Title
Frequency Control Requirements for Reliable Interconnection Frequency Response

Authors
Eto, JH
Undrill, J
Roberts, C
et al.

Publication Date
2018-02-13

Peer reviewed

LBNL-2001103

Frequency Control Requirements for Reliable Interconnection Frequency Response

Authors:

Joseph H. Eto,[1] John Undrill,[2] Ciaran Roberts,[1] Peter Mackin,[3] and Jeffrey Ellis[3]

[1] Lawrence Berkeley National Laboratory
[2] John Undrill, LLC.
[3] Utility Systems Efficiencies, Inc.

Energy Analysis and Environmental Impacts Division

Lawrence Berkeley National Laboratory

February 2018

This work was supported by the Federal Energy Regulatory Commission, Office of Electric Reliability, under interagency Agreement #FERC-16-I-0105, and in accordance with the terms of Lawrence Berkeley National Laboratory' Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy.

Frequency Control Requirements for Reliable Interconnection Frequency Response

Prepared for the
Office of Electric Reliability
Federal Energy Regulatory Commission

Principal Investigator
Joseph H. Eto
Lawrence Berkeley National Laboratory

Project Team
John Undrill, John Undrill, LLC
Ciaran Roberts, Lawrence Berkeley National Laboratory
Peter Mackin and Jeffrey Ellis, Utility System Efficiencies, Inc.

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R4000
Berkeley CA 94720-8136

LBNL-2001103

February 2018

The work described in this study was funded by the Federal Energy Regulatory Commission's Office of Electric Reliability, under interagency Agreement #FERC-16-I-0105, and in accordance with the terms of Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy.

Acknowledgments

The work described in this study was funded by the Federal Energy Regulatory Commission (FERC), Office of Electric Reliability, under interagency Agreement #FERC-16-I-0105, and in accordance with the terms of Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. The authors acknowledge project management provided by Eddy Lim, FERC's Office of Electric Reliability.

The authors thank Hongming Zhang and Terry Baker at Peak Reliability (PeakRC), Julia Matevosyan at the Electric Reliability Council of Texas (ERCOT), and Raja Thappetaobula at the Midwest Independent System Operator (MISO) for providing information on each of the three U.S. interconnections that was used as a basis for aspects of the operating conditions that we studied using simulation tools.

The authors thank Dmitry Kosterev, Bonneville Power Administration, Sydney Niemeyer, NRG Energy (retired), and Julia Matevosyan and Sandip Sharma, Electric Reliability Council of Texas for review comments on an early draft of this report.

All opinions, errors, and omissions remain the responsibility of the authors. All reference URLs were accurate as of February 2018.

Table of Contents

Acknowledgments i

Table of Contents ii

Table of Figures iv

List of Tables vi

Acronyms and Abbreviations vii

Executive Summary viii

Review of Frequency Control Concepts ix

Analysis Approach x

Study Findings xi

Observations xvii

Recommendations xxi

1. Introduction 1

2. Review of Frequency Control Concepts and Terminology 4

2.1 System Frequency Reflects the Balance between Generation and Load 4

2.2 Power System Frequency is Controlled by Resources that Adjust Their Output to Oppose Deviations in Frequency from 60 Hz 5

2.3 The Sudden Loss of a Large Amount of Generation is the Most Significant Threat to the Reliable Management of Interconnection Frequency 9

2.4 The Sequence of Frequency Control Actions Following a Sudden Loss of Generation 10

2.5 The Criterion that Determines whether Interconnection Frequency Control is Reliable is Avoiding Triggering Under-Frequency Load Shedding 12

2.6 The Rate of Frequency Decline Determines the Requirements for Reliable Interconnection Frequency Response 14

2.7 Summary 17

3. Background 19

3.1 LBNL's 2010 Study 19

3.2 Industry Activities Following the Publication of LBNL's 2010 Study 21

3.3 Summary 31

4. Analysis Approach 32

4.1 Overview of the Simulation Tool and System Modeling Approach Used for this Study 33

4.2 Modeling of Interconnection Frequency Response Design Criteria 34

4.3 Representation of System Inertia 35

4.4 Specification of Generation Providing Primary Frequency Response 38

4.5 Review of Generation Modeling Procedures 46

4.6 Treatment of Load Sensitivity 48

4.7 Comparison of Simulation Results with Industry-Developed Interconnection Models 48

4.8 Summary 51

5. Frequency Control Findings 53

5.1 Reserves Held to Provide Primary Frequency Control Must Exceed the Expected Loss of Generation 53

5.2 Primary Frequency Response Must be Delivered Quickly, which Requires Many Participating Generators 55

5.3 For a Given Loss of Generation, System Inertia and the Timing of Primary Frequency Response Determine How Frequency is Arrested 57

5.4 Primary Frequency Response Must be Sustained until Secondary Frequency Response can Replace it 58

5.5 Plant Load Controllers that Operate in Pre-Selected Load Mode without Frequency Bias will Withdraw and Not Sustain Primary Frequency Response 61

5.6 Plant Load Controllers Operated in Pre-Selected Load Mode with Frequency Bias will Sustain Primary Frequency Response 62

5.7 Gas Turbines May Not be able to Sustain Provision of Primary Frequency Response Following Large Loss-of-Generation Events 63

5.8 "Synthetic Inertia" Controls on Electronically Coupled Wind Generation Appear not to Sustain Primary Frequency Response 65

5.9 Fast Demand Response Provides Robust Primary Frequency Response, but Currently is Inflexible 65

5.10 Smaller Deadbands on Turbine-Governors Increase how Quickly Delivery of Primary Frequency Response will Begin 67

5.11 Load Sensitivity Currently Complements Primary Frequency Response, but this Sensitivity may be Going Away 69

5.12 Summary 70

6. Observations and Recommendations 75

6.1 Observations Regarding the Three U.S. Interconnections 75

6.2 Recommendations 80

References 85

Appendix A. Glossary of Terms 88

Table of Figures

Figure ES - 1. The Sequential Actions and Impacts on System Frequency of Primary, Secondary, and Tertiary Frequency Control x

Figure ES - 2. Frequency is Arrested when the Amount of Primary Frequency Response Delivered Equals the Amount of Generation Lost (2%) xii

Figure ES - 3. System Inertia and the Speed of Primary Frequency Response Determine the Nadir at Which Frequency is Arrested xiii

Figure ES - 4. Failure to Sustain Sufficient Primary Frequency Response Will Trigger UFLS xiv

Figure ES - 5. The Relative Impacts of Generation Loss versus System Inertia on Frequency Nadir xviii

Figure ES - 6. Frequency of the Eastern Interconnection during the First 199 Seconds Following the Loss of 4,500 MW of Generation—A Comparison of Recorded Data with Results from a Simulation of the Event xx

Figure ES - 7. Comparison of Selected U.S. and International Grid Codes Related to Frequency Response xxii

Figure 1. North American Electric Reliability Council Interconnections 2

Figure 2. The Concept of Power System Frequency Explained Using the Analogy of Water Level in a Container 5

Figure 3. The Sequential Actions and Impacts on System Frequency of Primary, Secondary, and Tertiary Frequency Control 10

Figure 4. The Effect of System Inertia on ROCOF, for a Fixed Amount of Generation Loss 15

Figure 5. The Effect of the Amount of Generation Lost on ROCOF, for a Fixed System Inertia 15

Figure 6. The Analytical Relationships among System Inertia, Generation Loss, Rate of Change of Frequency (ROCOF), and Frequency Nadir, for a Fixed Amount of Primary Frequency Control... 17

Figure 7. Frequency Response Performance Metrics 20

Figure 8. Frequency of the Eastern Interconnection during the First 199 Seconds Following the Loss of 4,500 MW of Generation: A Comparison of Recorded Data with Results from a Simulation of the Event 24

Figure 9. Eastern Interconnection Generator Responses to NERC Survey 25

Figure 10. Western Interconnection Generator Responses to NERC Survey 25

Figure 11. Texas Interconnection Generator Responses to NERC Survey 25

Figure 12. Representative Inertia Constants by Unit Type 27

Figure 13. Eastern Interconnection Inertia vs. Net Load (August-September 2016) 28

Figure 14. WECC Inertia vs. Net Load (June-September 2016) 28

Figure 15. Texas Interconnection Inertia vs. Net Load (July-September 2016) 29

Figure 16. Response comparison between simulations using new base case and May 13, 2012 event.... 30

Figure 17. Estimated Effective Inertia Constant of the Generation Contributing Inertia in the Western Interconnection (August 2016 to March 2017) 37

Figure 18. Estimated Effective Inertia Constant of Generation Contributing Inertia in the Texas Interconnection (July to December 2016) 37

Figure 19. Generation Modeling Schema for Study of Interconnection Frequency Response 39
Figure 20. WECC Spinning Reserves (January 2016 through March 2017)..41
Figure 21. Texas Interconnection Unloaded Capacity Reserves: 20 Operating Hours during July 2016 through December 2016..41
Figure 22. Turbine-Governor Performance 43
Figure 23. Generation Modeling Schema and Composition of the Responsive Fraction of Generation for the Two Systems Used to Study Interconnection Frequency Response 44
Figure 24. Snapshots of the Composition of Generator Headroom from Five Separate Hours of Dispatch for Texas Interconnection when System Load was Low and Renewable Generation was High between July 2016 and December 2016 45
Figure 25. Snapshots of the Composition of Generation for 5 Separate Hours of Dispatch for the Western Interconnection When System Load was Low and the Contribution of Renewable Energy Generation was High between August 2017 and March 2017..45
Figure 26. Illustration of the Relationships among Simulation Analysis Modeling Parameters.................47
Figure 27. Comparison of Simulations of System Model Developed for this Study to the Industry-developed System Model for the Texas Interconnection...49
Figure 28. Comparison of Simulations of System Model Developed for this Study to the Industry-developed System Model for the Western Interconnection..49
Figure 29. Comparison of Simulations of System Model Developed for this Study to the Industry-developed System Model for the Eastern Interconnection and Measured System Performance 50
Figure 30. Frequency is Arrested when the Amount of Primary Frequency Response Delivered Equals the Amount of Generation Lost (2%) 54
Figure 31. The initial contributions of primary frequency response from many generators is required for reliable interconnection frequency response...56
Figure 32. System Inertia and the Speed of Primary Frequency Response Determine Nadir at Which Frequency is Arrested 57
Figure 33. System Inertia and the Speed of Primary Frequency Response is Determine How Frequency is Arrested 59
Figure 34. Combinations of Responsive, and Responsive and Sustaining Generation Required to Avoid Triggering UFLS 60
Figure 35. Turbine-Governor Primary Frequency Response When Operated With and Without Pre-selected Load Mode...62
Figure 36. Operation in Pre-selected Load Mode with Frequency Bias Will Sustain Primary Frequency Response When Frequency is Less than Nominal 63
Figure 37. Exhaust Gas Temperature Controls on Gas Turbines Will Decrease Primary Frequency Response if Frequency Remains Depressed 64
Figure 38. Fast Demand Response Can Augment Primary Frequency Response Delivered by Generation 66
Figure 39. The Effect of Deadbands on Interconnection Frequency Response...68
Figure 40. Increased Load Damping Supports Reliable Interconnection Frequency Response 70
Figure 41. The Relative Impacts of Generation Loss versus System Inertia on Frequency Nadir...............76

Figure 42. Combinations of Responsive and Responsive and Sustaining Generation Required to Avoid Triggering UFLS Following Loss of 2% and 4% of Generation 76
Figure 43. Comparison of Selected U.S. and International Grid Codes Related to Frequency Response .. 82

List of Tables

Table ES - 1. Interconnection Frequency Response Design Criteria at Times of Minimum System Load xviii
Table 1. Interconnection Frequency Response Obligations 22
Table 2. Interconnection Frequency Response Design Criteria 35
Table 3. Interconnection Generation Characteristics Based on Recent Industry System Planning Models 39
Table 4. Interconnection Generation Dispatch Based on Recent Industry System Planning Models 40

Acronyms and Abbreviations

ACE	Area control error
AGC	Automatic generation control
BA	Balancing Authority
CAISO	California Independent System Operator
CCGT	Combined-cycle gas turbine
CCST	Combined-cycle steam turbine
DOE	United States Department of Energy
Efrac	Electronically coupled fraction of generation
ERCOT	Electric Reliability Council of Texas
ERSWG	Essential Reliability Services Working Group
FERC	Federal Energy Regulatory Commission
GE	General Electric
GT	Gas turbine
GW	Gigawatt
Hydro	Hydro-electric turbine
Hz	Hertz
ISO	Independent system operator
LBNL	Lawrence Berkeley National Laboratory
mHz	millihertz
MW	Megawatt
NERC	North American Electric Reliability Corporation
Nfrac	Non-responsive fraction of generation
NREL	National Renewable Energy Laboratory
PFR_t	Primary frequency response
Rfrac	Responsive fraction
ROCOF	Rate of change of frequency
Sfrac	Sustaining fraction
Steam	Steam turbine
UFLS	Under-frequency load shedding
WECC	Western Electricity Coordinating Council

Executive Summary

The reliability of interconnected electric power systems depends on controlling power system frequency so that it remains within pre-established, safe operating bounds. Reliability is threatened when a large electric generator(s) disconnects from the power system because the loss of generation causes an immediate decline in power system frequency. If the loss of generation is large enough and the remaining, still-connected generators do not respond and rapidly arrest the decline in frequency, power system frequency may decline below established, safe operating bounds and trigger automatic, emergency load shedding to avoid a cascading blackout.

The collective ability of the power system to respond to such events is called interconnection frequency response. Advance planning is required to operate the power system in a manner that ensures reliable frequency response at all times because generation-loss events are always unpredictable even though they occur relatively often.[1]

The Federal Energy Regulatory Commission (FERC) has tasked Lawrence Berkeley National Laboratory (LBNL) to conduct this study to support ongoing FERC and industry efforts to ensure reliable interconnection frequency response for the three major interconnections in the United States: the Western, Eastern, and Texas Interconnections.2 The purpose of this study is to support policymaker and industry understanding of the physical requirements for reliable interconnection frequency response by building upon an initial study conducted by LBNL for FERC in 2010.3 Improved understanding is especially timely now for several reasons.

First, industry experience with the frequency response-related requirements in the North American Electric Reliability Corporation (NERC) Reliability Standard BAL-003-1.1 (Frequency Response and Frequency Bias Setting), which mandates an interconnection-wide frequency response obligation is nascent. Increased understanding of the physical requirements for reliable interconnection frequency response will support industry efforts to comply effectively and in a timely way. Understanding of these requirements will also support possible future efforts to revise BAL-003-1.1 as well as supporting standards and other related activities (e.g., generator interconnection requirements).

Second, industry and policymakers are currently grappling with the reliability implications of changes in the composition of the generation fleet. Deeper understanding of reliable interconnection frequency

---

[1] Given the large number of generators in the three U.S. interconnections, generation-loss events of varying sizes take place routinely, on a weekly, if not more frequent, basis. The very largest events, however, are considerably less frequent and rarely take place more than once a year.

[2] Throughout this report, we refer to the interconnection operated by the Electric Reliability Council of Texas (ERCOT) as the "Texas Interconnection."

[3] In 2010, FERC commissioned LBNL to study the use of frequency response metrics for assessing the reliable integration of variable renewable sources of electricity generation. LBNL prepared a technical report supported by five stand-alone technical appendices (Eto et al. 2010, Undrill 2010, Martinez et al. 2010, Illian 2010, Mackin et al. 2010, and Coughlin, Eto 2010).3 All reports are available at: https://certs.lbl.gov/project/integration-variable-renewable-generation. This study will be referred to in this report as "LBNL's 2010 Study."

response will enable industry to focus on the requirements that they must manage—rapid and sustained primary frequency response—and thereby help guide appropriate focus on related issues, such as how reductions in system inertia increase these requirements.

Finally, industry recognizes the important role that generator turbine-governors currently play in interconnection frequency performance. Expanded understanding of frequency response will enable industry to focus on the most important factors for ensuring that turbine-governors contribute to reliable interconnection frequency response: the speed at which the fleet first delivers and then sustains primary frequency response. Greater understanding will, in turn, support industry discussion of related issues for reliable interconnection frequency response, such as the value of small deadbands.[4]

Review of Frequency Control Concepts

The sudden, unplanned loss of a large amount of generation is an important, periodic challenge for reliable management of interconnection frequency. While unpredictable, these events are commonly the result of a mechanical or electrical problem within a large generating plant that causes the plant to disconnect itself automatically from the grid. Loss of a large amount of generation causes an immediate decline in system frequency that is felt throughout an interconnection. If no corrective actions are taken, frequency declines until the power system collapses, and a cascading, widespread blackout ensues.

Four physical factors determine whether an interconnection will respond reliably—i.e., without triggering emergency, interconnection-coordinated, under-frequency load shedding (UFLS)—to a sudden loss of generation (see Figure ES - 1):

1. The size of the generation-loss event;
2. The interconnection's inertia, which, in combination with the amount of generation lost, determines the initial rate of decline of frequency following an event (i.e., the rate of change of frequency [ROCOF]);
3. The speed with which other on-line resources[5] respond to arrest and stabilize frequency (i.e., provide primary frequency control);[6] and
4. The means by which other generators respond subsequently to restore frequency to its original scheduled value and to restore reserves to their original state of readiness (i.e., provide secondary and tertiary frequency control).

Reliable interconnection frequency response requires that frequency be arrested and stabilized above the highest set-point for UFLS. This report describes the relationships on which interconnection

[4] A deadband on a turbine-governor determines the size of frequency deviation at which delivery of primary frequency will begin.
[5] Non-generation-based resources for primary frequency response, such as the Texas Interconnection's reliance on a nearly instantaneous form of demand response, are also reviewed in this report.
[6] This report refers to primary frequency *control* as an objective and primary frequency *response* as the means by which resources contribute to this objective.

frequency response requirements are based and the considerations that must be addressed to ensure that these requirements are met.

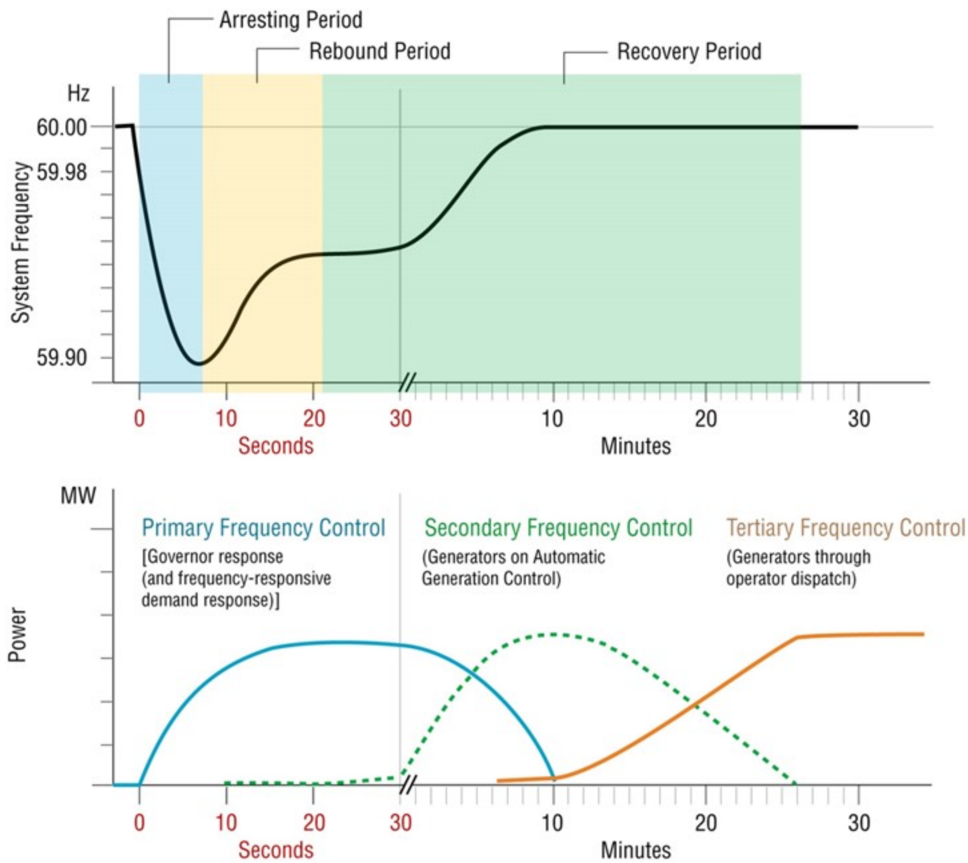


Figure ES - 1. The Sequential Actions and Impacts on System Frequency of Primary, Secondary, and Tertiary Frequency Control

Source: Eto, et al. (2010): Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation

Analysis Approach

We developed a highly flexible modeling approach to illustrate key relationships and interactions among the factors that influence interconnection frequency response. The approach aggregates generators according to whether they do or do not; (1) respond to frequency deviations (i.e., provide primary frequency response); (2) sustain primary frequency response and; (3) contribute inertia to the interconnection.

We implemented the modeling approach by conducting dynamic simulations using the General Electric (GE) Positive Sequence Load Flow tool, known as PSLF—the same commercially available tool that is currently in wide use by industry to conduct, among other things, production-grade studies of frequency response. By using a commercially available tool, we were able to study the performance of turbine-governors and plant load controllers for different types of generators (e.g., combined-cycle, hydro, and steam) using the same models of these generators that are used by industry to conduct reliability studies for planning and operations.[7]

The models we developed allowed us to examine: (1) the interconnection requirements for primary frequency response; (2) the headroom available on generators, which establishes an upper bound on the amount of primary frequency response; (3) the rate at which turbine-governors deliver primary frequency response from this headroom; and (4) plant-specific control settings or operating factors that limit or withdraw primary frequency response early (i.e., before frequency has been stabilized). We also examined fast demand response, governor deadband settings, and load sensitivity (sometimes called load damping),8 which also contribute to frequency response. We directly compared the performance of our simplified study model to the production-grade models developed by industry for each interconnection to demonstrate that we can meaningfully capture the important features of frequency response as predicted by industry models.

This simplified approach does not consider certain complexities of the transmission system, such as the transient or dynamic behavior of power flows, propagation of the generation-loss event, transmission losses, and deliverability, among others, which can be the focus of studies involving the detailed models used by industry.

Study Findings

Our findings are organized into three broad groups: (1) confirmation of fundamental, but potentially not widely understood factors that determine the initial requirements for resources held to provide primary frequency control; (2) the importance of equal attention to ensuring primary frequency response is sustained, including illustrations of various means by which primary frequency response may not be sustained and therefore must either be modified to sustain response or augmented by other resources that will sustain their response; and (3) findings related to other primary frequency control topics, including fast demand response, governor deadband settings, and load sensitivity.

1. *Reserves held to provide primary frequency control must exceed the expected loss of generation.*

Interconnection frequency reflects the balance between generation and load. The rapid decline in frequency following a loss of generation results from the sudden imbalance between generation and load. The decline is arrested once the balance between generation and load has been re-established. See Figure ES - 2. This is not a new or novel finding of this study; however confirmation and illustration of it as a fundamental principle forms the basis for all subsequent findings in this study. Furthermore, prudence dictates that the total primary frequency response capacity held on-line should exceed the

[7] We emphasize, however, that the modeling approach we developed was intentionally simplified in order to focus on the interactions among the central physical factors influencing and resource performance characteristics required for reliable interconnection frequency response. It is not intended to replicate all aspects of, nor hence displace the need for interconnection- and system-level modeling conducted by industry.
[8] The majority of our simulations assumed no load sensitivity in order to focus attention on the relationship between primary frequency control provided by active sources, such as generators, and interconnection frequency response.

size of the design generation-loss event to acknowledge uncertainty in the actual performance of the fleet.

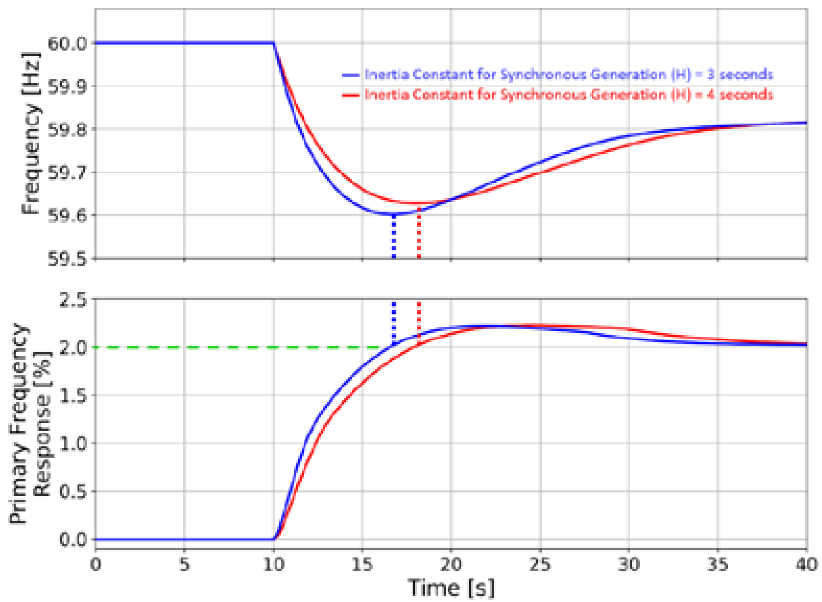


Figure ES - 2. Frequency is Arrested when the Amount of Primary Frequency Response Delivered Equals the Amount of Generation Lost (2%)

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

2. Primary frequency response must be delivered quickly, which requires many participating generators.

The reserve for primary frequency response must be allocated among generators[9] with recognition of the amounts of primary response that each type of generator can produce in the few seconds that are available to arrest the decline of frequency. This recognition can best be achieved, and in practice can only be achieved, by allocating reserves across a number of generators, such that each needs to make a small contribution to the required cumulative response. This finding, too, is generally understood, but it may not be widely appreciated.

> *"…it is prudent to ensure to the extent technically practical, that all generators [should] have the capability to provide primary frequency response."*

As a consequence, it is prudent to ensure to the extent technically practical, that all generators—both those that are directly coupled and those that are electronically coupled to the grid—have the capability to provide primary frequency response. Ensuring this capability provides maximum flexibility to grid operators to assign primary frequency response duty to generators as appropriate for the current grid operating conditions. Exceptions should be considered only on a case-by-case basis.

[9] The principal focus of this study is on primary frequency response provided by generation resources. Separate findings on primary frequency response by non-generation resources and on the sympathetic, but changing, role of load sensitivity appear at the end of this section.

3. *For a given loss of generation, system inertia and the timing of primary frequency response determine how frequency is arrested.*

The key factors determining the nadir of frequency are (a) the effective inertia constant of the system, which determines the initial rate of decline of frequency; and (b) the rate at which generation is increased by primary frequency response. Lower system inertia will require faster primary frequency response. Understanding the expected dynamic performance of the reserves that are held to respond, therefore, becomes of greater importance. For example, if the reserves held are quick to respond, they may be adequate for a wide range of possible future generation loss scenarios and corresponding system inertias. If they are slow to respond, they may require augmentation by faster responding reserves. See Figure ES - 3.

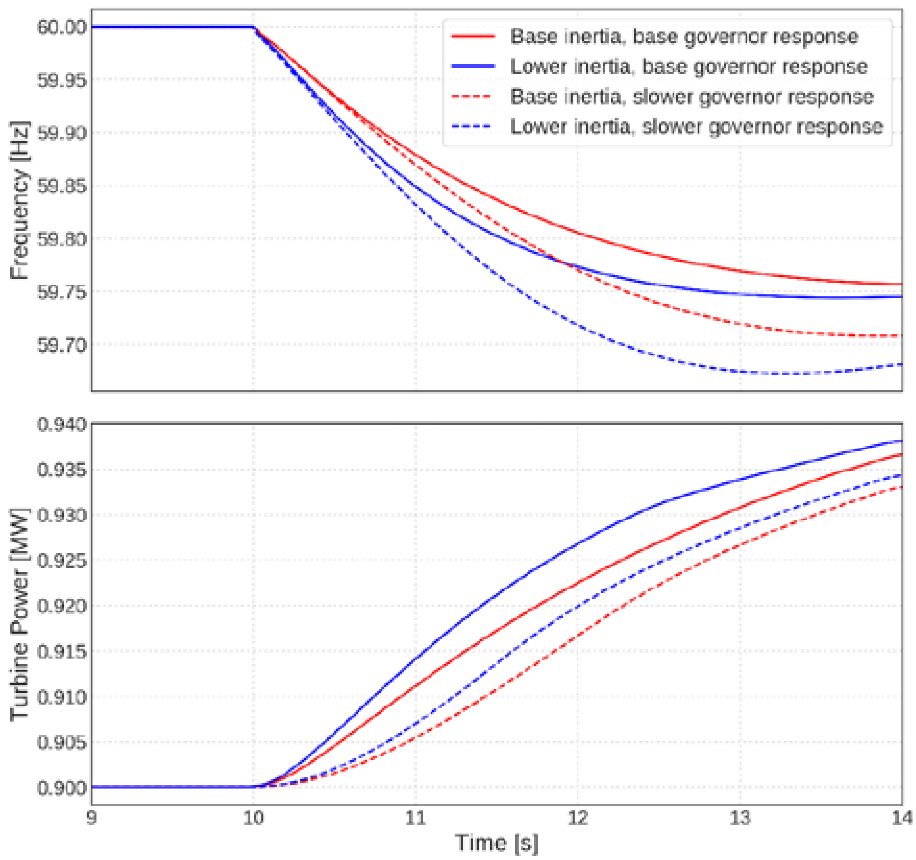


Figure ES - 3. System Inertia and the Speed of Primary Frequency Response Determine the Nadir at Which Frequency is Arrested

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

4. *Primary frequency response must be sustained until secondary frequency response can replace it.*

Much attention has been devoted to ensuring adequate primary frequency response over the initial seconds following the loss of generation. Due attention should also be devoted to ensuring primary frequency response is sustained during the period when frequency is stabilized following the formation of the nadir. During this period, primary frequency response is required in order to stabilize frequency. It must, therefore, be sustained until secondary frequency response can replace it. If, during this period, primary frequency response is not sustained and is reduced to less than the amount of generation lost, frequency will again decline and UFLS will be triggered. See Figure ES - 4. This point is less well appreciated but of equal importance for reliable interconnection frequency response.

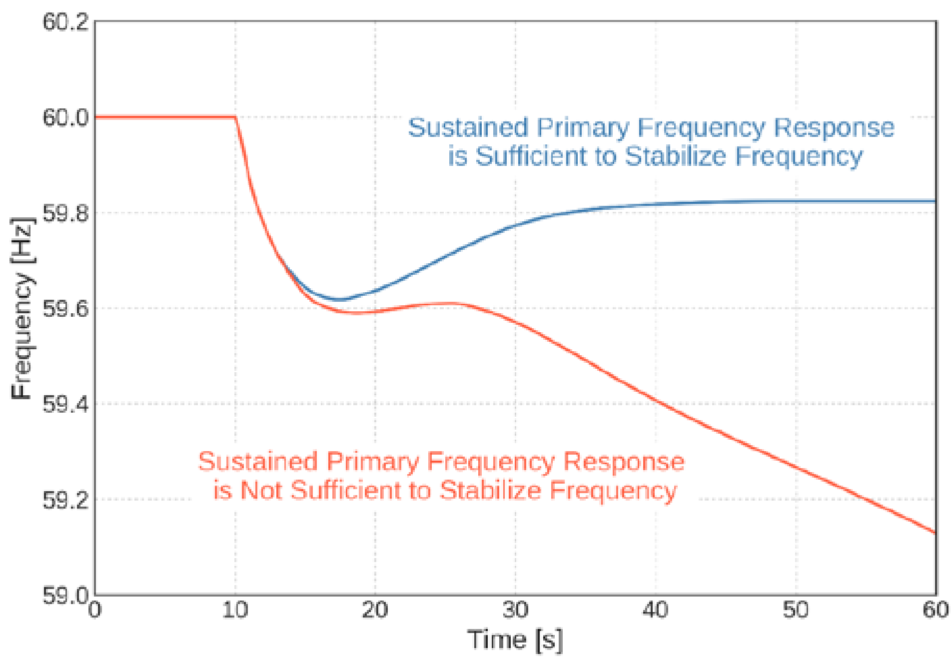


Figure ES - 4. Failure to Sustain Sufficient Primary Frequency Response Will Trigger UFLS

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

There are several means by which primary frequency response may not be sustained. The first is through withdrawal of primary frequency response by the actions of plant load controllers, which override and reset the actions of the turbine-governor responding to frequency deviations. We describe this finding first because it is an action that is directed by the plant owner/operator. As such, it is one that can be corrected by a plant owner/operator, which we discuss in Finding 5. The second is through actions stemming from inherent dynamic characteristics of turbine generators. One example is exhaust gas temperature limits on gas turbines, which are intrinsic to the current design of these types of turbines. Unlike plant-level controllers, these actions cannot be overridden or corrected. We conclude by discussing what we have observed in published information on wind turbines providing what is called “synthetic inertia.”

The bottom line is that, if primary frequency response from some sources will not be sustained, primary frequency response from additional sources will be required. The requirement is to stabilize frequency until primary frequency response can be replaced by secondary frequency response.

5. *Plant load controllers operated in pre-selected load mode without frequency bias will withdraw and not sustain primary frequency response.*

Plant load controllers establish the ranges around which turbine-governors operate in response to frequency. A logic commonly followed by these controllers seeks to maintain generation output in accordance with a pre-determined schedule. Consequently, when generation output increases because the turbine-governor responds to a decrease in interconnection frequency, the plant load controller overrides the turbine-governor and automatically restores output to the scheduled value. This results in primary frequency response being withdrawn, which negatively affects restoration of interconnection frequency.

6. *Plant load controllers operated in pre-selected load mode with frequency bias will sustain primary frequency response.*

The early withdrawal of primary frequency response by plant load controllers can be prevented by specifying a control logic that seeks to operate the generator at the scheduled value only when the frequency of the interconnection is at its normal operating value of 60 Hertz (Hz). That is, when frequency deviates significantly from 60 Hz, for example because of loss of generation on the interconnection, the plant load controller does not override the turbine-governor but instead allows the turbine-governor to continue delivering primary frequency response until system frequency returns to the nominal value. This control logic supports the restoration of interconnection frequency following a loss of generation.

7. *Gas turbines may not be able to sustain primary frequency response following large loss-of-generation events.*

Gas turbines are among the fastest responders that contribute to arresting the decline in system frequency following the sudden loss of generation. They can readily increase their output by a few percent of rated capacity within a handful of seconds (5 to 8 seconds). As a result, they are excellent initial sources of primary frequency response. However, if an under-frequency event calls for maximum output, this maximum will be less than would be reached when running at nominal frequency. Unlike the withdrawal of response by plant load controls, reduction of output is an action of the protection system of the turbine and cannot be deactivated at the discretion of the plant operator. There is feedback between these controls and system frequency that can be detrimental to reliable interconnection frequency response. That is, if, as exhaust gas temperature controls reduce turbine output, and system frequency continues to decline, then the temperature limit controls will further reduce turbine output. It is therefore essential to recognize this dependence of gas turbine maximum output on frequency and ensure that response is available from sources that will sustain or increase their response during the comparatively longer periods when system frequency may be depressed following large loss-of-generation events.

8. "Synthetic inertia" controls on electronically coupled wind generation appear not to sustain primary frequency response.

Inverter-based controls on electronically coupled generation sources (such as wind turbines, solar photovoltaics, and batteries) can provide sustained primary frequency response through a droop relationship in the same manner that turbine-governors operate in conventional power plants. In cases of low frequency, this requires reserving headroom from which the response is drawn. As an alternative, "synthetic inertia" controls are said to provide a form of primary frequency response without reserving headroom.

So-called "Synthetic inertia" controls on electronically coupled wind generation can have a similar impact as described above, in which primary frequency response is delivered but terminated before frequency is stabilized. This type of frequency response comes from the extraction of kinetic energy from spinning wind turbine blades. That is, the turbine blades slow down. Based on published information, the response, however, appears to be one that is not sustained and is, in effect, withdrawn within five to ten seconds. If this is unavoidable, then other sources of sustaining primary frequency response will be required that continue to stabilize frequency until secondary frequency response can replace them.

9. Fast demand response provides robust primary frequency response, but currently is inflexible.

Removing load immediately affects interconnection frequency and is therefore an effective strategy to restore the balance between load and generation after generation is lost. However, the amount of load shed—as well as the frequency at which it is shed and the time delay beforehand—must be established carefully in advance. If the amount of load shed is greater than the amount of generation that was lost, an over-frequency situation can result, which may also pose a severe challenge to system reliability.

10. Smaller deadbands on turbine-governors increase how quickly delivery of primary frequency response will begin.

Deadband settings on turbine-governors determine at what frequency deviation a generator will begin delivering primary frequency response. Smaller deadbands mean that a generator will respond to smaller generation-loss events than a generator with a larger deadband. Moreover, for larger generation-loss events, generators with smaller deadbands will begin responding sooner than generators with larger deadbands. Operating with smaller deadbands therefore will improve interconnection frequency response compared to operating with larger deadbands. Importantly, unequal deadbands among generators mean that those with smaller deadbands will begin responding sooner and more often than generators with larger deadbands. In the extreme, if a generator operates with a very large deadband (300 or 500 mHz), then the generator will only respond to the very largest generation-loss events (and may do so too late to avoid triggering UFLS). Such extreme deadband settings undermine the goal of providing frequency response because the turbine-governor will not respond to the vast majority of generation-loss events, which are smaller in size.

11. Load sensitivity currently complements primary frequency response, but this sensitivity may be going away.

Although a portion of load has traditionally reduced consumption autonomously in proportion to a decline in interconnection frequency and, hence, augmented the frequency response that generators provide, the characteristics of load are changing.[10] In particular, load that is electronically coupled to the grid using power electronic interfaces is increasingly common. These loads include variable-frequency drives on motors, fans, and pumps. These electronically controlled forms of load can work to the detriment of frequency control because they generally act to prevent the power drawn from declining as frequency declines. Directly coupled motors "slow down" when frequency declines and reduce power consumption, and thereby work in concert with primary frequency response delivered by generators. By not slowing down and not reducing power consumption, electronically coupled motors no longer contribute or support primary frequency response delivered by generators. This impact of electronic controllers on interconnection frequency response, however, is not a given. Electronic controllers can be programmed to support reliable interconnection frequency response.

Observations

Texas Interconnection

Frequency response is a significant issue for the Texas Interconnection, and has been managed proactively by ERCOT for some time. On a percentage basis, the design generation-loss event[11] for the Texas Interconnection—especially at times of minimum system load is nearly three times larger than the design generation-loss event at minimum system load for the Western Interconnection, and more than five times larger than the comparable event for the Eastern interconnection. (See Table ES - 1.) As a result, the design generation-loss event in the Texas Interconnection results in a much sharper and more rapid decline in frequency compared to the other two interconnections. (See Figure ES - 5.) These ROCOF values present a challenge for the frequency response obtainable from conventional power plants. Thus, ERCOT's reliance on a fast form of demand response is important, if not critical, for ensuring reliable interconnection frequency response.[12]

---

[10] Because of these considerations related to load, our initial set of simulations removed the effects of load in augmenting provision of primary frequency control. This enabled us to obtain direct insight into the role of generators in responding to frequency excursions and ensured that our findings would be conservative.
[11] The design generation-loss event is expressed as a percentage of the total system load at the time of the event.
[12] Future forms of fast demand response could be more flexible than the current form of demand response relied upon by ERCOT. One form, illustrated in Section 5.9, might involve shedding smaller blocks of load at different frequency set points. Another form, alluded to in Section 5.11, might involve load that could be varied continuously in response to frequency changes in a manner analogous to droop control of turbine-governors.

Table ES - 1. Interconnection Frequency Response Design Criteria at Times of Minimum System Load

Design Criteria:	Eastern Interconnection	Western Interconnection	Texas Interconnection
Generation-Loss Event	4.5 GW	2.7 GW	2.7 GW
Minimum Load – 2015	210 GW	64 GW	24 GW
Gen. Loss Event/Min Load	2.1 %	4.1 %	11.3 %

Source: Developed by LBNL from NERC (2017b): 2017 Frequency Response Annual Analysis; and Matevosyan (2016): Inertia Data

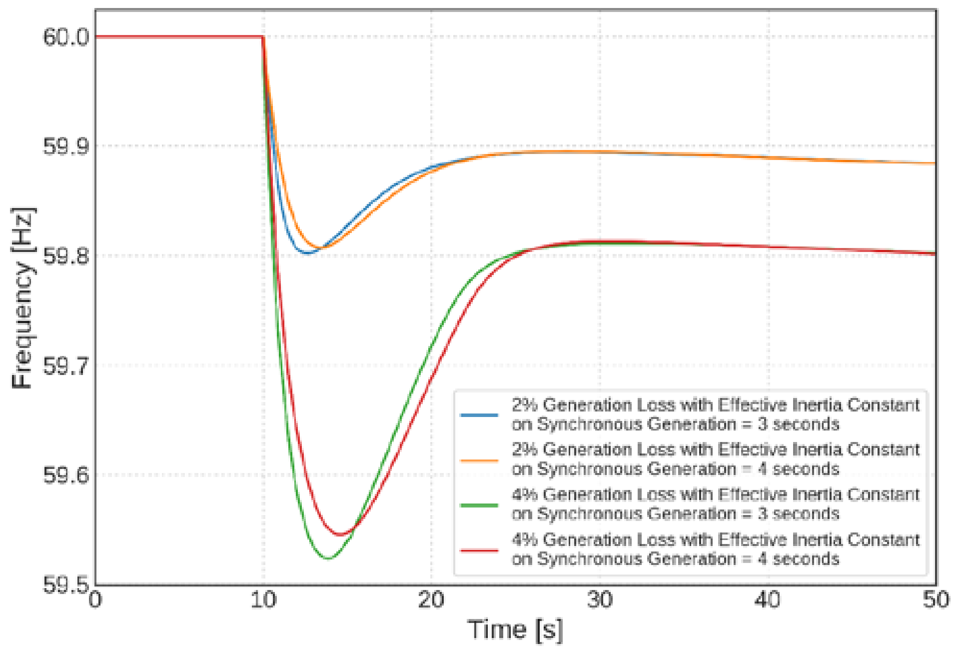


Figure ES - 5. The Relative Impacts of Generation Loss versus System Inertia on Frequency Nadir

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

Texas will likely lead the three U.S. interconnections in addressing the challenge of synchronous generating resources that provide primary frequency control being retired and replaced with generation from non-synchronous resources, such as wind and solar. One obvious solution is to find ways to engage the primary frequency control capabilities of the large and growing amount of wind and solar generation in the Texas Interconnection. However, doing so will require addressing the commercial arrangements that currently create strong financial incentives for wind and solar to operate without headroom. As noted above, "synthetic inertia," which, based on the literature, is a form of primary frequency control provided by wind generation without reserving headroom and is then quickly withdrawn is not a substitute for sustained primary frequency response.

Western Interconnection

Frequency response is also a long-studied issue in the Western Interconnection. The Western Electricity Coordinating Council (WECC)-specific version of the NERC BAL standard (BAL-002.WECC-2a) states that the Western Interconnection's spinning reserve requirements must be met by generation that is on-line and now clarifies that this generation must respond autonomously and automatically to changes in frequency. In other words, all spinning reserves in WECC must provide primary frequency response. In addition, the Western Interconnection has routine practices to update and validate the models used to support, among other things interconnection frequency response studies.

In the Western Interconnection, the geographic distribution of the reserves relied on for primary frequency control means that the transmission system plays a critical role in delivering frequency response. As discussed in LBNL's 2010 Study, reliance on the long-distance transmission system poses a risk to reliable interconnection frequency response. The risk stems from the need to ensure that sufficient reserve capability is available on the transmission lines to reliably deliver primary frequency response to the area where generation was lost. Failure to maintain sufficient transmission reserve capability increases the risk that primary frequency response cannot be delivered and that UFLS will be triggered. This risk will increase if older, thermal-based reserves that provide primary frequency control located in the Southwest retire, leading to even more reliance on hydro-based reserves in the Northwest.

Retiring generation will of course be replaced with newer forms of generation. The character (and location) of this new generation will affect interconnection frequency control. For example, if older thermal generation is replaced by combined-cycle,[13] wind, or solar generation that responds to frequency, primary frequency control capability in the Southwest could be maintained and could increase. The rules and incentives for generators to install, maintain, and make available primary frequency control capability will determine the outcome.

More recently, the Western Interconnection has experienced a new type of generation-loss event involving electronically coupled, inverter-based generators (solar PV). WECC and NERC are actively studying the implications of these events both for frequency response and other aspects of system control, including the means to address them (NERC 2017a).

Eastern Interconnection

Primary frequency control is a known issue in the Eastern Interconnection. This study's findings clarify that the comparatively large size of the Eastern Interconnection relative to the number of generation-loss events explains the lower ROCOF values observed for these events compared to the ROCOF values observed in other interconnections after losses of comparable amounts of generation. This study also clarifies how the large number of generators providing primary frequency response also contributes to

[13] As an aside, the inertia of the interconnection would increase if older thermal generation is replaced with combined-cycle gas plants. For a given size generation plant, the inertia of a combined-cycle gas plant is higher than that of nuclear-, coal-, or gas-fired steam plants. See Figure 12.

measured frequency response of the interconnection.

Two aspects of primary frequency control are currently the focus of attention in the Eastern Interconnection. First, the characteristic "lazy L" shape of frequency response in the Eastern Interconnection is widely recognized as being driven by withdrawal of primary frequency response by plant load controllers. See Figure ES - 6. As noted in Section 3 of this report, this explanation has been corroborated in modeling studies conducted by both NERC staff and GE. Our study has shown that withdrawal can have a detrimental effect on interconnection frequency response and that, when it is caused by plant load controls, it is can be remedied. Therefore, it is important that industry monitor and, as appropriate, implement interconnection and region-specific operating policies and procedures that prevent excessive withdrawal of primary frequency response.

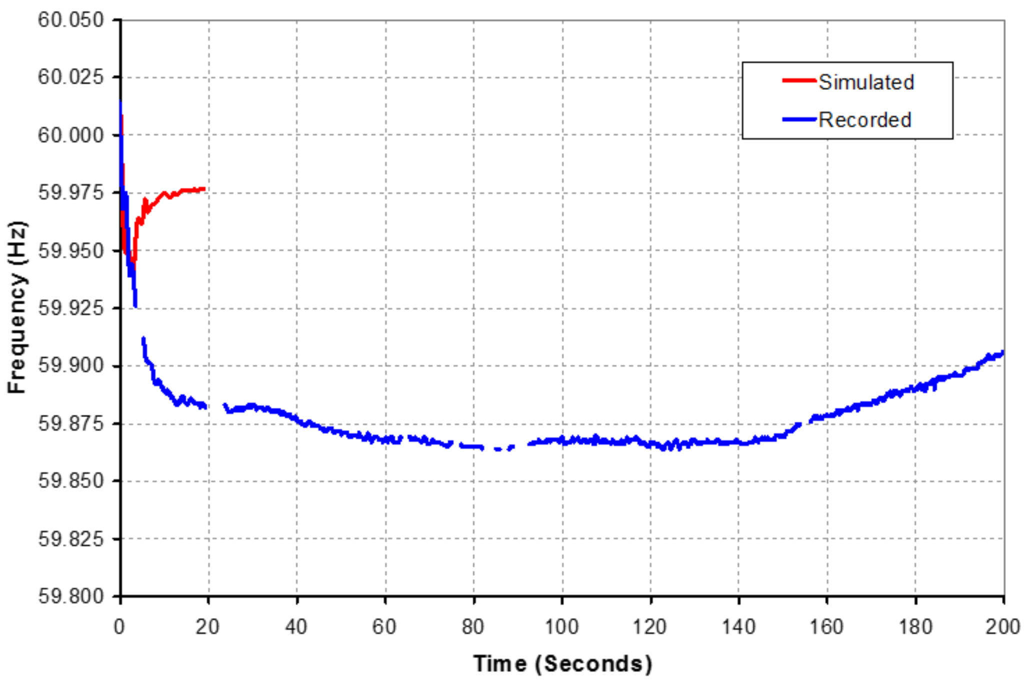


Figure ES - 6. Frequency of the Eastern Interconnection during the First 199 Seconds Following the Loss of 4,500 MW of Generation—A Comparison of Recorded Data with Results from a Simulation of the Event

Source: Eto, et al. (2010): Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation

Second, the importance of primary frequency response withdrawal for the Eastern Interconnection is also a motivation to continue to improve the ability of the interconnection's dynamic planning models to replicate and explain the interconnection's observed frequency response. As noted first in LBNL's 2010 Study and found again in Undrill, et al. (2018), the current Eastern Interconnection planning models do not reproduce the measured performance of the interconnection during a design generation-loss event (which is used to evaluate the adequacy of primary frequency control within the interconnection). We note that NERC staff is already working with planners in the Eastern

Interconnection to improve the quality of frequency response modeling (NERC 2017c).

Well-calibrated planning models are essential for assessing current performance and guiding modifications to interconnection agreements and operating policies to ensure continued, reliable interconnection frequency response. Continuous updating and ongoing calibration are essential for building confidence in applications of the models to study future scenarios involving changes in the mix of generation as well as load.

Recommendations

1. *Focused attention should be directed to understanding the aggregate frequency control performance required of the fleet of resources that must be kept on-line at all times to respond to generation-loss events. This will involve collection, maintenance, and validation of the data necessary for accurate planning and operating studies as well as collection of comprehensive data to measure trends in interconnection frequency control.*

The dynamic simulation tools and system models that the interconnections use to study frequency response must be based on accurate, up-to-date information about the actual characteristics of generators and load. This information should track not only interconnection loading, inertia, design generation-loss event, and highest set-point for UFLS, but also generator headroom, turbine-governor performance characteristics, and the number and location of resources for primary frequency control. Data are needed on the performance characteristics of non-traditional, non-governor-based resources for primary frequency response that indicate how much primary frequency response is available and how rapidly the response can be delivered. In the case of fast demand response, such as ERCOT's Load Resources (an element of its Response Reserves Service), it is important to study the size of load blocks and the triggering conditions for them. Performance measures should apply equally to traditional and non-traditional resources. For all resources, this should entail explicit performance measures that assess the factors that might withdraw primary frequency response early or cause it to not be sustained.[14] Simulation-based or other forms of study should consider the full period over which primary frequency response must be sustained—which may be as long as several minutes—and determine the rate at which non-sustaining response must be replaced in order to ensure reliable interconnection frequency response. Studies should examine worst-case situations involving either or both times of low system inertia and times when reserves of primary frequency control may be low.

Routine, comprehensive measurement of interconnection frequency control performance is essential for tracking trends.[15] This will require ongoing updating and verification of the performance of generators and other resources for primary frequency response as well as the conditions of the interconnection during which these resources are called upon. To the extent feasible, measurements should form the basis for the information used to model and plan for procurement and dispatch of resources that provide primary frequency response. As an example, performance measurements during

[14] NERC's ERSWG has developed and is currently tracking measures that seek to address this issue. See NERC 2015a.
[15] In fact, NERC has become compiling and now regularly publishing this information. See, for example, NERC 2017b.

actual events should be the bases for establishing limits on procuring primary frequency response.[16] This includes ensuring that modeling assumptions regarding primary frequency response capability are reflective of actual dispatch and power plant operating practices. In addition to tracking traditional measures of frequency response (such as interconnection frequency response), this process should document the conditions under which these measurements are made, such as the state of the power system (its loading and inertia, and whether load, and hence generation, is increasing or decreasing at the time generation is lost) and the size and location of generation-loss events relative to the performance of the primary frequency response resources, including the extent to which they sustain provision of primary frequency response.

2. *International practices should be reviewed as options for U.S. grid operators to consider adopting to ensure continued reliable interconnection frequency response.*

As the fleet of U.S. generation and the characteristics of load change, we must assess our approaches to frequency control to ensure that they continue to support reliable interconnection frequency response. Gaps, conflicts, and disincentives must be identified, analyzed, and addressed as appropriate.

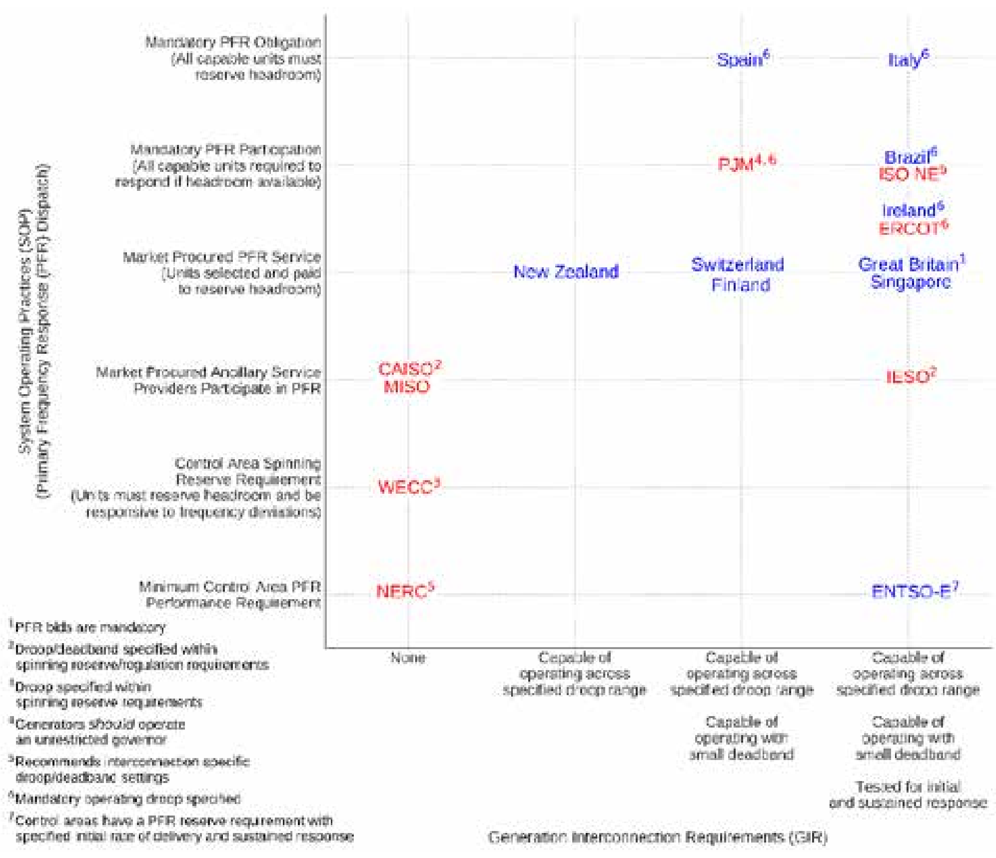


Figure ES - 7. Comparison of Selected U.S. and International Grid Codes Related to Frequency Response

Source: Roberts (2018): Review of International Grid Codes

---

16 ERCOT, in fact, routinely conducts these measurements.

Our review of international frequency control practices spans a range of approaches that represent functioning alternatives to or variants of current U.S. approaches. Because of the demonstrated success of these approaches in other power systems, they should be reviewed, analyzed in current and expected future operating conditions in the United States, and then given due consideration for adoption, as is, or in modified form. See Figure ES - 7.

3. *All generators, to the extent feasible, should be capable of providing sustained primary frequency response.*

Reliable interconnection frequency response requires participation by many generators. Ensuring that as many generators as is technically feasible are capable of providing sustaining response provides maximum flexibility to grid operators to assign primary frequency response duty as appropriate for current grid operating conditions.

Moreover, reliability of the interconnections is enhanced by enabling this capability on all generators capable of providing sustained primary frequency response. Doing so increases the pool of responding generators and reserves of primary frequency response, and thereby reduces the risk of unforeseen shortages of primary frequency response. It is recognized that some generators will not contribute if they are already dispatched at maximum capacity and hence do not have headroom available.

4. *Barriers to adding a frequency bias[17] to plant load controllers should be evaluated and addressed.*

This study describes the detrimental effects of early withdrawal of primary frequency response by plant load controllers. We also describe how early withdrawal of primary frequency response by plant load controllers can be prevented by introducing a frequency bias to the control logic of pre-selected load mode controls. We also recognize that some U.S. grid operators already require or have performance requirements that support the use of these controls. Still, others in the United States do not.

> *"...we recommend education and outreach as a minimum first step toward encouraging wider adoption of this control approach. In addition, it is important to understand and address any financial disincentives that would reinforce current practices."*

Anecdotally, we perceive that awareness of the efficacy of this alternative control logic is limited within the generator community. Accordingly, we recommend continued but expanded education and outreach to foster wider adoption of this control approach.[18] In addition, it is important to understand and address any financial disincentives that would reinforce current practices.

[17] This use of the term "frequency bias" is distinct from the use of this same term in the Area Control Error equation that guides automatic generation control, which is a form of secondary frequency control.
[18] NERC (2015a) is a good, initial example of this approach.

5. *The contributions of non-traditional resources for primary frequency control (demand response, energy storage, and other forms of electronically coupled loads and generation, including wind and solar photovoltaic) should be studied and incorporated, as appropriate, into future operations.*

One future change in the makeup of the generation fleet is that traditional resources for primary frequency response may retire and be replaced by non-traditional resources, including demand response, energy storage, and other forms of electronically coupled loads and generation such as wind and solar photovoltaics. The performance characteristics of non-traditional resources are not widely known or fully understood. Future frequency response-related operating and planning policies and decisions should be based on up-to-date, accurate information about the performance and potential contributions of these resources.[19] Research, development, and demonstration are also needed to improve the performance capabilities of these new sources and to support timely industry adoption.

6. *Factors that negatively influence the sensitivity of loads to frequency should be studied and addressed.*

Load sensitivity historically complemented primary frequency response from generators. However, this sensitivity appears to be disappearing as newer forms of load are electronically coupled to the grid using power electronic interfaces, which currently do not reduce power consumption when frequency deviates from nominal. These forms of load include variable-frequency drives on motors, fans, and pumps. Better information is needed on how the frequency support provided by load changes over the course of a day and seasonally.

Still, no inherent technical limitations prevent power electronic interfaces from supporting primary frequency response by generators. In many instances, a simple firmware upgrade of power electronics controls is all that is required. The technical and commercial reasons that current controls do not provide primary frequency response should be understood and, where appropriate and feasible, modified or adjusted so that future loads will work in concert with and support primary frequency response from generators.

[19] NERC's Inverter Based Resource Performance Task Force may be one source for this information.

1. Introduction

The reliability of interconnected electric power systems depends on controlling power system frequency so that it remains within pre-established, safe operating bounds. Reliability is threatened when a large electric generator(s) disconnects from the power system because the loss of generation causes an immediate decline in power system frequency. If the loss of generation is large enough and the remaining, still-connected generators do not respond and rapidly arrest the decline in frequency, power system frequency may decline below established, safe operating bounds and trigger automatic, emergency load shedding to avoid a cascading blackout.

The collective ability of the power system to respond to such events is called interconnection frequency response. Advance planning is required to operate the power system in a manner that ensures reliable frequency response at all times because generation-loss events are always unpredictable even though they occur relatively often.[20]

The Federal Energy Regulatory Commission (FERC) has tasked Lawrence Berkeley National Laboratory (LBNL) to conduct this study to support ongoing FERC and industry efforts to ensure reliable interconnection frequency response for the three major interconnections in the United States: the Western, Eastern, and Texas Interconnections.[21] See Figure 1. The purpose of this study is to support policymaker and industry understanding of the physical requirements for reliable interconnection frequency response by building upon an initial study conducted by LBNL for FERC in 2010.[22]

This report is organized as follows:

In Section 2, we introduce basic frequency control concepts and the terminology used throughout this report. This discussion expands on the material presented in LBNL's 2010 Study by addressing the relationship among three key technical concepts: (1) System inertia and the relative size of generation-loss events; (2) the dynamic response characteristics of the primary controls of on-line resources; and (3) the actions of secondary controls, including withdrawal of primary frequency response.

In Section 3, we review the principal interconnection frequency response-related findings from LBNL's 2010 Study and summarize industry activities that have taken place since the publication of that Study. Viewing ongoing industry activities in light of LBNL's 2010 Study findings provide the basis and motivation for the analysis and discussion presented in the current report.

[20] Given the large number of generators in the three U.S. interconnections, generation-loss events of varying sizes take place routinely, on a weekly, if not more frequent, basis. The very largest events, however, are considerably less frequent and rarely take place more than once a year.

[21] Throughout this report, we refer to the interconnection operated by the Electric Reliability Council of Texas (ERCOT) as the "Texas Interconnection."

[22] In 2010, FERC commissioned LBNL to study the use of frequency response metrics for assessing the reliable integration of variable renewable sources of electricity generation. LBNL prepared a technical report supported by five stand-alone technical appendices (Eto et al. 2010, Undrill 2010, Martinez et al. 2010, Illian 2010, Mackin et al. 2010, and Coughlin, Eto 2010).22 All reports are available at: https://certs.lbl.gov/project/integration-variable-renewable-generation. This study will be referred to in this report as "LBNL's 2010 Study."

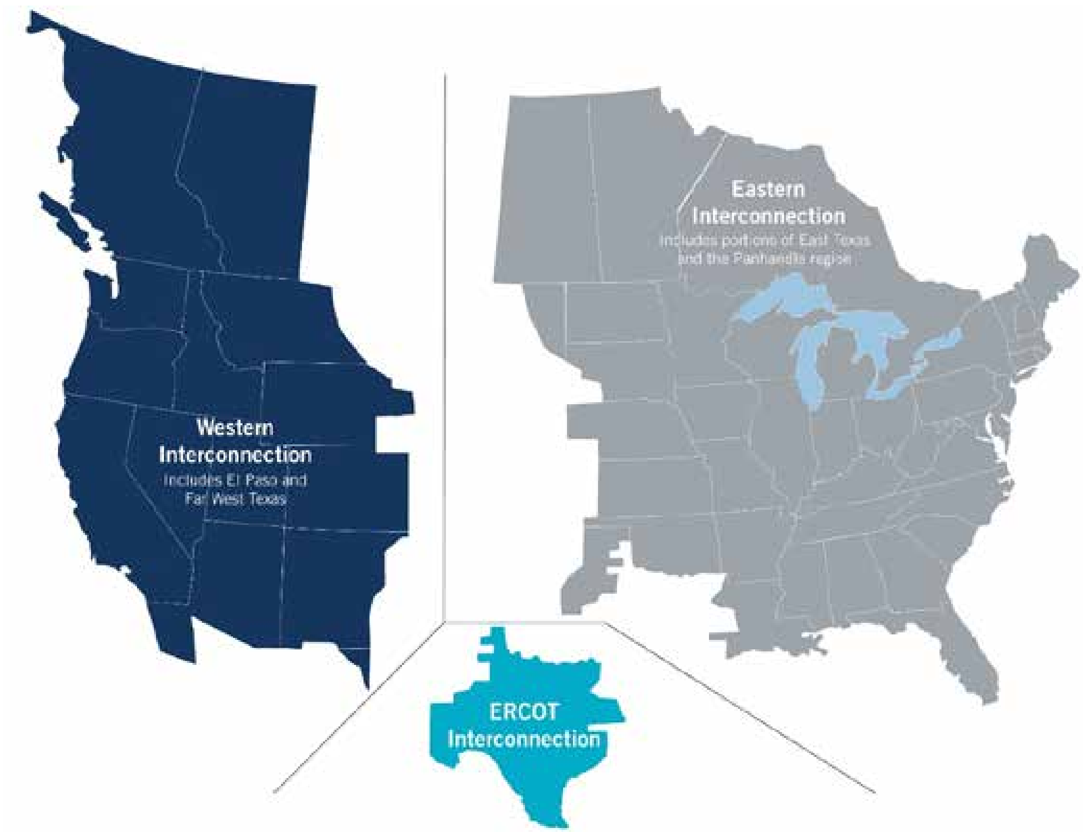


Figure 1. North American Electric Reliability Council Interconnections

Source: http://www.ercot.com/content/wcm/landing_pages/89373/ERCOT-Internconnection_Branded.jpg

In Section 4, we describe the modeling approach we developed to examine how the technical concepts discussed in Section 2 lead to the requirements for reliable frequency response. We use the modeling approach to explore how the requirements can be met by various amounts, types, and combinations of generator technologies. The modeling approach allows us to study systemically some important but previously not well understood aspects of frequency control. This study entails understanding factors that jointly determine whether frequency response will be sustained after an initial rapid decline in frequency has been arrested. These factors include turbine-generator-specific limits on the headroom accessible by generators that provide primary frequency response and plant-specific controls that limit or withdraw frequency response early. To our knowledge, these factors have not been studied and reported systematically by industry. To establish the basis and boundaries of our modeling work, we reference measured performance and system modeling information developed by industry for each of the three U.S. interconnections.

In Section 5, we present technical findings regarding the requirements for reliable interconnection frequency response and factors that can affect the quality of frequency control. Each finding is illustrated with results from the simulation studies we conducted using the modeling approach described in Section 4.

In Section 6, we apply our findings and describe the frequency control requirements and issues facing each of the three U.S. interconnections. We conclude Section 6 with recommendations for actions and future study.

The main report is accompanied by three standalone supporting technical reports. The first technical report reviews international grid codes related to frequency control, including an initial comparison of these codes to their counterparts in the United States (Roberts 2018). The second technical report provides complete technical documentation for the modeling approach described in Section 4 (Undrill 2018). It also presents the simulation results that form the basis for the findings in Section 5. The third technical report compares aspects of the modeling approach used in this study to results from the system models and modeling approaches currently used by planners in each of the three U.S. interconnections to conduct studies of, among other things, frequency response (Undrill et al. 2018).

2. Review of Frequency Control Concepts and Terminology

This section introduces the frequency control concepts and terminology used throughout this report. We first provide details and insights regarding the two physical factors that determine the requirements for reliable interconnection frequency response: interconnection system inertia and the size of the generation-loss events that an interconnection is designed and operated to withstand. Second, we describe in detail the principal means by which frequency control requirements are currently met through the actions of generator turbine-governors. Third, and of special importance for this report, we review the ways that reliability can be compromised by secondary plant load controls when they prematurely override turbine-governor responses.

In addition to the LBNL's 2010 Study, this section draws on textbook references and recent technical reports (Kirchmeyer 1959; Cohn 1971; NERC 2009; Undrill 2010; and Undrill 2018). Throughout this section, technical terms that are defined in the glossary at the end of this report are denoted in *italics* when they are first introduced.

2.1 System Frequency Reflects the Balance between Generation and Load

Frequency *refers to the number of cycles of an alternating current voltage waveform per unit of time (seconds). Frequency in North America is normally 60 cycles per second or 60 Hertz (Hz).*

The instantaneous balance between generation and load is directly reflected in an interconnected electric power system's *frequency*. Reliable power system operation depends on controlling frequency within predetermined boundaries above and below a scheduled value. In North America this value is normally 60 cycles per second or 60 Hertz (Hz). Failure to maintain frequency within these boundaries can disrupt the operation of customers' equipment, initiate disconnection of power plant equipment to prevent it from being damaged, and lead to widespread blackouts.

Figure 2 illustrates how the relationship between generation and load determines the frequency of an electric power system using the analogy of water following into and out of a container. If generation and load are exactly in balance (water inflow and outflow are equal), frequency is stable at 60 Hz. If generation begins to exceed load (inflow begins to exceed outflow), frequency will rise above 60 Hz. If load begins to exceed generation (outflow begins to exceed inflow), frequency will fall below 60 Hz. If, in this last example, generation is not increased (to match the outflow), then frequency (water level) will fall until the power system collapses (the water in the container is depleted).

Controlling frequency to remain at or very close to a scheduled value is challenging because load varies continuously following well-understood patterns. In addition, generation sometimes varies abruptly as a result of unplanned events such as the sudden loss of a generator. Both situations change the balance between load and generation, causing frequency to deviate from its scheduled value.

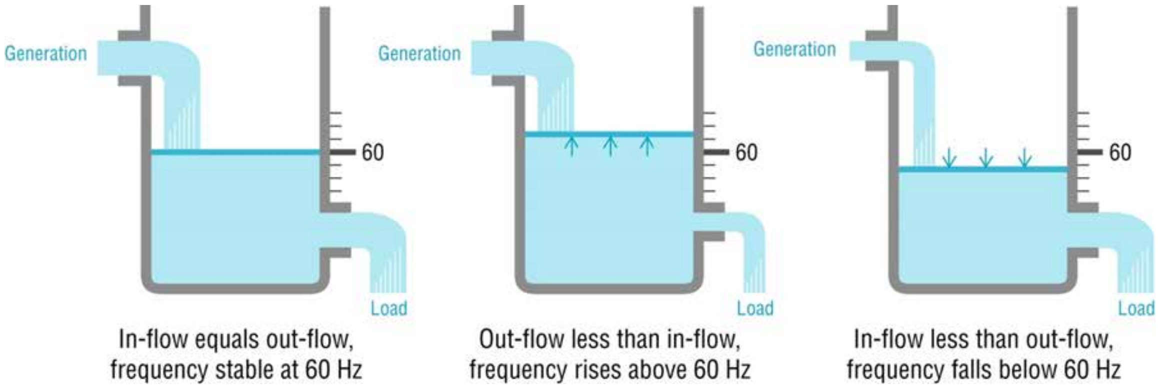


Figure 2. The Concept of Power System Frequency Explained Using the Analogy of Water Level in a Container

Source: Eto et al. (2010): Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation

"The objective of power system frequency control is to ensure that frequency remains within predetermined boundaries around the scheduled value at all times."

Power system operators are responsible for ensuring that adequate resources are available to respond to imbalances and restore frequency to its scheduled value, both when those imbalances are expected and when they are unexpected, and especially when they are large. That is, the objective of power system frequency control is to ensure that frequency remains within predetermined boundaries around the scheduled value at all times. Failure to maintain frequency within this range is a threat to reliability because it can initiate a widespread cascading blackout.

2.2 Power System Frequency is Controlled by Resources that Adjust Their Output to Oppose Deviations in Frequency from 60 Hz

Power system operators control frequency mainly through automatic and manual adjustments to the generators' output;[23] the goal of these adjustments is to maintain the balance between generation and load and manage frequency to a scheduled value (normally 60 Hz as explained above). When frequency is above the scheduled value, operators rely on generators to decrease their output. When frequency is below the scheduled value, operators rely on generators to increase their output. The generators' actions are sometimes referred to as "opposing the change in frequency" (i.e., generator actions are meant to counter an upward or downward shift to restore frequency to the scheduled value).

Generator output adjustments to control frequency are defined as primary and secondary frequency control. The amount by which a generator changes its output to provide these forms of control is called primary or secondary frequency response, respectively.

[23] Specialized forms of demand response sometimes provide frequency control (notably, primary frequency control within the Texas Interconnection). For ease of exposition, the introductory discussion in this section focuses on the dominant role that generators play as the source of frequency control actions to ensure reliability on today's power grid. However, the concepts introduced here apply with equal generality to demand-side resources that can respond in time frames comparable to those of generation resources.

Primary frequency control consists of automatic, autonomous, and rapid (within seconds) changes in a generator's output to oppose changes in frequency. Primary frequency control is especially important during the period immediately following the unexpected loss of a large generator. This is because primary frequency response must be initiated within seconds. Failure to rapidly oppose and arrest the change in frequency will trigger under-frequency load shedding (or UFLS, as described in Section 2.5) and can lead to a blackout.

To participate in primary frequency control, a generator must be "on line" (i.e., synchronized with the interconnection) and dispatched so that it can change its output or maneuver to oppose the change in frequency. To illustrate: for a generator to be able to respond to a loss-of-generation event, it must be dispatched at less than its maximum operating capability and thus have capacity available to provide for frequency control. The term *headroom* is used to describe the difference between a generator's current operating point and the generator's maximum operating capability. The headroom available determines the total amount of power a generator is able to deliver to oppose a decline in frequency. If a generator is dispatched at its maximum operating point, then it has no headroom available from which it can deliver primary frequency response to oppose a decline in frequency. Note that headroom, as defined here, represents an upper bound on the amount of increased generation available to respond to a sudden decline in frequency. Whether the full amount of available headroom will, in fact, contribute to opposing a frequency decline is determined by considerations we discuss in Section 2.5.

> *"A governor acts in a manner that is analogous to the 'cruise-control' of an automobile; its actions are both automatic and autonomous. That is, they do not depend on external commands or on the actions of other generators)."*

Generators participate in primary frequency control through the actions of a *governor*. A governor acts in a manner that is analogous to the "cruise-control" of an automobile; its actions are both automatic and autonomous. That is, they do not depend on external commands or on the actions of other generators. When the frequency of an interconnection falls below 60 Hz (i.e., the interconnection "slows down"), the governor increases the generator's power output. Conversely, when the frequency of an interconnection increases above 60 Hz (i.e., the interconnection "speeds up"), the governor reduces the generator's power output.

The control logic followed by a governor is specified using a concept called *droop* or droop curve. The increase (and decrease) in generator output is directly proportional to the deviation in frequency from 60 Hz. This means that the farther below 60 Hz the frequency falls, the greater the increase in output directed by the governor. It also means that as frequency is restored towards 60 Hz, the governor will direct proportionally lower increases in output. Once frequency is restored to 60 Hz, the governor will withdraw its response fully and direct no change in output with respect to its present set-point.

The electric power industry generally operates governors with four or five-percent droop. A droop of five percent means that, to oppose a change in interconnection frequency, a unit's output will change in proportion to the total capacity of the unit divided by five percent of the nominal frequency of 60

Hz.[24] Ultimately, the rate at which a generator increases (or decreases) its output in response to frequency changes is determined by the mechanical properties of the generator's turbine. Therefore, the term *turbine-governor* is used to clarify the relationship between a governor and the type of turbine it controls. For example, a governor on a hydro-electric turbine (a hydro turbine-governor) will increase the generator's output more slowly than a governor on steam-electric turbine (a steam turbine-governor) because the working fluid in a hydro-electric turbine (water in liquid form) has a density many times greater than the working fluid in a steam-electric turbine (water in a gaseous form).

Throughout this report, we will refer to the "quality" of primary frequency response when discussing the speed or rate at which this response is delivered.[25] Sometimes, we will also refer to the "quantity" of primary frequency response when discussing the amount of power (MW) delivered. Both the droop setting and, as noted earlier, generator headroom are relevant to discussions of these aspects of primary frequency response. In Section 2.5, we will discuss practical limits to the amount of headroom that can be drawn upon as primary frequency response.

It is also important to note that, historically, delivery of primary frequency response from generators has been augmented by load responses (i.e., reduction of electricity consumption in proportion to the decline in interconnection frequency). This characteristic of loads is called *load sensitivity,* or sometimes *load damping*. Load sensitivity depends on the devices consuming electricity at the time of the generation-loss event, so its contribution will vary depending on time of day, week, or season when the generation-loss event takes place. The concern today is that load sensitivity has always been beneficial to interconnection frequency response because it supports or augments primary frequency response from generators. Today, there is evidence that load's contributions are declining as end-use equipment, such as motor drives, is increasingly electronically coupled, rather than direct-coupled, to the grid.[26]

Secondary frequency control consists of control actions that are slower than primary frequency control. Secondary control actions change a generator's output to restore or maintain reserves that were expended or held to deliver primary frequency response. As with primary frequency control, a generator providing secondary frequency control must have headroom so that it can maneuver (change its output) to deliver secondary frequency response.

Secondary frequency response is delivered more slowly than primary frequency response because secondary frequency control actions are directed through commands that are external to the control actions of turbine-governors. These commands change the operating set-point of the turbine-governor.

---

[24] For example, a 300-megawatt (MW) unit with a 5 percent droop will change its output proportionally at 100 MW/Hz (=300 MW/(.05*60 Hz)). If system frequency falls to 59.8 Hz, the governor will direct an increase in generator output of 20 MW (= (60 – 59.8 Hz) * 100 MW/Hz). If system frequency falls to 59.6 Hz, the governor will direct an increase in generator output of 40 MW (= (60 – 59.6 Hz) * 100 MW/Hz).

[25] LBNL's 2010 study developed new metrics to describe these aspects of primary frequency response. See Section 3.1.

[26] This topic is explored further in Section 5.11.

Secondary frequency control directions are sent to the turbine-governor through a variety of generator controls that we refer to collectively as *plant load controls* (or *controllers*). For example, *automatic generation control* (AGC) is an automated form of secondary frequency control.[27] AGC sends signals from a centralized control system every 2 to 6 seconds to adjust a generator's output. These signals typically come from a Balancing Authority to the plant load controller of a generator via an even higher level control system that supervises multiple generators. See Text Box.

Secondary Frequency Control

This report describes, in a generalized manner, interactions among primary and secondary forms of control for reliable management of interconnection frequency. We focus primarily on the distinction between the turbine-governor and the plant load controller. The turbine-governor implements all changes in a generator's output—both automatic (primary control) and externally directed (secondary control). The plant load controller is the source of all externally directed, secondary control actions. These controls and their interactions are required to manage a generator's output.

Generating stations that consist of multiple generators may also have higher-level controls that coordinate and direct the actions of the plant load controls on individual generators. Automatic generation control (or AGC), as described in the text, is an example of secondary frequency control and illustrates how these higher-level controls are sometimes implemented.

AGC signals sent to a generating station initiate an increase or decrease in the amount of generation as specified by contract. Receipt of an AGC signal by a generating station triggers pre-programmed decisions regarding which generators within the station will be directed to fulfill the requirement. The amounts can and often will vary by generator, with some contributing little or no generation and others fulfilling the bulk of the requirement.

The term "secondary control" refers to all generator-control objectives that are implemented by plant load controllers changing the operating set-point of a turbine-governor. Secondary frequency control is only one kind of secondary control exercised by plant load controllers; it is initiated to meet the system-wide objective of interconnection frequency management. In fact, the principal purpose of plant load controllers is to adjust a generator's output in response to local objectives, namely the generation of electricity following a commercial schedule.

Local objectives can conflict with and override system-wide objectives. For example, following a loss-of-generation event, primary frequency control will immediately increase a generator's output to oppose the sudden decline in interconnection frequency. If the plant load controller then directs the generator to reduce its output again in order to return to the previously scheduled output, this will override and withdraw the primary frequency response delivered by the generator. This action of the plant load controller following local control objectives will be detrimental to interconnection frequency response.[28]

[27] Some would argue that the purpose of AGC is not to manage frequency, per se, but instead to manage flows across inter-ties in addition to using frequency deviations as a control signal to address generation imbalances among the balancing authorities within an interconnection.
[28] The impact of these conflicts on interconnection frequency control is illustrated with simulation results in Section 5.5.

2.3 The Sudden Loss of a Large Amount of Generation is the Most Significant Threat to the Reliable Management of Interconnection Frequency

Imbalances caused by losses of a large amount of generation are a special concern for the reliable management of interconnection frequency because these events are sudden and unpredictable. The effect of the loss of a large amount of generation is felt within a very few seconds throughout an interconnection as an immediate decline in system frequency. If no corrective actions are taken, frequency will decline until the power system collapses and a cascading, widespread blackout ensues.

Unpredictable generation-loss events take place with some regularity. The events are commonly the result of a mechanical or electrical problem within a large generating plant that causes the plant to disconnect itself automatically from the grid. How often these events occur depends on the size of, and therefore the number of generators within, an interconnection. In the very large Eastern Interconnection, events are recorded almost daily. In the much smaller Texas Interconnection, events are recorded on average about once every week. The majority of these events are relatively small when compared to the size of the interconnection. The largest events, which pose the greatest threats to reliability, are very infrequent. The largest single loss-of-generation event in the Eastern Interconnection took place more than 10 years ago in 2007.[29]

To respond to generation-loss events that can occur at any time, system operators must maintain generation reserves for primary and secondary frequency control at all times. From this perspective, the principal objective of interconnection frequency control is to ensure the continued delivery of electricity following these events.

"Power system operations planners routinely conduct studies to ensure that their frequency-control reserves can always arrest and restore frequency following the unplanned loss of generation."

Power system operations planners routinely conduct studies to ensure that their frequency-control reserves can always arrest and restore frequency following the unplanned loss of generation. Because these generation-loss events are unpredictable, conservatism guides these studies and they therefore focus on worst-case conditions. This entails studying the interconnection's frequency response to severe hypothetical, yet plausible, generation-loss events. In the Western and Texas Interconnections, these events are based on the largest N-2 loss of resource event. In the Eastern Interconnection, it is based on the worst (largest) generation-loss event ever recorded in the interconnection, which was the aforementioned one in 2007. See NERC (2012).

[29] The August 14, 2003 U.S.-Canada blackout, by contrast, involved many distinct, loss-of-generation events spread over a wide area.

2.4 The Sequence of Frequency Control Actions Following a Sudden Loss of Generation

To understand the requirements for reliable interconnection frequency response, it is important to understand the sequence of frequency control actions needed to respond to and recover from a sudden loss of generation. Figure 3 illustrates the evolution of interconnection frequency following a sudden loss of generation (top panel) and the three types or stages of frequency control actions that are sequentially engaged to manage or control frequency during this period (bottom panel).

In the first stage, the free-fall of frequency must be arrested quickly to prevent the system from collapsing. This stage is labeled the "arresting period." In the second stage, frequency is stabilized, typically at or above the value at which it was arrested but still at less than the original scheduled value. This stage is labeled the "rebound period." In the third stage, frequency is restored to its scheduled value, and the reserves held to provide primary and secondary frequency control are restored. This stage is labeled the "recovery period."

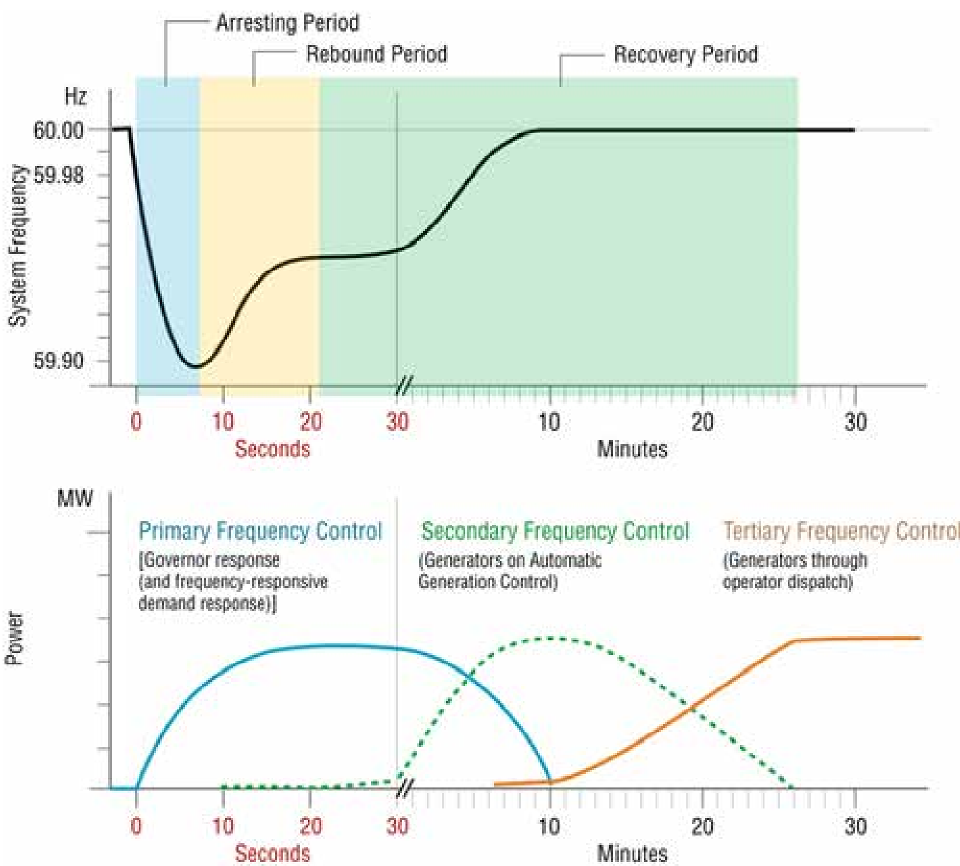


Figure 3. The Sequential Actions and Impacts on System Frequency of Primary, Secondary, and Tertiary Frequency Control

Note: Load sensitivity is not shown in this figure. See discussion in Section 2.2

Source: Eto, et al. (2010): Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation

In the first stage, as noted previously, primary frequency control is the only action capable of opposing and arresting decline of frequency quickly enough to prevent a blackout. The reserves held for primary frequency control are determined conservatively and must exceed the minimum amount required to arrest frequency.[30] As illustrated in the bottom panel of Figure 3, frequency is arrested before the reserves for primary frequency control are fully deployed. The point at which frequency decline is arrested is called the *frequency nadir*.

In the second stage, the generators participating in primary frequency control reach the increased output levels directed by their governors according to their droop. The collective primary frequency response from these generators establishes a new balance between load and generation. The point at which frequency is stabilized is called the *settling frequency*.

Frequency is always stabilized at a value lower than the original scheduled frequency. This is an expected and necessary consequence of primary frequency response delivered via droop. Recall that droop changes generator output in proportion to the deviation of frequency from the original scheduled value.[31] Therefore, primary frequency response from turbine-governors alone cannot restore frequency to its original scheduled value because governor-directed changes in generator output depend on the deviation of frequency from that original scheduled value.

Thus, after primary frequency control has arrested and stabilized frequency, the goal of secondary frequency control is to return frequency to the scheduled value by means of AGC and other manual secondary frequency control actions directed by system operators—such as deploying contingency reserves, including demand response, and establishing emergency interchange schedules. As noted earlier, secondary frequency control actions take longer to initiate than primary frequency control actions that are initiated automatically once frequency deviates from 60 Hz. For example, in the case of AGC, secondary frequency control is initiated by external automated commands sent every 2 to 6 seconds. As with primary frequency response, once changes in generation output are initiated, they are determined by the rate at which turbine-governors can direct them. In the case of operator-directed secondary frequency actions, the commands take even longer to initiate.

As a result, secondary frequency response does not contribute materially to the control of frequency until 30 seconds or more following a loss of generation. In fact, secondary response can take anywhere from about 5 to 15 minutes (and sometimes more) to complete the restoration of frequency to its original, scheduled value. All through this period, delivery of primary frequency response continues until frequency is restored to the original scheduled value. It is therefore critical to understand and

[30] The reason that reserves of primary frequency response must exceed the amount of generation represented by the design generation-loss event is discussed in Section 5.1.

[31] In fact, subtle exchange also takes place during this stage among generators participating in primary frequency control because of the droop settings on governors and the speed with which turbine-governors can change the output of a generator. During the rebound period, faster responding generators will decrease their output because the deviation in frequency from 60 Hz is less than the deviation at the point frequency is arrested. This decrease is offset by increased generation from slower responding generators that have not yet reached the maximum generation increase that their droop calls for.

recognize that the sustained delivery of primary frequency response is essential throughout the recovery period to ensure system reliability. Sustained delivery is essential for stabilizing frequency during the time required for secondary frequency control to restore frequency to the original scheduled value.

Tertiary frequency control *consists of centrally coordinated actions to re-dispatch generation that take place on longer time scales than AGC (i.e., minutes to tens of minutes, rather than every 2 to 6 seconds).*

Tertiary frequency control consists of centrally coordinated actions to re-dispatch generation that take place on longer time scales than AGC (i.e., minutes to tens of minutes, rather than every 2 to 6 seconds). It is a "manual" form of what we have called secondary control that is implemented by plant-level controllers. The goal of tertiary control actions is to restore the reserves that were used to deliver secondary frequency response during the recovery period. Restoring these reserves completes the repositioning of the power system so that it is now prepared to respond to a subsequent loss-of-generation event.[32] Tertiary frequency control actions involve coordinated changes in generating unit loading and commitment (e.g., dispatching one generator down to restore its reserve capability while simultaneously dispatching another generator up to replace the power provided by the first generator, all the while maintaining system frequency) and establishing new interchange schedules, which may commit other units in other balancing areas.

2.5 The Criterion that Determines whether Interconnection Frequency Control is Reliable is Avoiding Triggering Under-Frequency Load Shedding

Under-Frequency Load-Shedding (UFLS) *is an extreme measure to arrest frequency decline that disconnects large, pre-set groups of customers at predetermined frequency set points.*

If the amount of generation reserved for primary frequency control is unable to arrest the decline in frequency (that is, frequency continues to decline because an imbalance continues between generation and load), an extreme measure to arrest frequency, called *under-frequency load-shedding (UFLS)* is deployed. UFLS involves automatically disconnecting large, pre-set groups of customers at predetermined frequency set-points. It coordinated on an interconnection-wide basis.[33]

Interconnection-coordinated UFLS is a blunt, drastic form of emergency frequency control.[34] It is intended to prevent damage to generators from the extreme frequency excursions that result when the integrity of the interconnected power system has been so severely compromised that portions of the

[32] As noted in the earlier discussion of the "self-healing" property of droop control, return of frequency to the scheduled value fully restores the reserves of primary frequency control.
[33] UFLS is distinct from other, less drastic forms of non-voluntary load shedding that involve fewer customers and that serve more localized reliability objectives. UFLS is also distinct from demand response, in which customers voluntarily offer to shed load and are compensated financially for providing this service to the grid operator.
[34] As discussed, governor actions are self-limiting because they inject (or withdraw) power to oppose changes in frequency only to the extent that frequency has deviated from the scheduled value. Loads disconnected through UFLS must be reconnected through specialized, operator-directed procedures.

system may be operating as electrical "islands" separate from one another.[35] In these situations, the purpose of UFLS is to restore the balance between load and generation (by removing load) before frequency declines even further to a point at which generators disconnect automatically to prevent themselves from being damaged. Once generators disconnect, the imbalance will be even larger, and a more widespread blackout is likely to ensue.

UFLS can also have unintended consequences. For example, if the amount of load dropped is greater than the amount of generation that was lost, frequency will quickly rise and exceed the scheduled value. When this happens, other generators may disconnect, either automatically to protect themselves or for other reasons because frequency is now too high. Frequency will then start to decline again, and a more widespread blackout may ensue.

> *"…the criterion that defines reliable interconnection frequency control is the arrest of frequency prior to the highest frequency set-point for UFLS…"*

UFLS is an emergency operating measure that is to be avoided in routine operations. It is only used when there are no alternatives left to arrest rapidly declining frequency and prevent a blackout. As a result, primary frequency control from generators and other resources is expected to be the principal means to prevent blackouts following the sudden loss of generation. Therefore, the criterion that defines reliable interconnection frequency control is the arrest of frequency prior to the highest frequency set-point for UFLS (and subsequent restoration of frequency to the scheduled value).

The highest set-point for UFLS also has an important implication or consequence for delivery of primary frequency response via turbine-governors operated with a droop setting. Recall from Section 2.2 that the droop setting establishes a relationship between the amount by which a turbine-governor increases (or decreases) generator output and the amount or size of frequency deviation from the scheduled value. If the scheduled frequency is 60 Hz and the highest set-point for UFLS is 59.5 Hz, then the maximum deviation of frequency from the scheduled value before UFLS is triggered is 0.5 Hz (= 60 Hz – 59.5 Hz). If the droop setting on a turbine-governor is 5 percent then a frequency deviation of 0.5 Hz will cause the turbine-governor to increase generator output by no more than about 17 percent or 1/6 of its total rated capacity (= 0.5 Hz / (0.05*60 Hz). Thus, even if more operating headroom were available (for example, if a generator is operating at 50 percent of rated capacity), the total amount that can be delivered is limited by the droop in conjunction with the highest set-point for UFLS. In this example, no more than 17 percent of the total rated capacity that may be available on the generator

[35] See, for example, the description contained in the final report on the 2003 U.S.-Canada blackout: "…automatic [UFLS] is designed for use in extreme conditions to stabilize the balance between generation and load after an electrical island has been formed, dropping enough load to allow frequency to stabilize within the island." (U.S.-Canada Power System Outage Task Force 2004). See, also the introduction to Reliability Standard PRC-006-3, which states: "To establish design and documentation requirements for automatic under-frequency load shedding (UFLS) programs to arrest declining frequency, assist recovery of frequency following under-frequency events and provide last resort system preservation measures." (NERC 2017g).

can be delivered as primary frequency response.[36]

2.6 The Rate of Frequency Decline Determines the Requirements for Reliable Interconnection Frequency Response

Rate of Change of Frequency (ROCOF) *is determined by: (1) the inertia of the interconnection at the time the generator is lost; and (2) the amount of generation that is lost (the size of the generation-loss event). Inertia is a technical term that describes the ability of a power system to resist changes in frequency.*

Focusing now on the "arresting" period, the requirements for primary frequency control can be understood by focusing on the rate at which frequency declines following a sudden loss of generation. The rate of change of frequency (change in Hz per second) is often abbreviated in the literature as *ROCOF.* ROCOF can be understood as the basis for establishing the amount of time that can pass before frequency reaches the highest set-point that triggers UFLS. A higher ROCOF means that frequency is falling more quickly, and a lower ROCOF means that frequency is falling more slowly. The requirement for primary frequency control is to arrest frequency before UFLS is triggered by injecting an amount of power[37] that is at least equal to the amount of generation that was lost.

ROCOF is determined by: (1) the *inertia* of the interconnection at the time a generator is lost; and (2) the amount of generation that is lost (i.e., the size of the generation-loss event). Inertia is a technical term that describes the ability of a power system to resist changes in frequency and is measured in megawatt (MW)-seconds[38]. Inertia is an inherent property or characteristic of each generator and element of load that is on-line and coupled directly to the interconnection. The inertia of the interconnection is the sum of the combined inertias of all such connected generators and loads.

"Neither Inertia nor the size of the generation-loss event, considered individually, provides information on ROCOF. Both factors must be considered jointly."

To a first approximation, ROCOF is determined by the amount of generation lost divided by twice the inertia of the interconnection.[39] It is of utmost importance to recognize that system inertia and the amount of generation lost directly interact with one another. Neither Inertia nor the size of the generation-loss event, considered individually, provides information on ROCOF. Both

36 Bear in mind that this discussion does not consider limits on the rate at which primary frequency response is delivered, which is determined by characteristics of the turbine generator. Hence, while in this simplified example 17 percent is the total capacity available, the amount of primary frequency response that can be delivered in time to arrest frequency may be less than this total.

37 Note that the balance between load and generation can also be re-established by augmenting injection of power with withdrawal of load, either through load sensitivity or fast demand response in the form of contracted load shedding. Both load sensitivity and fast demand response will be examined in this study.

38 Often, system inertia is discussed using the concept of effective system inertia and the inertia of an individual generator is expressed using an inertia constant. In both instances, inertia is normalized and expressed in units of seconds. Expressing inertia in this fashion enables ready comparisons among power systems and generators of different sizes.

39 For the purpose of this illustrative discussion, we ignore the effects of *load sensitivity*, which refers to comparatively small changes in the load that are caused by the change in system frequency. In fact, sensitivity of load to voltage can have a secondary effect on ROCOF. See also discussion at end of Section 2.2.

factors must be considered jointly. Consequently, studying ROCOF requires understanding how these two underlying factors have jointly contributed to an event.

For example, if we hold fixed the amount of generation lost, frequency will fall faster in a power system that has lower system inertia than in a power system that has higher system inertia. See Figure 4. A power system with more inertia is more resistant to the change in frequency caused by the loss of a given amount of generation. Similarly, if we hold system inertia fixed, frequency will fall faster when more generation is lost than when less generation is lost. See Figure 5. For a power system of a given size, loss of a greater percentage of total generation causes frequency to fall faster than loss of a smaller percentage of total generation.

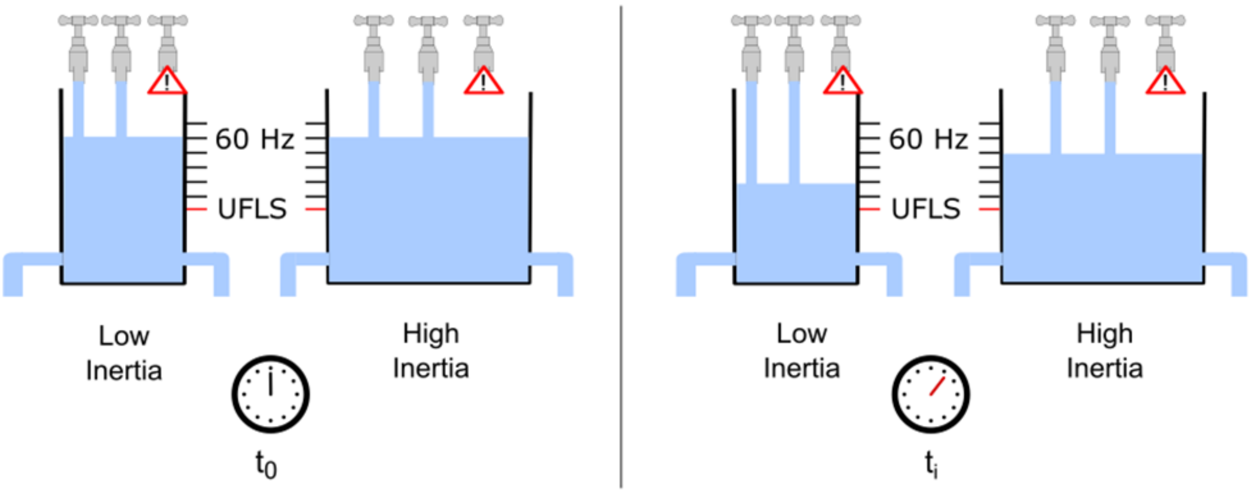


Figure 4. The Effect of System Inertia on ROCOF, for a Fixed Amount of Generation Loss

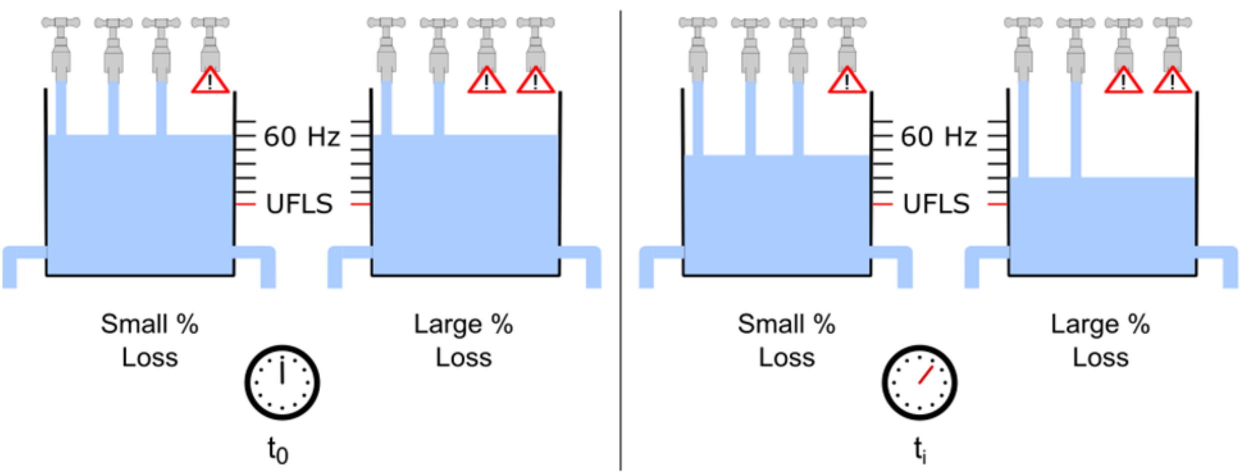


Figure 5. The Effect of the Amount of Generation Lost on ROCOF, for a Fixed System Inertia

The relationships described above are critical to understanding the requirements for reliable frequency control. For example, when normalized to account for the differences among the sizes of the Eastern, Western, and Texas Interconnections, the effective system inertias of each are, in fact, relatively close in value to one another (roughly 3 to 4 seconds).[40] This should not come as a surprise because, despite large differences in the total number of generators in these interconnections, the composition of generators is similar.[41]

The difference in the sizes of the interconnections, however, also means that a given amount of generation lost represents a very different percentage of the total system load in each interconnection. These differences, along with whatever differences may exist in effective system inertia, together establish the requirements for primary frequency control.

To illustrate, the loss of a very large generating station (say, 2 gigawatts [GW]) at a time of high demand in the Texas Interconnection represents about 3 percent of the total system load of 70 GW. By contrast, the loss of this same amount of generation at a time of high demand in the Eastern Interconnection represents only about 0.3 percent of total system load (approximately 600 GW). Therefore, at the time of peak demand and assuming roughly comparable effective system inertias for the two interconnections, loss of 2 GW of generation will cause frequency to fall roughly 10 times faster in the Texas Interconnection than in the Eastern Interconnection. As a result, successful primary frequency control (i.e., that avoids triggering UFLS) requires delivering 2 GW roughly 10 times faster in the Texas Interconnection than in the Eastern Interconnection.[42]

> *"There is a direct relationship between the amount of generation lost and the size and inertia of an interconnection in determining the rate at which frequency will decline immediately following the loss of generation."*

There is a direct relationship between the amount of generation lost and the size and inertia of an interconnection in determining the rate at which frequency will decline immediately following the loss of generation. See Figure 6. The right-hand panel of Figure 6 is a nomogram that depicts the approximate relationship among effective system inertia (X-axis), generation loss as a percentage of total system load (curved lines), and immediate ROCOF following the loss of this generation (Y-axis).[43] The left-hand panel of Figure 6 depicts two views of the decline in frequency (i.e., ROCOF) during the period of time immediately following the loss of generation. The

---

[40] The relationship between system inertia and amount of generation lost is easiest to understand by re-expressing each quantity as a percentage of a "normalizing" factor that is related to the total size of the interconnection, as follows: (1) effective system inertia is total interconnection inertia divided by the MVA base of the total directly connected generation; and (2) effective generation loss is the amount of generation lost divided by total system load.

[41] The aspects of generators that determine their contribution to system inertia depend on the types of turbines used to drive them (e.g., steam turbines, combustion turbines, hydro-electric turbines, etc.), not on the types of fuels consumed (e.g., nuclear, coal, natural gas, and fuel oil, which can all be used to run a steam turbine). These differences will be explored in detail using the modeling tool described in Section 4. See also Figure 12.

[42] This discussion is continued in Section 4.2 with examples drawn directly from each of the three U.S. interconnections. See Table 2.

[43] The relationship between system inertia constant, ROCOF and generation loss, requires normalizing generation loss to the connected capacity used in calculating the system inertia constant. Therefore, the actual ROCOF will vary dependent on connected unloaded capacity of the system in question. The nomogram in Figure 6 was created for the case where connected MVA equals generation and therefore there may exist minor discrepancies between the predicted ROCOF in the nomogram and the ROCOF from the simulation.

inset panel shows the immediate rate of decline in frequency that corresponds to the relationships depicted in the right-hand panel. The larger panel shows that an equivalent amount of primary frequency response will arrest frequency at a lower value and an earlier point in time when frequency falls more rapidly.

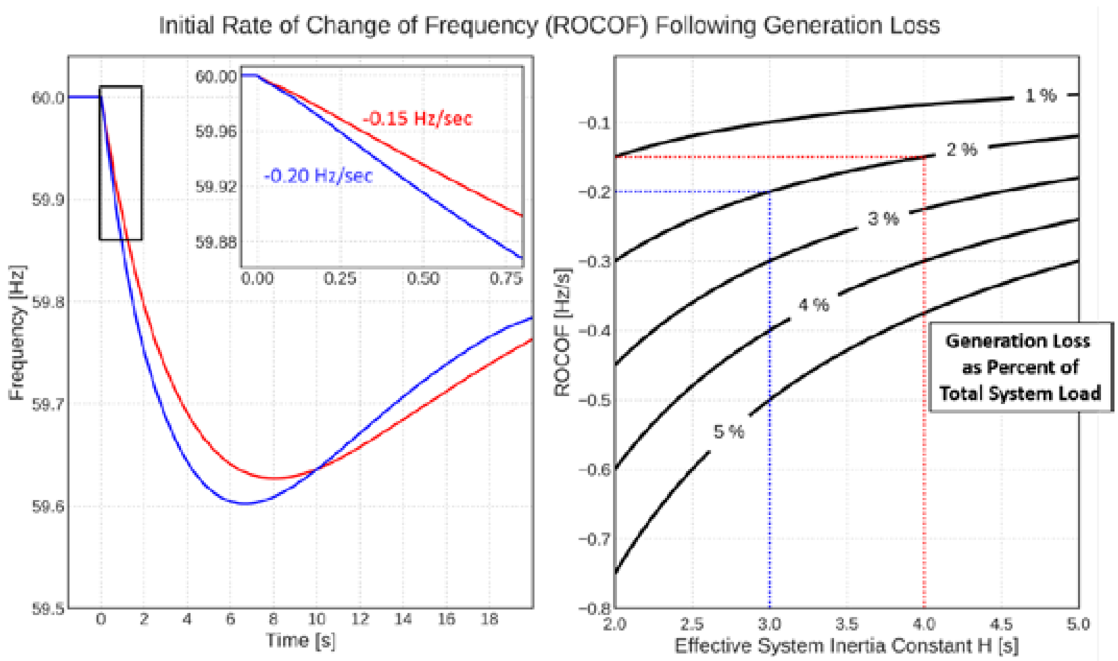


Figure 6. The Analytical Relationships among System Inertia, Generation Loss, Rate of Change of Frequency (ROCOF), and Frequency Nadir, for a Fixed Amount of Primary Frequency Control

2.7 Summary

Imbalances caused by the sudden and unpredictable loss of a large amount of generation pose a challenge for reliable management of interconnection frequency. Loss of a large amount of generation causes an immediate decline in system frequency that is felt throughout an interconnection. If no corrective actions are taken, frequency declines until the power system collapses, and a cascading, widespread blackout ensues.

Four physical factors determine whether an interconnection will respond reliably (i.e., without triggering emergency, interconnection-coordinated UFLS) to a sudden loss of generation:

1. The size of the generation-loss event;
2. The interconnection's inertia, which, in combination with the amount of generation lost, determines the initial rate of decline of frequency following an event (i.e., the rate of change of frequency, or ROCOF);
3. The speed with which other on-line generators respond to arrest and stabilize frequency (i.e.,

provide primary frequency control);[44] and

4. The means by which other generators respond subsequently to restore frequency to its original scheduled value and to restore reserves to their original state of readiness (i.e., provide secondary and tertiary frequency control).

Reliable interconnection frequency response requires that frequency be arrested and stabilized above the highest set-point for UFLS. This report describes the relationships on which interconnection frequency response requirements are based, and the considerations that must be addressed to ensure that these requirements are met.

[44] Non-generation-based resources for primary frequency response, such as ERCOT's reliance on Load Resources, are also reviewed in this report.

3. Background

This section first reviews the principal outcomes and findings from LBNL's 2010 Study regarding interconnection frequency response. It then summarizes relevant industry activities that have taken place since the publication of the 2010 Study. Seen in the light of the 2010 Study, these industry activities, many of which are ongoing, provide both the basis and motivation for this report's analysis and discussion of frequency response. This section presumes knowledge of basic frequency response concepts.[45]

3.1 LBNL's 2010 Study

LBNL's 2010 Study developed and demonstrated the use of three new metrics to assess the performance of primary frequency control resources in responding to generation-loss events (See Figure 7):

Frequency nadir. Industry practices to measure frequency response were, at the time of LBNL's 2010 Study, based on settling frequency—which was the only characteristic of interconnection frequency response that could be measured reliably using then-current grid monitoring technologies. However the nadir of frequency determines whether frequency response has been adequate. Newer grid monitoring technologies, such as phasor measurement units (see, for example, BPA 2017), though not yet widely deployed, can now measure frequency nadir.

Nadir-based Frequency Response. This metric is a direct extension of industry's traditional measure of frequency response. This metric is calculated based on the frequency nadir rather than the settling frequency, to establish a relationship between the size of the generation-loss event and the amount of frequency response delivered from on-line generators responding to this event.

Primary Frequency Response (PFR_t). This is a direct measure of the amount of primary frequency response delivered from on-line generators and loads responding to a generation-loss event. It is time-indexed because, as discussed in Section 2, the amount of frequency response delivered is not immediate but evolves at varying rates over time.

"For a given loss-of-generation event, wide variations in system inertia have a minor impact on the nadir compared to similarly wide variations in the speed and quantity of primary frequency response."

LBNL's 2010 Study found that:

1. For a given loss-of-generation event, wide variations in system inertia have a minor impact on the nadir compared to similarly wide variations in the speed and quantity of primary frequency response.
2. Then-current U.S. spinning reserve policies, while intended originally to address interconnection frequency response, did not, in fact, explicitly require primary frequency control (although some system operators, such as ERCOT, mandate primary frequency control

[45] See Section 2 for a review of these concepts.

capability in their generator interconnection requirements).

3. Periods of high wind generation during times of minimum system load (which is lower than the "light load" conditions that are considered in many studies) and minimum operating reserves, such as those required by spinning reserve standards, are appropriate operating conditions during which to study interconnection frequency response because at these times the amount of on-line generation available for primary frequency control may be at a minimum. In other words, these are among the most challenging conditions under which to ensure that the interconnection can reliably arrest and restore frequency following the sudden loss of generation.
4. The Eastern, Western, and Texas Interconnections were, in 2010, expected to have reliable frequency response with the amount of wind generation and supporting transmission expected by 2012.
5. Routine studies by the industry of Eastern Interconnection frequency response did not reproduce the interconnection's recorded frequency response following the interconnection's largest historic generation-loss event.

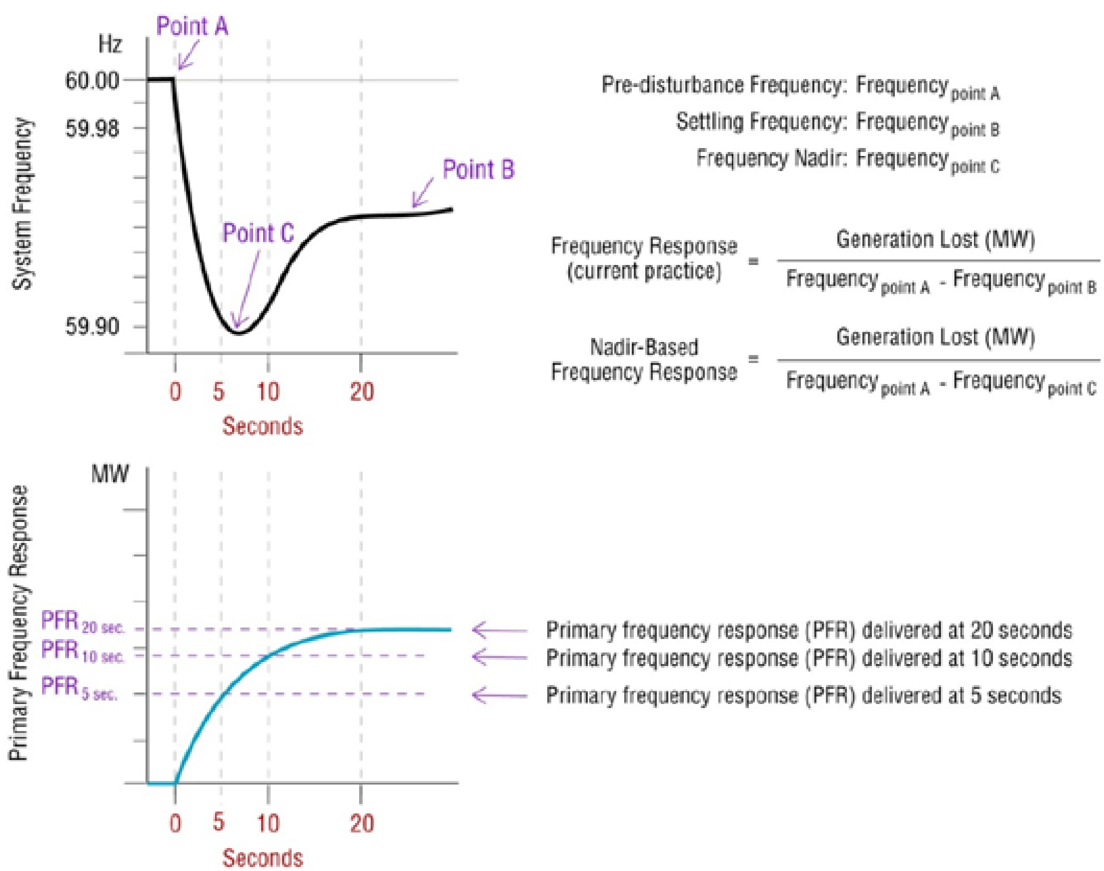


Figure 7. Frequency Response Performance Metrics

Source: Eto, et al. (2010): Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation

LBNL's 2010 Study also found the following, but these findings were not studied in detail:

1. Rapid wind ramps are not comparable to the generation-loss events that must be studied when evaluating interconnection frequency response.
2. Under-deployment of resources for secondary frequency control will reduce the reserves held for primary frequency control.
3. If primary frequency response cannot be delivered reliably from generation located behind a transmission constraint, then that generation will not contribute to interconnection frequency response.

3.2 Industry Activities Following the Publication of LBNL's 2010 Study

LBNL's 2010 Study was published during important transitions in how U.S. interconnections were planned and operated to ensure reliable frequency response. The Study reinforced and, in some instances, helped to accelerate many industry activities that were then in progress. It also at least partially inspired other new activities addressing interconnection frequency response and integration of variable renewable generation. The current report seeks to support these ongoing activities.

3.2.1 Mandatory Reliability Rules for Interconnection Frequency Response

Perhaps the most important industry development since the publication of LBNL's 2010 Study was the revision to the mandatory reliability standards for interconnection frequency response. FERC's 2014 Order No. 794 ratified revisions to North American Electric Reliability Corporation (NERC) Reliability Standard BAL-003-1.1.[46] This obligated each responsible entity (a Balancing Authority or collection of Balancing Authorities known as a frequency response sharing group) to provide frequency response in support of the interconnection.[47]

Reliability Standard BAL-003-1.1 contains three elements related to interconnection frequency response.[48] (See NERC 2015d.) The first element is determination of an interconnection-wide frequency response obligation. This involves both specifying the design criteria for acceptable frequency response and describing factors that must be taken into account when calculating acceptable frequency response on an interconnection-wide basis. See Table 1.

46 BAL-003-1.1, which is a subsequent version of BAL-003.1, was approved later by delegated letter order. However, the differences between the two versions are minor.

47 FERC first approved frequency response-related revisions to BAL-003 in Order No. 693 (in 2007). However, these revisions did not mandate provision of frequency response in support of the interconnection.

48 NERC standards for the United States can be found here: http://www.nerc.com/pa/stand/Pages/ReliabilityStandardsUnitedStates.aspx?jurisdiction=Unitedpercent20States

Table 1. Interconnection Frequency Response Obligations

	Eastern	Western	ERCOT	Québec	Units
Starting Frequency	59.974	59.966	59.968	59.967	Hz
Minimum Frequency Limit	59.500	59.500	59.300	58.500	Hz
Base Delta Frequency	0.474	0.467	0.667	1.468	Hz
CB_R[26]	1.111	1.670	1.648	1.550	Ratio
Delta Frequency (DF_{CBR})[27]	0.427	0.280	0.405	0.947	Hz
BC'_{ADJ}[28]	0.007	N/A	N/A	N/A	Hz
Max. Allowable Delta Frequency	0.420	0.280	0.405	0.947	Hz

Source: NERC (2017b): 2017 Frequency Response Annual Analysis

The second element involves assigning balancing authorities (or frequency response reserve sharing groups) responsibility for a portion of the interconnection-wide frequency response obligation. The assignments are made by expressing total generation plus load within each responsible entity as a fraction of the total generation and load within the interconnection.

The third element entails assessing each responsible entity's past frequency response performance in relation to the entity's frequency response obligation. This involves selecting past frequency response events and calculating a median frequency response for the interconnection as a whole and then for each responsible entity individually. Notably, the standard does not require generators to contribute to primary frequency control.

In support of BAL-003-1.1, the NERC Resources Subcommittee regularly analyzes interconnection frequency response events to: (1) refine the methods for measuring and evaluating interconnection frequency response; (2) monitor trends in each interconnection's frequency response; and (3) gain experience with and establish a basis for future modification to BAL-003-1.1's frequency response requirements.[49] (See, for example, NERC 2017b.)

The frequency response obligations articulated in BAL-003 are also supported by two interconnection-specific frequency response-related standards. BAL-001-TRE for ERCOT requires that, with limited exceptions, all generators must (a) have governors capable of providing frequency response; (b) have specific maximum droop and deadband settings; and (c) meet minimum performance measures during identified frequency events (NERC 2017f).[50] BAL-002-WECC-2a for the Western Interconnection clarifies that units providing spinning reserve must respond immediately and automatically to frequency deviations through turbine-governors or other control systems. (See NERC 2017e.)

---

[49] NERC Resources Subcommittee agendas, presentations, and minutes can be found here: http://www.nerc.com/comm/OC/Pages/Resources-Subcommittee.aspx.

[50] This is an expansion of a practice common among some U.S. grid operators to require primary frequency response capability on only new generator interconnections. See Section 3.2.2.

In 2016, FERC issued a Notice of Inquiry[51] on essential reliability services and the evolving bulk-power system that focused on primary frequency response (FERC 2016a). The notice requested comments on several possible actions: (1) modifications to the pro forma large and small generator interconnection agreements mandating primary frequency response requirements for new resources; (2) new primary frequency response requirements for existing resources; and (3) requirement to provide and compensate for primary frequency response.

3.2.2 Supporting Frequency Response-Related Requirements and Approaches

Some operating regions, such as Independent System Operators (ISOs) and Regional Transmission Operators (RTOs), have also adopted region-specific operating requirements and approaches that supplement NERC BAL standards and further contribute to reliable interconnection frequency response. We briefly summarize a few notable examples below.[52]

In 2014, the Independent System Operator New England (ISO-NE) began requiring that generators above a minimum size (10 MW), including renewable resources, operate with a functioning governor (ISO NE 2014). The requirements further direct operation with a specified deadband and droop setting, and that primary frequency response that is delivered during frequency response events will not be inhibited by secondary control actions directed by plant load controllers.

In 2016, the California ISO (CAISO) supplemented efforts that reinforce the BAL–002-WECC-2a regional standard by establishing a maximum permissible deadband and prohibiting the withdrawal of primary frequency response during frequency response events (FERC 2016b). In addition, formal verification of a generator's ability to adhere to these requirements is required prior to and as a condition of participation in CAISO's spinning reserve market.

PJM Interconnection also has rules that address frequency response. For example, PJM Operating Manual 14D indicates that generators "should" operate with unrestricted governors and are "requested" to operate with a not-to-exceed deadband and specified droop (PJM 2017).

3.2.3 Related Industry Activities Focused on Ensuring Reliable Interconnection Frequency Response

Since the publication of LBNL's 2010 Study, and in parallel with the development and roll-out of BAL-003-1.1, a number of related industry efforts have contributed to our understanding of interconnection frequency response. Several leading examples of these efforts are summarized below with a focus on the motivation they provide for topics addressed in the current study.

One of the findings from LBNL's 2010 Study was that the system models then available to analyze the frequency response of the Eastern Interconnection did not reproduce the recorded frequency response

[51] FERC docket RM16-6-000.
[52] A more comprehensive survey of U.S. and international practices is provided in a separate, standalone technical report prepared for this project. See Roberts 2018.

of the interconnection following a large generation-loss event. Notably, the amount of generation lost during this event is now the design event that is used to establish the frequency response obligation for the Eastern Interconnection under BAL-003-1.1. A key feature of this event, and of the Eastern Interconnection's frequency response to other events, is formation of the nadir at a frequency lower than that predicted by these models. Of equal importance is the apparent early withdrawal of primary frequency response by secondary control actions leading to what has been termed the "lazy L." See Figure 8.

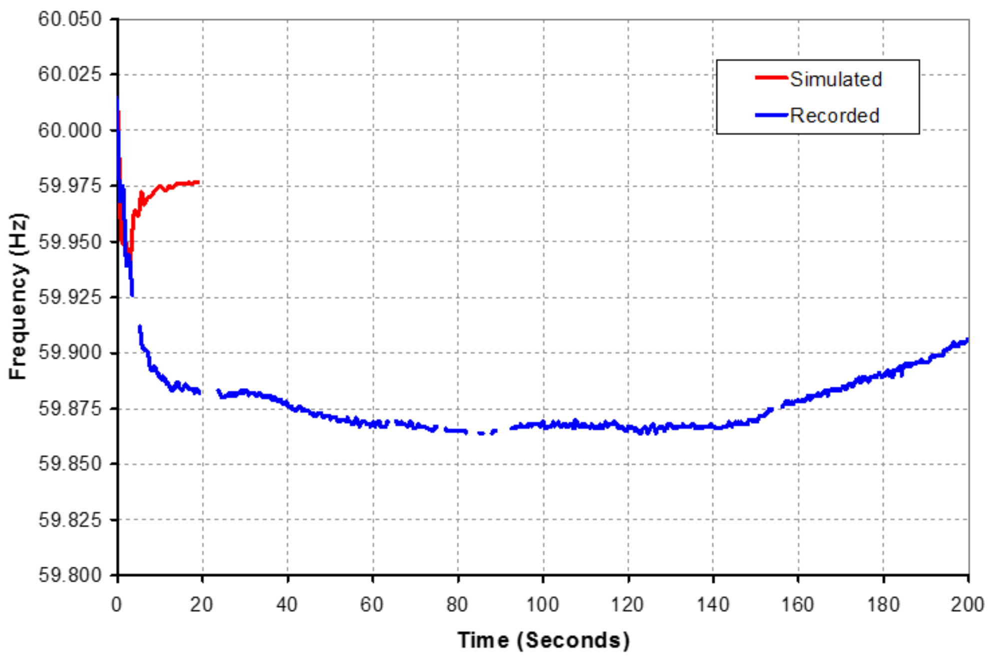


Figure 8. Frequency of the Eastern Interconnection during the First 199 Seconds Following the Loss of 4,500 MW of Generation: A Comparison of Recorded Data with Results from a Simulation of the Event

Source: Eto, et al. (2010): Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation

Several industry initiatives have begun to address this finding. First, in 2010, NERC, as part of its Frequency Response Initiative, surveyed generator turbine-governors (NERC 2012). The survey collected information on the number of units equipped with governors and, among these, which ones had active governors, droop, sustaining response, and deadbands. See Figure 9 through Figure 11. A key finding of the survey was that some generators were operating their turbine-governors with deadbands that were so large that they effectively prevented generators from contributing to primary frequency control. Another key finding was that some generators were withdrawing primary frequency response early as a result of secondary control actions by plant-level controllers. These findings led NERC to issue an Industry Advisory in 2015 recommending against these practices (NERC 2015a). Later in 2015, NERC also approved a Primary Frequency Control Guideline that recommends settings for turbine-governors and discourages generators from withdrawing primary frequency response during frequency response events (2015b). Neither is a mandatory requirement, however.

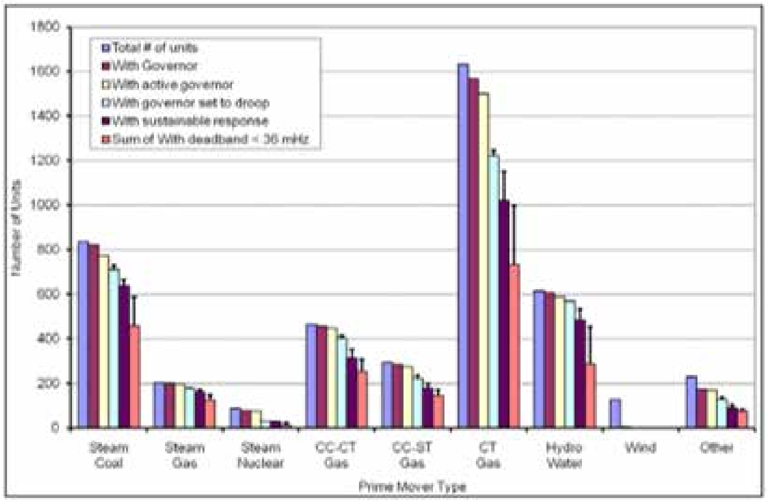


Figure 9. Eastern Interconnection Generator Responses to NERC Survey

Source: NERC (2012): Frequency Response Initiative: The Reliability Role of Frequency Response

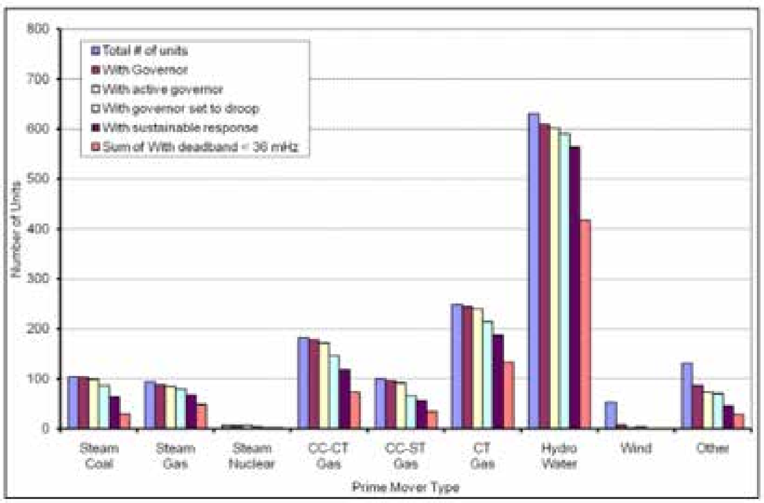


Figure 10. Western Interconnection Generator Responses to NERC Survey

Source: NERC (2012): Frequency Response Initiative: The Reliability Role of Frequency Response

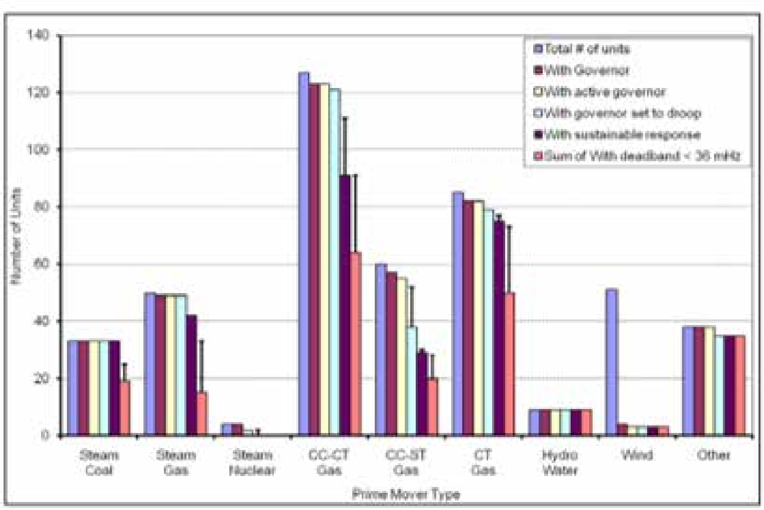


Figure 11. Texas Interconnection Generator Responses to NERC Survey

Source: NERC (2012): Frequency Response Initiative: The Reliability Role of Frequency Response

Separately, the information collected by NERC has been augmented and incorporated to varying degrees into the model databases used in updated versions of the production-grade simulation tools that industry uses to study, among other things, Eastern Interconnection frequency response. Related to this, at the time this report was being written, NERC staff was also responding to a FERC directive, issued as part of the approval for BAL-003-1.1, to further enhance the production-grade simulation models used to study frequency response in the Eastern Interconnection. See NERC (2017d).

The Western Electric Coordinating Council (WECC) has given special emphasis to system modeling, and especially model validation, since at least 1996. See, for example, WECC (2014a) and WECC (2014b). WECC studies show good correspondence between dynamic simulations and actual system events. More recently, WECC and NERC staff have worked collaboratively to understand the system frequency impacts of fault-induced tripping and momentary cessation of solar photovoltaic generation (NERC 2017a).

Within the Western Interconnection, Bonneville Power Administration (BPA) has been active in monitoring and assessing frequency response events. For example, BPA runs a frequency event detection software application in real time (BPA 2017). Events are triggered by information provided by BPA's synchrophasor-based grid monitoring network, which is a source of time-stamp, high time-resolution measurements of grid conditions, including frequency (Kosterev et al. 2014). Among other things, the application uses the rate of change of frequency to locate the source of frequency disturbances.[53]

As noted in LBNL's 2010 Study, the Texas Interconnection has long had an active focus on frequency response. In 2011, ERCOT staff published a peer-reviewed article discussing interconnection frequency response and introducing an on-line method for calculating interconnection inertia to enhance operator situational awareness. (Sharma et al. 2011). In 2013, ERCOT staff began an initiative to restructure the ancillary services they procure in their wholesale market to ensure adequate frequency response (ERCOT 2013). The initiative sought to parse the traditional rapid-response reserve product into a series of independent products, each tailored to addressing frequency response reserve requirements as a function of the specific time or point within the response to a generation-loss event at which the resource would be deployed or expected to perform. Ultimately, ERCOT did not approve the proposed redefinitions. More recently, ERCOT has begun monitoring system inertia in operations.

In 2013, NERC established an Essential Reliability Services Task Force; in 2015 the Task Force was converted to a standing Essential Reliability Services Working Group (ERSWG). The working group, among other things, studies reliability topics related to the changing composition of the generation fleet, such as retirement of coal-fired generators and introduction of variable renewable generators (NERC 2017c). For example, if coal-fired generators are retired and replaced with wind generators, total

[53] Personal communication from Dmitry Kosterev, Bonneville Power Administration (BPA), to LBNL on January 11, 2018.

system inertia would decrease. But, if they are replaced with combined-cycle power plants, total system inertia would increase.[54] See Figure 12.

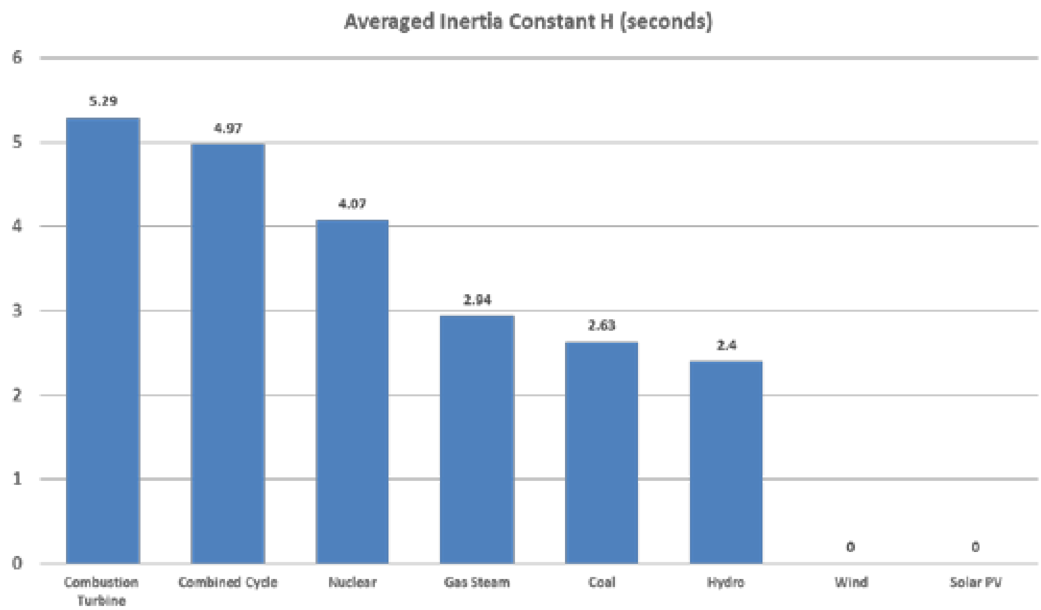


Figure 12. Representative Inertia Constants by Unit Type

Source: Sharma (2016): Renewable Integration at ERCOT

A major focus of ERSWG study efforts has been interconnection frequency response. See, for example, NERC 2015a. Recent ERSWG meetings have featured presentations that estimate total system inertia of each interconnection based on the actual dispatch of generating resources.[55] See Figure 13 through Figure 15. A number of presentations have focused on methods to estimate ROCOF immediately following the loss of generation and assess aspects of interconnection frequency response (such as the nadir and the time at which frequency is arrested) and aspects related to the extent to which primary frequency response is sustained.

---

54 The aspects of generators that determine their contribution to system inertia depend on the types of turbines used to drive them (e.g., steam turbines, combustion turbines, hydro-electric turbines, etc.), not on the types of fuels consumed (e.g., nuclear, coal, natural gas, and fuel oil, which can all be used to run a steam turbine).

55 NERC Essential Reliability Services Working Group work products, agendas, presentations, and minutes can be found here: http://www.nerc.com/comm/Other/Pages/Essential-Reliability-Services-Task-Force-(ERSTF).aspx

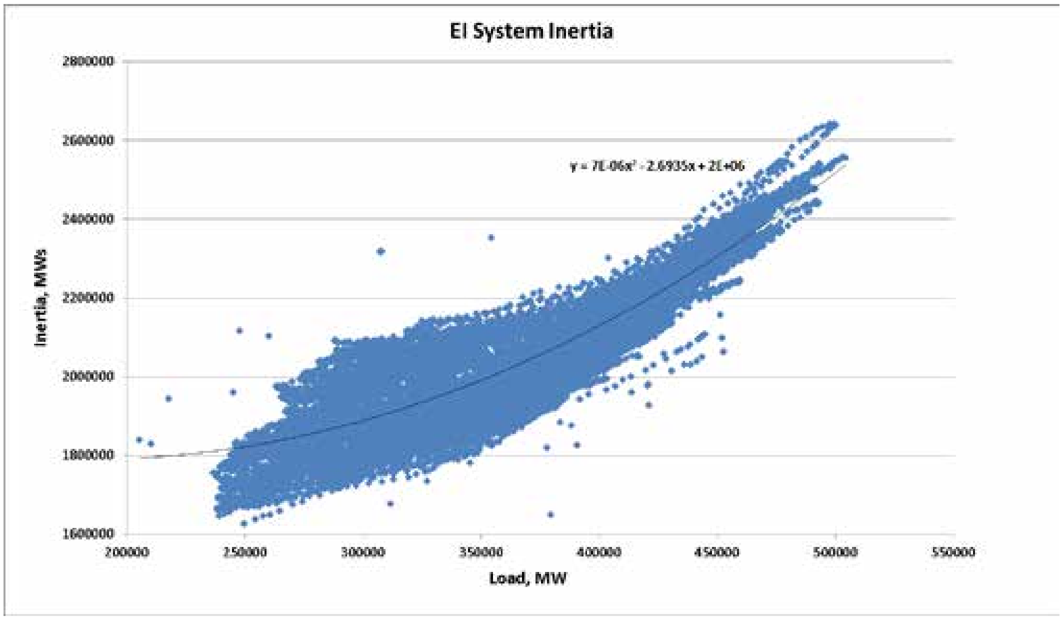


Figure 13. Eastern Interconnection Inertia vs. Net Load (August-September 2016)

Source: Matevosyan, J. (2016): Inertia Data

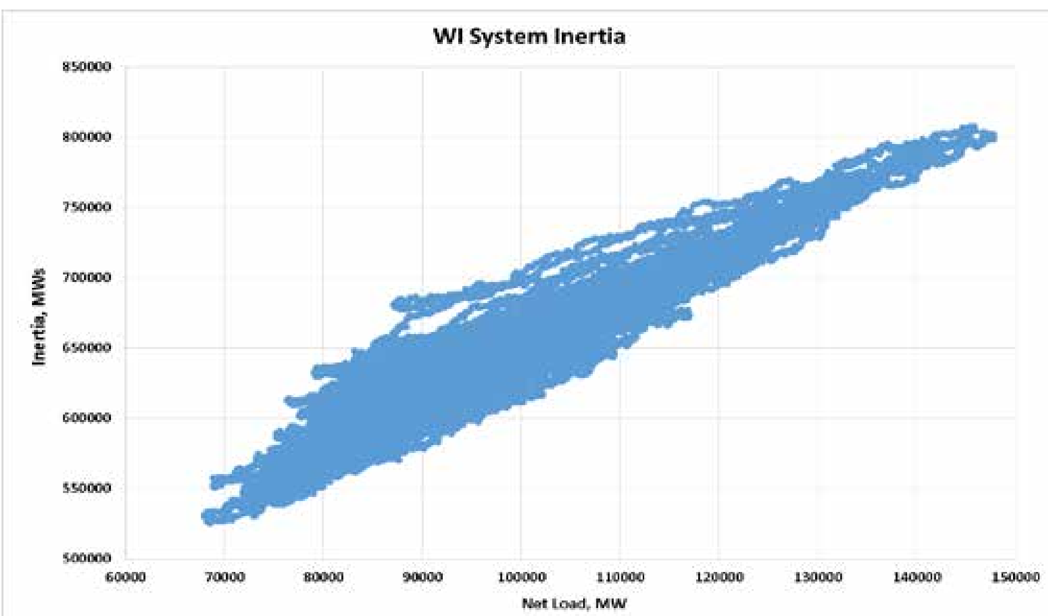


Figure 14. WECC Inertia vs. Net Load (June-September 2016)

Source: Matevosyan, J. (2016): Inertia Data

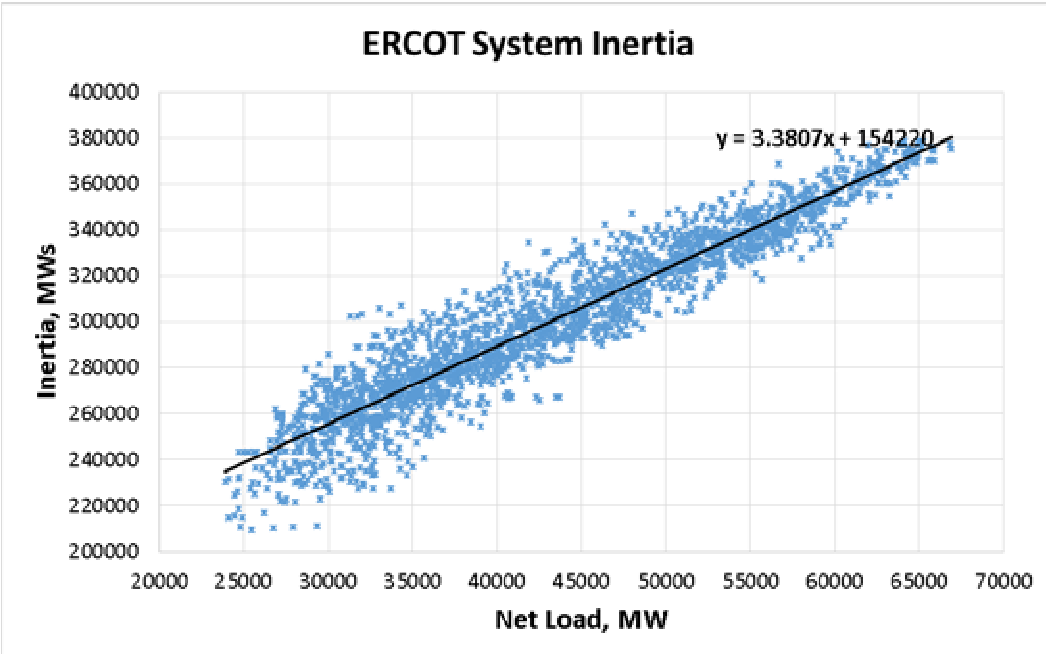


Figure 15. Texas Interconnection Inertia vs. Net Load (July-September 2016)

Source: Matevosyan, J. (2016): Inertia Data

3.2.4 Treatment of Frequency Response Issues in Renewable Integration Studies

Prior to LBNL's 2010 Study, industry and academic studies of renewable integration focused primarily on the use of production-cost modeling tools to study balancing requirements for the conventional generation needed to ensure resource adequacy in response to time-varying production from renewable generation. Some studies had also begun to examine requirements for regulation services.[56] However, no studies of renewable generation had addressed the topic of interconnection frequency response.

[56] LBNL 2010 study found that the assumptions relied on by these studies regarding the statistics describing the distribution of wind power production were optimistic compared to measured wind generation data and that the effect of this optimism would be to understate requirements for regulation. Moreover, the study showed that under-procurement of resources for regulation posed previously unrecognized risks for interconnection frequency response.

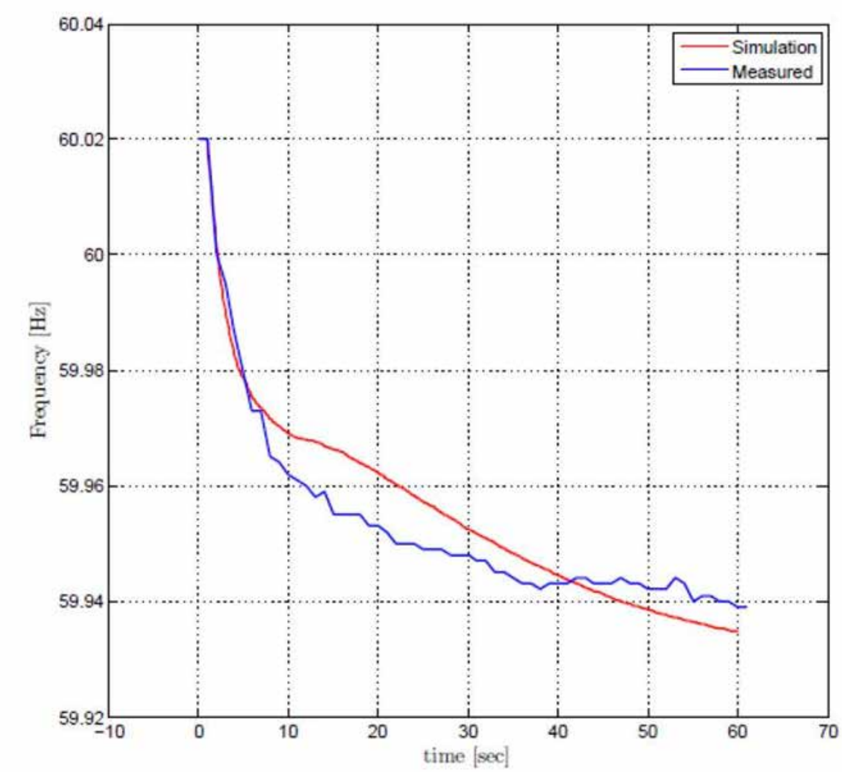


Figure 16. Response comparison between simulations using new base case and May 13, 2012 event

Source: Miller et al. (2013): Eastern Frequency Response Study

Since the publication of the LBNL's 2010 Study, several nationally recognized studies have been published examining interconnection frequency response issues raised in relation to the integration of variable renewable generation. The most well known of these studies are those that have been conducted by General Electric (GE) Consulting for CAISO and for the U.S. Department of Energy's (DOE's) National Renewable Energy Laboratory (NREL). Collectively, these studies have made a number of contributions to the study of frequency response. Some reference and use the frequency response metrics proposed by LBNL (Miller et al. 2011). Some seek to identify and address the sources of inaccuracies in the production models used by industry to study Eastern Interconnection frequency response (Miller at al. 2014). See Figure 16. In addition, these studies have simulated the performance of a proprietary frequency response product that can be provided by one manufacturer's wind generators.[57]

However, none of these studies of renewable energy integration has considered the very strict operating scenarios identified by the LBNL 2010 Study as posing the greatest challenges for interconnection frequency response. These include operation during times of minimum system load, which is lower than the "light load" cases that have been considered in the studies, and during conditions of minimum operating reserves, such as those required in spinning reserve requirements

[57] This feature, called "synthetic inertia" is discussed in Section 5.8.

(e.g., BAL-002-WECC-2a). Therefore, the findings of these studies regarding frequency response may be optimistic in relation to what might take place under these more challenging conditions.

3.3 Summary

The purpose of this study is to support policymaker and industry understanding of the physical requirements for reliable interconnection frequency response. Improved understanding is especially timely now for several reasons.

First, industry experience with the frequency response-related requirements in the NERC Reliability Standard BAL-003-1.1, which mandates an interconnection-wide frequency response obligation, is nascent. Increased understanding of the physical requirements for reliable interconnection frequency response will support industry efforts to comply effectively and in a timely way. Understanding of these requirements will also support possible future efforts to revise BAL-003-1.1, as well as supporting standards and other related activities (e.g., generator interconnection requirements).

Second, industry and policymakers are currently grappling with the reliability implications of changes in the composition of the generation fleet. Deeper understanding of reliable interconnection frequency response will enable industry to focus on the requirements that they must manage—rapid and sustained primary frequency response—and thereby help guide appropriate focus on related issues, such as how reductions in system inertia increase these requirements.

Finally, industry recognizes the important role that generator turbine-governors currently play in interconnection frequency performance. Expanded understanding of frequency response will enable industry to focus on the most important factors for ensuring that turbine-governors contribute to reliable interconnection frequency response: the speed at which the fleet first delivers and then sustains primary frequency response. Greater understanding will, in turn, support industry discussion of related issues for reliable interconnection frequency response, such as the value of smaller deadbands.

4. Analysis Approach

In this section, we describe the application of a production-grade modeling tool to examine how the technical concepts discussed in Section 2 lead to the requirements for reliable frequency response. We then use the modeling approach to explore how the requirements can be met by various amounts, types, and combinations of generator technologies. The modeling approach allows us to study systematically several important but previously not well-understood aspects of frequency control. This includes first understanding the factors that determine whether frequency will be arrested quickly after generation is lost. It also involves understanding whether frequency will be stabilized until primary frequency response can be replaced by slower secondary frequency response. These factors include turbine-generator-specific limits on the headroom accessible by generators that provide primary frequency response and plant-specific and other controls that may limit or withdraw frequency response early.

This study's dynamic simulations were conducted using GE's Positive Sequence Load Flow tool, known as PSLF—the same commercially available tool that is currently in wide use by industry to conduct, among other things, production-grade studies of frequency response. By using this tool, we were able to study the performance of turbine-governors and plant load controllers for different types of generators (e.g., combined-cycle, hydro, and steam) using the same models of these generators that are used by industry to conduct mandatory reliability planning and operations studies.

Together, these features allowed us to study systematically, and in much greater detail than our earlier study, the interacting factors that jointly determine the frequency response of an interconnection. These factors are: (1) the interconnection requirements for primary frequency response; (2) the headroom available on generators, which establishes an upper bound on the amount of primary frequency response; (3) the rate at which turbine-governors deliver primary frequency response from this headroom; and (4) plant-specific control settings or operating factors that limit or withdraw primary frequency response early (i.e., before frequency has been stabilized). We also examined fast demand response, governor deadband settings, and load sensitivity (sometimes called *load damping*),[58] which also contribute to frequency response.

The structure and organization of our modeling framework, including the starting points and boundaries for the simulations we conducted, is based on information drawn from each of the three U.S. interconnections. Information on historic dispatch was provided by representatives from each of the three interconnections. Additional information was taken directly from the system models that are currently used by planners within each of the interconnections to conduct mandatory reliability studies.

Although our simulations were conducted using the same dynamic simulation tool used by industry, the system models we develop are less detailed than those used by industry in order to focus attention on

[58] The majority of our simulations assumed no load sensitivity in order to focus attention on the relationship between primary frequency control provided by active sources, such as generators, and interconnection frequency response.

key relationships and interactions that affect frequency response. Section 4.7 compares simulation results from our simplified models to those from the highly detailed interconnection system models used routinely by industry planners in order to confirm that our models have accurately captured the most important characteristics and aggregate features of interconnection frequency response that are found using industry's system models. At the same time, we emphasize that the models we developed are not intended to replace the need for or duplicate the exact results produced by industry's system models.

Full documentation for the system models we developed and summaries of all the simulations we conducted are provided in Undrill (2018). In addition, a more in-depth discussion of our analysis approach compared to the analyses conducted with the system models used by industry planners in the each of the interconnections is provided in Undrill et al. (2018).

4.1 Overview of the Simulation Tool and System Modeling Approach Used for this Study

GE's PSLF is the simulation tool used to conduct this study.[59] PSLF is used widely in industry, especially in the Western Interconnection, for mandatory reliability studies of dynamic phenomena such as interconnection frequency response.

The system modeling approach used in our study illustrates key relationships and interactions by focusing on the performance characteristics of a small number of representative types of generators. This involves distinguishing, first, between generators that do and do not contribute inertia to the interconnection. Second, it involves distinguishing between generators that do and do not provide primary frequency response. For generators that provide primary frequency response, it involves specifying the headroom available, the turbine-governor type, the droop of these turbine-governors, and the operation of plant load controls and other forms of control, which together determine how primary frequency response is delivered and whether it is sustained.

An organizing principle of the modeling approach is to rely on dimensionless quantities, namely percentages, to illustrate key relationships among the above factors. For example, in discussing system inertia, we present information on the percentage of on-line generation capacity that contributes inertia to the power system. One hundred minus this percentage is the fraction of generation capacity that does not contribute inertia to the power system. Similarly, as illustrated in Section 2, we express the design generation-loss event as a percentage of the total system load at the time of the event.

The abstractions and simplifications used in this study are intended to illustrate key relationships and interactions. They are not intended to replace production-grade studies containing the full details of each generator operating within the interconnection for the purpose of analyzing dynamic behavior to support operational decision making.

---

[59] See http://www.geenergyconsulting.com/practice-area/software-products/pslf-re-envisioned.

For example, the modeling approach developed for this study was designed specifically to focus on the interconnection- and generator-specific characteristics that jointly determine the reliability of an interconnection's frequency response. Other important aspects of interconnection and generator performance are not studied or modeled. These include the interactions between generators and the transmission system, such as synchronous stability and voltage control. Similarly, the actions of system protective devices and remedial action schemes/special protection systems are not considered. Hence, in contrast to the approach taken in LBNL's 2010 Study, we do not study transmission bottlenecks that prevent the delivery of primary frequency response from generators.

Finally, we emphasize that this study does not examine interactions between primary frequency control and system-directed secondary frequency control (e.g., AGC). As discussed in the LBNL 2010 Study, these interactions cannot be studied with dynamic simulation tools alone, because these tools were developed to study automatic actions taken by power system elements, such as generator turbine-governors and plant load controllers following pre-set control directions.[60] The actions of system-directed, secondary frequency control, such as AGC, depend on dispatch decisions and actions taken by human operators. These actions can be simulated, but they cannot be studied solely using dynamic simulation tools. Informed judgement regarding the many influences (human, mechanized, and "acts of nature") that affect the way AGC is used must be applied in concert with dynamic simulation tools. Thus, we do not examine whether or how secondary frequency control via AGC or manual dispatch restores frequency and replaces reserves of primary frequency response.

4.2 Modeling of Interconnection Frequency Response Design Criteria

The generation-loss events and the requirement for reliable interconnection frequency response (arrest and stabilize frequency above the highest set-point for UFLS) examined in this study follow the design guidelines prescribed in NERC BAL-003-1.1. See Table 1 in Section 3. Table 2 repeats these guidelines and adds recent information from each of the interconnections that guided our simulation studies.

Table 2 presents recent information on both the peak and minimum loads of each of the interconnections and expresses the design generation-loss event as a percentage of these values. This presentation illustrates the significance of the analysis approach first introduced in our 2010 Study, which is to examine frequency response during periods of minimum interconnection load. As noted in LBNL's 2010 Study, minimum load can represent a particularly challenging operating period because the reserves available to provide primary frequency control in the time required may be at a minimum. As discussed in Section 2, minimum load is also a time when a given loss of generation will create a significant challenge for these reserves because the loss-of-generation event represents a much larger percentage of the total system load. Consequently, evaluating frequency response at times of minimum

[60] As an aside, the majority of dynamic simulation studies conducted by industry today consider only the first 10 to 20 seconds following a loss-of-generation event. Our dynamic simulations extend this time horizon significantly to examine how early withdrawal of primary frequency response affects interconnection frequency response.

system load is consistent with the conservative principle of focusing on worst-case operating conditions.[61]

Table 2. Interconnection Frequency Response Design Criteria

Design Criteria	Eastern Interconnection	Western Interconnection	Texas Interconnection
Generation-Loss Event	4.5 GW	2.7 GW	2.7 GW
Peak Load – 2015	546 GW	162 GW	70 GW
Minimum Load – 2015	210 GW	64 GW	24 GW
Gen. Loss Event/Peak System Load	0.8 %	1.6 %	4.0 %
Gen. Loss Event/Min System Load	2.1 %	4.1 %	11.3 %
Highest Set-Point for Interconnection-Coordinated, Under-Frequency Load	59.5 Hz	59.5 Hz	59.3 Hz

Source: Developed by LBNL from NERC (2017b): 2017 Frequency Response Annual Analysis; and Matevosyan (2016): Inertia Data

This study's simulations focused on generation-loss events representing either two or four percent of the total system load of the power system that we model. Loss of two percent of generation is consistent with the design generation-loss event studied for the Eastern Interconnection at the time of minimum interconnection load. Loss of four percent of generation is consistent with the design generation-loss event studied for the Western Interconnection at the time of minimum interconnection load and for the Texas Interconnection at the time of interconnection peak load.[62]

4.3 Representation of System Inertia

"…the proportion of on-line or connected generation that does not contribute inertia to the power system the "electronically coupled fraction of generation" or "Efrac." Following this nomenclature, the proportion of generation that does contribute inertia is all of the remaining generation (or 1 - Efrac)."

To represent system inertia in our model, we first classified generators as one of two types: generators that contribute inertia and generators that do not contribute inertia. As discussed in LBNL's 2010 Study, the principal types of generation that do not contribute inertia to the power system are variable renewable generation from the majority of wind generators (Type 3 and Type 4) and all solar photovoltaic generators. These types of generation do not contribute inertia because they are electronically coupled to the interconnection. For ease of exposition, we label the proportion of on-line or connected generation that does not contribute inertia to the power system the "electronically coupled fraction of generation"

[61] We recognize that times of minimum system are not the only times when worst-case condition may be encountered. Times when system load is not at a minimum but system inertia may be low due the significant generation from sources that do not contribute inertia (e.g., Type 3 and Type 4 wind turbines), can also create increased requirement for primary frequency control

[62] For reasons that will be described in greater detail in Section 5, the design generation-loss event for the ERCOT Interconnection at the time of interconnection minimum load presents a significant operating challenge from the standpoint of frequency response. ERCOT addresses this challenge with a specialized form of primary frequency control involving fast demand response.

or "*Efrac*." Following this nomenclature, the proportion of generation that does contribute inertia is all of the remaining generation (or 1 - *Efrac*).

To study the effect of changes in system inertia on the requirements for primary frequency control, we vary both the percent of generation that contributes inertia and the inertia of this percentage. To illustrate: assume that the entire fleet of generation contributes inertia (*Efrac* = 0 percent) and that the inertia constant of every generator in this fleet is four seconds. If half of this generation fleet is then replaced by generation that does not contribute inertia (*Efrac* = 50 percent), the effective system inertia of this modified fleet is two seconds.

We selected two values of inertia constants for inertia contributing portion of the fleet for our study: three and four seconds.

As noted in Section 3, the NERC ERSWG has begun developing and presenting information on the observed total system inertia of each of the interconnections over time. We were able to work with operating entities for the Western and Texas Interconnections to obtain information from recent historical operating periods, which enabled us to estimate the effective inertia constant of the portion of the fleet contributing inertia during selected operating periods.[63]

Figure 17 presents an effective inertia constant for synchronously connected generation estimated by LBNL from hourly dispatch information provided by PeakRC for the Western Interconnection for the period from August 2016 through March 2017. Figure 17 shows that over this period a majority of estimated values lie between three and four seconds.

Figure 18 presents hourly generation production information provided by ERCOT for the period from July through December 2016. The figure shows how our estimate of total system inertia varies for selected dispatch hours compared to the dispatch of wind generation (which does not contribute inertia). For a handful of operating hours (shown as red dots), we obtained additional information that allowed us to estimate the effective inertia of synchronous generation. Figure 18 shows that effective inertia constant for this portion of fleet, which contributes inertia to the Texas Interconnection, also tends to lie between three and four seconds.

[63] PeakRC provided inertia, in MW-seconds, and capacity, in MVA, broken down by fuel types for a period of 8 months. ERCOT provided inertia, in MW-seconds, and capacity, in MVA, broken down by fuel types for 20 distinct operating points. To determine the effective inertia constant for the connected synchronous generation, we summed these quantities across all fuel types, excluding wind and solar, to develop an estimate of total system inertia and total on-line inertia-contributing capacity. We then divided our estimates of total system inertia by total on-line capacity, which we label effective inertia constant for connected synchronous generation. We emphasize that did not review the underlying information provided for our calculations.

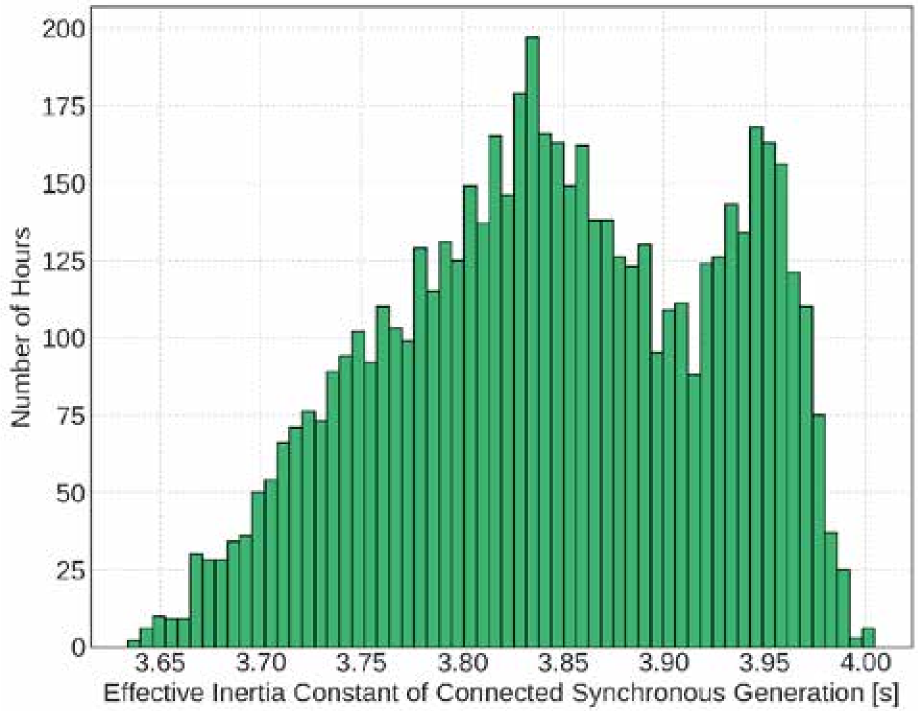


Figure 17. Estimated Effective Inertia Constant of the Generation Contributing Inertia in the Western Interconnection (August 2016 to March 2017)

Source: Estimated by LBNL from information provided by PeakRC

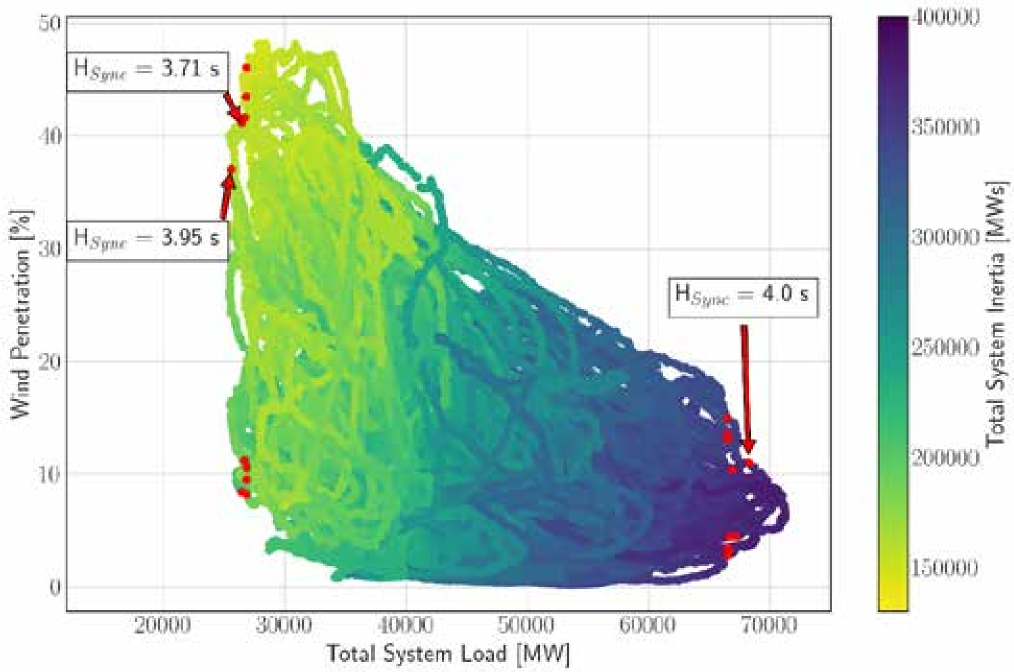


Figure 18. Estimated Effective Inertia Constant of Generation Contributing Inertia in the Texas Interconnection (July to December 2016)

Source: Estimated by LBNL from information provided by ERCOT

As a final source of guidance, we obtained the production-grade system models developed by all three interconnections to study dynamic phenomena, such as frequency response. In contrast to the information we obtained from ERCOT and WECC, which spans many operating periods, the system models represent the dispatch of generation for only a single operating point, such as a lightly loaded winter morning hour.

4.4 Specification of Generation Providing Primary Frequency Response

As discussed in Section 2, the sole means by which interconnections ensure reliable operation following the sudden unplanned loss of a generator is through primary frequency response delivered by generation (or equivalent resources such as demand or storage) with headroom. Accordingly, this study focused on the key factors that affect delivery of primary frequency response. We approached this by conducting a structured series of parametric simulations involving selective adjustments to each of these key factors individually and then in conjunction with one another. The key factors include (a) the fraction of generators that respond to changes in frequency; (b) the headroom from which response can be provided by these generators; (c) the rate at which technology-specific turbine-governors deliver primary frequency response from this headroom; and (d) whether primary frequency response is sustained, limited, or withdrawn early as a result of the actions of plant-load secondary or other forms of control.

4.4.1 Frequency Responsive Generation

As noted in LBNL's 2010 Study, not all generators participate in primary frequency control. To simplify our exposition of concepts, we limit the portion of generation on-line that is able to contribute to primary frequency control to the portion that also contributes inertia (1 - *Efrac*). The fraction of generation that contributes frequency response is equal to or less than (1 - *Efrac*). This simplification is consistent with current operating practices in the United States. In much of the United States today, electronically coupled variable renewable generation does not routinely participate in primary frequency control in response to the sudden loss of generation. The main reason is that variable renewable generation is normally operated at its maximum output, so there is no headroom available from which primary frequency response could be delivered for a generation-loss event.[64]

"The portion that participates in primary frequency control is the responsive fraction or 'Rfrac,' and the portion that does not participate is the non-responsive fraction or 'Nfrac.'"

Turning now to generators that contribute inertia, we divide them into two categories: those that do and those that do not participate in primary frequency control. The portion that participates in primary frequency control is the responsive fraction or "*Rfrac*," and the portion that does not participate is the non-responsive fraction or "*Nfrac*." Total generation is equal to the sum of *Rfrac*, *Nfrac*, and *Efrac* (that is, *Rfrac* + *Nfrac* + *Efrac* = 1). See Figure 19.

[64] Importantly, this situation is changing. Several grid operators now require all new wind generators to have frequency response capability. The Texas Interconnection requires all wind generators to have frequency response capability. When wind generators that have frequency response capability must curtail their output, for example, when there are transmission constraints, the wind generator will provide upward primary frequency response when frequency is depressed.

Required information on the responsive fraction of each interconnection's generation fleet includes how the output of each generator is controlled and how each generator is dispatched. For example, as noted in our discussion of variable generation, a generator may be capable in principle of participating in primary frequency control, but if it is dispatched at its maximum operating point (e.g., a baseload generating plant or a wind generating plant), it does not have headroom available from which primary frequency response could be delivered. Similarly, a generator operating with headroom whose turbine-governor controls are blocked or operated with a very large deadband (e.g., > 300 mHz) will not participate in primary frequency control.

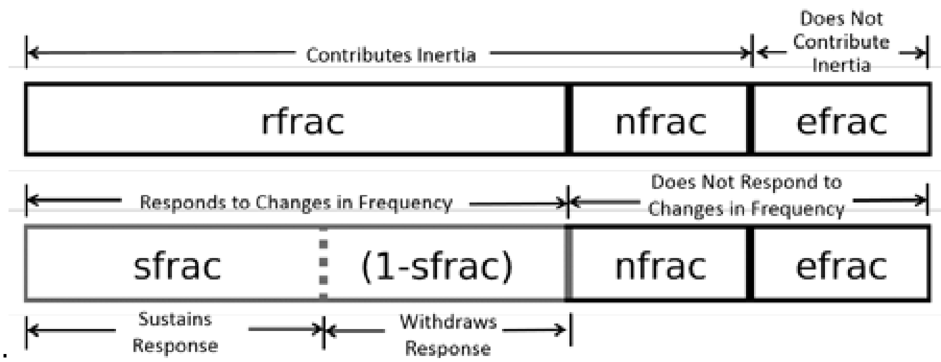


Figure 19. Generation Modeling Schema for Study of Interconnection Frequency Response

Explicit assumptions about the control of each generator turbine-governor and about generator dispatch are necessary for the production-grade dynamic system modeling studies conducted by planners for each of the three U.S. interconnections. Accordingly, as starting points for our simulations, we reviewed the dispatch point and generator turbine-governor settings contained in recent models from each interconnection. This enabled us to understand the amount and types of generation that participate in primary frequency control (i.e., generator turbine-governor settings) as represented in these system models. See Table 3 and Table 4.

Table 3. Interconnection Generation Characteristics Based on Recent Industry System Planning Models

Interconnection and Planning Model Case Name	Total Synchronous Generation	Total Real Power (GW)	*Efrac* *Nfrac* *Rfrac*	*Sfrac*	Design Event/ Total Gen
Eastern *MMWG_2016SUM_2015*	878 GVA*	666	*Efrac* = 1 % *Nfrac* = 32% *Rfrac* = 67%	66%	0.7%
Western *WECC 18LW2*	169 GVA	99	*Efrac* = 4% *Nfrac* = 17% *Rfrac* = 79%	57%	2.8%
Texas *NT2018_2015*	57 GVA	43	*Efrac* = 30% *Nfrac* = 5% *Rfrac* = 65%	52%	6.3%

**GVA = gigavolts ampere*

Source: Developed by LBNL from Undrill, et al. (2018): Relating the Microcosm Simulations to Full-Scale Grid Simulations

Table 4. Interconnection Generation Dispatch Based on Recent Industry System Planning Models

	Eastern Interconnection *MMWG_2016SUM_2015*		Western Interconnection *WECC 18LW2*		Texas Interconnection *NT2018_2015*	
	GW	%	GW	%	GW	%
Steam	399	60%	20	20%	14	32%
Combined Cycle	66	10%	20	20%	7	16%
Hydro	19	3%	40	40%	0.9	2%
Simple Cycle	33	5%	2	2%	7	16%
Nuclear	66	10%	2	2%	2	4%
Nonresponsive	71	11%	12	12%	0	0%
Electronic	8	1%	4	4%	13	30%
Total	**666**	**100%**	**100**	**100%**	**44**	**100%**

Source: Developed by LBNL from Undrill, et al. (2018): Relating the Microcosm Simulations to Full-Scale Grid Simulations

Before proceeding, it is also useful to review how the historic dispatch information we received from the three interconnections supports the development of these assumptions, as well as the ranges in assumptions we will study using our system model.

To a first approximation, information on operating headroom is available from the historic dispatch information provided by the Western and Texas Interconnections. See Figure 20 and Figure 21. Figure 20 presents hourly spinning reserves, as calculated by PeakRC. Figure 21 presents unloaded, on-line capacity, which is the difference between on-line generation capacity (in MVA) and generation output (in MW) for 20 selected hours from the historic period.[65] These hours were selected because they represented times when either wind generation was high and/or system load was low. Note that both of these estimates should be greater than the headroom derived from the interconnection system models because they include both generation we label *Rfrac* and *Nfrac*, while the headroom derived from the interconnection system models represents headroom on only *Rfrac*.

There are two points to note: First, historic dispatch practices in the Western and Texas Interconnections vary considerably. Western Interconnection spinning reserve margins range from approximately 8 to 16 percent of total generation; the Texas Interconnection's unloaded capacity reserves were much larger. Second, the headroom available in the WECC system model was somewhat higher than Western Interconnection historic dispatch practices. The headroom available in the ERCOT system model was within the range we observed in Texas Interconnection historic dispatch practices.

The corroboration between system models and historic dispatch practices is necessarily inexact because the information on historic dispatch practices covers many hours of operation, while the system models are developed for dispatch at a single point in time (such as a single, lightly loaded early winter morning hour). More importantly, the information we obtained on historic dispatch practices did

[65] Unloaded capacity that might be available from nuclear generators is not included in this calculation because the purpose of the calculation is to estimate the headroom that is available on the responsive portion of generation, *Rfrac*, which does not include nuclear generation.

not contain important information on the operation of individual generators. For example, the information we obtained does not identify which specific generators were responsive to frequency deviations (i.e., *Rfrac*) nor consequently which of these generators could sustain provision of primary frequency response (*Sfrac*). Hence, information on historic dispatch practices should be expected to lead to headroom estimates that exceed what might be found in a system model. Of course, this does not explain why the headroom assumption in the WECC system model exceeds that found in historic dispatch practices in this interconnection.

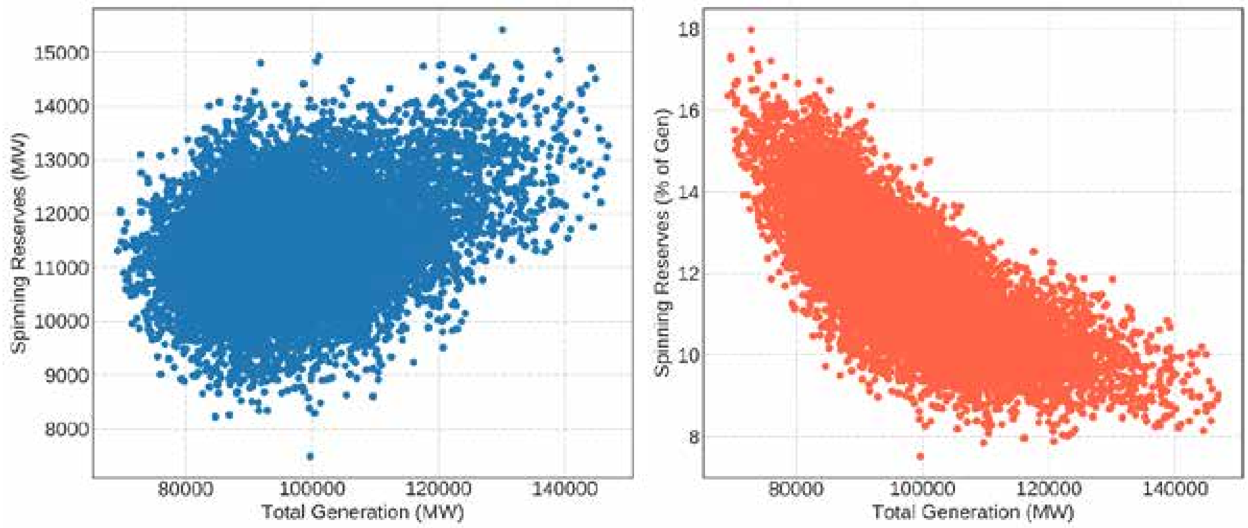


Figure 20. WECC Spinning Reserves (January 2016 through March 2017)

Source: Communication from PeakRC

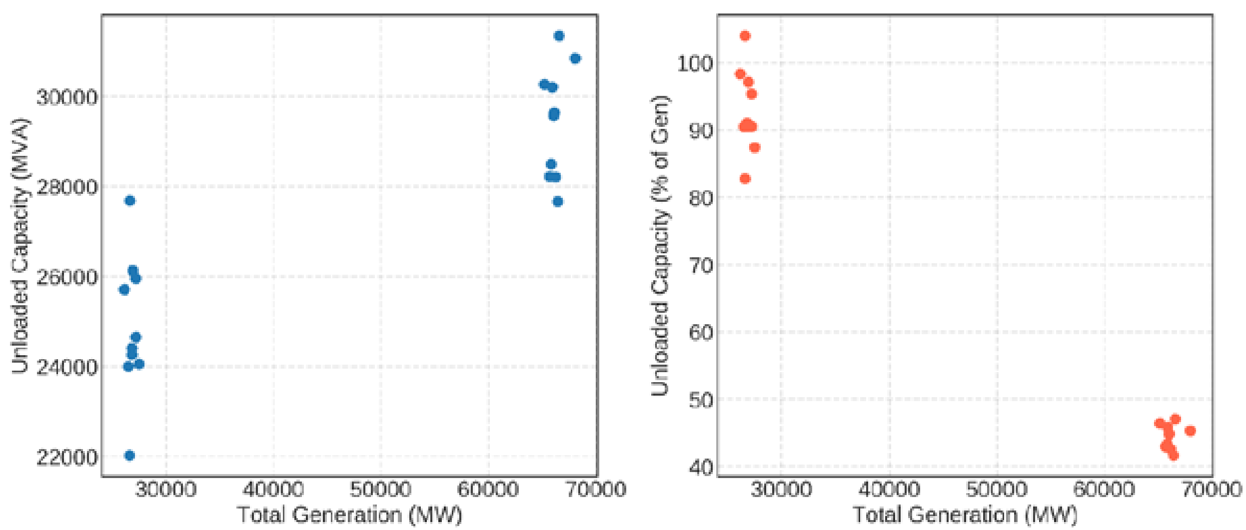


Figure 21. Texas Interconnection Unloaded Capacity Reserves: 20 Operating Hours during July 2016 through December 2016

Source: LBNL calculations based on information provided by ERCOT (note unloaded nuclear units are not included in LBNL's calculation)

To examine limiting cases for primary frequency control, we assume that the headroom available on responsive units is nine percent in all of our simulations. Thus, *Rfrac,* along with this assumption, define an upper bound on the primary frequency control that is available to respond to generation-loss events. It is an upper bound because, as discussed in Section 2.5, the amount of primary frequency response that can be delivered from this headroom is limited by the highest set-point for UFLS in conjunction with the droop setting on the turbine-governors.

It is important to recognize the significance of our assumption regarding headroom. On the one hand, it may be deemed conservative because historic dispatch practices (and the system models we reviewed) suggest that more headroom may be available. On the other hand, it may be deemed aggressive because, for example, WECC regional reliability standards require three percent spinning reserve calculated based on forecast daily peak load.[66] [67] Regardless, we emphasize that it has been selected in order to illustrate the reliability risks posed by not having enough primary frequency response.

4.4.2 Technology-specific Turbine-Governor Models and the Relative Proportions of these Technologies that Respond to Frequency

The responsive fraction of generation encompasses a wide variety of generator types and their associated turbine-governors. For brevity and ease of illustrating key turbine-governing concepts, we group each responsive generator into one of four technologies: steam turbines (Steam), combined-cycle gas turbines (CCGT), combined-cycle steam turbines (CCST), and hydro turbines (Hydro). We assign a single technology-specific turbine-governor model from the PSLF library to each type.[68] The droop setting on all turbine-governors is four percent, which is consistent with industry practice.

Each turbine-governor model has a characteristic rate at which it can increase a generator's output as a function of time. Figure 22 depicts the rates at which the four turbine-generators respond, expressed as a decimal percentage increase above an initial starting value, to facilitate direct comparisons. Both stand-alone steam turbines and gas turbines in a combined-cycle plant can increase their outputs rapidly, but the steam turbines in a combined-cycle plant and hydro turbines will increase their outputs

[66] More precisely, the WECC regional reliability standard, BAL-002-WECC-2a, requires contingency reserves equal to the greater of either the loss of the single largest contingency or the sum of three percent of hourly integrated Load plus three percent of hourly integrated generation (Requirement R1). Requirement R2 then states that responsible entities must keep at least half of its minimum amount of contingency reserve identified in Requirement R1, as operating reserve–spinning that meets both of the following reserve characteristics: Reserve that is immediately and automatically responsive to frequency deviations through the action of a governor or other control system; and reserve that is capable of fully responding within ten minutes.

[67] Note, however, that our assumption of nine percent headroom is calculated relative to the responsive fraction of generation, which is generally less than the basis used to calculate WECC's three percent spinning reserve requirement because responsive generation is always less than forecasted daily peak load.

[68] Although the PSLF library contains many turbine-governors, including options for users to create their own turbine-governor models, the turbine-governor models selected for this study represent the most up-to-date and flexible models currently available. As a result, they are widely specified in the system models we obtained for each interconnection and, hence, are broadly representative of the turbine-types examined in this study.

more gradually.[69] The rate at which a steam turbine in a combined-cycle plant can increase its output is determined by the relatively slow dynamic response of the heat recovery steam generator. The steam turbine output follows the gas turbine output, but does so with substantial lag. The primary frequency response delivered from a fleet of generators is the aggregated sum of the contributions of each generator.

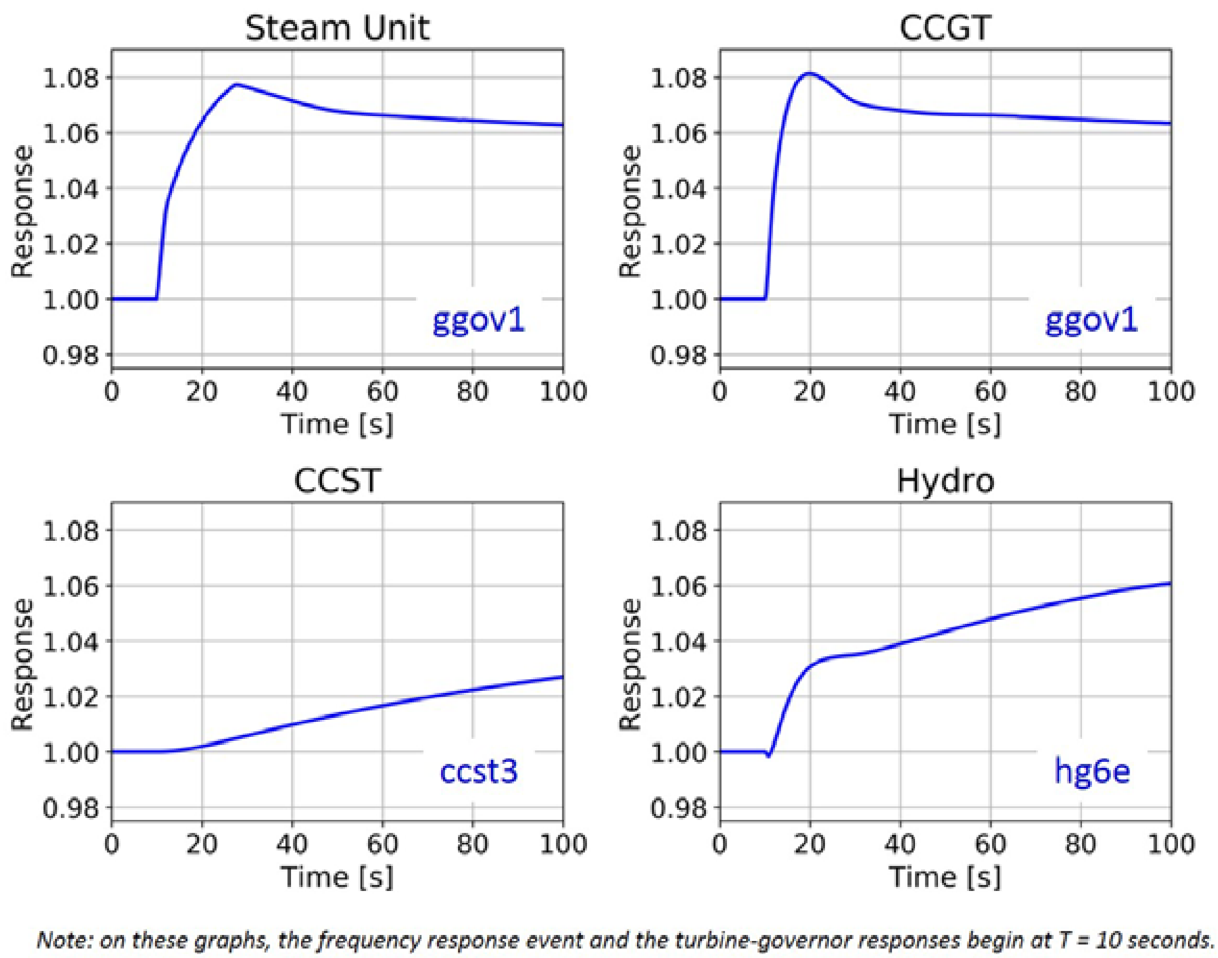


Figure 22. Turbine-Governor Performance

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

We modeled two compositions of frequency-responsive turbine-governors. The first, called the “Thermal” system, is dominated by combined-cycle and steam plants. The second, called the “Thermal-Hydro” system, features lower proportional output from combined-cycle and steam plants that is replaced by greater proportional output from hydro plants. See Figure 23.

69 The figure also indicates the name of the turbine-governor model specified in the PSLF library to represent the behavior of each generator type. Note that the same turbine-governor model is used to represent the behavior of both the steam turbine plant and the gas-turbine plant. Differences in the behaviors of these two turbine-governors are specified through the adjustment of modeling parameters that are associated with each turbine-governor model.

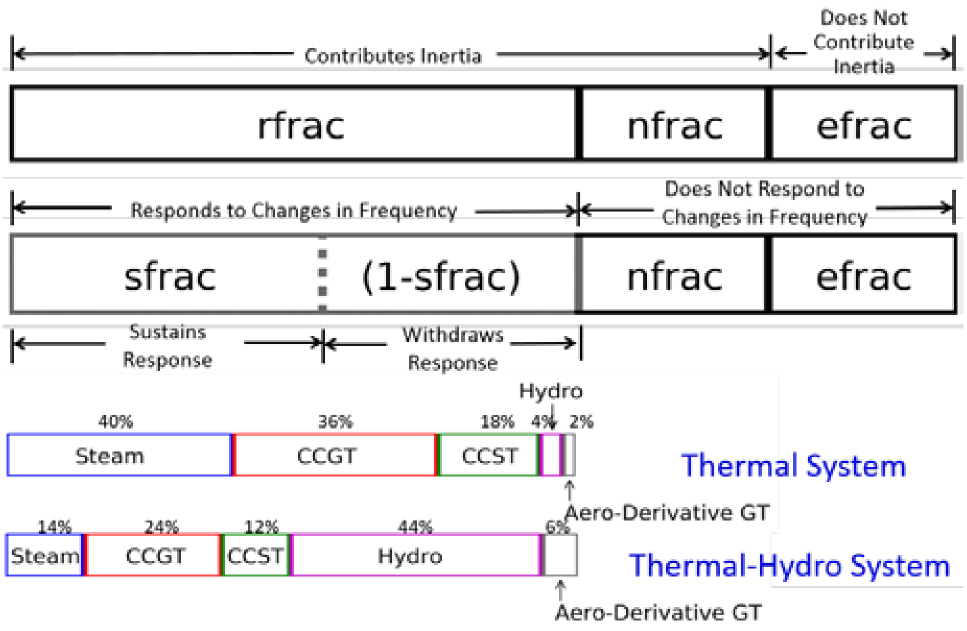


Figure 23. Generation Modeling Schema and Composition of the Responsive Fraction of Generation for the Two Systems Used to Study Interconnection Frequency Response

The thermal system is broadly representative of the composition of generation in the Eastern and the Texas Interconnections. The thermal-hydro system is broadly representative of composition of generation in Western Interconnection. The proportions were developed, in part, based on review of historical operating dispatch records for the Texas and Western Interconnections mentioned earlier. See Figure 24 and Figure 25, respectively. These figures illustrate the amount of headroom by types of generation from selected hours of dispatch for each interconnection. Figure 24 is developed from five separate hours of dispatch in the Texas Interconnection when total system load was close to minimum and wind generation was high relative to other hours when total system load was low. Figure 25 is developed from five similar hours of dispatch in the Western Interconnection when system loads were low and renewable generation was high relative to other hours when total system load was low. We caution, again, that our reviews of historic dispatch can provide no more than approximate information to guide the composition of the thermal and thermal-hydro systems because, as mentioned, information in the dispatch records does not indicate which on-line generators with headroom were, in fact, responsive to frequency.

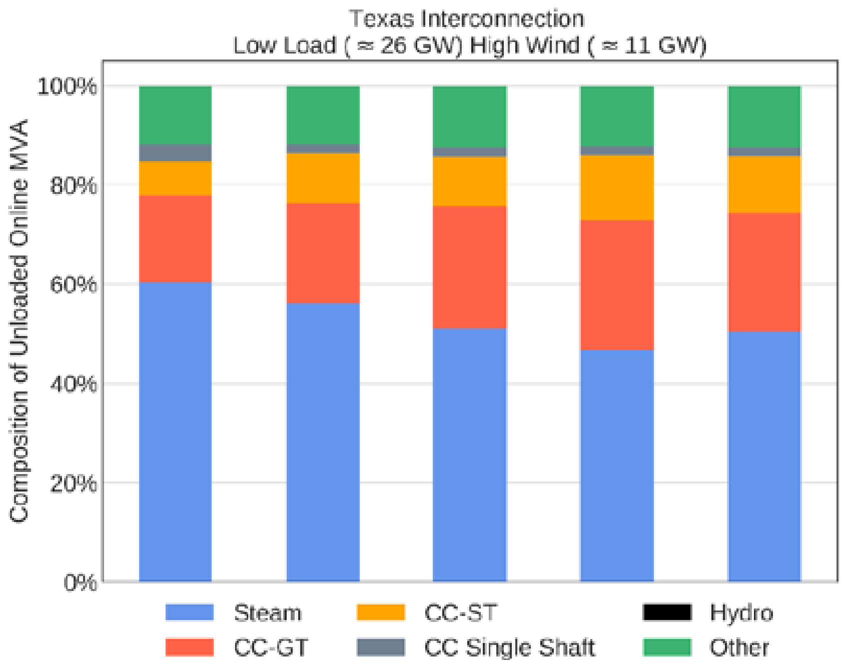


Figure 24. Snapshots of the Composition of Generator Headroom from Five Separate Hours of Dispatch for Texas Interconnection when System Load was Low and Renewable Generation was High between July 2016 and December 2016

Source: LBNL calculations based on information provided by ERCOT

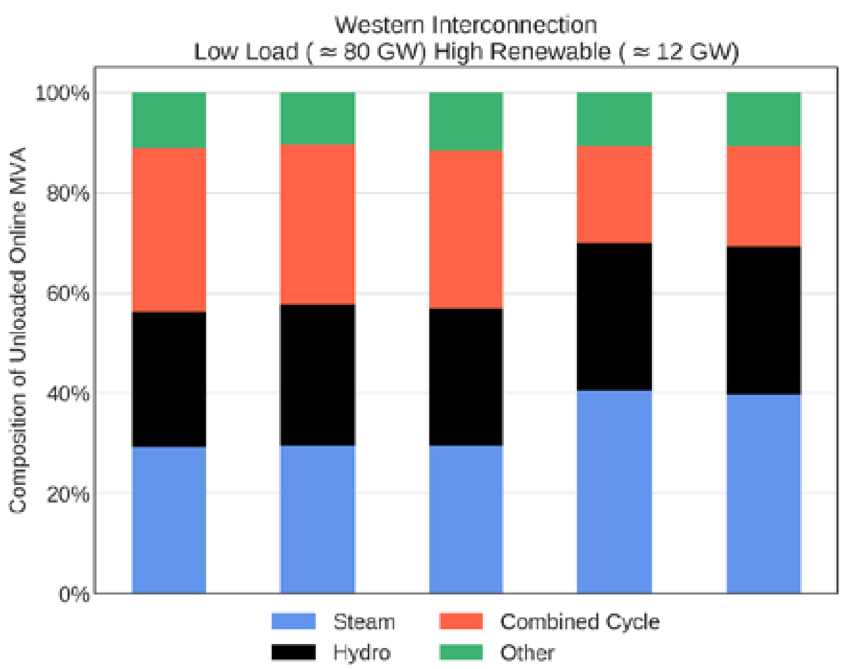


Figure 25. Snapshots of the Composition of Generation for 5 Separate Hours of Dispatch for the Western Interconnection When System Load was Low and the Contribution of Renewable Energy Generation was High between August 2017 and March 2017

Source: LBNL calculations based on information provided by PeakRC

4.4.3 Plant Load Controllers and Withdrawal of Primary Frequency Response

As noted in Section 2 and observed in studies of recorded interconnection frequency response events (especially in the Eastern Interconnection), withdrawal of primary frequency response inhibits and delays an interconnection's ability to restore frequency following the loss of generation. In this study, we model this phenomenon by modifying the plant load control settings that supervise the four types of turbine-governors described in the previous subsection.

Specifically, we distinguish between two populations of generators that are responsive to frequency (i.e., two separate populations that comprise *Rfrac*): those that sustain primary frequency response, called the "sustaining fraction" (*Sfrac*), and those that provide but then withdraw primary frequency response (1 - *Sfrac*).

For the population that withdraws frequency response (1 - *Sfrac*), we consider several types of plant load controls. The majority of simulations are conducted with plant load controls that simply restore generator output to its original, pre-event set-point. This is called pre-selected load mode (or MW-set-point control). The effects of other types of plant load control approaches are also examined. One type models the effects of pre-selected load mode with frequency bias. The effects of these plant load controllers on primary frequency response are illustrated and discussed in detail in Section 5.

Finally, and distinct from *Sfrac,* temperature limits are applied to gas turbines. These limits are not a form of secondary control. They can be understood as a reduction of the headroom that is available for provision of primary frequency response. Their effect is to limit the ability of gas turbines to provide or sustain primary frequency response while frequency remains depressed.

4.5 Review of Generation Modeling Procedures

This section gives an example of how the structured set of generator characteristics we have specified for our modeling approach is used to conduct the simulation-based sensitivity analysis of primary frequency response presented in Section 5. See Figure 26.

In this example, 10 percent of generation is electronically coupled to the power system and therefore does not contribute inertia (*Efrac* = 10 percent). Returning to the example presented above, if the effective system inertia of the portion of this power system (1 - *Efrac* = 90 percent) that does contribute inertia is 4 seconds, then the effective system inertia of this entire power system is, to a first approximation, about 3.6 seconds (4 seconds * 0.9).[70]

Turning next to the 90 percent of generation that contributes inertia, 10 percent of total generation contributes inertia but does not respond to frequency (*Nfrac* = 10 percent). The remaining 80 percent both contributes inertia and is responsive to changes in frequency (*Rfrac* = 1 - *Efrac* - *Nfrac*). Following

[70] Note that this is not an exact calculation because the headroom on the responsive fraction of generation is not considered. Moreover, these examples implicitly describe capacity in MW, whereas capacity in MVA is required for the calculation of a systems inertia constant and connected MVA is always greater than MW capacity.

our assumption of 9 percent headroom on the responsive fraction, the total headroom available is eight percent of total generation (80 percent responsive * 9 percent headroom).

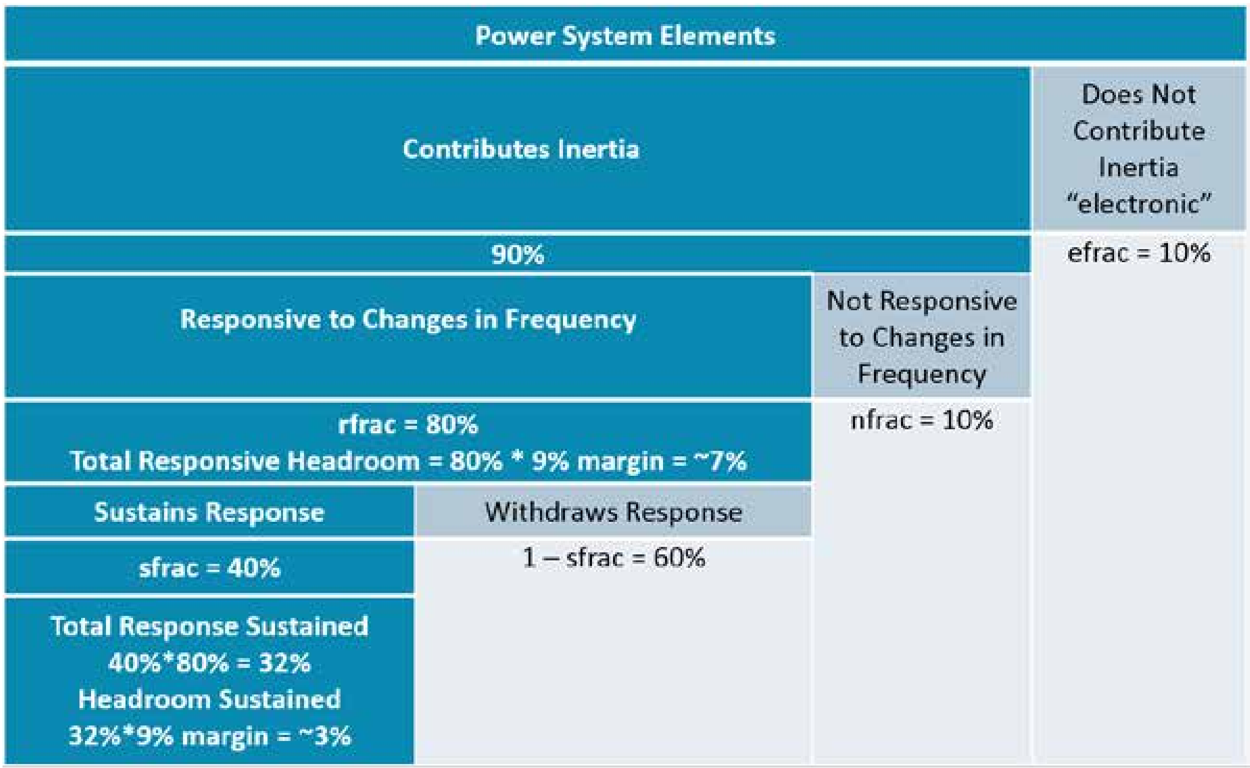


Figure 26. Illustration of the Relationships among Simulation Analysis Modeling Parameters

Hence, the total headroom available to deliver primary frequency response can, in principle, arrest frequency following a four-percent generation-loss event. However, two additional considerations will determine whether delivery of primary frequency response from this available headroom will be successful in arresting frequency above the highest set-point for UFLS. The first is whether or not primary frequency response be delivered fast enough to arrest frequency decline before it triggers UFLS. As discussed in Section 2.5, the amount of primary frequency response that can be delivered from a given amount of headroom is limited by both the droop setting on the turbine-governors, which establishes a target or amount by which output is increased, and, more importantly, the rate at which turbine-governors increase generator output toward this target. The second consideration is whether or not primary frequency response be sustained. If too much is not sustained, the resulting imbalance between generation and load will cause frequency to resume its decline. How much will be sustained is determined by the settings on plant load controllers and other forms of turbine controls, including temperature limits on gas turbines.

In this example, we assume only 40 percent of primary frequency response will be sustained (*Sfrac* = 40 percent) and that the balance, 60 percent, will be withdrawn by the actions of plant-level controls. Consequently, the total sustained primary frequency response is approximately 3 percent (40 percent sustained * 8 percent total primary frequency response available). By inspection, even delivery of all available headroom to provide primary frequency response (which is not feasible because of the first consideration listed above) will not arrest frequency following the loss of four percent of generation.

4.6 Treatment of Load Sensitivity

As discussed in Section 2.2, some portion of load responds to changes in interconnection frequency. Historically, load damping or load sensitivity supported the delivery of primary frequency response from generators. However, as the composition of load changes, load sensitivity is also changing. Many types of loads are coupled electronically to the grid and have controls that either do not respond to oppose changes in interconnection frequency, or respond by increasing power consumption and thereby accelerating the decline in interconnection frequency.

For the majority of simulations conducted for this study, we make the conservative assumption that load is not responsive to changes in interconnection frequency. By making this assumption, we focus exclusively on the role of generators in providing primary frequency response. We present selective sensitivities in Section 5 where we relax this assumption and show the relative impact of increased load sensitivity on the ability of primary frequency response to arrest frequency following a loss of generation.

Section 5 also considers examples of demand response in which pre-determined relatively small (compared to UFLS) blocks of load are disconnected, through voluntary agreements with customers, when frequency falls to certain triggering values that are above the highest set-points for UFLS.

4.7 Comparison of Simulation Results with Industry-Developed Interconnection Models

Before turning to the frequency control insights we obtained using our simplified model, it is instructive to compare the simulation results that our model produces with those produced by the highly detailed, production-grade industry models of each interconnection. The goal of this comparison is not to formally validate or calibrate the simplified model. Rather, the goal is to confirm that the simplified model we developed provides frequency response insights that are credible and, more importantly, relevant to the current and potential future operating conditions of the three U.S. interconnections.

As noted throughout this section, the structural elements of our simplified model are intended to capture the principal factors that determine interconnection frequency response. The findings presented in Section 5 are a result of varying many of these factors individually and in conjunction with one another. For the purpose of comparing models, however, we review simulations prepared using values we found or derived through our examination of the interconnection models we obtained from industry (See Table 3).

Figure 27 through Figure 29, taken from Undrill, et al. (2018), compare frequency response estimated using the detailed system models obtained from industry to multiple simulations using the simplified model developed for this study for the Texas, Western, and Eastern Interconnections, respectively. The design generation-loss event required by the BAL-003-1.1 standard for each interconnection is simulated.

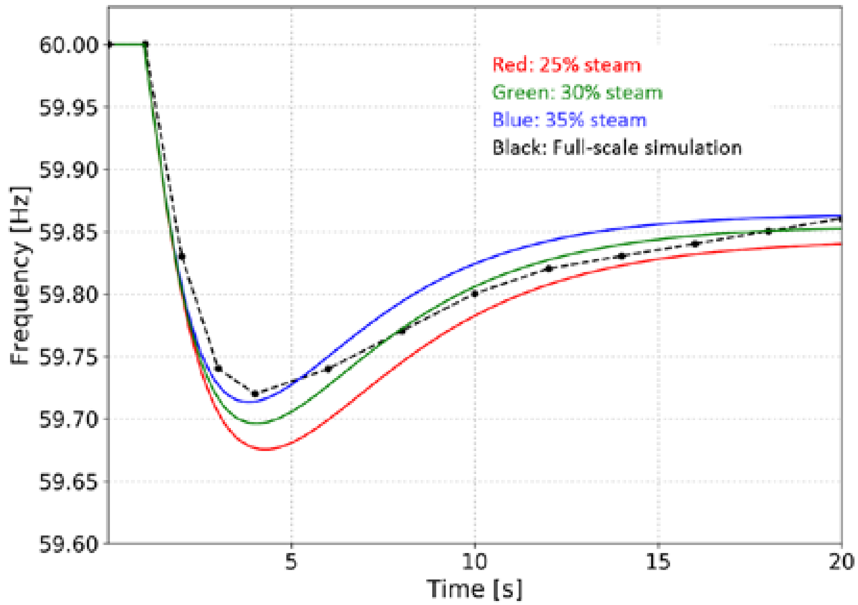


Figure 27. Comparison of Simulations of System Model Developed for this Study to the Industry-developed System Model for the Texas Interconnection

Source: Developed by LBNL from Undrill et al. (2018): Relating the Microcosm Simulations to Full-Scale Grid Simulations

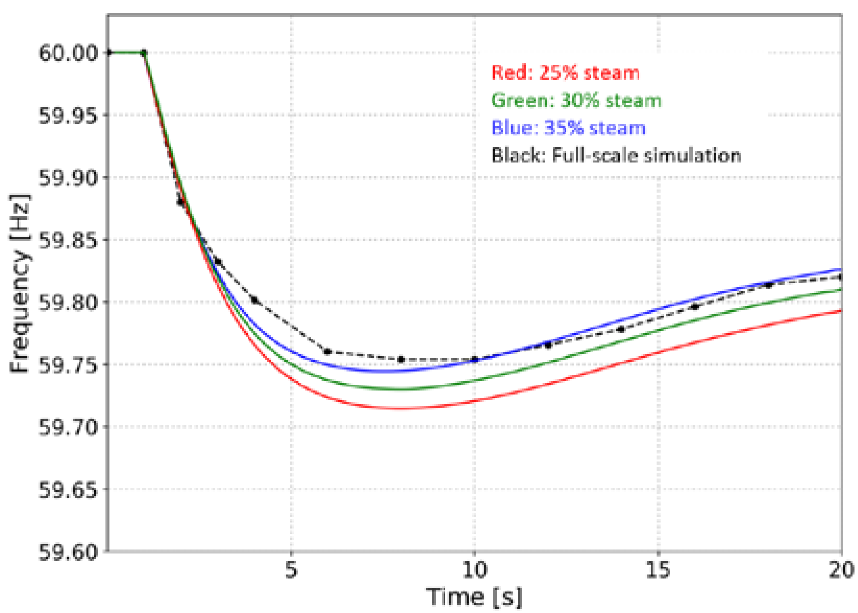


Figure 28. Comparison of Simulations of System Model Developed for this Study to the Industry-developed System Model for the Western Interconnection

Source: Developed by LBNL from Undrill et al. (2018): Relating the Microcosm Simulations to Full-Scale Grid Simulations

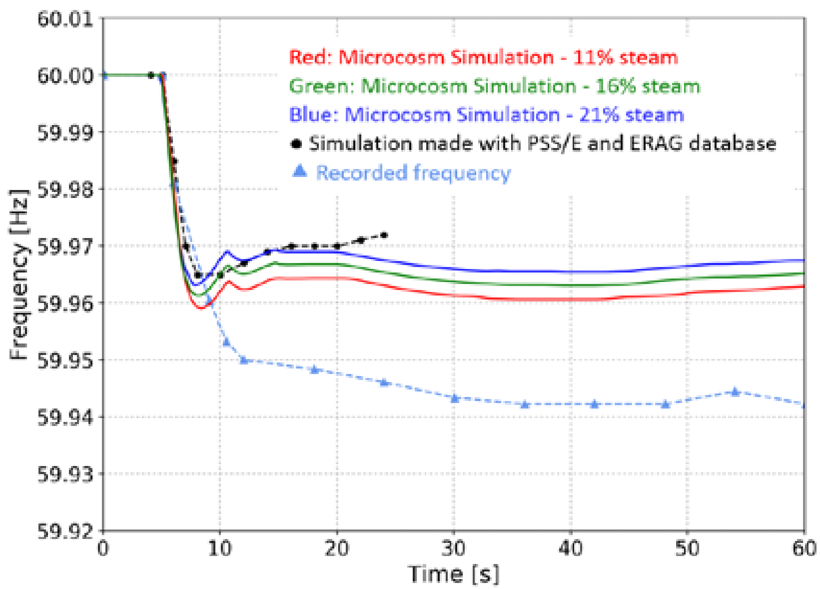


Figure 29. Comparison of Simulations of System Model Developed for this Study to the Industry-developed System Model for the Eastern Interconnection and Measured System Performance

Source: Developed by LBNL from Undrill et al. (2018): Relating the Microcosm Simulations to Full-Scale Grid Simulations

Frequency traces are plotted in red, green, and blue for three different simulations using the simplified model. The dashed black line is the frequency trace predicted using the industry system model. The dashed blue line with triangles is the recorded behavior of the Eastern Interconnection to the large generation-loss event.

For all three interconnections, the simulations conducted using the simplified model developed for this study show good agreement, with two key, anchoring features of simulations conducted using the detailed models developed by industry. First, the initial ROCOF is virtually identical between the two models. This means that the salient features of interconnection inertia in conjunction with the size of the generation-loss event relative to the size of the interconnection have been accurately captured and reproduced. Second, all of the frequency traces are well on their way toward convergence by the end of the simulations. This means that the fraction of generation participating in frequency response in conjunction with the droop specified for the turbine-governors has also been accurately captured and reproduced.

The principal differences between the two models appear during the intervening period between the initial decline in frequency and the final settling frequency. The differences include the frequency and time at which the nadir is formed. These differences are modest and of secondary importance for the purpose served by these comparisons, which is to demonstrate that our model can realistically reproduce the principal salient features of interconnection frequency response predicted using large-scale models. That is, these differences are reflective of subtle differences in the composition of turbine-governors in the responsive fraction of the fleet and, more importantly, on their control settings.

Some perspective on these subtleties can be seen by examining the sensitivities developed using our simplified model, which are also found in Figure 27 through Figure 29. These sensitivities involve modest changes in the relative amounts of responsive steam turbines while holding *Rfrac* constant. These changes move the frequency curves both closer in and farther from alignment with the simulations produced using industry's models.

We maintain that these differences are of secondary importance for the purpose of this study because resolving them would involve further "tuning" other input assumptions. Further tuning of such assumptions is not particularly meaningful because, in fact, there are more "tunable" parameters than there are measured data to support these efforts. For example, the information available in the interconnection models we obtained, especially those for the Eastern and the Texas Interconnections do not provide definitive information on the amount of generation that contributes primary response. Hence the derivations of *Rfrac* presented in Table 3 are only approximate, which is why we have presented sensitivities around the base values we developed. Consequently, we maintain that the comparisons have adequately established the usefulness of the simplified model for studying the fundamental features and overall performance characteristics of interconnection frequency response.

Finally, it is important to note that comparison of simulations to one another should not be confused with comparisons of simulations to reality. Figure 29 shows that the industry-developed model of the Eastern Interconnection does not reproduce the recorded behavior of the Interconnection to a large loss-of-generation events, which is shown with blue triangles. This finding is consistent with that found by LBNL in the 2010 Study. Undrill, et al. (2018) also presents another set of simulations conducted using the simplified model that accurately captures the recorded behavior of the Eastern Interconnection by reducing *Rfrac* and decreasing *Sfrac*. This additional set of simulations further supports the usefulness of the simplified model we have developed for use in studying interconnection frequency response.

4.8 Summary

We developed a highly flexible modeling approach to illustrate key relationships and interactions among the factors that influence interconnection frequency response. The approach aggregated generators according to whether they do or do not (1) respond to frequency deviations (i.e., provide primary frequency response); (2) sustain primary frequency response and; (3) contribute inertia to the interconnection.

We implemented the modeling approach by conducting dynamic simulations using GE's PSLF tool—the same commercially available tool that is currently in wide use by industry to conduct, among other things, production-grade studies of frequency response. By using a commercially available tool, we were able to study the performance of turbine-governors and plant load controllers for different types of generators (e.g., combined-cycle, hydro, and steam) using the same models of these generators that are used by industry to conduct reliability studies for planning and operations.[71]

The models we developed allow us to examine: (1) the interconnection requirements for primary frequency response; (2) the headroom available on generators, which establishes an upper bound on the amount of primary frequency response; (3) the rate at which turbine-governors deliver primary frequency response from this headroom; and (4) plant-specific control settings or operating factors that limit or withdraw primary frequency response early (i.e., before frequency has been stabilized). We also examine fast demand response, governor deadband settings, and load sensitivity (sometimes called *load damping*),[72] which also contribute to frequency response. We compared the performance of our simplified study model to the production-grade models developed by industry for each interconnection to confirm that we can meaningfully capture the relevant features of frequency response as predicted by industry models.

---

[71] We emphasize, however, that the modeling approach we developed was intentionally simplified in order to focus on the interactions among the central physical factors influencing and resource performance characteristics required for reliable interconnection frequency response. It is not intended to replicate all aspects of, nor displace the need for interconnection- and system-level modeling conducted by industry.

[72] The majority of our simulations assumed no load sensitivity in order to focus attention on the relationship between primary frequency control provided by active sources, such as generators, and interconnection frequency response.

5. Frequency Control Findings

This section presents frequency control findings obtained using the tool described in Section 4. The findings are based on results from parametric simulations that varied the factors that influence interconnection frequency response either individually (holding all other assumptions fixed) or jointly (varying two or more factors in the same simulation). The findings are organized into three broad groups. First, we present basic findings on the factors that determine the requirements for, and initial adequacy of, primary frequency response (Sections 5.1–5.3). Second, we present findings that illustrate the reliability risk posed by non-sustaining primary frequency response and examine three independent means by which these risks arise, including a method to prevent the predominant action by which primary frequency response is not sustained; this method entails specifying an alternate control logic for plant load controllers (Sections 5.4–5.5). Third, we present findings on additional frequency response topics including fast demand response, governor deadband settings, and load sensitivity (Sections 5.9–5.11).

5.1 Reserves Held to Provide Primary Frequency Control Must Exceed the Expected Loss of Generation

As introduced in Section 2, interconnection frequency is a reflection of the balance between generation and load. The rapid decline in frequency following a loss of generation results from the sudden imbalance that this loss creates between generation and load. It follows that frequency is arrested or stabilized once the balance between generation and load has been re-established. In other words, the frequency nadir is formed when the amount of power or generation delivered through primary frequency response equals the amount of generation that was lost.[73] This is not a new or novel finding of this study; however confirmation and illustration of it as a fundamental principle forms the basis for all subsequent findings in this study.

Figure 30 illustrates this relationship by showing how frequency evolves (upper panel) and primary frequency response is delivered (lower panel) over time. Two different systems were simulated. One specifies an effective inertia constant of four seconds for all generation that contributes inertia (indicated in red). The other specifies an effective inertia constant of three seconds for all generation that contributes inertia (indicated in blue). Both simulations were conducted assuming the loss of generation equal to two percent of total system load and with the same quantity and quality of reserves held to provide primary frequency control.

[73] For the purposes of this discussion, the effect of load sensitivity can be considered a contributor to primary frequency response. In fact, we assumed no load sensitivity in our initial simulations in order to focus attention on the relationship between primary frequency control provided by active sources, such as generators, and interconnection frequency response. Load sensitivity is examined in Section 5.11.

As discussed in Section 2, frequency declines faster (i.e., ROCOF is higher) in a system with lower-inertia generators.[74] Once frequency begins to decline, turbine-governors immediately begin increasing generator output. Because the reserves held to provide primary frequency control and the droop settings on the turbine-governors are identical, the rate at which total generation increases is nearly identical in both systems. Yet, because frequency is declining faster for the system with lower inertia, the droop setting seeks to increase generator output at a slightly faster rate.

Though it may be difficult to observe visually, frequency is, in fact, falling faster than generation is increasing in the system with lower inertia. The net result is that the nadir is formed earlier but at a lower value than in the system with higher inertia. Nevertheless, in both systems, the nadir is only formed when the amount of primary frequency response delivered is equal to the amount of generation lost (two percent). See the horizontal, dashed green line in the lower panel of Figure 30.

"If the total amount of primary frequency control held in reserve is less than the amount of generation lost, frequency will not be arrested prior to the triggering of UFLS."

It also follows that, if the total amount of primary frequency control held in reserve is less than the amount of generation lost, frequency cannot be arrested prior to the triggering of UFLS. (See, also, discussion in Section 2.4). Hence, prudence further dictates that the total primary frequency response capacity held on-line should exceed the size of the design generation-loss event to acknowledge uncertainty in the actual performance of the fleet.

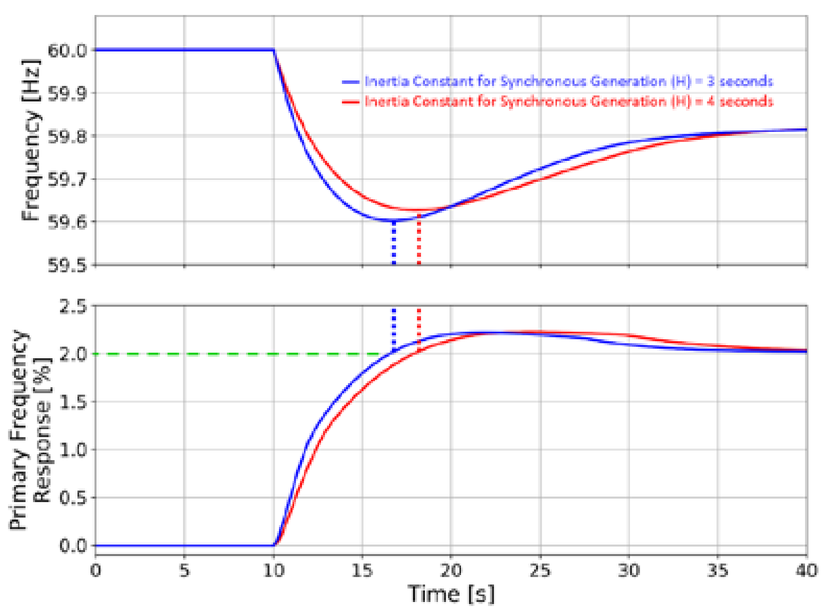


Figure 30. Frequency is Arrested when the Amount of Primary Frequency Response Delivered Equals the Amount of Generation Lost (2%)

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

[74] In fact, these two simulations are the same simulations shown in Figure 6 in Section 2.6. The visual scaling has been changed for ease of presentation.

5.2 Primary Frequency Response Must be Delivered Quickly, which Requires Many Participating Generators

If turbine-governors could deliver primary frequency response instantaneously, then the reserves held for primary frequency control would simply need to equal the amount of generation represented by the design generation-loss event.[75] Yet, as shown in Figure 22, the rate at which a turbine-governor increases generator output is not instantaneous; the rates differ depending on the type of generator. Steam turbines in stand-alone generators and the gas turbines in a combined-cycle generator can increase output rapidly. Steam turbines in combined-cycle generators and hydro generators increase output more gradually. The amount of primary frequency response reserves allocated to a given generator should not exceed the amount it can produce in the time available to arrest frequency decline. This finding, too, is generally well understood, but it may not be widely appreciated; hence, we illustrate it here.

ROCOF can be thought of as roughly establishing the time available for delivery of primary frequency response equal to the amount of generation lost to arrest frequency prior to frequency reaching the highest set-point for UFLS. For example, given a ROCOF of approximately 0.15 to 0.2 Hz/sec, which could result from a comparatively large generation-loss event (see Figure 6), if the loss is not opposed by primary frequency response, the highest UFLS set-point in the Western and Eastern Interconnections (59.5 Hz) will be triggered in roughly three seconds or less.

Three seconds is not enough time for even the fastest turbine-governor on a conventional generator to deliver all of the reserves it may be holding for primary frequency control. Referring again to Figure 22, it takes more than five seconds for the turbine-governor on a steam plant to deliver the full reserves held for primary frequency control. The turbine-governors on other types of generators can take even longer to deliver their frequency-responsive reserves (see, for example, hydro).

Consequently, for primary frequency response, the interconnection can only rely on the initial, modest increase in generation that can be produced by an individual participating generator. See Figure 31. The key to ensuring that enough total response will be produced fast enough is to rely on many generators, each contributing only the initial portion of the reserve capability that they can deliver in the time available for primary frequency control. The total amount contributed by each is in proportion to the size of the generator and is determined by the rate at which the turbine-governor can increase the generator's output.

[75] The findings on fast demand response in Section 5.9 explore the implications of a frequency control approach that might address load-generation imbalances nearly instantaneously.

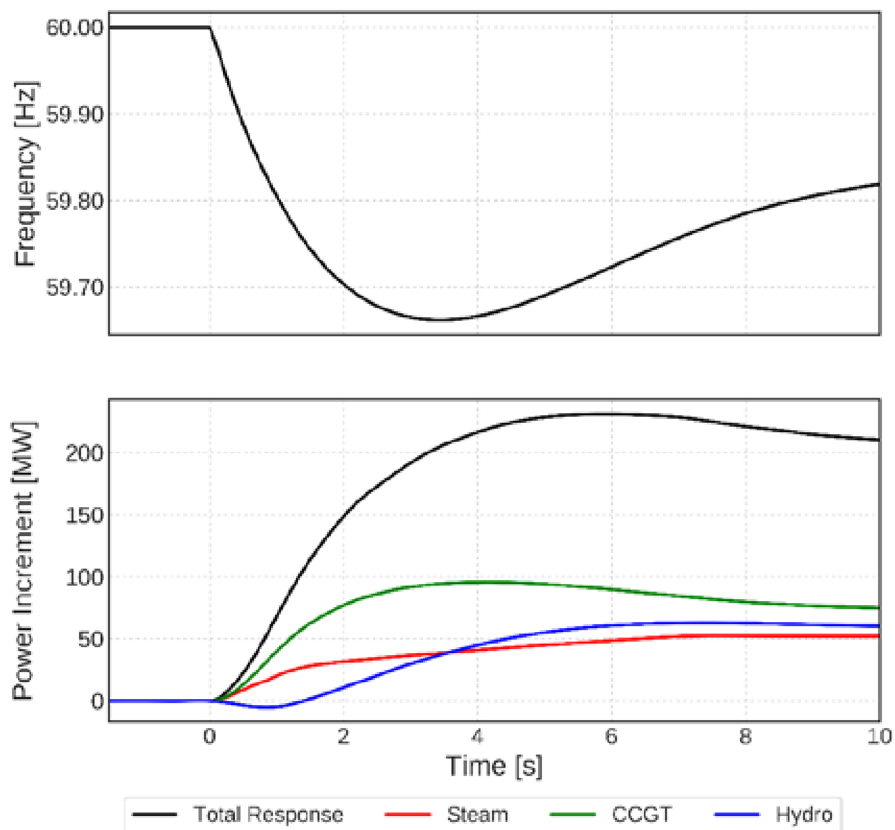


Figure 31. The initial contributions of primary frequency response from many generators is required for reliable interconnection frequency response

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

The risk to reliability is, therefore, reduced by drawing on these reserves from a larger, rather than smaller, number of generators. In the extreme, relying on too few generators to provide more response than they are each capable of providing in the time available means that the additional reserves they are carrying are moot. These additional reserves cannot be deployed in time to prevent declining frequency from triggering UFLS. Relying, instead, on many generators to contribute only a small portion of the total primary frequency response required by the interconnection (as determined by their capability to provide this response in a timely manner) is inherently less risky than relying on fewer generators to contribute a larger portion of the total required.

"…all generators [should] have the capability to provide primary frequency response."

As a consequence it is prudent to ensure, to the extent technically practical, that all generators—both those that are directly coupled and those that are electronically coupled to the grid—have the capability to provide primary frequency response. Ensuring this capability provides maximum flexibility to grid operators to assign primary frequency response duty to (or to procure primary frequency response from) generators as appropriate for the current grid operating conditions. Exceptions should be considered only on a case-by-case basis.

5.3 For a Given Loss of Generation, System Inertia and the Timing of Primary Frequency Response Determine How Frequency is Arrested

By simple summation, the amount of primary frequency response that must be delivered must be at least equal to the amount of generation lost. That the available amount of primary frequency response is enough to cover the loss is not sufficient, however. The timing with which the primary frequency response is delivered is critical.

Hence, for a given loss of generation, the key factors determining the nadir of frequency are (a) the total inertia of the system, which determines the initial rate of decline of frequency; and (b) the rate at which generation is increased by primary frequency response.

Figure 32 illustrates the relative effects of system inertia and the timing of primary frequency response on the arrest of frequency decline. In the top panel, the solid curves show the trajectory of frequency in a base condition (solid red curve) and with 22 percent of the rotating generation fleet exchanged (solid blue curve) for renewable generation, which reduces the system inertia by 22 percent. The effective inertia constant of the system with respect to the total rotating-plus-renewable capacity is reduced from 3.6 to 2.8 second.

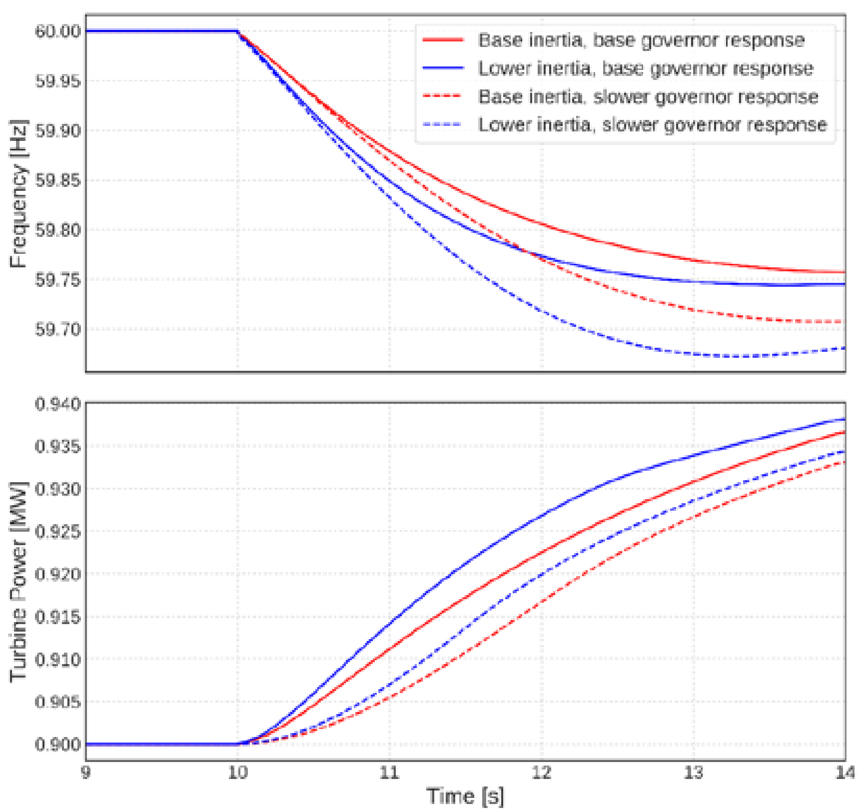


Figure 32. System Inertia and the Speed of Primary Frequency Response Determine Nadir at Which Frequency is Arrested

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

The dashed curves show the effect of the same reduction in system inertia when the dynamic response of the steam and gas turbine governors is slightly slowed down. The increase in generation by steam turbines is illustrated in the lower panel.

Comparing both of the solid curves and both of the dashed curves shows that the quite minor one-second delay in the primary response has as much effect on the frequency nadir as does the 22 percent change in inertia.

Thus, for a given loss of generation, understanding how changes in the composition of the generation fleet will affect the total inertia of the interconnections is important because it directly influences the requirements for primary frequency control and how reserves held for this control will perform during a generation-loss event.

Lower system inertia will require faster primary frequency response. Understanding the expected performance of the reserves that are held to respond, thus, becomes of greater importance. For example, if the reserves held are quick to respond, they may be adequate for a wide range of possible future generation-loss scenarios and corresponding system inertias. If, they are slow to respond, they may require augmentation by faster responding reserves.

5.4 Primary Frequency Response Must be Sustained until Secondary Frequency Response can Replace it

Figure 33 shows results from one of the simulations presented in Figure 30. Figure 33 also shows results from a second simulation in which the sustaining fraction of responsive generation has been reduced. Together, these simulations illustrate the importance of sustaining primary frequency response until secondary frequency response can replace it.

In the initial phase of the event, the frequency trends for the two simulations are nearly identical because the same amount of primary frequency response has been delivered. However, even as the nadir is being formed, the effect of a lower amount of sustaining primary frequency response can be observed, leading to a lower apparent settling frequency. As the event progresses, the non-sustaining portion continues to reduce the amount of primary frequency response delivered. At some point, frequency again begins to decline because the total amount of primary frequency response delivered is now less than the amount of generation lost. Moreover, at this point, there are also no additional reserves available to provide primary frequency control. As result, the decline in frequency would continue until UFLS is triggered.

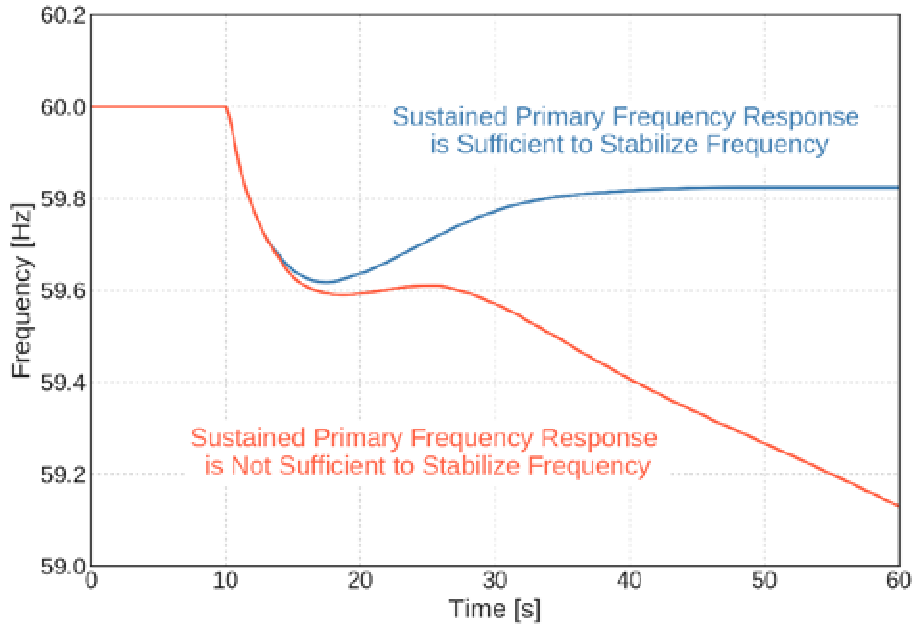


Figure 33. System Inertia and the Speed of Primary Frequency Response is Determine How Frequency is Arrested

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

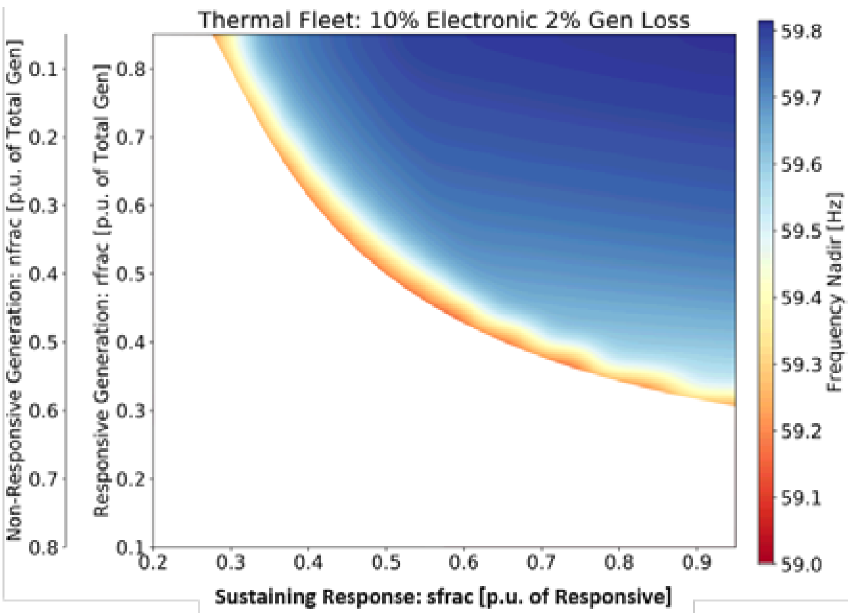
These simulations illustrate an important limitation of the majority of modeling studies of frequency response. Most modeling studies end the simulation period approximately 15 to 20 seconds after the generation is lost. This practice means that the full effect of non-sustaining primary frequency response is not seen in the simulation results because the full impact of non-sustained response on frequency control is not seen until after 20 seconds into a simulation.

"…if it is anticipated that primary frequency response will not be sustained, then the reserves held to provide sustaining primary frequency control will need to be increased."

The above example indicates that, if it is anticipated that primary frequency response will not be sustained, then the reserves held to provide sustaining primary frequency control will need to be increased. That is, if some providers of primary frequency response are not expected to sustain their participation shortly after a frequency response event has begun, then their value as reserves must be discounted. Specifically, it means that additional sources of sustaining primary frequency control must be kept on-line. In other words, the requirements for both immediate and sustained primary frequency response should be thought of as ones that must be met through a portfolio of sources, each contributing according to their individual physical characteristics.

Figure 34 illustrates a means of visualizing the interaction between total responsive generation (*Rfrac*) on the Y-axis and sustaining responsive generation (*Sfrac*) on the X-axis. This example was developed assuming a generation-loss event of two percent and an inertia constant of four seconds for the portion of generation contributing inertia (*Rfrac* + *Nfrac*). In addition, 10 percent of the fleet is electronic, contributing no inertia and no frequency response. The blue regions of the figure indicate the combinations of total responsive and sustaining responsive generation that will arrest frequency above the highest set-point for UFLS (59.5 Hz).

Looking at one extreme, if roughly 40 percent of the generation is responsive to frequency (*Rfrac* = 40 percent), then roughly 70 percent of this response must be sustained (*Sfrac* = 70 percent) to arrest frequency above 59.5 Hz. Similarly, looking at the other extreme, if roughly 70 percent of the generation is responsive to frequency (*Rfrac* = 70 percent), then roughly 40 percent of this response must be sustained (*Sfrac* = 40 percent) to arrest frequency above 59.5 Hz. This examples indicate that roughly 28 percent of generation (28 percent = 40 percent * 70 percent or 70 percent * 40 percent, respectively) must be responsive to frequency and also able to sustain primary frequency response.



Note: White areas indicate combinations of responsive and sustaining generation that will not arrest frequency above 59Hz

Figure 34. Combinations of Responsive, and Responsive and Sustaining Generation Required to Avoid Triggering UFLS

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

There are several means by which primary frequency response may not be sustained. The first is through withdrawal of primary frequency response by the actions of plant-level controllers, which override and reset the actions of the turbine-governor responding to frequency deviations. We describe this finding first because it is an action that is directed by the plant owner/operator. As such, it is one that can be corrected by a plant owner/operator, which we discuss in Section 5.5. The second is through actions stemming from inherent physical characteristics of turbine generators. One example is exhaust gas temperature limits on gas turbines, which are intrinsic to the current design of these types of turbines. Unlike plant-level controllers, these actions cannot be overridden or corrected. We conclude by discussing what we have observed in published information on wind turbines providing what is called "synthetic inertia."

The bottom line, however, is that if primary frequency response from some sources will not be sustained, primary frequency response from additional sources may be required. The requirement is to sustain primary frequency response until it can be replaced by secondary frequency response.

5.5 Plant Load Controllers that Operate in Pre-Selected Load Mode without Frequency Bias will Withdraw and Not Sustain Primary Frequency Response

As discussed in Section 2, primary frequency control is executed locally and autonomously by turbine-governors. The set-points that establish the ranges around which turbine-governors operate are, in turn, directed by secondary controls or plant load controllers.

Secondary controls at the plant level can operate in various modes, some of which are noted in Undrill (2018). One mode of operation, pre-selected load mode, which is widely used in the United States (though not in the Texas Interconnection), acts to maintain generator output at a present level and thus withdraws the change in output that the primary control has made in response to change of frequency. The purpose is to meet commercial obligations, which direct the sale of electricity generated by the plant according to a pre-determined schedule. Consequently, when generation output increases as the turbine-governor responds to a decrease in interconnection frequency (a frequency response event), the plant load controller responds by automatically resetting and restoring output to the scheduled value. In other words, in this situation the plant load controller will cause the pre-determined, scheduled operation to override the primary frequency response.

"...early withdrawal of primary frequency response requires holding additional reserves of primary frequency control that will sustain their response."

Figure 35 compares the effects of plant load controllers that follow this control logic to those that do not, for each of the turbine-governor types modeled in our study. See Figure 22. The figure shows that plant load controllers act rapidly to withdraw primary frequency response. Withdrawal begins approximately 5 seconds after the initial delivery of primary frequency response. In every instance, primary frequency response is withdrawn prior to reaching its full, potential output. This is the reason that, in Figure 34, the nadir of frequency is lower in the simulation in which less response is sustained, compared to the simulation in which more response is sustained.

As noted above, early withdrawal of primary frequency response requires holding additional reserves of primary frequency control that will sustain their response, because early withdrawal means that the withdrawn reserves will not support full restoration of frequency. Consequently, reserves that will sustain their response must be in place, either instead of (or in addition to) those that will not sustain their response.

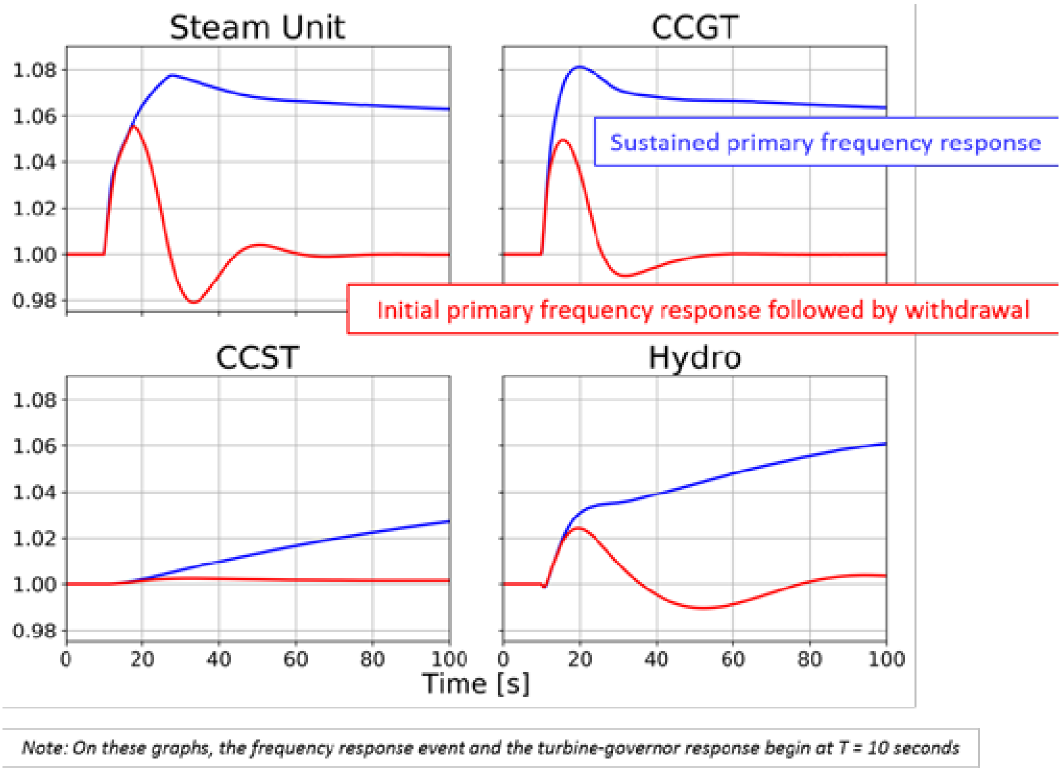


Figure 35. Turbine-Governor Primary Frequency Response When Operated With and Without Pre-selected Load Mode

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

5.6 Plant Load Controllers Operated in Pre-Selected Load Mode with Frequency Bias will Sustain Primary Frequency Response

The early withdrawal of primary frequency response by plant load controllers seeking to return generation to a MW set-point can be prevented by specifying a different control logic for the plant load controller than that described in Section 5.5. This control logic operates the generator at the MW set-point only when the frequency of the interconnection is at (or very close to) its normal operating value of 60 Hz. When frequency deviates significantly from 60 Hz (e.g., because of a loss-of-generation event), the plant load controller does not override the turbine-governor but instead allows the turbine-governor to continue delivering primary frequency response until interconnection frequency returns to the scheduled value. This control logic is called “pre-selected load mode with frequency bias.”

The top panel of Figure 36 compares the effect on interconnection frequency of pre-selected load mode control to the effect of two forms of pre-selected load mode plus frequency bias control. Increasing the amount of frequency bias (expressed as a percentage of governor droop) increases the amount of primary frequency response that is sustained by each turbine-governor. Increasing the amount of primary frequency response that is sustained stabilizes the system at progressively higher settling frequencies. The lower two panels in Figure 36 illustrate how these controls modify the behavior of turbine-governors for steam turbines and combined-cycle gas turbines, respectively.

> *"...we recommend... that industry review and address barriers to implementation of pre-selected load mode with frequency bias..."*

Because of the above findings, we recommend, in Section 6 of this report, that industry review and address barriers to implementation of pre-selected load mode control with frequency bias as a replacement for pre-selected load mode control without frequency bias.

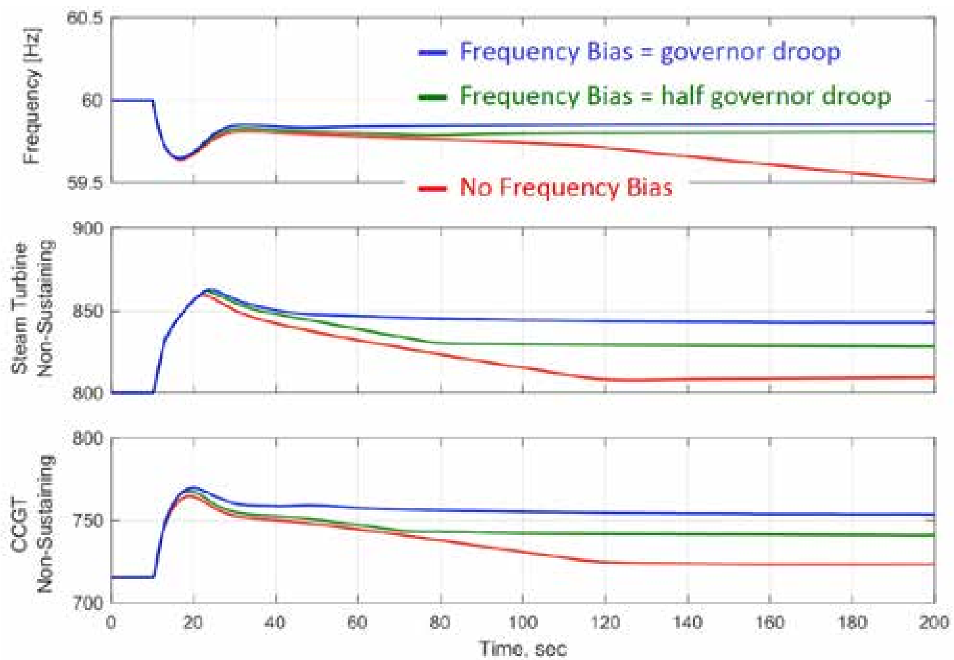


Figure 36. Operation in Pre-selected Load Mode with Frequency Bias Will Sustain Primary Frequency Response When Frequency is Less than Nominal

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

5.7 Gas Turbines May Not be able to Sustain Provision of Primary Frequency Response Following Large Loss-of-Generation Events

Gas turbines are among the fastest sources of primary frequency response. They are valuable contributors to the arrest of system frequency following the sudden loss of generation. They can readily increase their output by a few percent within a handful of seconds (5 to 8 seconds). As a result, they are excellent initial sources of primary frequency response. However, if an under-frequency event calls for maximum output, this output may not be sustainable due to the actions of a protection system of the turbine. Unlike the withdrawal of response by plant load-controls, reduction of output by this means cannot be deactivated at the discretion of the plant operator.

At less than nominal frequency, the gas turbine rotates more slowly and moves less air into/through the combustion process. Burning the same amount of fuel with less air means exhaust gas temperature will increase. If exhaust gas temperatures exceed a pre-set limit, the gas turbine will reduce output automatically in order to protect the turbine from damage.

Moreover, there is feedback between the exhaust gas temperature control and system frequency that can be detrimental to reliable interconnection frequency response. If, as the exhaust gas temperature controls reduce turbine output, system frequency continues to be depressed or decline, then the temperature limit controls will further reduce turbine output.

Figure 37 illustrates this effect. The lower panel shows the control actions directed by the turbine-governor (red) and the exhaust gas temperature controller (blue). Initially, the turbine-governor, responding to the decline in interconnection frequency, directs increased fuel flow. Once the gas turbine has reached its limiting temperature, the exhaust gas temperature controller overrides the turbine-governor and directs progressively lower levels of fuel flow. The top panel shows the impact of these control actions on interconnection frequency.

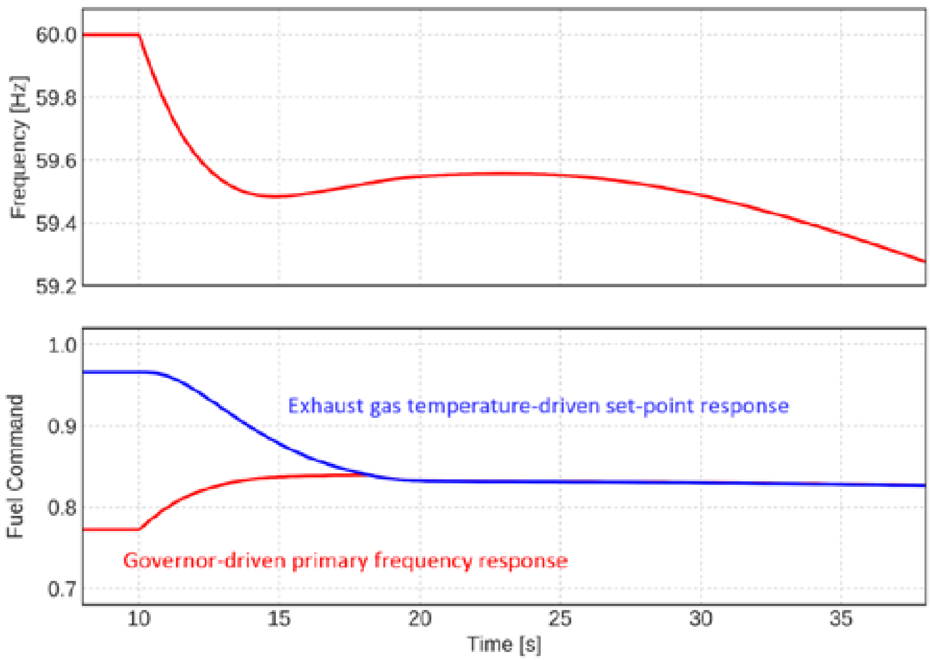


Figure 37. Exhaust Gas Temperature Controls on Gas Turbines Will Decrease Primary Frequency Response if Frequency Remains Depressed

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

As noted, the effect of these controls cannot be overridden; they are intrinsic to the design of protection for the turbine. This reduction is better thought of as a reduction in the headroom or primary frequency response capability of the gas turbine, rather than a form of withdrawal of primary frequency response. It is, in this regard, fundamentally different from the actions of plant load controllers discussed in the previous two findings. From an operator's perspective, it represents a derating of the capability of the generator.[76]

[76] Some grid operators, notably, ERCOT do, in fact, formally derate gas turbine capability.

It is therefore essential to supplement primary frequency response from gas turbines operated at maximum output with response from other sources that will sustain or increase their response during the comparatively longer periods when system frequency may be depressed following large loss-of-generation events.

5.8 "Synthetic Inertia" Controls on Electronically Coupled Wind Generation Appear not to Sustain Primary Frequency Response

Generally speaking, inverter-based controls on electronically coupled generation resources (such as wind turbines, solar photovoltaics, and batteries) are capable of providing sustained primary frequency response through a droop-relationship in the same way that turbine-governors operate in conventional power plants. Headroom must be reserved from which primary frequency response can be drawn. In contrast, "synthetic inertia" is described as a means for providing primary frequency response without reserving headroom.

Synthetic inertia controls on electronically coupled wind generation can have a similar impact as described previously, in which primary frequency response is delivered but terminated before frequency is stabilized. For this discussion, we rely on information in the published literature. Because of the proprietary nature of the controls involved, we did not attempt to develop simulation models to replicate their behavior. However, the principles related to the issues raised by non-sustaining primary frequency response discussed previously apply equally to these controls.

This type of response comes from the extraction of kinetic energy from spinning wind turbine blades. That is, the turbine blades slow down. Based on published literature, however, the response is not sustained and is withdrawn within five to ten seconds. See, for example, GE (2010). If this is unavoidable, then other sources of sustaining primary frequency response will be required that can continue stabilizing frequency until secondary frequency response can replace it.

5.9 Fast Demand Response Provides Robust Primary Frequency Response, but Currently is Inflexible

LBNL's 2010 Study reviewed the Texas Interconnection's unique form of frequency response that relies on a fast form of demand response, which involves voluntary load shedding procured from customers.[77] The review of the factors affecting ROCOF in Section 2, the discussion of design generation-loss events in Section 3, and our simulations give us insight into the logic that guides practices in the Texas Interconnection.

In Section 2, we presented the analytical relationships between the size of generation loss and system inertia that establish ROCOF (See Figure 6). In Section 4, we presented information on the design

[77] This form of contracted, voluntary load shedding is entirely distinct from interconnection-coordinated load shedding, which is involuntary and is only used in emergencies when primary frequency response, alone, has not been able to arrest the decline in frequency.

generation-loss event for the Texas Interconnection, which at times of minimum system load exceeds 10 percent (see Table 2). Viewed together, these elements explain that ROCOF can be high for the Texas Interconnection. Consequently, the Texas Interconnection is an interconnection where fast demand response is a highly effective complement to primary frequency response provided by generators. This form of demand response is referred to by ERCOT, as Load Resources in the Responsive Reserve Service market.

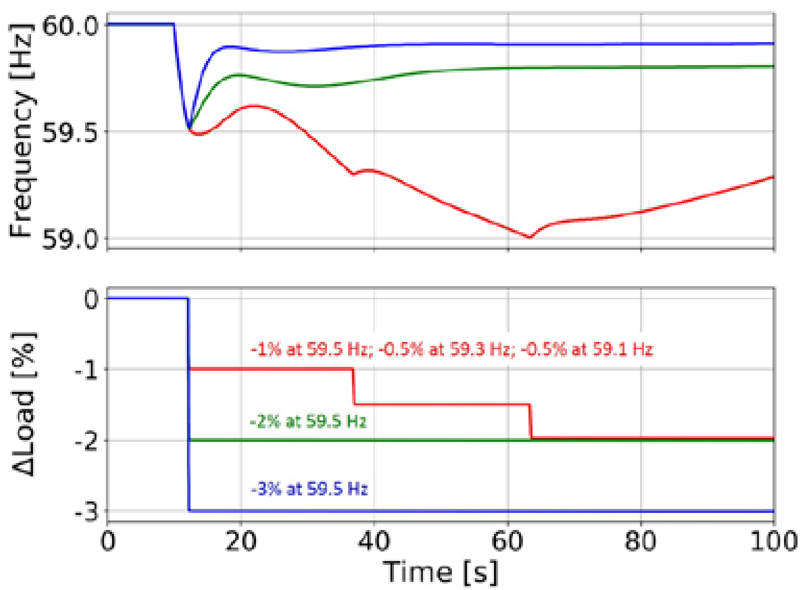


Figure 38. Fast Demand Response Can Augment Primary Frequency Response Delivered by Generation

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

Figure 38 illustrates why this form of fast demand response can be a successful complement to primary frequency control that relies only on generators. The lower panel depicts three demand response load shedding strategies, and the upper panel depicts the behavior of interconnection frequency corresponding to each strategy. For all three strategies, the design generation-loss event is 4 percent, and demand response is always complemented with the same amount of responsive and sustaining generation.

The first strategy, shown in red, involves dropping load in successive increments (1 percent at 59.5 Hz, an additional 0.5 percent at 59.3 Hz, and a final 0.5 percent at 59.1 Hz). Following the red line in the upper panel, we can see that dropping the first block of load (at 59.5 Hz) impacts interconnection frequency immediately; it is a sudden action to restore the balance between interconnection load and the remaining connected generation. However, in this instance, it is not sufficient to sustain frequency at a stable, higher value after the drop has been arrested. This is because of continued, early withdrawal of primary frequency response by a subset of generators. Dropping the second block (at 59.3 Hz) is similarly insufficient to reverse the decline in frequency. Only after the third block is dropped (at 59.1 Hz) does frequency begin to climb.

The second strategy, shown in green, drops all two percent of load at 59.5 Hz. This strategy is immediately effective in supporting frequency stabilization following the formation of the nadir. The third strategy, shown in blue, is even more aggressive. It involves dropping three percent of load at 59.5 Hz. It is also highly effective in stabilizing frequency following formation of the nadir.

These examples also illustrate two important inflexibilities associated with fast demand response that is achieved through load shedding, which are not associated with delivery of primary frequency response through turbine-governor control of generators. First, this form of fast demand response, once deployed, cannot be restored except through manual commands. In contrast, turbine-governors using droop control automatically restore the reserves held to provide primary frequency control, so they can be re-deployed immediately should another generation-loss event take place. That is, droop continuously adjusts the generator's output as a function of interconnection frequency. Generation increases as frequency declines and decreases as frequency increases back toward the scheduled value. When frequency is restored to the scheduled value, generation output returns to its original, pre-event value.

Second, the amount of load that is dropped and the frequency at which it is dropped, including time delay, must be specified carefully in advance. If the amount of load dropped is greater than the amount of generation that is lost, an over-frequency situation will result that poses a different but also potentially severe challenge for system reliability. This is the reason that the triggering frequency for ERCOT's Load Resources is set at 59.7 Hz and the maximum amount of load that is allowed to provide Responsive Reserve Service is limited to less than the total ERCOT procures. The Load Resources are only deployed in response to large generation-loss events. Evaluating the trade-offs between the amount of load to shed—and the frequency at which to shed it—requires careful study to ensure that this form of fast demand response complements, and does not compromise, primary frequency response delivered by generation.

Because of the above findings we recommend, in Section 6 of this report, that non-traditional forms of primary frequency response (such as fast demand response and storage) should be studied and incorporated as appropriate into operating practices for reliable interconnection frequency response.

5.10 Smaller Deadbands on Turbine-Governors Increase how Quickly Delivery of Primary Frequency Response will Begin

Industry is currently actively discussing the efficacy and role of governor deadbands for interconnection frequency control. These discussions tend to focus on secondary frequency control or AGC, but sometimes concern is expressed about the potential detrimental effect of deadbands on interconnection frequency response.

A generator with a smaller deadband will respond to smaller generation-loss events than a generator with a larger deadband. Moreover, for larger generation-loss events, generators with smaller deadbands will respond sooner than generators with larger deadbands. Operating with smaller deadbands therefore will improve interconnection frequency response compared to operating with larger deadbands.

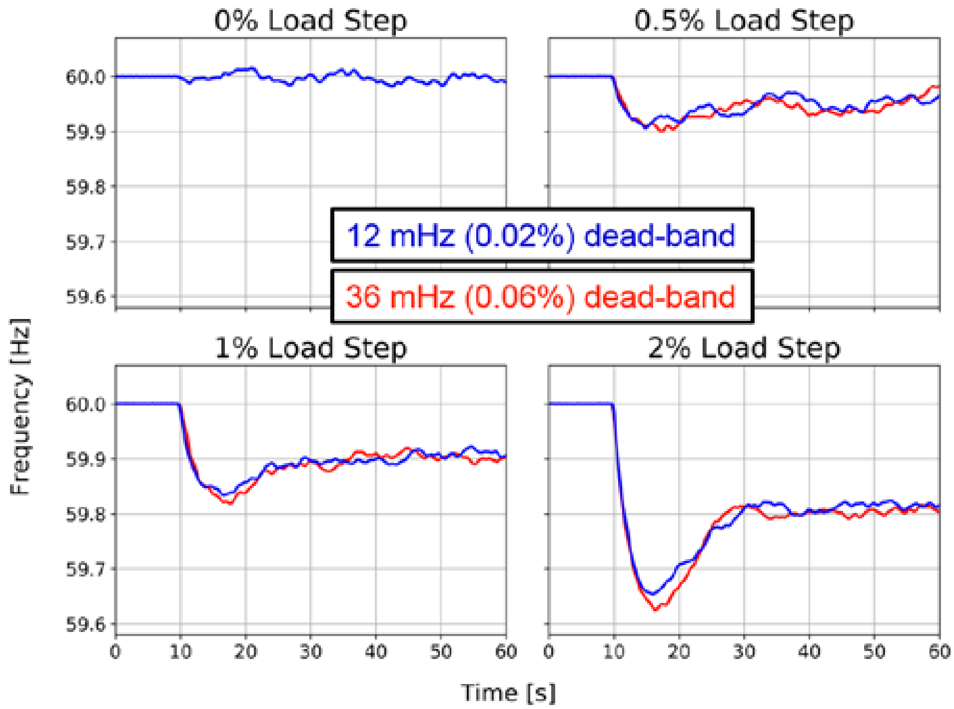


Figure 39. The Effect of Deadbands on Interconnection Frequency Response

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

Figure 39 shows how turbine-governors with frequency deadbands of 12 mHz (or 0.02 percent) and 36 mHz (or 0.06 percent) provide primary frequency response following generation-loss events ranging from 0 to 2 percent of total system load. As the generation-loss events increase in size, the faster response provided by generators operating with smaller deadbands becomes more evident. Operating with a smaller deadband leads to formation of the nadir sooner and at a higher frequency.

Figure 39 also illustrates the importance of ensuring that deadbands among generators should be as nearly equal as is feasible. Unequal deadbands among generators mean that those with smaller deadbands will respond sooner and more often than generators with larger deadbands. This inequity may be especially intolerable in settings where primary frequency response is procured via market mechanisms.

"…extreme deadband settings undermine the goal of providing frequency response…"

In the extreme, if a generator operates with a very large deadband (approximately 300 or 500 mHz), then the generator will only respond to the very largest generation-loss events (and will likely do so too late to avoid triggering UFLS). Such extreme deadband settings undermine the goal of providing frequency response because the turbine-governor will not respond to the vast majority of generation-loss events, which are smaller in size.

5.11 Load Sensitivity Currently Complements Primary Frequency Response, but this Sensitivity may be Going Away

As discussed in Section 2, a portion of load has traditionally reduced consumption autonomously in proportion to a decline in interconnection frequency and therefore augments primary frequency control by generators. Our simulations, by design, removed these effects of load to obtain direct insights into the role of generators in responding to frequency excursions and to inject a measure of conservatism into our findings.

Today, this characteristic of load is changing. The proliferation of power electronic interfaces between loads and the grid has introduced a means by which loads can be controlled actively in response to changes in interconnection frequency and voltage. These loads include variable-frequency drives on motors, fans, and pumps.

Currently, the principal objective of these interfaces is to maintain constant power in the face of interconnection frequency changes. This means that, as frequency declines, these loads will draw increased power from the grid. Thus, today, these electronically controlled forms of load can work to the detriment of frequency control because, in effect, they increase the amount of primary frequency response that must be delivered.

However, this situation is not a given. The electronic controllers on loads can also be programmed to respond to frequency in a different way. Figure 40 illustrates a potential future situation in which loads retain, and in fact increase, their former positive, symbiotic relationship with primary control of interconnection frequency. The figure shows first (in red) the pessimistic assumption we made in our simulations of no load damping or sensitivity to frequency. It then shows (in green and blue) progressively greater dependencies between load and interconnection frequency. As the dependency increases, interconnection frequency response is enhanced; the nadir is higher, and frequency recovers faster to a stable, settling value because load is augmenting or supporting primary frequency control from generators.

Because of this finding, we recommend, in Section 6 of this report, that the factors that are negatively influencing the sensitivity of loads to frequency should be better understood and addressed.

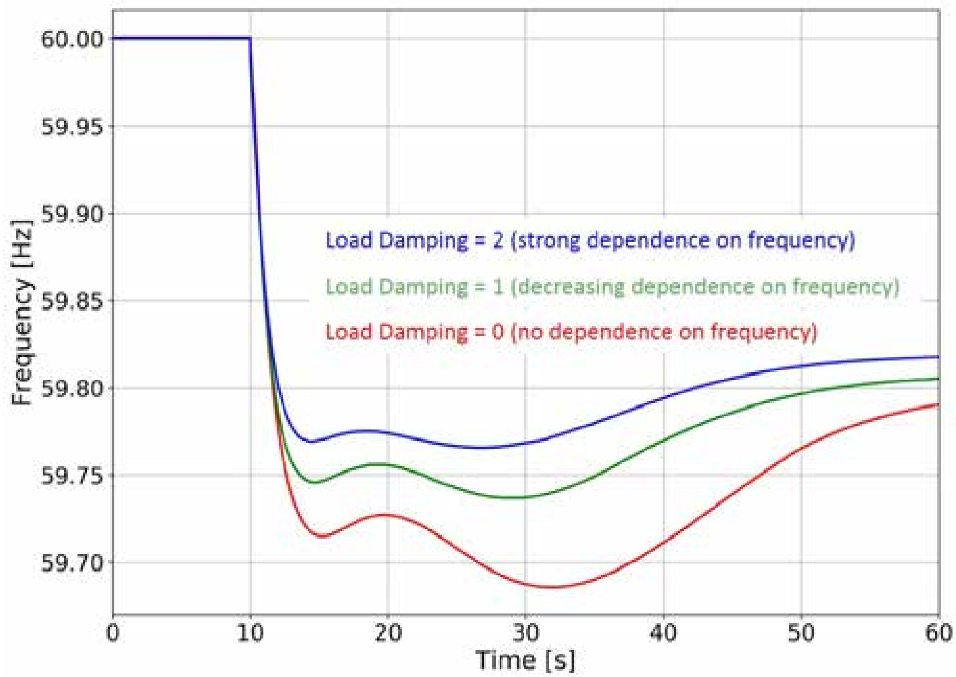


Figure 40. Increased Load Damping Supports Reliable Interconnection Frequency Response

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

5.12 Summary

Our findings are organized into three broad groups: (1) confirmation of fundamental, but potentially not widely understood factors that determine the initial requirements for resources held to provide primary frequency control; (2) the importance of equal attention to ensuring primary frequency response is sustained, including illustrations of various means by which primary frequency response may not be sustained and therefore must either be modified to sustain response or augmented by other resources that will sustain their response; and (3) findings related to other primary frequency control topics, including fast demand response, governor deadband settings, and load sensitivity.

1. Reserves held to provide primary frequency control must exceed the expected loss of generation.

Interconnection frequency reflects the balance between generation and load. The rapid decline in frequency following a loss of generation results from the sudden imbalance between generation and load. The decline is arrested once the balance between generation and load has been re-established. See Figure 30. This is not a new or novel finding of this study; however confirmation and illustration of it as a fundamental principle forms the basis for all subsequent findings in this study. Furthermore, prudence dictates that the total primary frequency response capacity held on-line should exceed the size of the design generation-loss event to acknowledge uncertainty in the actual performance of the fleet.

2. Primary frequency response must be delivered quickly, which requires many participating generators.

The reserve for primary frequency response must be allocated among generators[78] with recognition of the amounts of primary response that each type of generator can produce in the few seconds that are available to arrest the decline of frequency. This recognition can best be achieved, and in practice can only be achieved, by allocating reserves across a number of generators, such that each needs to make a small contribution to the required cumulative response. This finding, too, is generally understood, but it may not be widely appreciated.

> *"…it is prudent to ensure to the extent technically practical, that all generators—both those that are directly coupled and those that are electronically coupled to the grid—have the capability to provide primary frequency response."*

As a consequence, it is prudent to ensure to the extent technically practical, that all generators—both those that are directly coupled and those that are electronically coupled to the grid—have the capability to provide primary frequency response. Ensuring this capability provides maximum flexibility to grid operators to assign primary frequency response duty to generators as appropriate for the current grid operating conditions. Exceptions should be considered only on a case-by-case basis.

3. For a given loss of generation, system inertia and the timing of primary frequency response determine how frequency is arrested.

The key factors determining the nadir of frequency are (a) the effective inertia constant of the system, which determines the initial rate of decline of frequency; and (b) the rate at which generation is increased by primary frequency response. Lower system inertia will require faster primary frequency response. Understanding the expected dynamic performance of the reserves that are held to respond, therefore, becomes of greater importance. For example, if the reserves held are quick to respond, they may be adequate for a wide range of possible future generation loss scenarios and corresponding system inertias. If they are slow to respond, they may require augmentation by faster responding reserves. See Figure 32.

4. Primary frequency response must be sustained until secondary frequency response can replace it.

Much attention has been devoted to ensuring adequate primary frequency response over the initial seconds following the loss of generation. Due attention should also be devoted to ensuring primary frequency response is sustained during the period when frequency is stabilized following the formation of the nadir. During this period, primary frequency response is required in order to stabilize frequency. It must, therefore, be sustained until secondary frequency response can replace it. If, during this period, primary frequency response is not sustained and is reduced to less than the amount of generation lost,

[78] The principal focus of this study is on primary frequency response provided by generation resources. Separate findings on primary frequency response by non-generation resources and on the sympathetic, but changing, role of load sensitivity appear at the end of this section.

frequency will again decline and UFLS will be triggered. See Figure 33. This point is less well appreciated but of equal importance for reliable interconnection frequency response.

There are several means by which primary frequency response may not be sustained. The first is through withdrawal of primary frequency response by the actions of plant load controllers, which override and reset the actions of the turbine-governor responding to frequency deviations. We describe this finding first because it is an action that is directed by the plant owner/operator. As such, it is one that can be corrected by a plant owner/operator, which we discuss in Finding 5. The second is through actions stemming from inherent dynamic characteristics of turbine generators. One example is exhaust gas temperature limits on gas turbines, which are intrinsic to the current design of these types of turbines. Unlike plant-level controllers, these actions cannot be overridden or corrected. We conclude by discussing what we have observed in published information on wind turbines providing what is called “synthetic inertia.”

The bottom line is that, if primary frequency response from some sources will not be sustained, primary frequency response from additional sources will be required. The requirement is to stabilize frequency until primary frequency response can be replaced by secondary frequency response.

5. *Plant load controllers operated in pre-selected load mode without frequency bias will withdraw and not sustain primary frequency response.*

Plant load controllers establish the ranges around which turbine-governors operate in response to frequency. A logic commonly followed by these controllers seeks to maintain generation output in accordance with a pre-determined schedule. Consequently, when generation output increases because the turbine-governor responds to a decrease in interconnection frequency, the plant load controller overrides the turbine-governor and automatically restores output to the scheduled value. This results in primary frequency response being withdrawn, which negatively affects restoration of interconnection frequency.

6. *Plant load controllers operated in pre-selected load mode with frequency bias will sustain primary frequency response.*

The early withdrawal of primary frequency response by plant load controllers can be prevented by specifying a control logic that seeks to operate the generator at the scheduled value only when the frequency of the interconnection is at its normal operating value of 60 Hertz (Hz). That is, when frequency deviates significantly from 60 Hz, for example because of loss of generation on the interconnection, the plant load controller does not override the turbine-governor but instead allows the turbine-governor to continue delivering primary frequency response until system frequency returns to the nominal value. This control logic supports the restoration of interconnection frequency following a loss of generation.

7. Gas turbines may not be able to sustain primary frequency response following large loss-of-generation events.

Gas turbines are among the fastest responders that contribute to arresting the decline in system frequency following the sudden loss of generation. They can readily increase their output by a few percent of rated capacity within a handful of seconds (5 to 8 seconds). As a result, they are excellent initial sources of primary frequency response. However, if an under-frequency event calls for maximum output, this maximum will be less than would be reached when running at nominal frequency. Unlike the withdrawal of response by plant load controls, reduction of output is an action of the protection system of the turbine and cannot be deactivated at the discretion of the plant operator. There is feedback between these controls and system frequency that can be detrimental to reliable interconnection frequency response. That is, if, as exhaust gas temperature controls reduce turbine output, and system frequency continues to decline, then the temperature limit controls will further reduce turbine output. It is therefore essential to recognize this dependence of gas turbine maximum output on frequency and ensure that response is available from sources that will sustain or increase their response during the comparatively longer periods when system frequency may be depressed following large loss-of-generation events.

8. "Synthetic inertia" controls on electronically coupled wind generation appear not to sustain primary frequency response.

Inverter-based controls on electronically coupled generation sources (such as wind turbines, solar photovoltaics, and batteries) can provide sustained primary frequency response through a droop relationship in the same manner that turbine-governors operate in conventional power plants. In cases of low frequency, this requires reserving headroom from which the response is drawn. As an alternative, "synthetic inertia" controls are said to provide a form of primary frequency response without reserving headroom.

So-called "Synthetic inertia" controls on electronically coupled wind generation can have a similar impact as described above, in which primary frequency response is delivered but terminated before frequency is stabilized. This type of frequency response comes from the extraction of kinetic energy from spinning wind turbine blades. That is, the turbine blades slow down. Based on published information, the response, however, appears to be one that is not sustained and is, in effect, withdrawn within five to ten seconds. If this is unavoidable, then other sources of sustaining primary frequency response will be required that continue to stabilize frequency until secondary frequency response can replace them.

9. Fast demand response provides robust primary frequency response, but currently is inflexible.

Removing load immediately affects interconnection frequency and is therefore an effective strategy to restore the balance between load and generation after generation is lost. However, the amount of load shed—as well as the frequency at which it is shed and the time delay beforehand—must be established carefully in advance. If the amount of load shed is greater than the amount of generation that was lost, an over-frequency situation can result, which may pose a severe challenge to system reliability.

10. Smaller deadbands on turbine-governors increase how quickly delivery of primary frequency response will begin.

Deadband settings on turbine-governors determine at what frequency deviation a generator will begin delivering primary frequency response. Smaller deadbands mean that a generator will respond to smaller generation-loss events than a generator with a larger deadband. Moreover, for larger generation-loss events, generators with smaller deadbands will begin responding sooner than generators with larger deadbands. Operation with smaller deadbands therefore will improve interconnection frequency response compared to operation with larger deadbands. Importantly, unequal deadbands among generators mean that those with smaller deadbands will begin responding sooner and more often than generators with larger deadbands. In the extreme, if a generator operates with a very large deadband (300 or 500 mHz), then the generator will only respond to the very largest generation-loss events (and may do so too late to avoid triggering UFLS). Such extreme deadband settings undermine the goal of providing frequency response because the turbine-governor will not respond to the vast majority of generation-loss events, which are smaller in size.

11. Load sensitivity currently complements primary frequency response, but this sensitivity may be going away.

Although a portion of load has traditionally reduced consumption autonomously in proportion to a decline in interconnection frequency and, hence, augmented the frequency response that generators provide, the characteristics of load are changing.[79] In particular, load that is electronically coupled to the grid using power electronic interfaces is increasingly common. These loads include variable-frequency drives on motors, fans, and pumps. These electronically controlled forms of load can work to the detriment of frequency control because they generally act to prevent the power drawn from declining as frequency declines. Directly coupled motors "slow down" when frequency declines and reduce power consumption, and thereby work in concert with primary frequency response delivered by generators. By not slowing down and not reducing power consumption, electronically coupled motors no longer contribute or support primary frequency response delivered by generators. This impact of electronic controllers on interconnection frequency response, however, is not a given. Electronic controllers can be programmed to support reliable interconnection frequency response.

[79] Because of these considerations related to load, our initial set of simulations removed the effects of load in augmenting provision of primary frequency control. This enabled us to obtain direct insight into the role of generators in responding to frequency excursions and ensured that our findings would be conservative.

6. Observations and Recommendations

This section integrates the simulation findings presented in Section 5 as a basis for observations on frequency control issues specific to each of the three U.S. interconnections. Next, it offers cross-cutting recommendations for industry to consider in evaluating options for maintaining reliable interconnection frequency response in the future.

6.1 Observations Regarding the Three U.S. Interconnections

We present observations regarding the three U.S. Interconnections in ascending order of interconnection size, starting with the Texas Interconnection and followed by the Western and Eastern Interconnections. As discussed in Section 2, and as explored via simulations in Section 5, smaller interconnections face greater frequency response challenges than larger interconnections because the design generation-loss event, on a percentage basis, is much larger.

6.1.1 Texas Interconnection

As noted in Section 3, ERCOT has an interconnection-specific NERC BAL standard (BAL-001-TRE-1), which requires primary frequency control capability for all generators, existing and new, with the exception of nuclear and some older wind generators. Within this standard are requirements for droop and deadband settings as well as minimum performance measures during identified measurable events. These are more inclusive requirements than are found in other current generator interconnection agreements in other interconnections, which apply only to new generators. In addition, as discussed in Section 5.9 (and noted in the LBNL 2010 Study), ERCOT relies on a unique form of frequency control involving fast demand response (called Load Resources procured in the Responsive Reserves Service ancillary service market). More recently, ERCOT has pioneered new methods for tracking system inertia and explored market alternatives for procurement of frequency responsive reserves. The information presented in this study provides ample grounding for and insights into the Texas Interconnection's focus.

"…the design generation-loss event in the Texas Interconnection results in very sharp and rapid decline in frequency…"

As seen in Table 1, on a percentage basis, the design generation-loss event for the Texas Interconnection, especially at times of minimum system load (11 percent), is nearly three times larger than the design generation-loss event at times of minimum system load for the Western Interconnection and more than five times larger than the design event for the Eastern interconnection. Therefore, the design generation-loss event in the Texas Interconnection results in a much sharper and more rapid decline in frequency (high ROCOF) compared the other two interconnections. (See Figure 41 and Figure 42.) Such high rates of frequency decline present a great challenge for the frequency response obtainable from available conventional power plants within the interconnection. Thus, ERCOT's Load Resources program is an important, if not critical, complement to primary frequency response from generators for ensuring reliable interconnection frequency response.

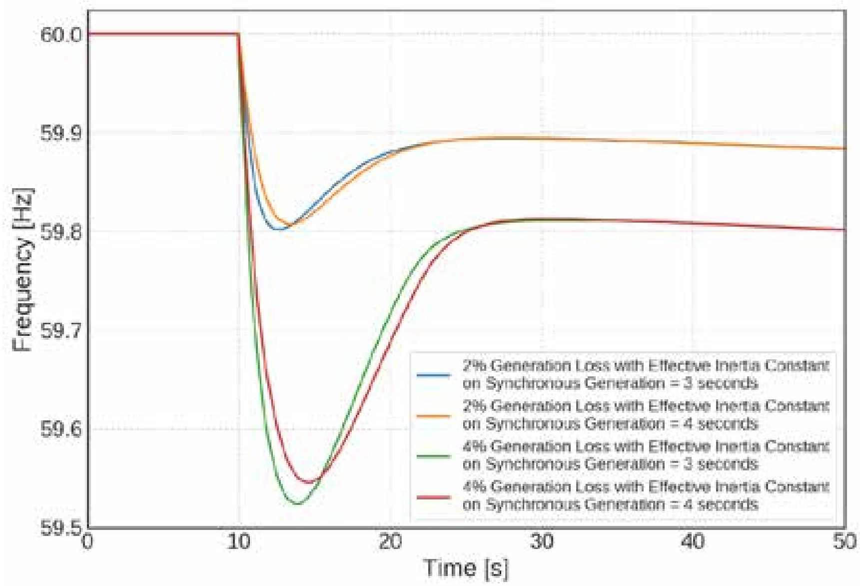
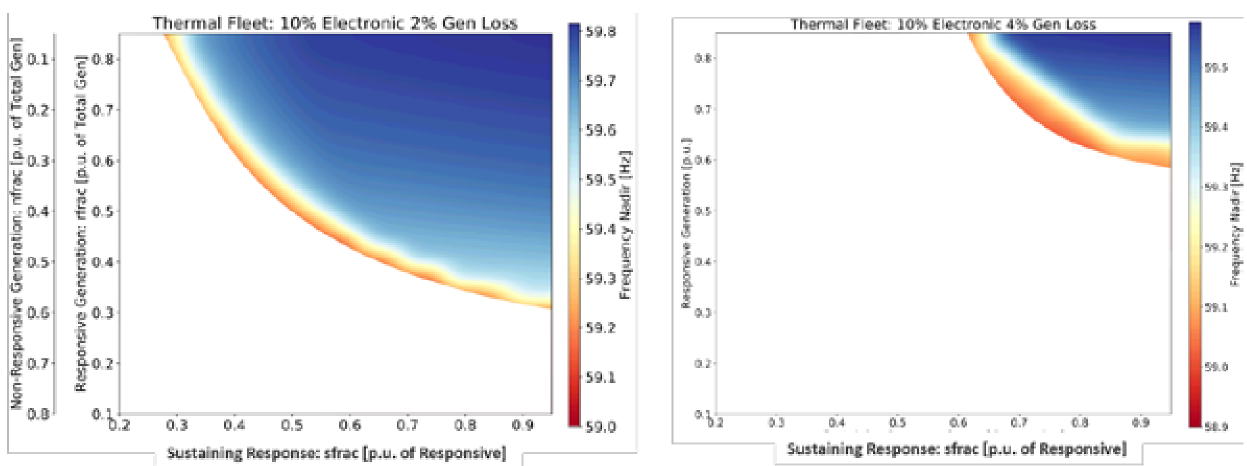


Figure 41. The Relative Impacts of Generation Loss versus System Inertia on Frequency Nadir

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency



Note: White areas indicate combinations of responsive and sustaining generation that will not arrest frequency above 59Hz

Figure 42. Combinations of Responsive and Responsive and Sustaining Generation Required to Avoid Triggering UFLS Following Loss of 2% and 4% of Generation

Source: Developed by LBNL from Undrill (2018): Primary Frequency Response and Control of Power System Frequency

In its present configuration, this form of demand response is an inflexible frequency response resource. Currently, Load Resources are deployed when frequency declines to 59.7 Hz, which means they respond only to the largest generation-loss events. Primary frequency response from conventional generation is relied on to manage smaller generation-loss events.[80]

[80] Future forms of fast demand response could be more flexible than the current form of demand response relied upon by ERCOT. One form, illustrated in Section 5.9, might involve shedding smaller blocks of load at different frequency set points. Another form, alluded to in Section 5.11, might involve load that could be varied continuously in response to frequency changes in a manner analogous to droop control of turbine-governors.

Texas will likely lead the three U.S. interconnections in addressing the challenge of the replacement of conventional generation that provides primary frequency control with wind and solar resources. Wind generation's percent of total generation is larger in the Texas Interconnection than it is in the Western and Eastern Interconnections. All reports suggest that the amount of wind and solar generation in the Texas Interconnection will increase in the future.

Section 5.3 clarifies that frequency control-related concerns regarding the fact that the latest technologies for wind and solar generation do not contribute inertia to the interconnection are misplaced. Very high levels of generation from resources that do not contribute inertia can be readily accommodated by ensuring adequate primary frequency control reserves.

The pressing concern is that, although a significant part of wind generation in the Texas Interconnection is equipped to respond to frequency, it is not normally dispatched to provide primary frequency control when frequency declines. These dispatch and market practices may create challenges in maintaining adequate reserves of primary frequency control from the remaining on-line generation resources that are dispatched for this purpose. These challenges will increase if these resources retire.

One obvious solution is to find ways to engage the primary frequency control capabilities of wind and solar generation. Doing so will require addressing the commercial arrangements that currently create strong financial incentives for wind and solar to operate without headroom. Section 5.8 clarifies that "synthetic inertia," in which a form of primary frequency control is provided by wind generation without reserving headroom, if is then quickly withdrawn, is not a substitute for sustained primary frequency response.

6.1.2 Western Interconnection

The WECC-specific version of NERC's BAL standard (BAL-002.WECC-2a) states explicitly that the interconnection's spinning reserve requirements must be met by generation that is on-line and capable of responding autonomously and automatically to changes in frequency. Previously, as noted in the LBNL 2010 Study, the spinning reserve requirement could be met by any means with on-line generation capable of providing full output within 10 minutes. The new requirement also directs a droop setting.

In addition, and perhaps equally important, the Western Interconnection has, since 1996, had a strong commitment to empirical validation and ongoing refinement and improvement of the dynamic simulation tools and models used to study interconnection frequency control among other things. See, for example, the ongoing activities of the WECC Model Validation Working Group and its supporting modeling Task Forces for load and renewable energy.[81] WECC is well-positioned to conduct scenario-based studies to explore the implications for interconnection frequency response of future changes in the generation and load mix that have been illustrated in this study.

[81] See https://www.wecc.biz/PCC/Pages/MVWG.aspx

Our review in Section 4.4.2 of the validated planning models prepared by WECC to represent a thermal-hydro fleet confirms that the interconnection relies extensively on hydro-based resources for primary frequency control. These resources are located primarily in the Northwestern portion of the interconnection. Likewise, the interconnection relies less on thermal-based resources as a proportion of its total reserves of primary frequency control. It is, furthermore, significant that thermal-based resources are located primarily in the Southwestern portion of the interconnection.

This geographic distribution of reserves relied on for primary frequency control means that the transmission system plays a critical role in delivering those reserves. For example, primary frequency response from the Northwest can only be delivered via the transmission system to address the interconnection's design generation-loss event, which is the loss of a pair of generators located in the Southwest. As discussed in LBNL's 2010 Study, reliance on the long-distance transmission system to deliver primary frequency response poses a risk to reliable interconnection frequency response because sufficient reserve capability must be available on the transmission lines to reliably deliver primary frequency response to the area that has become deficient because of a loss of generation. Failure to maintain sufficient transmission transfer-related reserve capability creates a risk that primary frequency response will not be delivered because it has tripped the line(s) carrying it and that UFLS will be triggered. Transmission losses must also be taken into account when relying on reserves of primary frequency response that are distant from the generation-loss event.

If older, thermal-based reserves for primary frequency control located in the Southwest of the interconnection retire, this will exacerbate the transmission-related risk to reliable primary frequency control because, all else being equal, these retirements would lead to even greater reliance on hydro-based reserves in the Northwest. Retirements of existing generation will need to be replaced with other, newer generation resources. The character and location of this generation will have implications for interconnection frequency control. These possibilities are best considered using the calibrated planning tools maintained by the WECC.

For example, if older thermal generation is replaced by combined-cycle,[82] wind, or solar generation that provided frequency response, primary frequency control capability in the Southwest could be maintained and could increase. The rules and incentives for generators to install, maintain, and make available primary frequency control capability will determine the outcome.

A new issue that has arisen recently in the Western Interconnection is a recognition that the characteristics of generation-loss events may be changing. In 2016, a fire caused a collector transmission line to trip off line. The line was fed radially by several large solar photovoltaic generating plants and by tripping, the generation was lost. The behavior of still-connected solar plants in other

[82] As an aside, but of secondary importance, the inertia of the interconnection would, in fact, increase if older thermal generation is replaced with combined-cycle gas plants. For a given size generation plant, the inertia of a combined-cycle gas plant is higher than that of nuclear-, coal-, or gas-fired steam plants. See Figure 12. It is important to note that the aspects of generators that determine their contribution to system inertia depend on the types of turbines used to drive them (e.g., steam turbines, combustion turbines, hydro-electric turbines, etc.), not on the types of fuels consumed (e.g., nuclear, coal, natural gas, and fuel oil, which can all be used to run a steam turbine).

areas was unexpected. Efforts are underway to better understand and respond to the causes of this unexpected behavior. See, for example, NERC (2017a). Here, too, experiences gained in the Western Interconnection through events such as these will provide valuable lessons for the other U.S. interconnections.

6.1.3 Eastern Interconnection

As noted in the LBNL 2010 Study, the recorded values of frequency response in the Eastern Interconnection are much greater than those of the Western and the Texas Interconnections. The LBNL 2010 Study concluded that the frequency response capability of the Eastern Interconnection was sufficient to maintain reliability with the increases in variable renewable generation that were projected at the time. The findings in the present study provide greater insight into this conclusion from the LBNL 2010 Study and point to specific areas for industry attention going forward.

> *"…the comparatively large size of the Eastern Interconnection relative to the generation-loss events it experiences explains the interconnection's lower ROCOF values…"*

This study's findings clarify that the comparatively large size of the Eastern Interconnection relative to the size of generation-loss events it experiences explains the interconnection's lower ROCOF values compared to those observed in other interconnections following losses of comparable amounts of generation.

More important, this study finds that the quantity and quality of primary frequency control are the most important factors to consider when assessing and planning for interconnection frequency control. By and large this importance stems from the simple recognition that, in fact, these are the sole aspects of interconnection frequency control that can be managed directly through operating policies and operator decisions. That is, quantity and quality of primary frequency control are determined directly and continuously by the combined effects of the plant control decisions made by power plant operators and the generation dispatch decisions made by grid operators. This means that these decisions are also the principal source of risk, as they can and will change over time. If un-checked, they could change in ways that are detrimental to reliable interconnection frequency response.[83]

For this reason, this study's findings validate industry's increased attention to frequency control, as discussed in Section 3. The original generator governor survey undertaken through the NERC Frequency Response Initiative (NERC 2012) and augmented recently by the NERC alert on generator governors (NERC 2015b) is fully consistent with and reinforced by this study's findings regarding the importance of the quantity and quality of primary frequency control.

> *"…the characteristic 'lazy L' shape of frequency response in the Eastern Interconnection is widely recognized as being driven by withdrawal of primary frequency response by plant level controllers."*

Two aspects of primary frequency control deserve specific attention in the Eastern Interconnection: First, the characteristic "lazy L" shape of frequency response in the Eastern Interconnection is widely recognized as being driven by withdrawal of primary frequency

[83] As noted in the LBNL 2010 study, currently these decisions have led to reliable interconnection frequency response.

response by plant load controllers. As noted in Section 3, this explanation has been corroborated in modeling studies by both NERC staff and GE. This study has shown how withdrawal can be detrimental to interconnection frequency response. It has also shown how a major reason for withdrawal—the settings on plant load controllers—can be changed to prevent withdrawal during frequency response events. Therefore, it is important that industry monitor and, as appropriate, implement interconnection and region-specific operating policies and procedures that prevent detrimental withdrawal of primary frequency response.

"As noted first in LBNL's 2010 study, and as observed now in this study… the Eastern Interconnection planning models currently developed and used by industry do not reproduce the measured performance of the interconnection to the design generation-loss event…"

Second, the recognition that withdrawal of primary frequency response is a material concern for the Eastern Interconnection provides additional motivation for efforts to improve the ability of the interconnection's dynamic planning models to replicate and explain the interconnection's observed frequency response. As noted first in LBNL's 2010 Study, and as observed now in this study (see Section 4.7), the Eastern Interconnection planning models currently developed and used by industry do not reproduce the performance of the interconnection that was recorded during the generation-loss event that is now used to establish interconnection frequency response obligations within the interconnection. We also note that NERC staff is already working with planners in the Eastern Interconnection to improve the quality of frequency response modeling (NERC 2017c).

Well-calibrated planning models are essential for assessing current performance and, when required, guiding modifications to interconnection agreements and operating policies to ensure continued, reliable interconnection frequency response. Continuous updating and ongoing calibration are essential for building confidence in application of the models to study future scenarios involving changes in the mix of generation and loads.

6.2 Recommendations

1. *Focused attention should be directed to understanding the aggregate frequency control performance required of the fleet of resources that must be kept on-line at all times to respond to generation-loss events. This will involve collection, maintenance, and validation of the data necessary for accurate planning and operating studies as well as collection of comprehensive data to measure trends in interconnection frequency control.*

The dynamic simulation tools and system models that the interconnections use to study frequency response must be based on accurate, up-to-date information about the actual characteristics of generators and load. This information should track not only interconnection loading, inertia, design generation-loss event, and highest set-point for UFLS, but also generator headroom, turbine-governor performance characteristics, and the number and location of resources for primary frequency control. Data are needed on the performance characteristics of non-traditional, non-governor-based resources for primary frequency response that indicate how much primary frequency response is available and how rapidly the response can be delivered. In the case of fast demand response, such as ERCOT's Load

Resources, it is important to study the size of load blocks and the triggering conditions for them. Performance measures should apply equally to traditional and non-traditional resources. For all resources, this should entail explicit performance measures that assess the factors that might withdraw primary frequency response early or cause it to not be sustained.[84] Simulation-based or other forms of study should consider the full period over which primary frequency response must be sustained, which may be as long as several minutes, and determine the rate at which non-sustaining response must be replaced in order to ensure reliable interconnection frequency response. Studies should examine worst-case situations involving either or both times of low system inertia and times when reserves of primary frequency control may be low.

Routine, comprehensive measurement of interconnection frequency control performance is essential for tracking trends.[85] This will require ongoing updating and verification of the performance of generators and other resources for primary frequency response as well as the conditions of the interconnection during which these resources are called upon. To the extent feasible, measurements should form the basis for the information used to model and plan for the procurement and dispatch of resources that provide primary frequency response. As an example, measurements recorded during actual events should serve as the basis for establishing limits on procuring primary frequency response.[86] This includes ensuring that modeling assumptions regarding primary frequency response capability are reflective of actual dispatch and power plant operating practices. In addition to tracking traditional measures of frequency response (such as interconnection frequency response), this process should document the conditions under which these measurements are made, such as the state of the power system (its loading and inertia, and whether load, and hence generation, is increasing or decreasing at the time generation is lost) and the size and location of generation-loss events relative to the performance of the primary frequency response resources, including the extent to which they sustain provision of primary frequency response.

2. *International practices should be reviewed as options for U.S. grid operators to consider adopting to ensure continued reliable interconnection frequency response.*

As the fleet of U.S. generation and the characteristics of load change, we must assess our approaches to frequency control to ensure that they continue to support reliable interconnection frequency response. Gaps, conflicts, and disincentives must be identified, analyzed, and addressed as appropriate.

Our review of international frequency control practices spans a range of approaches that represent functioning alternatives to or variants of current U.S. approaches. Because of the demonstrated success of these approaches in other power systems, they should be reviewed and analyzed in current and expected future operating conditions in the United States, and then given due consideration for adoption, as is, or in modified form. See Figure 43.

84 NERC's ERSWG has developed and is currently tracking measures that seek to address this issue. See NERC2015a.
85 In fact, NERC has begun compiling, and is now regularly publishing, this information. See, for example, NERC 2017b.
86 ERCOT, in fact, routinely conducts these measurements.

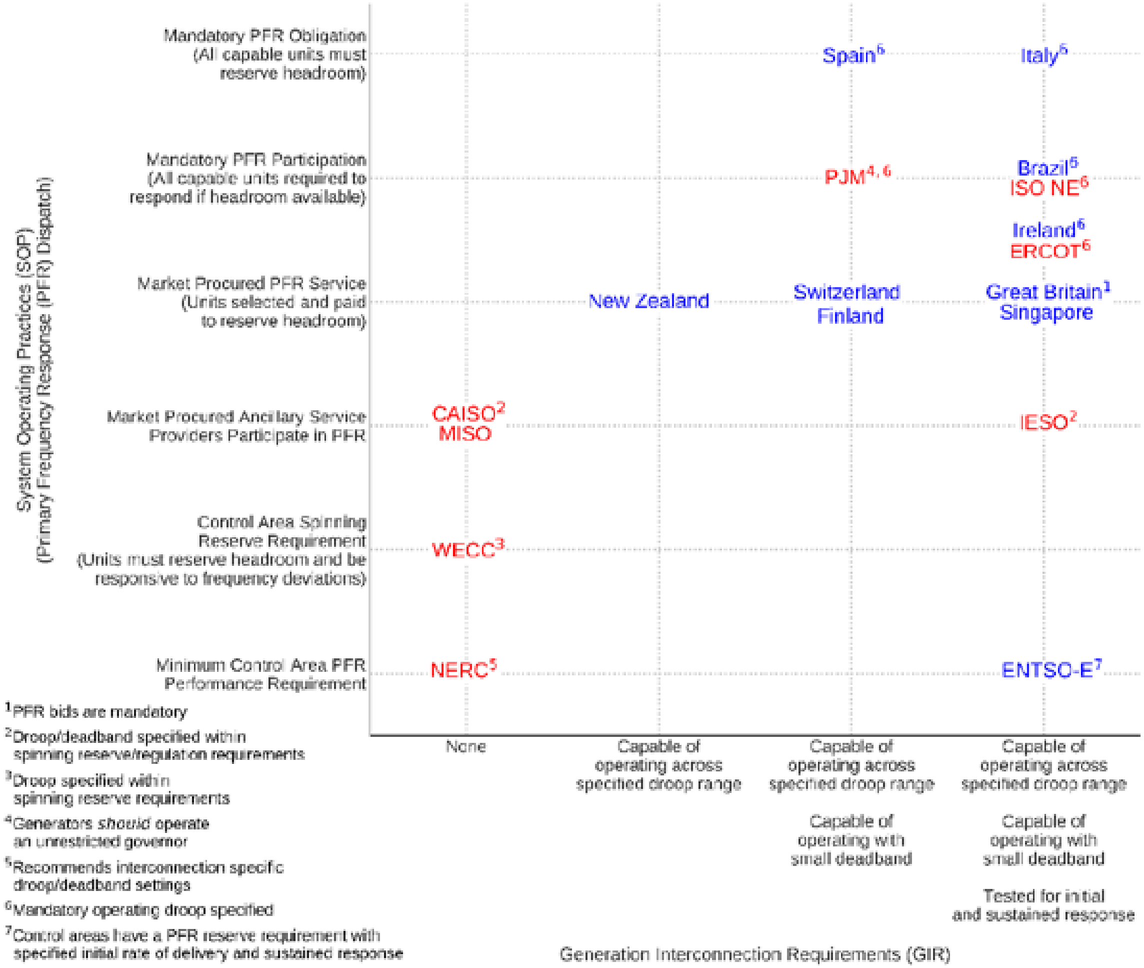


Figure 43. Comparison of Selected U.S. and International Grid Codes Related to Frequency Response

Source: Roberts (2018): Review of International Grid Codes

3. *All generators, to the extent feasible, should be capable of providing sustained primary frequency response.*

Reliable interconnection frequency response requires participation by many generators. Ensuring that as many generators as is technically feasible are capable of providing sustained response provides maximum flexibility to grid operators to assign primary frequency response duty as appropriate for current grid operating conditions.

Moreover, reliability of the interconnections is enhanced by enabling this capability on all generators capable of providing sustained primary frequency response. Doing so increases the pool of responding generators and reserves of primary frequency response, and thereby reduces the risk of unforeseen shortages of primary frequency response. It is recognized that some generators will not contribute if they are already dispatched at maximum capacity and hence do not have headroom available.

4. Barriers to adding a frequency bias[87] to plant load controllers should be evaluated and addressed.

This study describes the detrimental effects of early withdrawal of primary frequency response by plant load controllers. We also describe how early withdrawal by plant load controllers can be prevented by introducing a frequency bias to the control logic of pre-selected load mode controls. We also recognize that some U.S. grid operators already require or have performance requirements that support the use of these controls. Still, others in the United States do not.

Anecdotally, we perceive that awareness of the efficacy of this alternative control logic is limited within the generator community. Accordingly, we recommend continued but expanded education and outreach to foster wider adoption of this control approach.[88] In addition, it is important to understand and address any financial disincentives that would reinforce current practices.

5. The contributions of non-traditional resources for primary frequency control (demand response, energy storage, and other forms of electronically coupled loads and generation, including wind and solar photovoltaic) should be studied and incorporated, as appropriate, into future operations.

One future change in the makeup of the generation fleet is that traditional resources for primary frequency response may retire and be replaced by non-traditional resources, including demand response, energy storage, and other forms of electronically coupled loads and generation such as wind and solar photovoltaics. The performance characteristics of non-traditional resources are not widely known or fully understood. Future frequency response-related operating and planning policies and decisions should be based on up-to-date, accurate information about the performance and potential contributions of these resources.[89] Research, development, and demonstration are also needed to improve the performance capabilities of these new sources and to support timely industry adoption.

6. Factors that are negatively influencing the sensitivity of loads to frequency should be studied and addressed.

Load sensitivity historically complemented primary frequency response from generators. However, this sensitivity appears to be disappearing as newer forms of load are electronically coupled to the grid using power electronic interfaces, which currently do not reduce power consumption when frequency deviates from nominal. These forms of load include variable-frequency drives on motors, fans, and pumps. Better information is needed on how the frequency support provided by load changes over the course of a day and seasonally.

Still, no inherent technical limitations prevent power electronic interfaces from supporting primary frequency response by generators. In many instances, a simple firmware upgrade of power electronics controls is all that is required. The technical and commercial reasons that current controls do not

[87] This use of the term "frequency bias" is distinct from the use of this same term in the Area Control Error equation that guides automatic generation control, which is a form of secondary frequency control.
[88] NERC 2015a is a good initial example of this approach.
[89] NERC's Inverter Based Resource Performance Task Force may be one source for this information.

provide primary frequency response should be understood and, where appropriate and feasible, modified or addressed so that future loads will work in concert with and support primary frequency response from generators.

References

Bonneville Power Administration (BPA). (2017). *Synchrophasor Technology at BPA: From Wide-Area Monitoring to Wide-Area Control.* BPA. October.

Cohn, N. (1971). *Control of Generation and Power Flow on Interconnected Power Systems, Second Edition.* New York: John Wiley & Son, Inc.

Coughlin, K.C. and J.H. Eto. (2010). *Analysis of Wind Power and Load Data at Multiple Time Scales.* LBNL-4147E. Berkeley: Lawrence Berkeley National Laboratory. December.

Electric Reliability Council of Texas (ERCOT). (2013). *ERCOT Concept Paper: Future Ancillary Services in ERCOT.* Draft Version 1.1. November.

Eto, J., J. Undrill, P. Mackin, R. Daschmans, B. Williams, B. Haney, R. Hunt, J. Ellis, H. Illian, C. Martinez, M. O'Malley, K. Coughlin, and K. Hamachi-LaCommare. (2010). *Use of a Frequency Response Metric to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation.* LBNL-4142E. Berkeley: Lawrence Berkeley National Laboratory. December.

Federal Energy Regulatory Commission (FERC). (2016a). *Essential Reliability Services and the Evolving Bulk Power System – Primary Frequency Response*. Docket No, RM 16-6 issued Feb 16, 2016. https://www.ferc.gov/whats-new/comm-meet/2016/021816/E-2.pdf

FERC. (2016b). Order on Proposed Tariff Revision. 156 FERC ¶ 61,182. Docket Nos. ER 16 1483 000, 001. September 16. https://www.ferc.gov/CalendarFiles/20160916183725-ER16-1483-000.pdf

GE Energy. (2010). *Impact of Frequency Responsive Wind Plant Controls on Grid Performance*.

Illian, H.F. (2010). *Frequency Control Performance Measurement and Requirements.* LBNL-4145E. Berkeley: Lawrence Berkeley National Laboratory. December.

Independent System Operator New England (ISO-NE). (2014). *Operating Procedure No. 14: Technical Requirements for Generators, Demand Resources, Asset Related Demands and Alternative Technology Regulation Resources, Section 1*. ISO-NE.

Kirchmayer, L. (1959). *Economic Control of Interconnected Systems.* New York: John Wiley & Sons, Inc.

Kosterev, Dmitry, D. Davies, P. Etingov, A. Silverstein, J. Eto. (2014). "Using Synchrophasors for Frequency Response Analysis in the Western Interconnection." CIGRE 2014 Grid of the Future Symposium, Houston, TX. October. http://cigre-usnc.tamu.edu/wp-content/uploads/2015/06/Using-Synchrophasors-for-Frequency-Response-Analysis-in-the-Western-Interconnection.pdf

Mackin, P., R. Daschmans, B. Williams, B. Haney, R. Hunt, and J. Ellis. (2010). *Dynamic Simulations Studies of the Frequency Response of the Three U.S. Interconnections with Increased Wind Generation.* LBNL-4146E. Berkeley: Lawrence Berkeley National Laboratory. December. http://www.nerc.com/FilingsOrders/us/FERCOrdersRules/Dynamic_Simulation_Studies.pdf

Martinez, C. S. Xue, and M. Martinez. (2010). *Review of the Recent Frequency Performance of the Eastern, Western and ERCOT Interconnections.* LBNL-4144E. Berkeley: Lawrence Berkeley National Laboratory. December. http://www.nerc.com/FilingsOrders/us/FERCOrdersRules/Interconnection_Frequency_Performance.pdf

Matevosyan, J. (2016). *Inertia Data.* Presentation to NERC Resources Subcommittee. October 2016.

Miller, N.W., M. Shao, S. Pajic, and R. D'Aquila. (2014). *Western Wind and Solar Integration Study Phase 3—Frequency Response and Transient Stability.* Prepared by GE Energy Management for the National Renewable Energy Laboratory (NREL). NREL/SR-5D00-62906. December.

Miller, N.W., M. Shao, S. Pajic, and R. D'Aquila. (2013). *Eastern Frequency Response Study.* Prepared by GE Energy Management for the National Renewable Energy Laboratory (NREL). NREL/SR-5500-58077. May.

Miller, N.W., M. Shao, and S. Venkataraman. (2011). *California ISO (CAISO) Frequency Response Study.* Prepared by GE Energy for the California ISO. November. https://www.caiso.com/Documents/Report-FrequencyResponseStudy.pdf

North American Electricity Reliability Corporation (NERC). (2017a). *1,200 MW Fault Induced Solar Photovoltaic Resource Interruption Disturbance Report: Southern California 8/16/2016 Event*. NERC. June. http://www.nerc.com/pa/rrm/ea/1200_MW_Fault_Induced_Solar_Photovoltaic_Resource_/1200_MW_Fault_Induced_Solar_Photovoltaic_Resource_Interruption_Final.pdf

NERC (2017b). *2017 Frequency Response Annual Analysis*. NERC. November. http://www.nerc.com/comm/OC/BAL0031_Supporting_Documents_2017_DL/2017_FRAA_Final_20171113.pdf

NERC. (2017c). *Essential Reliability Services Working Group.* NERC. February. http://www.nerc.com/comm/Other/essntlrlbltysrvcstskfrcDL/ERSWG_Scope_February_8_2017.pdf

NERC. (2017d). *Informational Filing of the North American Electric Reliability Corporation Regarding the Light-Load Case Study of the Eastern Interconnection*. Docket No. RM13-11-000. http://www.nerc.com/FilingsOrders/us/NERC%20Filings%20to%20FERC%20DL/Eastern%20Interconnect%20Info%20Filing.pdf

NERC. (2017e). *Standard BAL-002-WECC-2a—Contingency Reserve*. Effective Date: January 24, 2017. Available from http://www.nerc.net/standardsreports/standardssummary.aspx.

NERC. (2017f). *Standard BAL-001-TRE-1—Primary Frequency Response in the ERCOT Region.* Effective Date: April 1, 2014. Available from http://www.nerc.net/standardsreports/standardssummary.aspx.

NERC. (2017g). *Standard PRC-006-3—Automatic Underfrequency Load Shedding*. Effective Date: October 1, 2017. Available from http://www.nerc.net/standardsreports/standardssummary.aspx.

NERC. (2015a). *Essential Reliability Task Force Measures Framework Report.* NERC. November. http://www.nerc.com/comm/Other/essntlrlbltysrvcstskfrcDL/ERSTF%20Framework%20Report%20-%20Final.pdf

NERC. (2015b). *Industry Alert: Generator Frequency Response*. http://www.nerc.com/pa/rrm/bpsa/Alerts%20DL/2015%20Alerts/NERC%20Alert%20A-2015-02-05-01%20Generator%20Governor%20Frequency%20Response.pdf

NERC. (2015c). *Reliability Guideline: Primary Frequency Control*. http://www.nerc.com/comm/OC/Reliability%20Guideline%20DL/Primary_Frequency_Control_final.pdf

NERC. (2015d). *Standard BAL-003-1.1—Frequency Response and Frequency Bias Setting.* Effective April 1, 2005, as amended though November 13, 2015. Available from http://www.nerc.net/standardsreports/standardssummary.aspx.

NERC. (2012). *Frequency Response Initiative: Report: The Reliability Role of Frequency Response*. NERC. October. http://www.nerc.com/docs/pc/FRI_Report_10-30-12_Master_w-appendices.pdf

NERC. (2009). *Balancing and Frequency Control (Part 1)*. Technical Concepts Document Prepared by the NERC Resources Subcommittee. NERC. November.

PJM. (2017). *PJM Manual 14D: Generator Operational Requirements Revision: 41*. June. http://www.pjm.com/~/media/documents/manuals/m14d.ashx

Roberts, C. (2018). *Review of International Grid Codes*. Berkeley: Lawrence Berkeley National Laboratory. February.

Sharma, S. (2016). *Renewable Integration at ERCOT.* Presentation dated October 2016.

Sharma, S., S. Huang, and N. Sarma. (2011). "System Inertial Frequency Response Estimation and the Impact of Renewable Resources in ERCOT Interconnection." *Proceedings of the IEEE Power and Energy Society General Meeting*, July. DOI: 10.1109/PES.2011.6038993

Undrill, J.M. (2018). *Primary Frequency Response and Control of Power System Frequency.* Berkeley: Lawrence Berkeley National Laboratory. February.

Undrill, J.M. (2010). P*ower and Frequency Control as it Relates to Wind-Powered Generation.* LBNL-4143E. Berkeley: Lawrence Berkeley National Laboratory. December. https://www.ferc.gov/CalendarFiles/20110120114503-Power-and-Frequency-Control.pdf

Undrill, J.M., P. Mackin, and J. Ellis. (2018). *Relating the Microcosm Simulations to Full Scale Grid Simulations.* Berkeley: Lawrence Berkeley National Laboratory. February.

Western Electricity Coordinating Council (WECC). (2014a). *Model Validation and System Performance Analysis for PDCI RAS Event that Occurred on May 30, 2013.* WECC. February. https://www.wecc.biz/Reliability/May%2030%202013%20Model%20Validation%20Report.pdf

WECC. (2014b). *Model Validation Report for May 16, 2014 RAS Event.* WECC. August. https://www.wecc.biz/Reliability/Model%20Validation%20Report%20for%20May%2016%202014%20RAS%20Event.pdf

Appendix A. Glossary of Terms

Automatic generation control (AGC)	An automated form of secondary frequency control that is used to oppose small deviations in system frequency around the scheduled value. AGC involves signals that are sent to plant-level controllers every 4 to 10 seconds to adjust a generator's output in order to return interconnection frequency to its scheduled value.
Frequency	In North America, this is normally 60 cycles per second or 60 Hertz (Hz).
Frequency control	Primary or secondary *frequency control* refers to the aggregate effect of the actions of all generators participating in the control of interconnection frequency taking a control action.
Frequency nadir	The point at which frequency decline is arrested.
Frequency response	The collective ability of the power system to respond to frequency excursions, such as those caused by the sudden unplanned loss of generation.
Governor	The means by which generators provide primary frequency response; a governor's actions are automatic (they do not depend on external commands) and autonomous (they do not depend on the actions of other generators).
Headroom	The difference between the current operating point of a generator or transmission system and its maximum operating capability. The headroom available at a generator establishes the maximum amount of power that generator theoretically could deliver to oppose a decline in frequency. However, the droop setting for the turbine-governor and the highest set point for UFLS will determine what portion of the available headroom will be able to deliver to contribute to primary frequency control.
Inertia	The ability of a power system to resist changes in frequency, measured in MW-seconds. Inertia is an inherent property or characteristic of each generator and element of load.
Load sensitivity	Loads that reduce their consumption of electricity in proportion to a decline in interconnection frequency, sometimes also called *load damping.*
Plant load controllers	External (to the turbine-governor) controls, taken together as a group.
Primary frequency response	*Primary frequency response* involves automatic, autonomous, and rapid (i.e., within seconds) changes in a generator's output to oppose sudden changes in frequency; it refers to the actions of individual generators.
Rate of change of frequency (ROCOF)	A measure of how quickly frequency changes following a sudden imbalance between generation and load. ROCOF is expressed in Hertz per second (Hz/second).
Secondary frequency response	Directed (i.e., external or supervisory to the autonomous actions of the turbine-governor), slower actions to change a generator's output to oppose changes in frequency.
Settling frequency	The point at which frequency is stabilized following formation of the nadir.

Tertiary frequency control	Refers to centrally coordinated actions (i.e., it is a "manual" form of what we have called secondary control) that operate on an even longer time scale (i.e., minutes to tens of minutes) than primary frequency response and secondary frequency control provided through AGC.
Under-frequency load-shedding (UFLS)	An extreme measure to arrest frequency decline that disconnects large, pre-set groups of customers at predetermined frequency set points.