woodruff. Applying Controller and Power Hardware-in-the-Loop Simulation in Designing and Prototyping Apparatuses for Future All Electric Ship. In *2007 IEEE Electric Ship Technologies Symposium*, pages 443–448, May 2007.
- [17] Kapil Upamanyu and G. Narayanan. Improved Accuracy, Modeling, and Stability Analysis of Power-Hardware-in-Loop Simulation With Open-Loop Inverter as Power Amplifier. *IEEE Transactions on Industrial Electronics*, 67(1):369–378, January 2020.
- [18] Nathan D. Marks, Wang Y. Kong, and Daniel S. Birt. Stability of a Switched Mode Power Amplifier Interface for Power Hardware-in-the-Loop. *IEEE Transactions on Industrial Electronics*, 65(11):8445–8454, November 2018.
- [19] A. Markou, V. Kleftakis, P. Kotsampopoulos, and N. Hatziargyriou. Improving existing methods for stable and more accurate Power Hardware-in-the-Loop experiments. In *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pages 496–502, June 2017.
- [20] Shengbo Wang, Binbin Li, Zigao Xu, Xiaodong Zhao, and Dianguo Xu. A Precise Stability Criterion for Power Hardware-in-the-Loop Simulation System. In *10th International Conference on Power Electronics and ECCE Asia*, pages 3265–3270, May 2019.
- [21] Efren Guillo-Sansano, Mazheruddin H Syed, Andrew J Roscoe, and Graeme M Burt. Assessment and Development of Stability Enhancing Methods for Dynamically Changing Power Hardware-in-the-Loop Simulations. In *25th Intl. Conference on Electricity Distribution*, page 5, June 2019.

- [22] Nathan D. Marks, Wang Y. Kong, and Daniel S. Birt. Interface Compensation for Power Hardware-in-the-Loop. In *2018 IEEE 27th International Symposium on Industrial Electronics (ISIE)*, pages 413–420, June 2018.
- [23] Olivier Tremblay, Handy Fortin-Blanchette, Richard Gagnon, and Yves Brissette. Contribution to stability analysis of power hardware-in-the-loop simulators. *IET Generation, Transmission & Distribution*, 11(12):3073–3079, 2017.
- [24] Ron Brandl. Operational Range of Several Interface Algorithms for Different Power Hardware-In-The-Loop Setups. *Energies*, 10, 2017.
- [25] A. Castro, Pavel Zuniga, F. A. Uribe, and E. Barocio. Phase compensation scheme to improve the accuracy of a power hardware-in-the-loop experiment based on a synchronous generator. In *2015 IEEE Eindhoven PowerTech*, pages 1–6, June 2015.
- [26] E. Guillo-Sansano, A. J. Roscoe, C. E. Jones, and G. M. Burt. A new control method for the power interface in power hardware-in-the-loop simulation to compensate for the time delay. In *49th Intl. Universities Power Engineering Conference (UPEC)*, pages 1–5, September 2014.
- [27] D Barakos, P Kotsampopoulos, A Vassilakis, V Kleftakis, and N Hatziargyriou. Methods for Stability and Accuracy evaluation of Power Hardware In the Loop Simulations. In *MedPower 2014*, page 6, 2014.
- [28] Wei Ren, Michael Steurer, and Thomas L. Baldwin. An Effective Method for Evaluating the Accuracy of Power Hardware-in-the-Loop Simulations. *IEEE Transactions on Industry Applications*, 45(4):1484–1490, July 2009.
- [29] Panos Kotsampopoulos, Dimitris Lagos, Nikos Hatziargyriou, M. Omar Faruque, Georg Lauss, Onyi Nzimako, Paul Forsyth, Michael Steurer, F. Ponci, A. Monti, V. Dinavahi, and Kai Strunz. A Benchmark System for Hardware-in-the-Loop Testing of Distributed Energy Resources. *IEEE Power and Energy Technology Systems Journal*, 5(3):94–103, September 2018.
- [30] P2004 - Hardware-in-the-Loop (HIL) Simulation Based Testing of Electric Power Apparatus and Controls.

- [31] Jing Wang, Blake Lundstrom, Ismael Mendoza, and Annabelle Pratt. Systematic Characterization of Power Hardware-in-the-Loop Evaluation Platform Stability. In *IEEE Energy Conversion Congress and Exposition (ECCE)*, pages 1068–1075, September 2019.
- [32] J.C. Doyle, K. Glover, P.P. Khargonekar, and B.A. Francis. State-space solutions to standard H2 and H-infinity control problems. *IEEE Transactions on Automatic Control*, 34(8):831–847, August 1989.
- [33] U.S. EIA. Electricity Explained: Use of Electricity, 2018.
- [34] Kaustav Basu, Vincent Debusschere, Seddik Bacha, Ujjwal Maulik, and Sanghamitra Bondyopadhyay. Nonintrusive Load Monitoring: A Temporal Multilabel Classification Approach. *IEEE Transactions on Industrial Informatics*, 11(1):262–270, February 2015.
- [35] Seyed Mostafa Tabatabaei, Scott Dick, and Wilsun Xu. Toward Non-Intrusive Load Monitoring via Multi-Label Classification. *IEEE Transactions on Smart Grid*, 8(1):26–40, January 2017.
- [36] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake. A real-time non-intrusive load monitoring system. In *2016 11th International Conference on Industrial and Information Systems (ICIIS)*, pages 850–855, December 2016.
- [37] B. Liu, Y. Yu, W. Luan, and B. Zeng. An unsupervised electrical appliance modeling framework for non-intrusive load monitoring. In *2017 IEEE Power Energy Society General Meeting*, pages 1–5, July 2017.
- [38] Michael A. Devlin and Barry P. Hayes. Non-Intrusive Load Monitoring and Classification of Activities of Daily Living Using Residential Smart Meter Data. *IEEE Transactions on Consumer Electronics*, 65(3):339–348, August 2019.
- [39] Hui He, Zixuan Liu, Runhai Jiao, and Guangwei Yan. A Novel Nonintrusive Load Monitoring Approach based on Linear-Chain Conditional Random Fields. *Energies*, 12(9):1797, May 2019.

- [40] Ding Li and Scott Dick. Whole-house Non-Intrusive Appliance Load Monitoring via multi-label classification. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2749–2755, Vancouver, BC, Canada, July 2016. IEEE.
- [41] Maria Kaselimi, Nikolaos Doulamis, Anastasios Doulamis, Athanasios Voulodimos, and Eftychios Protopapadakis. Bayesian-optimized bidirectional lstm regression model for non-intrusive load monitoring. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2747–2751. IEEE, 2019.
- [42] Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 55–64, 2015.
- [43] T Hassan, F Javed, and N Arshad. An Empirical Investigation of V-I Trajectory based Load Signatures for Non-Intrusive Load Monitoring. *IEEE Transactions on Smart Grid*, 5(2):10, 2014.
- [44] R. A. S. Fernandes, I. N. da Silva, and M. Oleskovicz. Load Profile Identification Interface for Consumer Online Monitoring Purposes in Smart Grids. *IEEE Transactions on Industrial Informatics*, 9(3):1507–1517, August 2013.
- [45] C. Duarte, P. Delmar, K. W. Goossen, K. Barner, and E. Gomez-Luna. Non-intrusive load monitoring based on switching voltage transients and wavelet transforms. In *2012 Future of Instrumentation International Workshop (FIIW) Proceedings*, pages 1–4, October 2012.
- [46] Jessie M. Gillis, Sami M. Alshareef, and Walid G. Morsi. Nonintrusive Load Monitoring Using Wavelet Design and Machine Learning. *IEEE Transactions on Smart Grid*, 7(1):320–328, January 2016.
- [47] D. He, W. Lin, N. Liu, R. G. Harley, and T. G. Habetler. Incorporating Non-Intrusive Load Monitoring Into Building Level Demand Response. *IEEE Transactions on Smart Grid*, 4(4):1870–1877, December 2013.
- [48] M. Nardello, M. Rossi, and D. Brunelli. An innovative cost-effective smart meter with embedded non intrusive load monitoring. In *2017 IEEE PES Innovative*

Smart Grid Technologies Conference Europe (ISGT-Europe), pages 1–6, September 2017.

- [49] Blake Lundstrom, S. Patel, S. Attree, and M. V. Salapaka. Fast Primary Frequency Response using Coordinated DER and Flexible Loads: Framework and Residential-scale Demonstration. In *IEEE Power and Energy Society General Meeting*, August 2018.
- [50] I Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.
- [51] S. Kashyap and A. K. Singh. Most suitable mother wavelet for measurement of power system harmonics using DWT in view of IEEE Standard 1459-2000. In *2008 13th International Conference on Harmonics and Quality of Power*, pages 1–6, September 2008.
- [52] Blake R Lundstrom, Govind Saraswat, and Murti V Salapaka. High-frequency, multiclass nonintrusive load monitoring for grid-interactive residential buildings. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2020.
- [53] Aurélien Geron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, 2019.
- [54] Ludmila I Kuncheva. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning*, 51:181–207, 2003.
- [55] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [56] In-Kwon Yeo and Richard A Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 2000.
- [57] François Chollet et al. Keras. <https://keras.io>, 2015.
- [58] Jean B. Ubertalli and T. B. Littler. Extending the Reach of Traditional Frequency Control for Fast Responses. In *2019 IEEE Milan PowerTech*, pages 1–6, June 2019.

- [59] Frequency Stability Evaluation Criteria for the Synchronous Zone of Continental Europe: Requirements and impacting factors. Technical report, ENTSO-E, March 2016.
- [60] Vahan Gevorgian, Yingchen Zhang, and Erik Ela. Investigating the Impacts of Wind Generation Participation in Interconnection Frequency Response. *IEEE Transactions on Sustainable Energy*, 6(3):1004–1012, July 2015.
- [61] Inertia: Basic Concepts and Impacts on the ERCOT Grid. Technical report, Electric Reliability Council of Texas (ERCOT), April 2018.
- [62] Weifeng Li, Pengwei Du, and Ning Lu. Design of a New Primary Frequency Control Market for Hosting Frequency Response Reserve Offers From Both Generators and Loads. *IEEE Transactions on Smart Grid*, 9(5):4883–4892, September 2018.
- [63] Vahan Gevorgian and Barbara O’Neill. Advanced Grid-Friendly Controls Demonstration Project for Utility-Scale PV Power Plants. Technical Report NREL/TP-5D00-65368, National Renewable Energy Laboratory, 2016.
- [64] Anderson F. Hoke, Mariko Shirazi, Sudipta Chakraborty, Eduard Muljadi, and Dragan Maksimovic. Rapid Active Power Control of Photovoltaic Systems for Grid Frequency Support. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 5(3):1154–1163, September 2017.
- [65] Young-Jin Kim, Gerard Del-Rosario-Calaf, and Leslie K. Norford. Analysis and Experimental Implementation of Grid Frequency Regulation Using Behind-the-Meter Batteries Compensating for Fast Load Demand Variations. *IEEE Transactions on Power Systems*, 32(1):484–498, January 2017.
- [66] D S Callaway and I A Hiskens. Achieving Controllability of Electric Loads. *Proceedings of the IEEE*, 99(1):184–199, January 2011.
- [67] Angel Molina-Garcia, François Bouffard, and Daniel S. Kirschen. Decentralized Demand-Side Contribution to Primary Frequency Control. *IEEE Transactions on Power Systems*, 26(1):411–419, February 2011.

- [68] *Annual Energy Outlook 2019*. U.S. Energy Information Administration, January 2019.
- [69] Electric Storage Participation in Markets Operated by Regional Transmission Organizations and Independent System Operators. Notice of Proposed Rulemaking, U.S. Federal Energy Regulatory Commission, November 2016.
- [70] Joe A. Short, David G. Infield, and Leon L. Freris. Stabilization of Grid Frequency Through Dynamic Demand Control. *IEEE Transactions on Power Systems*, 22(3):1284–1293, August 2007.
- [71] Donald J. Hammerstrom, Jerry Brous, David P. Chassin, Gale R. Horst, and et al. Pacific Northwest GridWiseTM Testbed Demonstration Projects; Part II. Grid FriendlyTM Appliance Project. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2007.
- [72] Hazem Karbouj, Zakir Rather, Damian Flynn, and Hassan Qzai. Non-synchronous fast frequency reserves in renewable energy integrated power systems: A critical review. *International Journal of Electrical Power & Energy Systems*, 106:488–501, 2019.
- [73] Jianming Lian, Jacob Hansen, Laurentiu D. Marinovici, and Karanjit Kalsi. Hierarchical decentralized control strategy for demand-side primary frequency response. In *IEEE Power and Energy Society General Meeting*, 2016.
- [74] Hassan W Qazi and Damian Flynn. Analysing the Impact of Large-scale Decentralised Demand Side Response on Frequency Stability. *International Journal of Electrical Power and Energy Systems*, September 2016.
- [75] Mohammad R. Vedady Moghadam, Richard T. B. Ma, and Rui Zhang. Distributed Frequency Control in Smart Grids via Randomized Demand Response. *IEEE Transactions on Smart Grid*, 5(6):2798–2809, November 2014.
- [76] Zhao Xu, Jacob Ostergaard, and Mikael Tøgeby. Demand as Frequency Controlled Reserve. *IEEE Transactions on Power Systems*, 26(3):1062–1071, August 2011.

- [77] Yu-Qing Bao, Yang Li, Ying-Yi Hong, and Beibei Wang. Design of a hybrid hierarchical demand response control scheme for the frequency control. *IET Generation, Transmission & Distribution*, 9(15):2303–2310, 2015.
- [78] Rajesh Bhana and Thomas J. Overbye. The Commitment of Interruptible Load to Ensure Adequate System Primary Frequency Response. *IEEE Transactions on Power Systems*, 31(3):2055–2063, May 2016.
- [79] Benjamin Biegel, Lars Henrik Hansen, Palle Andersen, and Jakob Stoustrup. Primary control by ON/OFF demand-side devices. *IEEE Transactions on Smart Grid*, 4(4):2061–2071, 2013.
- [80] Xiaoyu Wu, Jinghan He, Yin Xu, Jian Lu, Ning Lu, and Xiaojun Wang. Hierarchical Control of Residential HVAC Units for Primary Frequency Regulation. *IEEE Transactions on Smart Grid*, 9(4):3844–3856, July 2018.
- [81] Armin Ghasem Azar, Hamidreza Nazaripouya, Behnam Khaki, Chi-Cheng Chu, Rajit Gadh, and Rune Hylsberg Jacobsen. A Non-Cooperative Framework for Coordinating a Neighborhood of Distributed Prosumers. *IEEE Transactions on Industrial Informatics*, 15(5):2523–2534, May 2019.
- [82] Yinliang Xu, Wenxin Liu, and Jun Gong. Stable Multi-Agent-Based Load Shedding Algorithm for Power Systems. *IEEE Transactions on Power Systems*, 26(4):2006–2014, November 2011.
- [83] Changhong Zhao, Ufuk Topcu, and Steven H. Low. Optimal Load Control via Frequency Measurement and Neighborhood Area Communication. *IEEE Transactions on Power Systems*, 28(4):3576–3587, November 2013.
- [84] Jianqiang Hu, Jinde Cao, Josep M. Guerrero, Taiyou Yong, and Jie Yu. Improving Frequency Stability Based on Distributed Control of Multiple Load Aggregators. *IEEE Transactions on Smart Grid*, 8(4):1553–1567, July 2017.
- [85] Lei Tang and James McCalley. Two-Stage Load Control for Severe Under-Frequency Conditions. *IEEE Transactions on Power Systems*, 31(3):1943–1953, May 2016.

- [86] Alexander M. Prostejovsky, Mattia Marinelli, Michel Rezkalla, Mazheruddin H. Syed, and Efren Guillo-Sansano. Tuningless Load Frequency Control Through Active Engagement of Distributed Resources. *IEEE Transactions on Power Systems*, 33(3):2929–2939, May 2018.
- [87] Opal-RT Technologies. ePHASORSIM.
- [88] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In *SustKDD workshop on Data Mining Applications in Sustainability*, 2011.
- [89] Raspberry Pi Foundation. Raspberry Pi 3 Model B+.
- [90] iMatix. ZeroMQ Distributed Messaging.
- [91] *American National Standard for Electric Power Systems and Equipment—Voltage Ratings (60 Hz)*. ANSI Standard C84.1, 2016.
- [92] X. Xi, H. Geng, G. Yang, Y. Zhou, J. Lin, and Y. Song. Small-signal stability analysis and PHIL experiment validation of power system with WECS voltage/frequency ancillary control. In *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, pages 002544–002549, November 2015.
- [93] Lei Jiang, Suhuai Luo, and Jiaming Li. An Approach of Household Power Appliance Monitoring Based on Machine Learning. In *2012 Fifth International Conference on Intelligent Computation Technology and Automation*, pages 577–580, Zhangjiajie, Hunan, China, January 2012. IEEE.
- [94] Z. Wang and G. Zheng. Residential Appliances Identification and Monitoring by a Nonintrusive Method. *IEEE Transactions on Smart Grid*, 3(1):80–92, March 2012.
- [95] Y. Su, K. Lian, and H. Chang. Feature Selection of Non-intrusive Load Monitoring System Using STFT and Wavelet Transform. In *2011 IEEE 8th International Conference on e-Business Engineering*, pages 293–298, October 2011.

- [96] S. B. Leeb, S. R. Shaw, and J. L. Kirtley. Transient event detection in spectral envelope estimates for nonintrusive load monitoring. *IEEE Transactions on Power Delivery*, 10(3):1200–1210, July 1995.
- [97] A. I. Cole and A. Albicki. Data extraction for effective non-intrusive identification of residential power loads. volume 2, pages 812–815, May 1998.
- [98] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox. Nonintrusive Load Monitoring and Diagnostics in Power Systems. *IEEE Transactions on Instrumentation and Measurement*, 57(7):1445–1454, July 2008.
- [99] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle. Circuit-Level Load Monitoring for Household Energy Management. *IEEE Pervasive Computing*, 10(1):40–48, January 2011.
- [100] D. He, L. Du, Y. Yang, R. Harley, and T. Habetler. Front-End Electronic Circuit Topology Analysis for Model-Driven Classification and Monitoring of Appliance Loads in Smart Buildings. *IEEE Transactions on Smart Grid*, 3(4):2286–2293, December 2012.
- [101] K. Basu, V. Debusschere, and S. Bacha. Appliance usage prediction using a time series based classification approach. pages 1217–1222, October 2012.
- [102] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [103] Philip J. Douglass, Rodrigo Garcia-Valle, Preben Nyeng, Jacob Østergaard, and Mikael Tøgeby. Smart Demand for Frequency Regulation: Experimental Results. *IEEE Transactions on Smart Grid*, 4(3):1713–1720, September 2013.
- [104] D Kithany, N Akhtar, and H Liu. Major Home Appliance Market Report – 2017, December 2017.
- [105] Towfiq Rahman, Roland Harvey, Zhihua Qu, and Marwan A. Simaan. A distributed cooperative load control approach for ancillary services in smart grid. In *American Control Conference (ACC), 2017*, pages 1401–1406. IEEE, 2017.

- [106] NERC. Project 2017-01 Modifications to BAL-003-1.1.
- [107] Essential Reliability Services and the Evolving Bulk-Power System–Primary Frequency Response. Technical Report Order No. 842, Federal Energy Regulatory Commission, Washington, D.C., USA, February 2018.
- [108] Jian Meng. Dynamic frequency response from electric vehicles considering traveling behavior in the Great Britain power system. *Applied Energy*, page 14, 2016.
- [109] Seyedmahdi Izadkhast, Pablo Garcia-Gonzalez, and Pablo Frias. An Aggregate Model of Plug-In Electric Vehicles for Primary Frequency Control. *IEEE Transactions on Power Systems*, 30(3):1475–1482, May 2015.