Inertia and the Power Grid: A Guide Without the Spin

Paul Denholm,¹ Trieu Mai,¹ Rick Wallace Kenyon,² Ben Kroposki,¹ and Mark O’Malley¹

¹ National Renewable Energy Laboratory
² University of Colorado Boulder
Inertia and the Power Grid: A Guide Without the Spin

Paul Denholm,1 Trieu Mai,1 Rick Wallace Kenyon,2 Ben Kroposki,1 and Mark O’Malley1

1 National Renewable Energy Laboratory
2 University of Colorado Boulder

Suggested Citation
NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by GridLab. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.


Cover Photo by Dennis Schroeder: NREL 50710.

NREL prints on paper that contains recycled content.
Acknowledgments

The authors would like to thank the following individuals for their contributions. Editing and graphics support were provided by Devonie McCamey, Mike Meshek, Besiki Kazaishvili, and Joshua Bauer. Billy Roberts created the map. Helpful review and comments were provided by Mark Ahlstrom, Doug Arent, Sam Baldwin, Aaron Bloom, Jaquelin Cochran, Vahan Gevorgian, Patrick Gilman, Himanshu Jain, Julia Matevosjana, and Ric O’Connell. This work was inspired in part by conversations with Debbie Lew.

Finally, we gratefully acknowledge funding for this work from GridLab. This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.
List of Acronyms and Abbreviations

AC           alternating current
DC           direct current
DOE          U.S. Department of Energy
EI           Eastern Interconnection
ERCOT        Electric Reliability Council of Texas
FERC         Federal Energy Regulatory Commission
FFR          fast frequency response
GW           gigawatt
GWh          gigawatt-hour
GW•s         gigawatt-second
IBR          inverter-based resource
kW           kilowatt
kWh          kilowatt-hour
LR           load response
MISO         Midcontinent Independent System Operator
mph          miles per hour
MW           megawatt
MWh          megawatt-hour
MW•s         megawatt-second
NERC         North American Electric Reliability Corporation
NREL         National Renewable Energy Laboratory
PFR          primary frequency response
RPS          rotations per second
PV           photovoltaics
RoCoF        rate of change of frequency
RRS          responsive reserve service
UFLS         underfrequency load shedding
VG           variable generation
WI           Western Interconnection
Executive Summary

Inertia in power systems refers to the energy stored in large rotating generators and some industrial motors, which gives them the tendency to remain rotating. This stored energy can be particularly valuable when a large power plant fails, as it can temporarily make up for the power lost from the failed generator. This temporary response—which is typically available for a few seconds—allows the mechanical systems that control most power plants time to detect and respond to the failure.

Historically, in the U.S. power grid, inertia from conventional fossil, nuclear, and hydropower generators was abundant—and thus taken for granted in the planning and operations of the system. But as the grid evolves with increasing penetrations of inverter-based resources—e.g., wind, solar photovoltaics, and battery storage—that do not inherently provide inertia, questions have emerged about the need for inertia and its role in the future grid.

Intended to educate policymakers and other interested stakeholders, this report provides an overview of inertia’s role in maintaining a reliable power system, why inertia may decrease with increasing deployment of wind and solar generation, and how system reliability can be maintained in the evolving grid. An accompanying video (https://youtu.be/b9JN7kj1tso) further illustrates several key concepts.

Key takeaways from the full report include:

1. **Grid frequency, which is a measure of the balance of supply of electricity and demand, can drop if a large power plant or transmission fails.** Inertia resists this drop in frequency, giving the grid time to rebalance supply and demand.

2. **Inertia is only one of several grid services that help maintain power system reliability.** Understanding the role of inertia requires understanding the interplay of inertia and these other services, particularly primary frequency response, which is largely derived from relatively slow-responding mechanical systems.

3. **The importance of inertia to a power system depends on many factors, including the size of the grid and how quickly generators in the grid can detect and respond to imbalances.** A grid with slower generators needs more inertia to maintain reliability than a grid that can respond quickly.

4. **Using power electronics, inverter-based resources including wind, solar, and storage can quickly detect frequency deviations and respond to system imbalances.** Tapping into electronic-based resources for this “fast frequency response” can enable response rates many times faster than traditional mechanical response from conventional generators, thereby reducing the need for inertia.

5. **Replacing conventional generators with inverter-based resources, including wind, solar, and certain types of energy storage, has two counterbalancing effects.** First, these resources decrease the amount of inertia available. But second, these resources can reduce the amount of inertia actually needed—and thus address the first effect. In combination, this represents a paradigm shift in how we think about providing frequency response.
6. In the United States, the Texas grid (the Electric Reliability Council of Texas, or ERCOT) is the smallest of three main grids. **ERCOT’s relatively small size, combined with its large wind deployment, has required it to compensate for declining inertia by adopting several low-cost solutions, including allowing fast-responding noncritical loads to respond to changes in frequency.** This has enabled ERCOT to achieve increasingly high instantaneous wind penetrations—reaching a record of 58% in 2019—while maintaining reliability.

7. **In the Western and Eastern Interconnections, which are much larger than ERCOT, it is unlikely that any significant concerns related to maintaining frequency due to declining inertia will arise in the coming decade.** Moving forward, demonstrated solutions—including those used in ERCOT today—can allow these regions to add significant wind and solar while maintaining reliable operation.

8. **Ongoing research points to the possibility of maintaining grid frequency even in systems with very low or no inertia.** The development of new “grid-forming” inverters enable inverter-based resources to take a more active role in maintaining reliability and could be an integral technology for a purely inverter-based grid.

Although growth in inverter-based resources will reduce the amount of grid inertia, there are multiple solutions for maintaining or improving system reliability—so declines in inertia do not pose significant technical or economic barriers to significant growth in wind, solar, and storage to well beyond today’s levels for most of the United States.
Table of Contents
1 Introduction ........................................................................................................................................... 1
2 What Is Inertia and Its Role in Today’s Grid? .................................................................................... 2
3 How Does Inertia Help Maintain Reliability in Today’s Grid? .......................................................... 4
4 Five Factors That Determine Frequency Stability ............................................................................. 8
5 Why Does The Amount of Inertia in the Grid Change? .................................................................. 18
6 Alternatives to Conventional Frequency Response: Why Texas Went First ............................... 22
7 How Grid Operators Are Responding to Increased Penetration of Renewables and the Declining Role of Inertia .................................................................................................................... 25
8 How Low Can We Go? Maintaining Frequency with Decreased Inertia........................................ 32
9 Conclusions and Research Needs .................................................................................................... 36
References ................................................................................................................................................. 37

List of Figures
Figure 1. The concept of synchronous generators working together in a electrical grid ................. 3
Figure 2. Map of the North American power system, showing the four main interconnections ........ 3
Figure 3. Deployment of mechanical PFR with a flyball governor ......................................................... 6
Figure 4. A successful recovery from a contingency event ................................................................. 7
Figure 5. Source of power in a 1,000-MW contingency case with 29 identical 1,000-MW generators ...... 9
Figure 6. Decline in frequency in an example of a system with a 1,000-MW contingency event and 29 remaining identical 1,000-MW generators ................................................................. 11
Figure 7. Impact of grid size on frequency decline ......................................................................... 12
Figure 8. Impact of load on frequency decline after a contingency in our example case ................. 13
Figure 9. Impact of contingency size on frequency decline in our example case ............................ 14
Figure 10. Impact of UFLS on required response time .................................................................. 15
Figure 11. Primary frequency response from synchronous generators .......................................... 16
Figure 12. Frequency after a contingency in our example case with the addition of PFR ............... 17
Figure 13. Relationship between system dispatch and inertia in our example case .................... 19
Figure 14. Impact of IBRs on inertia (a) and frequency decline (b) in our example case with 30% wind penetration without changing operational practices ...................................................... 21
Figure 15. Decline in frequency in an ERCOT-like system before the addition of VG ................. 23
Figure 16. The impact of system size on frequency in systems resembling ERCOT and the WI at 40% VG penetration ........................................................................................................... 24
Figure 17. Use of stored kinetic energy in a wind turbine to provide FFR ......................................... 28
Figure 18. Wind and solar providing frequency response .................................................................. 30
Figure 19. Use of frequency response from wind to achieve an 80% instantaneous penetration in the WI .................................................................................................................................. 32
Figure 20. Grid-following and grid-forming Inverters .................................................................. 34

List of Tables
Table 1. Summary of Factors That Drive Frequency Stability ............................................................ 8
Table 2. Key Parameters of the Three Main U.S. Grids ...................................................................... 22
Table 3. Summary of Options to Maintain Frequency Stability .......................................................... 25
1 Introduction

Inertia—the tendency of an object in motion to remain in motion—has historically been an important source of reliability in the electric grid. Inertia from rotating electrical generators in fossil, nuclear, and hydroelectric power plants represents a source of stored energy that can be tapped for a few seconds to provide the grid time to respond to power plant or other system failures.

Historically, in the U.S. power grid, inertia from rotating generators has been abundant—and thus taken for granted in the planning and operation of the grid. However, today’s grid is evolving to include new sources of electricity generation—namely variable generation (VG) wind and solar, which do not use conventional generators and therefore do not inherently provide inertia. As VG penetration increases and conventional generators are displaced, the grid will need to be planned and operated differently in order to maintain reliability.

To inform interested stakeholders, this guide addresses a variety of questions about how inertia factors into the evolving power grid, and it describes the many proven or possible solutions that address the challenges of decreasing inertia. These questions include:

- What is inertia in the context of power systems?
- How does inertia work with other grid services to help maintain reliability?
- What technologies provide inertia?
- What impact does increasing VG penetration have on power system inertia?
- What changes can be made to maintain reliability as inertia declines?
2 What Is Inertia and Its Role in Today’s Grid?

Anyone who has driven a vehicle or ridden a bicycle is familiar with inertia—the tendency of an object in motion to remain in motion. For example, inertia is what keeps a vehicle moving when a driver stops pressing on the gas pedal. Spinning objects, like a wheel or power generator, have rotational inertia. In the grid, inertia refers to the kinetic energy stored in spinning generators.

In a power grid, inertia is derived from hundreds or thousands of generators that are synchronized, meaning they are all rotating in lock step at the same frequency (see Grid Frequency textbox). Figure 1 illustrates this conceptually, with all conventional (synchronous) generators connected via electromagnetic forces, represented by the chains. These electromagnetic chains mean all the individual generators that are online and spinning can work together to contribute to grid inertia.

The North American power system largely consists of four independent grids (shown in Figure 2) that are referred to as interconnections. All the synchronous generators within an interconnection are connected (via the electromagnetic “chains” shown in Figure 1) and contribute to the inertia within that interconnection. There are no synchronous connections across the boundaries of the interconnections themselves.\(^1\)

---

\(^1\) There are many other independent grids throughout North America on islands and various other locations not connected to the larger grids (EIA 2016).

\(^2\) There are small non-synchronous (direct current, or DC) connections between the interconnections that allow limited transfer of energy (but no inertia). They have essentially no impact on inertia or frequency response.

---

Grid Frequency

The U.S. power grid uses alternating current, meaning the flow of electricity rapidly switches directions across the individual wires that carry it from power plants to the point of use. The rate of this switching is referred to as the grid frequency, which is 60 cycles per second, or 60 Hertz (Hz) in the United States.

In today’s grid, frequency is determined by how fast conventional generators spin. Grid frequency is a measure of the health of the grid, as it reflects the ability of a grid to balance supply and demand. Today’s grid operating rules require tight tolerances on grid frequency.

The ability of a system to maintain nearly constant frequency under normal conditions and to recover from imbalance conditions that may result in changes in frequency is often referred to as frequency stability.
Figure 1. The concept of synchronous generators working together in a electrical grid

Figure 2. Map of the North American power system, showing the four main interconnections
3 How Does Inertia Help Maintain Reliability in Today’s Grid?

Inertia is a useful property whether one is riding a bicycle or operating a power grid. On a bicycle, inertia gives the rider a chance to stop pedaling and coast without falling over. In the grid, it gives the system operator a chance to respond to power plant failures (called contingencies). This is because inertia resists changes in frequency, giving other systems time to respond and rebalance supply and demand.

Grid operators in the United States may incur financial penalties if they do not maintain systems to keep frequency within tight tolerances. One of the challenges facing grid operators is maintaining grid frequency following a contingency event in which a large power plant (or associated transmission line) fails, meaning its supply of power is removed from the system very rapidly and without warning. The resulting imbalance will result in the grid slowing down and the frequency starting a steady decline, similar to how a vehicle slows when a driver eases up on the gas pedal.

When the grid frequency falls below a predefined level (59.5 Hz in most of the United States and 59.3 Hz in Texas), a portion of the customer load is disconnected (NERC 2019a). Such disconnections are known as underfrequency load shedding (UFLS) and are used to balance the remaining load with the remaining generation. A similar disconnection of generation occurs if the frequency exceeds a certain limit (e.g., 60.5 Hz in the Western Interconnection, or WI). The small change in frequency that results in UFLS is less than a 1% change from the base frequency. For instance, if the same principle were applied to a vehicle, it would mean driving 60.5 mph in a 60-mph zone would result in a speeding ticket.

3.1 Primary Frequency Response: Cruise Control for the Grid

To maintain grid frequency and avoid UFLS, grid operators use a variety of processes that can respond to events that might change frequency (Denholm, Sun, and Mai 2019). One of these key processes is primary frequency response (PFR). PFR detects changes in frequency and automatically—without action from the system operator—adjusts operations of online generators to maintain frequency within the desired range.

At a Glance: Inertia’s Role in Maintaining Grid Frequency

Grid frequency is used as an indicator of significant changes in either supply or demand. During normal grid operation, the supply of power from all the generators equals the demand for electricity, and the frequency remains constant. But just as a vehicle slows when you take your foot off the gas, if there is a loss of a power plant, the supply of power will drop almost instantaneously. However, the demand for electricity has not changed, so the same amount of power will still be extracted from the system.

That is where inertia comes in. Stored energy is extracted from the inertia of the spinning generators and can temporarily make up for the lost generator. This action will slow down the generators. Although it cannot be sustained for more than a few seconds, it provides time for the mechanical systems in the grid to detect the imbalance (as reflected in declining frequency) and tell power plants to speed up (or slow down).

---

3 Compliance is monitored and enforced in the United States by the North American Electric Reliability Corporation (NERC). https://www.nerc.com/Pages/default.aspx
Historically, primary frequency response was provided by a series of mechanical systems in power plants, as illustrated in Figure 3 and also demonstrated in the video (https://youtu.be/b9JN7kj1tso) at 1:33.

Figure 3a shows what happens when the generator speed falls below 60 Hz. In this example, the generator uses a spinning flyball governor, which is a device that is attached to a power plant shaft. The balls move in or out depending on the speed of the shaft, which in turn moves up or down a lever that is connected to a power plant valve.

If the grid frequency falls, the balls slow down and retract (move inward), moving a lever that opens a valve, allowing more steam into the turbine. This will act to increase power and therefore increase generator speed. Conversely, if the generator starts spinning too fast (more than 60 Hz), the flyballs will move outward, closing the valve and reducing the steam into the power plant (Figure 3b). This operation can work with a steam valve in a fossil-fueled plant or a water valve in a hydroelectric plant. To return to our car analogy, this whole process is essentially the cruise control for the power grid.

4 PFR is also sometimes referred to as governor response.
b) Frequency increases and governor closes valve, decreasing power

Figure 3. Deployment of mechanical PFR with a flyball governor

The governors historically used in synchronous generators shown in Figure 3 have largely been replaced with electronic sensors that rapidly measure frequency and send signals to the generator to change output when they detect a deviation. However, synchronous generators still rely on valves and other mechanical systems to change the supply of energy, and all these mechanical systems take time to respond. Because of the limited response rate from these mechanical systems, an adequate amount of inertia has been historically needed to slow the rate at which frequency drops and avoid UFLS.

3.2 Illustrating the Role of Inertia and Primary Frequency Response

Power system engineers often use frequency plots to examine how changes in grid characteristics can affect the ability of a system to recover from major imbalances. Figure 4 (next page) shows an example that plots the frequency of a power system as a function of time. In this example, it illustrates the sequence of events that follows a major contingency to successfully stop the decline of frequency and restore it to normal.

The far left of the figure shows the nearly constant frequency of a grid under normal operation (60 Hz), and then a large contingency event occurs at t=0 seconds. To accommodate the imbalance between supply and demand due to the drop in generation, the remaining online generators convert their rotational kinetic energy (inertia) into real power generation—referred to as inertial response. This process slows the generators and results in a drop in grid frequency. Inertial response provides time for the remaining online generators to detect changes in frequency and initiate PFR. As generator output from PFR increases, the net imbalance reaches zero and frequency stops declining. Ideally, this sequence of events will arrest the frequency decline before it reaches the solid orange line, which is the point of UFLS. PFR will then partially restore frequency. In subsequent seconds and minutes, grid operators will send signals to certain power plants to increase power to completely restore frequency to 60 Hz and return the
grid to normal operation. This action falls under the category of operating reserves and is not discussed in this guide.\footnote{For a discussion of operating reserves, see Denholm, Sun, and Mai (2019) and Ela et al. (2011).}

![Figure 4. A successful recovery from a contingency event](image)

Figure 4. A successful recovery from a contingency event

Overall, the ability of a power system to recover from a contingency event involves the interplay of various factors that determine how fast the frequency will fall and how quickly the system can respond.
4 Five Factors That Determine Frequency Stability

The addition of wind and solar to power grids both in the United States and internationally has raised concern about how much inertia is needed to maintain frequency stability—the ability of a power system to maintain steady frequency following an imbalance between supply and demand (Kundur et al. 2004). Any discussion of inertia must include the interplay of inertia and the many factors that determine the ability of the grid to successfully respond to a contingency event. Some of these factors are listed in Table 1, which introduces how changing each factor by itself can impact the ability of the system to balance the supply and demand after a contingency event. Each factor is then discussed in a subsequent subsection using a simple power system example.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Impact of Greater Amounta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generator inertia</td>
<td>Slows down frequency decline</td>
</tr>
<tr>
<td>Load inertia and load damping</td>
<td>Slows down frequency decline</td>
</tr>
<tr>
<td>Contingency size</td>
<td>Increases frequency decline</td>
</tr>
<tr>
<td>Underfrequency limits (UFLS settings)</td>
<td>Lower UFLS settings provides more time for overall response</td>
</tr>
<tr>
<td>Frequency response speed</td>
<td>Responds faster to a decline in frequency</td>
</tr>
</tbody>
</table>

a Assumes no other factors change

4.1 Generator Inertia: Only a Starting Point

Generator inertia is our starting point for examining how fast the system must respond to a contingency event. This section details how generator inertia resists changes in system frequency.

Under normal conditions, electricity demand is met by the constant injection of energy into the grid from many power plants. We will illustrate this using an example where 30,000 megawatts (MW) of demand (about the average demand in the state of Florida) is met by 30 1,000-MW generators. If one of these generators were to fail, the remaining generators online would only provide 29,000 MW. However, the loads on the system would still extract 30,000 MW of power from the system, with the extra 1,000 MW of power being extracted from the inertia of the remaining online generators.
In this example, we can estimate the rate at which frequency declines, which will begin to provide us some indication of the amount of time we have to replace the power from the lost generator and arrest the frequency decline.

At the moment the contingency occurs, each of the 29 remaining generators has stored inertia\(^6\) that can be extracted to provide extra power to the system, above and beyond the power provided by continuous injection of steam in the individual power plants.\(^7\) Here, if we assume the generators are all identical, each must provide an extra 34 MW of power from stored energy (1,000 MW divided by 29 generators).

Figure 5 illustrates the source of power from each generator in the post-contingency state. The constant injection of energy from fuel provides 1,000 MW, and 34 MW is drawn from stored energy, meaning each generator is providing 1,034 MW.

---

6 Regardless of its strict “physics” definition, the inertia of a power system is routinely defined as its stored kinetic energy with units of GW•s. See Units of Energy and Inertia textbox.

7 This example assumes the use of steam turbines, but generators could be any type, such as hydro-powered turbines.
As each generator uses up its inertia, it slows down. The relationship between rotational speed and energy allows us to calculate how much each generator will slow down and the corresponding decline in frequency.\(^8\)

In our example, we assume each generator rotating at 60 Hz has about 4 GW•s (1.1 MWh) of stored energy (in actuality, there is a large range of stored energy depending on generator type, as we discuss later). In the first second, each generator has to provide about 34 MW of extra power from its stored energy and so will need to give up about 34 MW•s, or a little under 1% of its stored energy.

Figure 6a plots the relationship between extracted energy and frequency for this example. The total system starts at 60 Hz with 115 GW•s of stored energy, and the load extracts about 1 GW•s after 1 second (due to the loss of 1 GW of generation), resulting in a frequency of about 59.7 Hz. If no other action occurs, this will be the frequency at 1 second.\(^9\)

Figure 6b shows the frequency as a function of time; this is similar to the plot in Figure 4, but here we are focusing on the first few seconds. In this example, the generator inertia provides about 2 seconds for the system to respond before it falls below 59.5 Hz. Assuming UFLS occurs at 59.5 Hz, this means the system has about 2 seconds to take corrective action.

---

\(^8\) The actual relationship between energy and frequency is given by the equation \(E=I \cdot \omega^2\) where \(I\) is the generator’s moment of inertia (a quantity determined by the amount and distribution of mass of the generator) and \(\omega\) is the rotational speed of the system.

\(^9\) This can also be expressed in terms of the “rate of change of frequency” or RoCoF, which is measured in Hz/second, which in this case is 0.26 Hz/second.
b) Frequency plot

Figure 6. Decline in frequency in an example of a system with a 1,000-MW contingency event and 29 remaining identical 1,000-MW generators

**Generator Inertia Constant**

A generator’s inertia constant represents how much stored energy it has per unit of rated capacity. This means the inertia constant represents how long the generator could generate at its rated power using only its stored rotational kinetic energy, so the inertia constant is measured in units of seconds.

A 1-GW generator with an inertia constant of 4 seconds could deliver 1 GW of power for 4 seconds (or has 4 GW·s of stored energy).

Typical power plants have inertia constants in the range of 2 to 7 seconds, with hydro plants having the lowest inertia, and gas plants having the highest inertia per unit of capacity (Eto et al. 2018).

The combination of inertia constant and total capacity of online generators determines the total inertia provided by the generators. Our simple example is a small system, particularly when compared to the two large U.S. grids (as discussed in Section 6.1). Grid size is a key factor in determining the total grid inertia and therefore how fast the frequency declines. Figure 7 provides an example where we double the grid size (resulting in twice as much load and twice as many generators) resulting in twice the effective grid inertia. However, the case still uses 1,000-MW

This very simplified case assumes a constant power draw of 30,000 MW and does not consider the reduction in load that results as the frequency declines (which is discussed in Section 4.2). It also assumes the inertia of each generator is the same. However, in real systems, inertia varies by generator size and type. For generators of the same type, a 200-MW generator would have roughly twice the inertia of a 100-MW generator. Inertia scales with generator size because generators with larger capacity have more physical mass in the turbine, generator, and other rotating machinery. But two equal-sized generators of different types may have different inertia due to the differences in the size and shape of the rotating equipment. This is reported as the “inertia constant” of a generator or generator type.10

---

10 Our example assumes an inertia constant of 4 seconds. For comparison, an estimate of the system-wide inertia constant for the WI is about 3.9 (Eto et al. 2018, Figure 17).
generators, so the contingency size does not increase. This leads to a large increase in the amount of time available for other systems to correct the imbalance.

![Figure 7. Impact of grid size on frequency decline](image)

Grid size is a critical factor, because inertia increases proportionally with grid size (larger grids have inherently more inertia); however, contingency size does not inherently scale with grid size. For example, the actual contingency size in the WI is about the same as that in ERCOT despite the WI being more than twice the size (based on electricity consumption). This will be particularly important when evaluating the impact of VG on real grids in the United States, as discussed in Section 6.

Finally, it is important to note that the amount of inertia available from a generator is independent of power output and depends only on whether it is online (committed) and spinning at grid frequency. For example, a committed synchronous generator rated at 1,000 MW provides the same amount of inertia when it is generating 600 MW as when it is generating 1,000 MW. As long as the generator is synchronized to the grid, the amount of inertia cannot be changed by any action taken by the generator operator.

### 4.2 Load Inertia and Damping: Small but Not Insignificant

The second element to consider is the response of actual load to changes in frequency. This involves two factors: the inertia of loads and the change in actual energy demand as a function of frequency.

Unlike an electric light, which shuts off instantly, an electric ceiling fan will continue to turn for some time after it is turned off. This represents inertia similar to that in electric generators. Certain types of motors add inertia to the grid.11

---

11 For a discussion of the types of loads that add inertia, see Omara (2012).
Another impact results from the actual change in electric demand that happens with changes in frequency. In our previous examples, we assumed the load remains constant after the contingency event. However, for some loads, including many motors used in industrial processes, the actual electricity demand will decrease at lower frequencies. This is analogous to the decreased amount of power needed to operate a vehicle at 55 mph compared to 60 mph. In the United States, this relationship is quantified in terms of a “damping constant” and a typical estimate is that a 1% decline in frequency will reduce load by 1%–2% (meaning a damping constant of between 1 and 2).\textsuperscript{12}

Figure 8 illustrates the impact of both load inertia and load damping on frequency decline compared to the base case shown earlier (assuming a damping constant of 1.3). The inertia in loads slightly slows the decline in frequency, while the impact of load damping is much greater, particularly as the frequency drops. For example, by the time the frequency reaches 59.6 Hz, the load has dropped by 300 MW, assuming a damping constant of 1.3.\textsuperscript{13} This means the imbalance has dropped from 1,000 MW to 700 MW. This drop will act to further slow the rate of frequency decline. The combined impact of load inertia and load damping is to add about 0.4 seconds to the time it takes the system to reach 59.5 Hz (from 1.8 seconds to 2.2 seconds).

\textbf{Figure 8. Impact of load on frequency decline after a contingency in our example case}

Figure 8 shows that the inherent and uncontrolled response of load has a relatively small but not insignificant increase in the amount of time the system has to respond and correct an imbalance. However, this impact is expected to lessen over time as older “inertia-providing” motors in industrial processes are replaced with motors powered by more efficient variable-speed controls (EPRI 2019). This trend will act to reduce both load inertia and load damping. However, load

\textsuperscript{12} The actual term used for this relationship vary and include load response, load damping, or load frequency sensitivity NERC (2012, NERC (2020).

\textsuperscript{13} The actual damping constant for loads in the United States is poorly understood. For additional discussion, see Fernández-Guillamón et al (2019)
can also be used in a controlled and managed way to provide rapid and accurate response to declining frequency, as discussed in Section 7.

4.3 Contingency Size: A Key Factor

In our simple example, a single generator failure leaves an initial imbalance of 1,000 MW. However, the failure could be larger or smaller depending on the mix of generators and transmission system.

Figure 9 shows the impact of contingency size on the decline in frequency. It includes the previous assumptions, including response from loads. The larger case assumes a contingency of 2,000 MW, which produces a much faster rate of frequency decline, resulting in the frequency dropping below 59.5 Hz in about 1 second (assuming the base generator inertia). Alternatively, if the contingency were 500 MW the frequency would drop more slowly, giving the system more time to respond. In this example, the system would have more than 6 seconds to respond.

![Figure 9. Impact of contingency size on frequency decline in our example case](image)

Figure 9 shows that contingency size can have a dramatic impact on the amount of time systems have to detect and correct the imbalance.

4.4 UFLS Settings: Keeping the Lights On By Turning Some of Them Off

The power system’s UFLS settings represent the final main element determining how much time is needed to respond. UFLS is initiated by circuit breakers that monitor frequency and automatically disconnect certain parts of the grid (rapidly and without warning) if the frequency drops below a certain setting. UFLS protocols actually use multiple settings that progressively shed more and more load as frequency drops lower and lower.\(^\text{14}\) The basic idea is that a relatively small amount of load is shed at some initial frequency, such as 59.5 Hz in much of the

\(^\text{14}\) For a more thorough discussion of an example UFLS sequence in ERCOT, see ERCOT (2017, Section 2.6).
United States. Hopefully this is enough to correct the imbalance with minimal impact to consumers, but if it is not, additional load shedding occurs until either the frequency decline is corrected, or in an extreme case, the entire grid is shut down.

The choice of initial UFLS setting impacts how much time the system has to respond, as illustrated in Figure 10. Our base case example from the previous section indicates that this level of inertia would require PFR and other systems to correct the imbalance in about 2.2 seconds before an UFLS event at 59.5 Hz. Reducing the UFLS setting to 59.3 Hz would increase this time by about 1.3 seconds, providing a total of 3.5 seconds to respond.

![Figure 10. Impact of UFLS on required response time](image)

The four factors discussed to this point describe how quickly the frequency will fall after a contingency event, and how much time the system has to respond to correct the resulting imbalance of supply and demand. Ultimately, the system will need to correct this imbalance, and the speed at which the system can do so is discussed next.

4.5 How Fast Can the System Respond? The Role of Traditional Generator Frequency Response

After a contingency event, PFR acts to increase power from the remaining generators and (temporarily) replace energy from the failed generator. Providing PFR from a generator requires it to have the necessary equipment (i.e., an active governor) and be operating at less than full output (i.e., providing headroom to increase output). The headroom requirement makes PFR very different from inertia (which is independent of its output). Only a generator that can increase output (and sustain that output for a period of time) can provide PFR. \(^\text{15}\) Upon a decline in frequency, generator governors detect this change and act to open valves and take other actions that increase the flow of fuel, steam, and/or water to generator turbines. This increases the power

\(^{15}\) For a discussion of length of time required, see FERC (2018) and NERC (2012).
produced, but this process takes time, much as it takes a time for a vehicle to accelerate after a driver presses the gas pedal.

Examples of PFR response rates are provided in Figure 11, with Figure 11a showing the measured PFR from an actual generator (Miller, Pajic, and Clark 2018), and Figure 11b showing curves used for two generator types in simulations. Though the response range is not linear, a typical range of response rates is about 0.3% per second for slower-responding units to 2% per second for fast units including certain gas turbines, meaning a 100-MW plant would be able to increase output by between 0.3 MW and 2.0 MW per second (assuming it was not already operating at maximum output).

Figure 12 shows the results of adding PFR to our example system, assuming an average PFR rate. It takes a little over 0.5 seconds for the generator output to begin to increase, and about 3.0 seconds for the PFR output to increase enough to arrest the decay. At this point, the PFR is greater than the contingency size, and the frequency begins to increase. The lowest frequency (nadir) of about 59.6 Hz is above the UFLS of 59.5, meaning this example would provide a reliable system from the perspective of maintaining frequency within tolerances after a contingency event.

---

16 At about 0.8% per second, which is between the rate of a fast-responding gas turbine and that of slower-responding steam plants or hydro units.
Historically, the combination of traditional inertia (from both generators and loads) plus PFR has been sufficient to address contingency events in most of the United States. But as the grid evolves with the addition of VG and other new technologies, system planners and operators are deploying new ways to maintain stable frequency even with declining amounts of conventional inertia.

Figure 12. Frequency after a contingency in our example case with the addition of PFR
5 Why Does the Amount of Inertia in the Grid Change?

5.1 The Impact of Variations in Electricity Demand

The demand for electricity varies as a function of time of day and season. As a result, power plants vary output and are turned on and off. This is illustrated in Figure 13a (next page), which shows both the changing load and the corresponding inertia that results from generators turning on and off.

As load increases, generators are turned on (committed) and increase output. This increases the amount of available inertia, which depends only on whether the generator is online and spinning at grid frequency (inertia from a generator is independent of power output). This results in the blocks of inertia observed in the figure. Inertia is at its lowest point in the middle of the night when many plants used to meet intermediate and peak load are turned off. Figure 13b shows the impact of varying amounts of inertia on frequency, assuming a contingency occurs at either 4 p.m., which is our base case, or at 4 a.m., when demand is much lower. Because the contingency size in both cases is the same, but the available inertia is lower during the period of lower demand, the frequency falls faster and just barely avoids UFLS.\footnote{This also means the period of greatest concern about frequency stability may be the opposite of when most other concerns regarding power system reliability occur (i.e., periods of highest demand).}

\begin{boxedtext}
\textbf{Inertia vs. Other Reserve Types: The Role of Headroom}

Inertial response is different from other operating reserve types that require power plants to vary output. A critical difference is the fact that inertial response from synchronous generators is inherent, uncontrolled, and independent of output level. Other reserve types require generators to operate at less than full output (hold headroom). This means that a 100-MW (rated) power plant operating at 60 MW can provide 40 MW of “upward” operating reserves, while the same plant operating at 100 MW cannot provide any. However, this plant will provide the same amount of inertia while operating at either output level or any level in between. As a result, the total inertia available from a plant is a function of only three factors: its size (power capacity), its inertia constant, and its rotational speed.
\end{boxedtext}
a) Load and inertia

b) Impact of a contingency at 4 a.m. and 4 p.m.

Figure 13. Relationship between system dispatch and inertia in our example case
5.2 The Added Challenge of Inverter-Based Resources

VG resources and battery storage are inverter-based resources (IBRs) that do not use synchronous generators to produce electricity. When generation from IBRs displaces significant amounts of generation from synchronous generators, the total amount of inertia during these periods will decrease, similar to the impact of reduced load observed in Figure 13.

Figure 14a illustrates the system dispatch in our example of a system where we have added enough wind to meet 30% of total demand on this day. The wind penetration on this day varies from 14% to 63%, reaching the maximum in the early morning. If we do not change how the system is operated, the amount of inertia will fall below that observed previously in Figure 13, and the frequency will fall below UFLS in the case of a contingency event during the overnight hours. This is illustrated in Figure 14b, which shows the frequency after a contingency at 4 a.m. or 4 p.m.

Inverters vs. Synchronous Generators

Most conventional power plants turn the spinning energy of a turbine into electricity via a synchronous generator, which inherently produces AC electricity. PV and batteries produce direct current (DC) electricity, which must be converted into AC for use by the grid. This converter is known as an inverter. Although wind turbines produce electricity via a spinning turbine, it is more efficient to use inverters as well. As a result, these technologies are collectively referred to as inverter-based resources.
These example cases show the motivation for addressing potential declines in inertia that can result from increased deployment of IBRs. Grid planners and operators have identified several options to address this concern and maintain reliable operation.
6 Alternatives to Conventional Frequency Response: Why Texas Went First

In response to IBR deployment that can reduce the availability of conventional inertia, grid planners and operators are changing sources of frequency response, including replacing traditional mechanical PFR with electronic-based alternatives. These changes are not only because of increased deployment of renewables; in some cases, new sources of frequency response may also be less costly or more efficient. In the United States, the most significant efforts have occurred in the ERCOT (Texas) grid.

6.1 The Texas Grid: A Big State, But a Small(er) Grid

Section 4 demonstrated why grid size (particularly relative to its contingency size) is a key factor for determining the rate at which frequency declines. Larger grids have inherently more inertia, and therefore time to respond.

Table 2 summarizes several key parameters for each of the three main interconnections.

<table>
<thead>
<tr>
<th>Interconnection</th>
<th>2018 System Size (GW) (Peak/Avg.)</th>
<th>Largest Contingency (MW)</th>
<th>Largest Contingency (% of Avg. Load)</th>
<th>Highest UFLS (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western</td>
<td>168/100</td>
<td>2,626</td>
<td>2.6</td>
<td>59.5</td>
</tr>
<tr>
<td>Eastern</td>
<td>556/354</td>
<td>4,500</td>
<td>1.3</td>
<td>59.5</td>
</tr>
<tr>
<td>ERCOT</td>
<td>73.5/43.0</td>
<td>2,750&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6.4</td>
<td>59.3</td>
</tr>
</tbody>
</table>

<sup>a</sup> Peak demand is sum of regional (noncoincident) peak demand from NERC (2019b).
<sup>b</sup> NERC (2017a)
<sup>c</sup> Increased to 2,805 MW in 2020 (ERCOT 2019)

The second column shows the size of each interconnection, based on both peak demand and annual energy (which for the WI and Eastern Interconnection, or EI, includes part of Canada). This demonstrates that on the basis of both peak ad average demand, ERCOT is less than half the size of the WI, and less than 15% of the size of the EI, meaning the WI and EI would have 2–7 times more available inertia than ERCOT, all else being equal. But this is not the case, as we describe below.

Section 4 also demonstrated the importance of contingency size. The third column in Table 2 lists the largest contingency that operators plan for in their reliability studies. In the WI and ERCOT these contingencies are two large, similarly sized nuclear plants. This means ERCOT must address about the same contingency size as the WI with less than half the inertia.

As a result, even without the impacts of VG deployment, ERCOT has led efforts to address the need for alternatives to conventional inertia and PFR. Figure 15 shows the decline in frequency that would result in a system that resembles ERCOT during a period of lower load without the addition of VG.<sup>18</sup> The green curve shows a case that depends only on conventional PFR and

---

18 Assuming 240 GW•s of inertia and typical PFR response times.
demonstrates that this case would have difficulty avoiding UFLS. Though ERCOT uses a lower UFLS setting (last column of Table 2) than the WI and EI, it still does not provide sufficient time for only mechanical-based PFR systems to address the imbalance created by the relatively large contingency. As a result, ERCOT has historically relied on loads as a source of frequency response (ERCOT calls this “load resources”).\textsuperscript{19} This consists of equipping certain large industrial loads with sensors that measure frequency and are programmed to disconnect automatically when the frequency drops to a certain level. In contrast with UFLS (which is typically involuntary), using loads for frequency response is voluntary and targeted toward specific non-critical loads, which are compensated for providing this service.\textsuperscript{20} In Figure 15, the blue curve shows the addition of 1,100 MW of load providing frequency response. These loads are designed to initiate a response when the frequency falls below 59.7 Hz and disconnects 0.5 seconds later.

![Figure 15. Decline in frequency in an ERCOT-like system before the addition of VG](image)

The use of loads as a source of frequency response allows ERCOT to maintain frequency while providing other benefits, such as potentially more efficient system operation, as discussed in Section 7.3.1.

### 6.2 How Wind in Texas Is Bigger: The Impact of Instantaneous Penetration

ERCOT has also led efforts to study and implement alternatives to conventional inertia because of its large deployment of wind. In 2019, ERCOT derived 20% of its energy from VG (mostly wind) and reached a maximum instantaneous penetration of 57.9% (ERCOT 2020). In contrast,

\textsuperscript{19} In addition, unlike the WI and EI, ERCOT requires all generators to provide PFR, and it has created a market for frequency-responsive reserves.

\textsuperscript{20} To participate in this service, the loads must go through qualification testing and provide data on availability and status to ERCOT on continuous basis in a manner similar to traditional generators.
the highest penetration of VG penetration in the WI in 2018 was 24% (Kroposki 2019). However, even if the WI were to achieve the same level of VG penetration, its larger size would result in more inherent inertia.

The impact of system size is illustrated in Figure 16, which shows the impact of a contingency in systems resembling ERCOT and the WI at an instantaneous VG penetration of 40%. The frequency in ERCOT falls much faster because of lower inertia, and it employs its load resources to respond to this frequency decline, as seen previously in Figure 15. The frequency in the WI drops at a much lower rate and does not need to employ load resources, as its normal PFR is sufficient to avoid UFLS.

Figure 16. The impact of system size on frequency in systems resembling ERCOT and the WI at 40% VG penetration

This example demonstrates why the WI and EI (which is even larger and has lower VG penetration) have not yet needed to make significant changes to how they maintain stable frequency. In contrast, ERCOT—based on both its size and VG penetration—has led efforts to study and implement approaches that address potential further declines in inertia that may occur with even greater deployment of VG. We explore some of these in the next section.

---

21 These very simplified simulations do not consider transmission limitations, and they assume the “average” PFR response rates used in Section 4. The simulations assume 1,400 MW of response from load resources in ERCOT. The contingency size in each region is based on values in Table 1. The ERCOT case assumes 1.1 GW of load resources, while the WI assumes 120 MW of load resources, which automatically trip in response to its largest contingency (NERC 2019a).
7 How Grid Operators Are Responding to Increased Penetration of Renewables and the Declining Role of Inertia

There are many possible—and proven—approaches to addressing the decline in inertial response in systems with increased VG penetration. Table 3 summarizes these options in three broad categories, discussed further in this section.

Table 3. Summary of Options to Maintain Frequency Stability

<table>
<thead>
<tr>
<th>Maintain Inertia (Section 7.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System operation</td>
</tr>
<tr>
<td>Synchronous renewable energy</td>
</tr>
<tr>
<td>Synchronous condensers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Provide More Response Time (Section 7.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced contingency size</td>
</tr>
<tr>
<td>Reduced UFLS settings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fast Frequency Response (Section 7.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load</td>
</tr>
<tr>
<td>Extracted wind kinetic energy</td>
</tr>
<tr>
<td>Dispatch of inverter-based resources</td>
</tr>
<tr>
<td>Additional or enhanced primary frequency response</td>
</tr>
</tbody>
</table>

7.1 Maintain Inertia

7.1.1 System Operation

Inertia can be maintained via operating the grid to ensure the mix of generators online exceeds critical inertia levels.\(^{22}\) Figure 13 showed how as both VG and load vary, power plants are turned on and off, which results in changes in the amount of inertia available. Historically, the amount of inertia was not monitored by system operators. However, ERCOT system operators now actively monitor inertia availability and can keep certain generators running to maintain inertia above critical levels. However, this action can have negative economic consequences by increasing the number of power plants online and operating at partial load, which reduces efficiency. It can also require curtailing the output of VG generators, reducing their economic benefit to the system. Still, this impact can be minimal if infrequent, and it can provide a useful temporary mechanism to allow grid operators to implement and test longer-term solutions discussed in subsequent sections, or to maintain grid stability if there are failures of those other systems.

\(^{22}\) ERCOT uses the term “critical inertia” to describe the minimum amount of inertia needed to avoid UFLS. As of early 2020, ERCOT has not needed to deploy this option, as normal operation has kept the system above critical inertia (ERCOT 2018c).
7.1.2 Synchronous Renewable Energy

Alternatively, options exist to deploy new sources of inertia. If the power system operator’s goal is to achieve high penetration of renewables (as opposed to only VG resources), several renewable technologies—including hydropower, geothermal, concentrating solar power, or biomass—use synchronous generators that can provide inertia. Other nonrenewable (but still low-carbon) resources, including nuclear and fossil plants with carbon capture, also use synchronous generators. Certain energy storage resources—pumped storage and compressed air—can also use synchronous generators or pumps that can add inertia to the system.

7.1.3 Synchronous Condensers

There are also non-generation alternatives to provide inertia. The most commonly discussed option is the deployment of synchronous condensers, which are synchronous motors/generators that draw energy from the grid to maintain a spinning mass and can inject power into the grid in the same manner as a synchronous generator. Historically, these have been installed to solve very localized grid issues (e.g., maintaining local voltage requirements). However, they could also provide a brute force solution to the potential need to maintain a minimum level of inertia (Kenyon et al. 2020). Synchronous condensers could be deployed by retrofitting generators from retiring plants or by equipping renewably fueled synchronous generators with clutches that would enable them to act as synchronous condensers when not generating. The cost of this option has not been compared to that of others.

7.2 Provide More Response Time

7.2.1 Reduced Contingency Size

Section 4.3 demonstrated that contingency size is a key driver of the rate of frequency decline and the need for inertia and other frequency response. As power systems evolve, contingency size may vary depending on the addition of new plants or the retirement of older plants. In addition, contingency size can be dynamically changed by reducing the output of the largest online plant, particularly during low inertia periods.

7.2.2 Reduced UFLS Settings

Decreasing UFLS settings can also add significant time for frequency response. ERCOT’s highest UFLS setting (59.3 Hz) is 0.2 Hz lower than that of the EI and WI, while the Quebec Interconnection (with a peak of 38 GW) uses initial UFLS setting of 58.5 Hz. Many smaller systems also have much lower requirements. More research is needed to understand the reliability and cost implications of lower UFLS settings.

Overall, several options can be used to provide alternatives to inertia in a high-VG system to slow the rate of frequency decline or otherwise provide more time for the system to respond after

---

23 For example, while ERCOT has installed synchronous condensers in part to respond to issues created by wind in West Texas, they were installed for reasons unrelated to inertia and frequency response (ERCOT 2018b).
24 Historically, synchronous condensers have lower amounts of usable inertia. The use of synchronous condensers for providing large amounts of inertia could require adding physical mass.
25 This depends on the type of resource that is the largest contingency and whether part-load dispatch is possible or economical. In several locations outside the U.S. the contingency size can be changed by downward dispatch of large units or transmission imports (ERCOT 2018a, O’Sullivan et al. 2012).
a fault. However, corrective action is still required to replace the supply of energy from the failed generator. If it is sufficiently fast, such corrective action can simultaneously replace traditional PFR and greatly reduce the need for inertial response—as described in the following section.

### 7.3 Fast Frequency Response

Fast frequency response (FFR) describes the ability of some resources, including inverter-based generators and load response, to increase the net supply of energy much faster than traditional mechanical-based PFR. This helps offset the impact of declining inertia, because the more-rapid decline in frequency that can result from lower inertia can be balanced by the more-rapid response from FFR. The multiple potential sources of FFR are discussed below.

#### 7.3.1 Loads

As discussed in Section 6, ERCOT has used loads as a source of frequency response for more than a decade (Newell et al. 2015).²⁶ Using loads provides some advantages over conventional PFR, including response time. Frequency can be measured, and mechanical relays can drop loads in a less than half a second (Matevosyan 2019).²⁷ This means load resources can essentially drop 100% of a targeted load in less than half a second, compared to the less than 2% per second response from most sources of PFR.

In addition, it can be less costly to incentivize industrial consumers to drop non-critical loads occasionally, as opposed to continuously providing PFR from conventional generators. Generators providing PFR must operate at less than full capacity, and generators running at less than full capacity typically are less efficient, meaning they incur additional fuel costs.²⁸

An example of a non-critical load is large cooling equipment; because load resources may only be required for a few minutes, dropping this load will not significantly impact temperature in a large cold-storage facility. Use of load resources in ERCOT has been historically rare—only 14 times between 2011 and May 2018 (ERCOT 2018a). Historically, load resources have been provided by larger industrial loads, as it is easier and less expensive to install and test the necessary equipment on a few large loads than many small loads.

As its markets have evolved, ERCOT has increased the amount of eligible load resources as well as the required response rate. Before 2018, load resources were allowed to provide up to half of ERCOT’s total frequency response requirement, and ERCOT now allows loads to provide up to 60% of its frequency-responsive services (Matevosyan 2019, ERCOT 2018c).

The use of frequency response from loads has been one of the main factors that has allowed and continues to allow ERCOT to increase its wind penetration, which has reached 58% on an instantaneous basis despite the associated decline in inertia. Other U.S. regions have yet to deploy significant load response because of limited need given their size and limited VG penetration.²⁹

²⁶ NERC also measures this for ERCOT and WI in terms of “credit for load resources” when calculating the contribution of loads to the PFR requirement (NERC 2019a).
²⁷ For additional discussion of the time to measure frequency, see Miller et al. (2017).
²⁸ For a discussion of the drivers of costs of operating reserves, see Hummon et al. (2013).
²⁹ The WI has 120 MW of load response, and the EI has none (NERC 2019a).
### 7.3.2 Extracted Wind Kinetic Energy

Modern wind turbines do not use synchronous generators and therefore do not provide inertia in the traditional sense (defined as inherently resisting changes in frequency). However, wind turbines do have kinetic energy in the rotating mass of the blades, shaft, and generator that can be extracted to rapidly inject real power into the grid. Though this service has not yet been widely deployed in the United States, it has been in use in the Quebec Interconnection since 2009 (Miller et al. 2017; Brisebois and Aubut 2011).

Provision of this service requires active sensing of grid frequency, so that when a decrease in frequency is sensed, the generator can be programmed to increase output to beyond what can be supported by steady-state wind speeds (Ela et al. 2014).

Figure 17 shows an example of a test that rapidly extracted kinetic energy from a 1.5-MW wind turbine. In this example, the turbine is operating under conditions that result in about 340 kW of steady-state generation (meaning fairly low wind conditions). In this example, the plant initiates a response about 0.4 seconds after a low frequency event. The plant then increases generation by about 70 kW in less than 0.5 seconds (in addition to the initial 0.4-second response time), or by about 10% of rated capacity per second. After this initial increase, the wind turbine will exhaust much of its stored kinetic energy, and its output will reduce. As with conventional inertia, this action can slow the frequency decay long enough for other mechanisms, including slower-responding mechanical PFR, to arrest and help restore frequency.

![Figure 17. Use of stored kinetic energy in a wind turbine to provide FFR](source: Gevorgian and Zhang 2016)

After a few seconds, the turbine output may fall below its initial output, as the turbine will need to use some of the collected wind energy to accelerate the turbine back to normal speed. This is similar to the way a synchronous generator will need to use energy to restore its kinetic energy after providing inertial response, with the restoration time depending on the amount of energy extracted and the wind conditions (Gevorgian, Zhang, and Ela 2015). Because the use of wind turbine kinetic energy has a limited response duration, other resources will need to increase...
output to restore the imbalance, including restoring the extracted wind kinetic energy. For both very rapid and sustained response, this type of FFR can be combined with FFR provided by upward dispatch of IBRs as discussed in Section 7.3.3.

### 7.3.3 Dispatch of Inverter-Based Resources

A second source of FFR that has been deployed at scale is response from IBRs, which include wind and solar as well as certain energy storage technologies. This option requires holding the generator at less than full output and using that headroom to increase output as needed, similar to the manner in which PFR is derived from conventional generators.

Both wind and solar are dispatchable resources in that they can reduce output from the maximum that is available based on instantaneous weather conditions, and they can then rapidly increase output to provide frequency support services. For wind, this requires reducing the output of the wind turbine, which is performed by changing the blade pitch angle and reducing the amount of energy extracted from the wind. For solar, the inverter can be commanded to reduce and then rapidly increase output via the use of power electronics. This operation reduces the amount of energy generated, so results in an economic tradeoff between providing energy and providing the ability to respond to imbalances.

Quebec has used wind to provide frequency-responsive reserves for more than a decade, and ERCOT has required wind generators to have frequency-responsive capability beginning in 2012 (Brisebois and Aubut 2011; Matevosyan 2019). In 2018, the Federal Energy Regulatory Commission required new utility-scale wind and PV plants to have frequency-responsive capabilities (FERC 2018).

With very high response rates, VG can provide significant amounts of FFR. After the time required to sense frequency and initiate a response, wind can increase output by as much as 25% per second, while PV can increase output over its full range in less than one second (>100%/second).30 This is much faster than conventional generators (as discussed in Section 4.5). Another benefit of FFR from wind and solar is their ability to operate over a large range of power plant output levels in contrast to thermal generators, which are restricted by minimum generation levels. Furthermore, the times when inertia is at its lowest due to VG penetration are precisely the times when large amounts of VG are available and likely to be operating in a curtailed state.

---

30 Response from IBRs requires time to measure frequency and then time to increase output (Miller et al. 2017). Actual response rates from wind are typically limited to reduce stress on the mechanical drivetrain (Wu et al. 2018; Chen et al. 2017). There has been limited testing of the upper bounds of PV response.
Energy storage is another inverter-based resource that can provide FFR. Some battery technologies can (1) increase output over full range in less than one second and (2) provide additional flexibility by rapidly switching between charging and discharging.

Overall, the rapid response from IBRs can mimic or even supersede traditional frequency-responsive reserves. However, it is important to note that because these services do not precisely match those from conventional generators, terminology is still in flux (see the text box at left, above).

Also evolving is the manner in which wind and solar can provide frequency-responsive services. The use of wind and solar to provide frequency response is illustrated in Figure 18. In it, the electromagnetic “chains” associated with synchronous generators have been replaced with sensors and software that controls the response of wind and solar to grid conditions. The response can be similar to that of synchronous generators, but the response of legacy generators is constrained by their physical characteristics and limits, and it is not necessarily perfectly aligned with the system’s needs. It is possible that FFR from VG could be optimized to respond based on grid conditions and needs, and could perform very differently from how synchronous generators provide frequency response (Loutan et al. 2017). Additional research is needed to understand how to best optimize provision of FFR from wind or solar to maximize its benefits to the grid.

Figure 18. Wind and solar providing frequency response
7.3.4 Additional or Enhanced PFR

Finally, traditional PFR from conventional synchronous generators can be enhanced via market mechanisms such as those that have been implemented in ERCOT, which is the only region in the United States with a market for frequency response. This option provides plant owners incentives to maintain or potentially enhance response rates, and it ensures the system dispatch captures the value of faster-responding synchronous generators when inertia is low. These markets can also incentivize deployment of newer, faster-responding synchronous technologies if frequency response provides a premium.31

---

31 As an example, Miller et al. (2018) discuss the possibility of increased response rate from a concentrating solar power plant using synchronous generators.
8 How Low Can We Go? Maintaining Frequency with Decreased Inertia

8.1 Studies of High VG Deployment Using Available Options to Address Declining Inertia

The measures discussed in Section 6 have allowed ERCOT to achieve 58% instantaneous penetration of wind and are expected to allow even greater penetration levels. The two larger U.S. grids have not yet needed to take significant action to address VG-driven declines in inertia, and studies have found that much greater penetrations are achievable while maintaining reliable operation with few changes.

The example provided in Figure 19 demonstrates the results of a study that examined a contingency event in the WI at up to 80% instantaneous penetrations of wind energy (Gevorgian, Zhang, and Ela 2015). In the base case for this study, the deployment of wind results in a drop in inertia, and the study assumes no additional frequency response measures, such as those already been deployed in ERCOT (e.g., the use of load resources or frequency response from wind and solar). Under these conditions, in the event of a contingency event, the mechanical response from PFR in the remaining synchronous generators is too slow to avoid UFLS at 59.5 Hz. However, the addition of frequency response from wind can address this concern.

Figure 19. Use of frequency response from wind to achieve an 80% instantaneous penetration in the WI
Figure 19 shows the use of FFR from wind, both by extracting stored kinetic energy from wind turbines (Section 7.3.2) and by increasing output from curtailed wind plants (Section 7.3.3). Each element was evaluated in isolation as well as combined. When simulating only the extraction of kinetic energy (not shown), the frequency still fell below UFLS, which reflects the need for additional energy to restore an imbalance. When providing only FFR, UFLS was just barely avoided. However, when providing both, the frequency decline was quite small. This shows that the effect of rapid response from stored kinetic energy, when combined with the slower response of increasing output, can effectively provide an alternative to the combination of inertia and PFR from conventional generators.

Overall, this and other studies point toward the potential to greatly increase VG penetration, particularly when employing the frequency response capabilities of VG (Tan et al. 2018; Liu et al. 2018). We are still in the early stages of studying very high VG penetrations, so analyses have yet to examine in detail scenarios in which entirely new means may be necessary to maintain stable frequency. However, researchers are considering the possibility of entirely VG-based systems, where VG resources are relied on to maintain frequency without the support of synchronous generators.

8.2 Beyond Inertia: Designing an Inverter-Based Grid

Zero-inertia microgrids have been in operation for decades, which demonstrates that inertia is not needed to operate an AC power system (Kroposki et al. 2017). Reliance on inertia is the result of the legacy use of synchronous generators. As interest grows in achieving even greater penetration of renewable energy in the grid, including paths toward 100% VG systems, there is discussion of how much (if any) inertia is actually needed to operate AC power systems. In scenarios with high annual percentages of VG, there would be meaningful periods where a system might operate at 100% instantaneous penetration of IBRs. This could result in a nearly zero-inertia power system—and would require additional research and new approaches to maintaining grid frequency when implemented at a large scale (Matevosyan et al. 2019).

In the current grid, VG inverters operate by tracking the existing grid voltage and frequency and injecting power following a reference waveform created by synchronous generators. These types of inverters are therefore known as grid-following, because they are literally following a reference waveform. Synchronous generators naturally form an AC voltage waveform (60 Hz in the United States) and self-synchronize with other synchronous generators (due to the forces represented by the chain in Figure 1, page 3). Figure 20a illustrates the concept of a grid-following inverter; it shows a synchronous generator creating a reference signal then followed by a VG inverter.
In the absence of grid-forming synchronous generators, inverters themselves will be responsible for establishing and maintaining this grid voltage and frequency. These inverters would thus be grid-forming. Because grid-forming inverters can create a waveform at a specified frequency, they can provide the basis of a synchronous AC power system.

Grid-forming inverters are already used in many zero-inertia microgrid systems (typically less than 10 MW) that do not use synchronous generators. These microgrid systems may employ one of the inverters to provide the master voltage and frequency waveform (grid-forming), with others acting in grid-following mode (Figure 20b).
A large zero-inertia power system would require inverters that could independently form a grid voltage and frequency—and maintain this voltage and frequency without an explicit communications network (Figure 20c). The field of grid-forming inverters is rapidly advancing (Lin et al. 2020; Ackerman et al. 2017). However, as discussed previously, the potential need for zero-inertia systems has yet to be explored in detail, particularly in grids with significant synchronous renewable or other low-carbon resources. The costs (or need) to develop a system that can reliably operate under near zero-inertia conditions has yet to be analyzed in detail, particularly in comparison to maintaining the current synchronous-based system with sufficient modification to accommodate very high levels of VG penetration.
9 Conclusions and Research Needs

Historically, nearly all grid capacity has been provided by synchronous generators. These rotating generators have stored kinetic energy, and this stored kinetic energy (inertia) provides the grid the time needed to respond to declines in system frequency that result from contingencies and other imbalances. This time is needed because response has traditionally been derived from a series of mechanical devices that may take a few seconds to respond to changes in frequency.

Replacing conventional generators with inverter-based resources, including wind, solar, and certain types of energy storage has two counterbalancing effects. First, these resources decrease the amount of inertia available. But second, these resources can respond much faster than conventional resources, reducing the amount of inertia actually needed—and thus addressing the first impact. In combination, this represents a paradigm shift in how we think about providing frequency response.

The combination of inertia and mechanical frequency response can be replaced to a large extent with electronic-based frequency response from inverter-based resources and fast response from loads, while maintaining system reliability. Given these solutions, reduced inertia is not an inherent barrier to increased deployment of wind and solar energy. Our reliance on inertia to date results largely from the legacy use of synchronous generators.

The level of renewable penetration achievable under this new paradigm has yet to be examined in detail, but the Texas grid has already met more than half of its energy demand with wind, and the other two large U.S. grids should face fewer challenges in achieving even greater levels of penetration, given their larger sizes. Well-understood and already deployed techniques and technologies can allow for substantially increased use of wind and solar in the U.S. grid without the negative impacts often associated with reduced inertia.

Further study will provide deeper insights into points at which new approaches might be needed to maintain system frequency at much greater levels of wind and solar deployment. In some regions, such as the western United States, the combination of fast-responding inverter-based generators, along with the use of renewables that employ synchronous generators, may allow extremely high levels of renewable generation without any fundamentally new approaches. Alternatively, smaller grids that must rely mostly on inverter-based renewables may need to deploy resources that transition away from the use of inertia as a resource to maintain system reliability.
References


EPRI (Electric Power Research Institute). 2019. Meeting the Challenges of Declining System Inertia, 3002015131 April 2019


Inertia: Basic Concepts and Impacts on the ERCOT Grid.


