A Market for Primary Frequency Response?
The Role of Renewables, Storage, and Demand

Working Paper by Thomas Lee

Table of Contents

Introduction

In the context of electric power grids, frequency response refers the ability of the bulk power system to autonomously adjust to sudden deviations from the regular alternating current (AC) frequency, which is 60 Hertz in North America. Frequency response is a crucial component of essential reliability services, which refer to a suite of tools available to the grid operator to affect the physical properties of electricity on the system. This suite of services is required for the bulk power system to operate reliably through expected and unexpected real-time disruptions. Discussions of essential reliability services commonly include frequency control, voltage control, and ramping.[1]

This paper will only focus its discussion on frequency response, which is effectively the power system's real-time ability to nearly instantaneously balance electrical supply-demand imbalances. However, the bulk power system's reliable operation also depends on the other aforementioned services not covered by this report. The focus here is on frequency response because it is an interconnection-wide property, compared to reactive power (related to voltage control) which is more local in nature and depends on lower voltage networks and distribution grids (Ela, Tuohy, et al. 2012).

There are existing ancillary service markets that cover ramping-related services (such as frequency regulation, contingency reserves, and load following). However, there are no common market mechanisms in North America to directly compensate frequency response, providing an opportunity to further explore the role of markets in providing this service. In addition, this paper seeks to explore the role that renewables, demand response, and energy storage can play in providing these services. Variable renewables now have the technical capability to provide frequency response and thereby contribute to essential reliability services. However, the most economically efficient solution may require creating a new market for primary frequency response, rather than a command-and-control requirement. Such a system would be a market for changes to supply or demand to bring the system into balance.

What is the Electric Grid's Frequency Response?

Power grids must always match electrical supply and demand. A stable AC frequency can be thought of as an equivalent physical condition for this constant supply-demand balance. Primary frequency response is defined as the ability of the electric grid to autonomously (i.e. without requiring any central dispatch or human intervention) adjust power output to counteract deviations in system frequency.

The costs of excessively deviating from a stable grid frequency are power outages, either in the form of 1) a blackout affecting the bulk power system, or 2) under-frequency load-shedding affecting a subset of customers. When normal operations (including primary frequency response) fails to adequately control a dropping grid frequency, "under-frequency load-shedding" involves "automatically disconnecting large, pre-set groups of customers at predetermined frequency set-points" (Eto, et al. 2018). Under-frequency load-shedding is employed as a "blunt, drastic" emergency line of defense in order to avoid the grid frequency reaching "a point at which generators disconnect automatically to prevent themselves from being damaged." Such an emergency of disconnecting generators would exacerbate the supply-demand imbalance and could lead to a widespread blackout (Eto, et al. 2018). In other words, the benefit of frequency management (including primary frequency response) is ultimately to prevent load outages on the bulk power system.

---

[1] Voltage control, which is closely related to reactive power, is important because industrial equipment and home appliances are designed to operate within certain voltage ranges, beyond which large deviations can cause degraded performance or equipment damage. Ramping refers to increasing or decreasing energy generation over the course of hours to follow major load shifts during the day (as opposed to frequency response's timescale of seconds to minutes), and are needed most during times of "morning ramp-up, afternoon ramp-down, and evening ramp-up" (NERC 2014).

Primary frequency response is not the same as secondary frequency regulation.[2] Frequency regulation refers to a central grid operator sending an automatic computer signal (called automatic generation control) that directs generators to inject or reduce power output; frequency regulation constitutes "secondary" frequency control and occurs after primary (i.e. instantaneous and autonomous) frequency control has already stabilized the initial frequency fluctuation to a certain level. Conceptually, the goal of frequency response is to limit the rate of the initial frequency deviation and stabilize system frequency to a manageable level, while frequency regulation's goal is to subsequently manage frequency back to the target level. This relationship is illustrated in Figure 1. Physically, primary frequency response occurs in a distributed and fully autonomous manner, while frequency regulation requires the central grid dispatcher's computer and is therefore subject to communications latency delay.

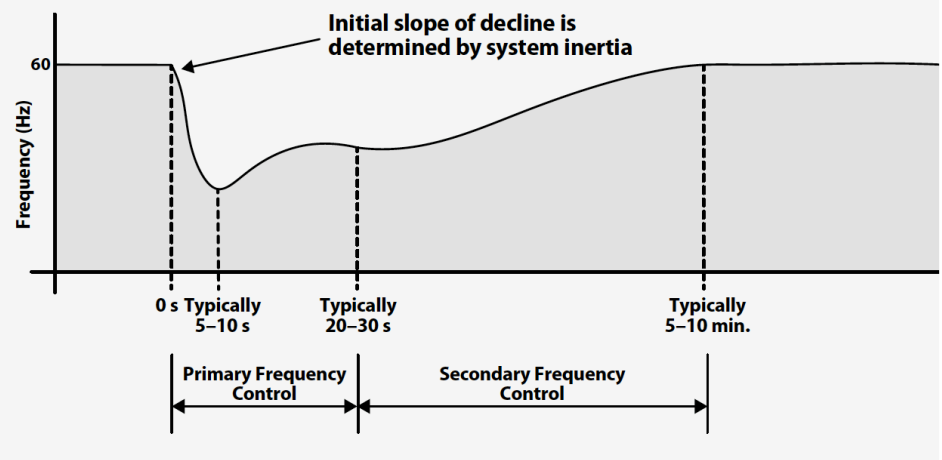

Figure 1: Time vs. frequency after a grid disturbance event causing a drop in grid frequency. Primary frequency control (or primary frequency response) is the autonomous stabilization immediately following the event, while secondary frequency control (or frequency regulation) is the comparatively slower recovery of grid frequency requiring the central operator to send communications signals to dispatch individual resources to adjust their power output. Source: Ela 2013.

Primary frequency response is usually measured in terms of the power output change in megawatts (MW) that results for every unit of frequency deviation (Hz). For example, a large power plant suddenly tripping offline reduces the total amount of available kinetic energy, leading the rotating generators on the system to start rotating less rapidly and thereby decreasing the AC frequency across the grid system.[3] Since a generator turbine's rotational velocity is directly coupled to the grid frequency, the generator's control systems can sense this frequency decline as an indicator of insufficient energy provision. The control system within each power plant, which usually has been in the form of a governor, can then automatically increase the plant's power output.[4] This process is autonomous because the

[2] This differentiation may be confusing because other jurisdictions like Europe use different sets of terminology.

[3] At a basic level, rotational kinetic energy (measured in joule or watt-second) is $KE = \frac{1}{2} I\omega^2$, where I is the rotational inertia or moment of inertia, and omega is the angular velocity (measured in radians per second). The moment of inertia is a property of the rotating object, and is a function of its mass and shape. Large, centralized, synchronous power generators provide their rotational inertia to the system. All else equal, lower kinetic energy leads to decreased angular velocity and thus decreased frequency (measured in times per second, or Hertz).

[4] Since the early days of the steam engine, the fly-ball governor is an archetypal example of a simple mechanical governor that exhibits frequency response (WECC Control Work Group 1998). As the engine's shaft spins more quickly, two masses or fly-balls undergo centrifugal motion and are pushed outwards, in turn moving a lever that reduces the throttle opening for steam input. As a result, the engine spins less quickly. This negative feedback reverses when the engine speed reduces too much.

governor does not have to wait for a central dispatcher to send a signal, thus bypassing communications system delays.

Figure 2 below is a stylized illustration of the frequency response behavior of a generator's governor. When the system frequency deviates from the nominal 60 Hz by larger than a preset amount, known as the "deadband," the governor will adjust power output. Past the deadband (the flat segment in the middle of the graph), the lower the system frequency is (towards the left of the graph), the more power output will increase (towards the top of the graph) to contribute to increasing frequency again. The higher the frequency (towards the right of the graph), the more nominal output will decrease (towards the bottom of the graph). In aggregate, the frequency response capabilities of individual generators, largely enabled through the use of governors, combine to form the entire system's frequency response behavior.

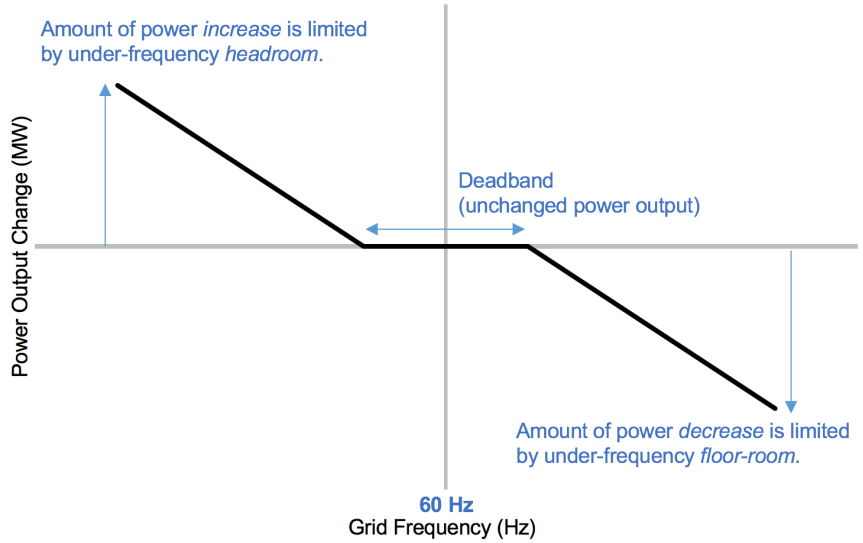

Figure 2: Stylized frequency response behavior showing relationship between grid frequency deviations and changes in power output, either of a generator or the whole grid.

Frequency Response Capability Versus Provision

As the stylized figure highlights, what matters is not only the power output's rate of change in response to frequency changes (MW / Hz) but also the total power over which this rate of change can be sustained (MW).[5] In other words, *capability* is different from actual *provision* of primary frequency response. When a generator installs necessary control equipment with properly enabled settings (as FERC now requires of essentially all new generators that can safely do so), it achieves the *capability* of frequency response in terms of how much MW of power output is adjusted for every Hz of grid frequency deviation. This capability measured in MW / 0.1Hz can be considered the "quality" of frequency response and corresponds to the slope of Figure 2's slanting segments.

Even after a generation plant has installed the equipment necessary for frequency response capability, whether it actually *provides* frequency response during a sudden low-frequency event requires generators to withhold "headroom," which is the difference between a generator's current power output level and its maximum possible level. Conversely, "floor-room" is required for the provision of frequency response during over-frequency events. Quantity matters, because for an under-frequency event if the

[5] As an analogy, in its revisions to its frequency regulation market PJM noted that batteries on the system had a fast rate of energy provision or withdrawal (i.e. power, measured in MW) but often did not have sufficient total energy over which this output can be sustained (MWh).

quantity of frequency response "is reduced to less than the amount of generation, frequency will again decline" (Eto, et al. 2018). The distinction between capability and provision is summarized in Table 1.

Table 1: Comparison of frequency response quality vs. quantity

Frequency Response Concept	**"Capability" or "Quality"**	**"Provision" or "Quantity"**
Units	MW / 0.1Hz	MW
Relation to Figure 2	Slope of slanting segments	How high and low the slanting segments extend
Depends on	Enabled equipment	Enabled equipment + headroom
Relative cost	Low	Potentially high

The Decline and Stabilization of Frequency Response Capability

In North American power systems,[6] frequency response has historically been degrading for the past two decades. In the Eastern Interconnection for example, frequency response during this period declined by an average every year of about 60 MW / 0.1 Hz (Gevorgian and Zhang 2016). Figure 3 shows an overall frequency response capability decline in the Eastern Interconnection, which has started since at least 1994 (Ingleson and Allen 2010). In Figure 3, "beta" refers to the interconnection's primary frequency response capability and is measured in units of power change per unit of frequency change, or MW / mHZ. Over this period, the decreasing amount of frequency response meant that the power grid was worryingly becoming less and less resilient in its ability to self-correct after frequency deviations.

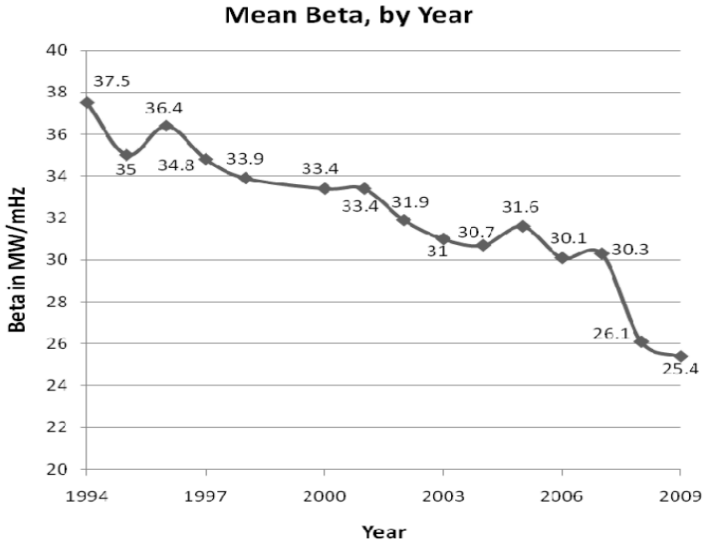

Figure 3: Historical trend of frequency response in the Eastern Interconnection up to 2009. Source: Ingleson and Allen 2010.

This decline in frequency response started before any significant deployment of variable wind or solar energy. Rather than renewables, the main culprits were disabled governor settings spurred by economic disincentives. For example, despite their physical capability to provide frequency response, in practice many synchronous generators[7] provided none or only a limited amount, via "excessive governor dead bands" or "blocked governors" (Gevorgian and Zhang 2016).

Why were generator governors not being used to effectively provide as much frequency response, even if they were technically capable? Economics has been a main driver for the decreasing provision of

[6] North American power systems include the Eastern, Western, Texas, and Quebec Interconnections.
[7] Synchronous generators are those that have physical masses rotating at the grid frequency through electromechanical coupling.

frequency response (i.e. providing below the actual physical capability of generators). For example, grid operators impose financial penalties for generators that deviate from their economic dispatch schedule, forming a strong disincentive for generators to allow output to vary according to system frequency (Ela, Tuohy, et al. 2012). More generally, having the ability to provide extra output during times of system under-frequency (such as after another large power plant suddenly trips offline) requires the generation plant to withhold some amount of backup power capacity, also known as headroom. Always leaving headroom means power plants are not generating at their physically available levels, leading to an opportunity cost of foregone energy compensation.

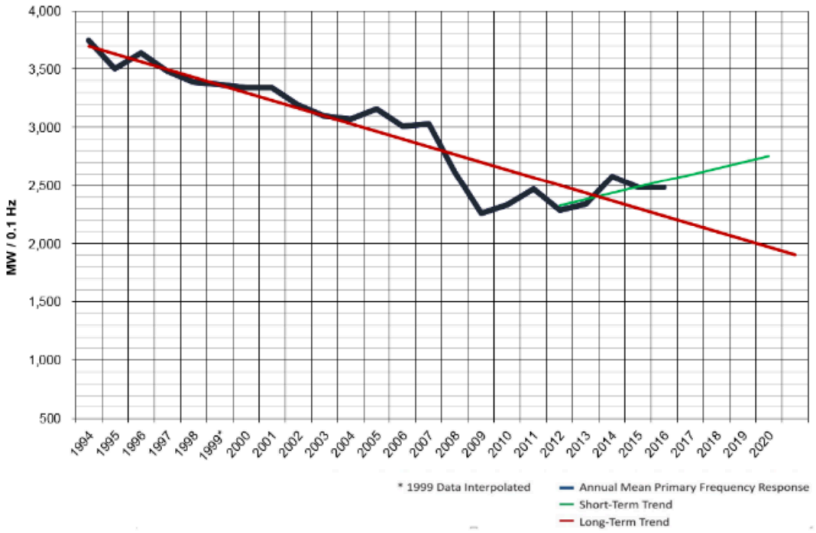

Figure 4: Trend of frequency response in the Eastern Interconnection up to 2016, showing more recent stabilization. Source: NERC 2017, Figure E.2.

Despite the historical downward trend, primary frequency response capability has actually been stabilizing and even recovering over the past several years. According to the North American Electric Reliability Council, "[t]hree of the four interconnections showed overall improvement while the Québec Interconnection frequency trend moved from "declining" to "stable" (NERC 2017). Figure 4 from NERC's report reproduces the period of substantial decline in the Eastern Interconnection, as seen in Figure 3, but also updates this plot with more recent data showing a potentially stabilizing short-term upward trend. Measuring frequency response as the energy output change per frequency deviation, Figure 4 has the same vertical axis as Figure 3 (except Figure 4's values are 100 times larger, since the units here are per 0.1 Hz rather than mHz). NERC attributes these recent improvements to deliberate operational improvements by grid and generation operators resulting from the "increased industry focus on frequency response in recent years" (NERC 2017). Currently, all interconnections have sufficient frequency response above target levels recently established by NERC (i.e. the Interconnection Frequency Response Obligations in BAL-003-1), as seen in Table 2.

Table 2: Comparison of obligated requirement vs. actual achieved primary frequency response (MW / 0.1Hz), for operating year 2016. Data source: NERC 2017.

Grid Interconnection	Interconnection Frequency Response Obligation	Actual Minimum	Actual Mean
Eastern	1,015	1,253	2,483
Western	858	902	1,545
Texas	381	404	807

Advances in Frequency Response Capability and Regulatory Requirements

Whether frequency response levels can continue improving, or at least remain stable, depends a lot on how the power system integrates higher amounts of more variable energy sources. As variable renewable energy resources provide increasing shares of generation on power grids, some stakeholders have raised concerns that frequency response may decline again. Concerned parties note that in the past, variable wind and solar plants usually were not able to inherently provide primary frequency response. For example, photovoltaic solar panels produce direct current (DC) energy, which must be converted to AC current using inverters. Similarly, even though wind turbines spin, the speed of their generators are usually variable and dependent on wind speeds, rather than always being synchronized to the grid frequency. Greater shares of renewables mean lower shares of conventional generation, which inherently have frequency response capability (even if not always fully utilized).

More recent technological advances enable variable renewable energy resources to possess primary frequency response capability. For example, with the use of power electronics[8] controlling the pitch of turbine blades, wind turbines can provide frequency response capability (Ela 2013). Solar photovoltaic (PV) plants can also provide frequency response capability with advanced power controls, as illustrated by a 2016 demonstration of a 300MW solar PV plant (Loutan, et al. 2017).[9] Sophisticated industry practitioners are aware of renewables' technical ability for frequency response. For example, the PJM Interconnection's 2017 report recognized that even though "legacy" nonsynchronous wind and solar generators are unable to provide frequency response, "newer non-synchronous generators, such as wind and solar, have the ability to provide frequency response with power electronics, which can be programmed to provide frequency response for very short periods using power electronics" (PJM Interconnection 2017). Still, it is important to remember that the actual provision of frequency response during an under-frequency event requires the wind or solar resource to be operating below its maximal capacity at that point in time.

In addition to renewables, non-generating resources can also provide primary frequency response. Electrical load by customers can be frequency-responsive. Simulations have demonstrated the technical ability of energy storage resources to provide primary frequency response and system inertia (Delille, Francois, and Malarange 2012). At the expense of higher upfront costs, energy storage systems can "provide frequency response significantly faster than the existing primary response," with laboratory experiments suggesting a time delay of about 80 milliseconds between frequency deviations and the response in the storage system's power output (Greenwood, et al. 2017).

Ela et al. (2012) explain that some loads will naturally draw less power when system frequency decreases. Intuitively, the feedback mechanism works like this: when bulk power system generation suddenly drops there is not enough power so generators spin less rapidly in aggregate; conversely, this reduced frequency means that certain inductive loads[10] (such as industrial motors) will spin less rapidly and draw less power usage. In practice, electric load facilities utilizing frequency-sensitive switches are currently being used as load resources in the Responsive Reserve Service program launched by the Electric Reliability Council of Texas (ERCOT). [11] In this program, participating sites install under-

---

[8] Power electronics are semiconductor electronics used for power conversion.

[9] This project also demonstrated the ability of the solar plant to follow an automatic generator control (AGC) signal, which is known as frequency regulation, as well as provide reactive power.

[10] An inductive load is characterized by the magnetic field produced when a changing current passes through it, as opposed to a capacitive load which stores charge in the form of an electric field. All electric motors are inductive.

[11] More info on ERCOTs program can be found at https://www.enernoc.com/resources/datasheets-brochures/faq-ercots-responsive-reserve-service-program or http://www.ercot.com/content/services/programs/load/Load%20Participation%20in%20the%20ERCOT%20Nodal%20Market_3.02.doc

frequency relays, which are triggered to automatically and instantaneously interrupt power when the system frequency drops below 59.7 Hz (EnerNOC 2017). Similarly, National Grid in the United Kingdom has a Frequency Control Demand Management program, where "electricity demand is automatically interrupted when the system frequency transgresses the low frequency relay setting on site" (National Grid 2015).[12]

In February 2018, the Federal Energy Regulatory Commission (FERC) finalized a new rule entitled "Essential Reliability Services and the Evolving Bulk-Power System—Primary Frequency Response" that requires all new electric generation plants connecting to the transmission grid—except nuclear and combined heat and power plants—to install equipment for primary frequency response capability (Federal Energy Regulatory Commission 2018). The rule requires each new generator interconnecting to the transmission system to have primary frequency response capability as a precondition for being connected to the grid system. The rule amends the Large Generator and Small Generator Interconnection Agreements to require all new generating plants to install and operate a "functioning governor or equivalent controls." As they have unique operating characteristics and safety requirements, nuclear generators are exempt from providing primary frequency response.

Headroom Provision—An Unresolved Issue

While the FERC rule establishes standards for installing and operating frequency response capability, it does not directly resolve the provision of sufficient headroom. FERC elucidated that its new rule does not mandate any "generic headroom requirement" nor require specific compensation schemes. Therefore, primary frequency response remains an issue that is not directly addressed by policy or market design. This might become problematic. In fact, the Electric Power Research Institute's comment to FERC notes that the "majority of the [stakeholder] comments centered around the idea that provision of primary frequency response service may be more important than resource capability by itself" (Electric Power Research Institute 2017). In particular, FERC states that the "greatest cost associated with providing primary frequency response results from maintaining headroom" (Federal Energy Regulatory Commission 2018).

How much frequency response headroom does the power grid actually need? In the absence of new directives, the North American electric grid has no explicit targeting of the necessary amount of frequency response headroom. Rather than set specific guidelines about the amount of headroom needed, NERC's standards for primary frequency response such as the Interconnection Frequency Response Obligation focus only on the capability aspect, in terms of the MW / 0.1Hz slope. From an economics perspective, the efficient level of a headroom reserve is where the marginal benefits equate the marginal costs. In electricity markets, an example of such an explicit matching of a supply curve (marginal costs) with a demand curve (marginal benefits) is the downward-sloping "variable resource requirement" demand curve used in PJM's forward capacity market (PJM 2017). Since the economic value of frequency response services is also based on the prevention of load outages, a similar demand curve concept could be applied to determine the necessary amount of headroom provision. From an engineering perspective, the necessary level of headroom should be related to the largest single generator online at a given time, since the most common cause of frequency events is the unexpected loss of generators (Eto, et al. 2018). In other words, a demand curve for frequency response headroom would be close to a vertical line: anything less than the largest-generator level is not a sufficient contingency, and a level that is much higher is likely excessive. In this sense, the necessary amount of headroom can be viewed as a megawatt

[12] More info on National Grid's program can be found at http://www2.nationalgrid.com/uk/services/balancing-services/frequency-response/frequency-control-by-demand-management/

requirement, not unlike the reserve requirement levels (rather than demand curves) common in most existing ancillary services markets (Ellison, et al. 2012).

Under current conditions, grid systems may not actually need any explicit targeting for headroom provision. This is because generation headroom naturally arises as a byproduct of economic dispatch, in addition to other existing targets of operating reserve margins like spinning reserves. Dispatch refers to the set of operational decisions about when to run each generator, and for what levels of energy generation. At any given point in time, not all of the generators that are online are fully utilized to generate energy or are otherwise withheld for other types of reserves. For a partially dispatched generator that is online at a certain time and also has frequency response equipment installed, its remaining online power capacity (in MW) naturally contributes to total system headroom.

An example of this relationship is shown in the Figure 5, based on energy market supply curve data from PJM. For a fixed set of generators supplying energy at varying marginal costs, the amount of implied headroom changes dynamically as the load demand changes. As load demand increases from zero (moving from left to right on the graph), additional new generators must be brought online to supply the additional energy; not all of these additional resources are completely committed and so the amount of headroom increases. Past a certain point, further increases in load start to lead to an overall decrease in headroom as the online generators become more fully utilized. This pattern where the lowest headroom occurs at the lowest load levels (towards the graph's left) is consistent with the fact that "the reserves available to provide primary frequency control" during times of very low load "may be at a minimum" because a fewer number of generators are online (Eto, et al. 2018).

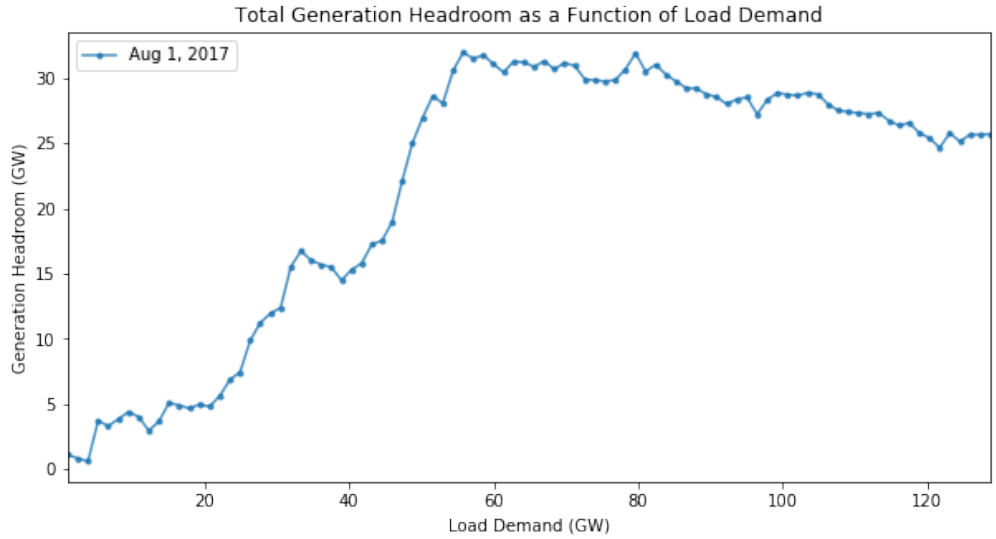

Figure 5: Estimated relationship between load demand and headroom using historical energy market supply data for an example day, measured as the total remaining capacity of online resources. Constructed using PJM's "Daily Energy Market Offer Data" available at http://www.pjm.com/markets-and-operations/energy/real-time/historical-bid-data/unit-bid.aspx. A more detailed explanation of the calculation can be found in the Appendix.

There are currently no systems that directly compensate generators for providing frequency responsive headroom in North America. A Sandia National Lab report explains that "[w]hile primary frequency control reserve is provided by all resources with autonomous governor response that are synchronized (and have the headroom to increase generation), none of the regions provides compensation for this reserve" (Ellison, et al. 2012).

A very closely related type of reliability reserve to primary frequency response is spinning reserves, which *are* compensated in regions with organized markets via ancillary service markets.[13] Spinning reserves, or synchronous reserves, are backup resources which must be online, synchronized to the grid frequency, and can be called on to make up a sudden generation shortfall. The major distinction with primary frequency response is that spinning reserves are intended to operate on a much slower timeframe because they are required to respond to the central grid operator's signals within ten minutes (PJM State & Member Training 2014).

At the same time, it is important to note that any generator that uses part of its power capacity as spinning reserves, if it has properly enabled primary frequency response equipment, also naturally contributes its remaining headroom toward the system's frequency response provision. In other words, generators that possess enabled frequency response equipment and also bid into spinning reserves are effectively choosing "to respond both to system operator commands [on the order of ten minutes] and autonomously to frequency deviations" (Ela, Tuohy, et al. 2012).

Of course, the Sandia report notes that *all* resources satisfying both the equipment and partially-dispatched criteria would provide headroom. Not all such resources with the frequency response capability necessarily choose to participate as spinning reserves; conversely, not all spinning reserve resources (which implicitly possess upward headroom) have primary frequency response capability. Therefore, while the combination of existing spinning reserve requirements plus FERC's new rule might strengthen total frequency response provision, there is no guarantee for sufficient headroom provision. How frequency response headroom changes in the future is unclear, especially as the generation fleet continues to evolve.

The Effect of Increasing Renewables on Headroom Provision

In the status quo, variable renewable resources generally do not provide frequency response headroom. In practice "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" (Eto, et al. 2018). The main reason for not providing headroom 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" (Eto, et al. 2018). An important exception is when there is excess variable wind or solar generation above the load (e.g. such as times where renewable energy is curtailed anyway, based on economic dispatch of energy generation), then the renewable units would naturally provide headroom without incurring any opportunity costs.

Generating renewables at their maximum output at each point in time is really a consequence of the principle of a supply curve or economic dispatch, which generally calls on using lower marginal cost resources first in order to maximize economic welfare. On the renewable project level, any curtailment regardless of reason decreases the volume of energy generation and in turn decreases cash flows (as well as federal production-linked tax credit and state renewable energy credits) and thus hurts attractiveness to investors and lenders (Cheszes 2012). On the system level, renewables are economically advantageous in terms of marginal energy costs, adding essentially zero marginal cost power into the system and pushing higher cost units out of the market, lowering clearing prices. Thus when renewables are curtailed there is the potential for energy market prices to rise, the costs of which are borne by all end-use electric consumers. Among the possible sources of frequency response headroom, curtailing output for headroom using wind and solar is the least cost-effective option. This is because the marginal cost for these renewable sources is zero, while other sources have fuel costs.

[13] Examples of ancillary service markets include non-spinning reserves, spinning reserves, and frequency regulation.

As a concrete illustration, suppose a hypothetical grid's generation resources consist of just two natural gas plants and a solar-plus-storage system, with their power capacities and associated marginal cost structures for a hypothetical hour in Table 3.[14] Assume over this particular hour the solar plant is known to generate a certain and constant 100 MW. Also assume all these plants have frequency response equipment installed and are capable of the exact same frequency response slope in terms of MW / Hz. If the load demand for a certain hour is a constant 700 MW, then what is the most economical configuration to provide 300 MW of upward generation headroom?

Table 3: Comparison of economic dispatch vs. headroom-sharing dispatch to achieve the same levels of load and headroom, for a hypothetical hour

Resource	Capacity (MW)	Marginal Cost ($ / MWh)	**Method 1 (Status Quo): Economic Dispatch** Generation (MWh)	Headroom (MW)	**Method 2: Headroom-Sharing (30% for Each)** Generation (MWh)	Headroom (MW)
Solar + Storage	100	0	100	0	70	30
Gas 1	500	20	500	0	350	150
Gas 2	200 200	25 30	100	300	280	120
System	1,000		700	300	700	300
			Clearing Price = $25 / MWh Load pays $17,500		**Clearing Price = $30 / MWh Load pays $21,000**	

The standard solution of economic dispatch ("Method 1") is to utilize all the lowest marginal cost resources first, before proceeding to generate electricity using higher marginal cost resources. In the hypothetical example, economic dispatch entails generating the maximum solar output of 100 MW during the hour, generating the full 500 MW of the first gas plant, and then generating the remaining 100 MW from the second gas plant. The remaining 300 MW of unused gas capacity of the second gas plant functions as headroom, which arises naturally as a byproduct of economic dispatch rather than any explicit targeting. The energy clearing price during this hypothetical hour is $25 / MWh (since the next marginal unit of generation would still come from the second gas plant's first output segment).

Another method ("Method 2") might be to equally share the provision of headroom across all generators, i.e. require all generators to withhold 30% of their power output as headroom, resulting in the same levels of energy generation and headroom as before. Compared to economic dispatch, this solution shifts generation from zero and cheaper resources towards a higher marginal cost output segment, thereby increasing the market clearing price for electricity generation to $30 / MWh. This solution is inefficient because it decreases total economic surplus, and particularly raises the cost of electricity borne by load customers to achieve the same level of energy and reserve services. In general, this example comparison between two dispatch allocations demonstrates that curtailing the energy generation of relatively lower marginal cost resources (such as the zero marginal cost of renewables generation) can increase total system costs. This is the reason why grid operators generally utilize economic dispatch.

In the above example, the economic dispatch solution was able to achieve headroom as a byproduct because there is a generator that is online but only partially dispatched. However, an influx of zero marginal cost renewable energy can push out these originally partially dispatched resources from being dispatched at all. According to Ellison et al. (2012), "[a]s the amount of power provided by variable

[14] For this illustration, ignore ancillary service markets, or assume the example values already account for these other reserve requirements being fulfilled.

renewable generation increases, the fraction of on-line generation capacity offering primary frequency control will decrease."

Table 4: Comparison of economic dispatch vs. headroom-sharing dispatch, for a hypothetical hour where economic dispatch leads to a lower level of headroom

Resource	Capacity (MW)	Marginal Cost ($ / MWh)	**Method 1 (Status Quo): Economic Dispatch** Generation (MWh)	Headroom (MW)	**Method 2: Headroom-Sharing (At Least 30% for Each)** Generation (MWh)	Headroom (MW)
Solar + Storage	300	0	300	0	210	90
Gas 1	500	20	400	100	350	150
Gas 2	200 200	25 30	***Not online***	0	140	260
System	1,200		700 **Clearing Price = $20 / MWh** **Load pays $14,000**	100	700 **Clearing Price = $30 / MWh** **Load pays $21,000**	500

Table 4 builds on the same example as before with the same cost structures, but now solar-plus-storage has grown to a certain and constant 300 MW output for this hypothetical hour. Under economic dispatch ("Method 1"), which is guaranteed to minimize system costs and maximize economic welfare, the clearing price now drops, because the higher marginal cost gas plant drops out from the dispatch. On the one hand, the addition of new renewables lowers energy costs for load customers, because the higher marginal cost plant no longer needs to run. On the other hand, the higher marginal cost plant going offline also means it no longer provides any headroom relevant for primary frequency response. Since primary frequency response is an autonomous, nearly instantaneous behavior, only generators that are already running can provide this service. Unlike in Table 3, Table 4 illustrates that deviating from economic dispatch by curtailing lower marginal costs resources ("Method 2") may be necessary to meet a certain level of headroom. With increasing renewables, there may be more binding tradeoffs between capturing the full economic cost savings from zero marginal cost renewables versus keeping higher cost resources partially running to maintain frequency responsive headroom.

The overall tradeoff between economic efficiency and headroom provision under high levels of renewables can be estimated using the aforementioned generator-level energy market offer and wind generation data published by PJM. These hypothetical calculations are intended to estimate the total amount of headroom available on an hourly resolution, so realistically the generators that have frequency response equipment installed and enabled are likely a subset of this amount. For reference, the actual amount of frequency responsive headroom in PJM at any point in time would be roughly equal to the calculated value multiplied by the percentage of frequency-response-capable generators, or around 66% for the Eastern Interconnection, as in Table 5 below (but this fraction can be time-varying and differs within the Eastern Interconnection).

Table 5: Fraction of generation capacity that responds to changes in frequency and sustains response, by interconnection. Source: Eto et al. 2018.

Interconnection	Eastern	Western	Texas
Percentage of Generation Capacity with Frequency Response Capability	66%	57%	52%

Figure 6 below shows a comparison between the natural byproduct headroom arising from economic dispatch for two cases using 2017 load and conventional generator supply curve data: the status quo versus a scenario with ten times as much wind generation capacity.[15] The orange line shows the baseline level of headroom at every hour of the year arising solely as a byproduct of economic dispatch. The blue line assumes that wind generation is increased to ten times the current levels (ignoring solar for simplicity), and the same method of economic dispatch is followed to always produce at the maximum available renewable capacity in each hour. Figure 6 illustrates how in certain hours with high wind penetration, the net load covered by conventional generators drops and so the associated natural headroom also decreases in accordance with the general pattern seen in Figure 5 above. The distribution of byproduct headroom levels becomes more negatively skewed in this high-wind generation case: the average level of total headroom decreases by about 7.5%, whereas the minimum headroom level over the whole year decreases by about 88%. The ten times growth case corresponds to about 27% of energy generation share; even greater amounts of renewable energy deployment would further erode headroom provision arising from economic dispatch.

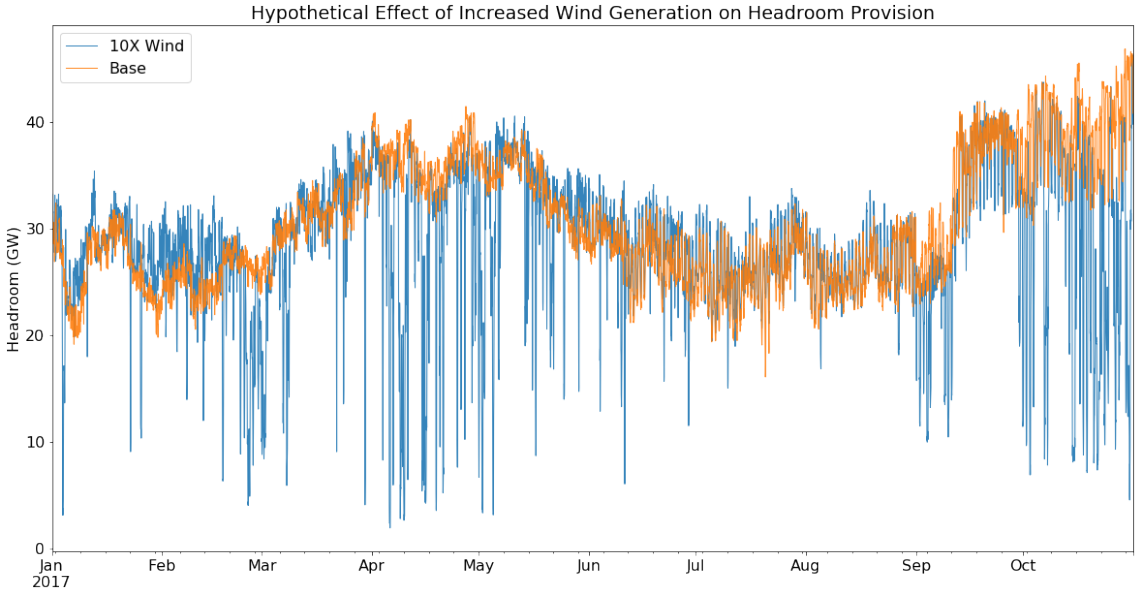

Figure 6: Estimated hourly total generation headroom in PJM, compared to hypothetical scenario with ten times more wind generation, for first ten months of 2017. Values are calculated assuming economic dispatch with no curtailment of wind generation. The baseline orange curve generally follows a similar annual seasonal shape as load demand, with a smaller peak during summer and a larger peak in winter. Constructed from PJM's "Daily Energy Market Offer Data" and hourly resolution wind generation data. The calculation method and simplifying assumptions are described in the Appendix..

While the hypothetical decrease in byproduct headroom provision may appear alarming, it is important to note that within this sample period for PJM the single largest online generator in any given hour is on the order of one to two GW. As mentioned previously, primary frequency response is intended to protect grid reliability against unexpected disruptions, most prominently those due to the sudden loss of a large generator. Over the course of this simulation with the high wind scenario, only 0.1% of the hours saw headroom provision dip below 3 GW, which corresponds to 2 GW of primary frequency responsive headroom if using the 66% figure from Table 5. In other words, the high wind generation case in the simulation still provides sufficient amounts of frequency response headroom even when abiding by the economic dispatch system of never curtailing renewable energy. This suggests that a much greater

[15] For reference, PJM had about 8.7 GW of nameplate wind capacity as of the end of 2017 (Monitoring Analytics 2018). This corresponded to about 2.7% of total energy generation throughout 2017. The “ten times” scenario would correspond to about 87 GW of wind in terms of nameplate capacity, and 27% of energy generation.

amount of renewable energy resources might be able to be integrated into the bulk power system, without jeopardizing system reliability and without deviating significantly from the status quo operational system of economic dispatch.

Further, although higher levels of renewables might raise the tradeoff between economic efficiency and headroom provision, greater renewables deployment might decrease the necessary amount of headroom due to a decentralization effect. Noting that the "chief reason PFR [primary frequency response] is needed, particularly upward PFR, is because of the sudden loss of individual large conventional generators," the American Wind Energy Association's comment to FERC suggests that the "need for PFR may decline as the market gradually transitions away from large conventional generators towards smaller generators"—where the sudden loss of any individual small wind or solar plant would be too small to cause a serious frequency disturbance (American Wind Energy Association 2016).[16] Still, increased penetration of distributed energy resources might make the system more susceptible to outages on the distribution system, a part of the overall electrical grid seen as most vulnerable.

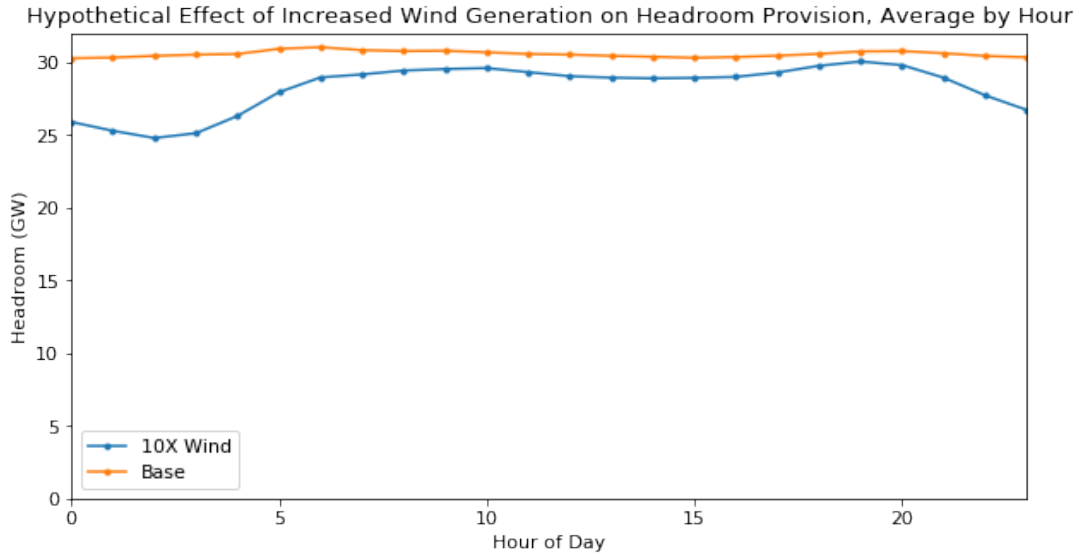

Figure 7: Average headroom by hour of the day assuming economic dispatch of resources (no renewables curtailment), showing that the impact of high wind generation levels on decreasing headroom tends to occur during off-peak hours.

Figure 7 further specifies that most of this expected decrease in natural headroom would tend to occur during off-peak or relatively lower load hours. This effect results from lower load hours corresponding to the left and upward-sloping portion of Figure 5's pattern, where increasing wind generally leads to lower net load requiring conventional generation, in turn lowering byproduct headroom provision. Another factor is that wind generation on average is usually higher during nighttime hours. Figure 7's pattern is consistent with the statement by Eto et al. (2018) about how it is more challenging to ensure sufficient primary frequency response during off-peak hours, where the challenge might become heightened by greater usage of renewables. A practical implication of this pattern is that this decrease in headroom could be mitigated by pairing wind generation with resources that might usually possess reserves during off-peak hours, such as energy storage. However, under the current market design, such resources have no incentive to provide headroom for the purpose of primary frequency response.

[16] Such an idea of resilience through decentralization is consistent with Amory Lovins's vision of the soft energy path, for example as explained in *Brittle Power*.

Table 6: Summary statistics of hypothetical headroom provision measured in GW in PJM at different levels of wind energy penetration (expressed as multiple of present day), assuming economic dispatch with no renewables curtailment.

Multiple of Wind Growth	1X (Baseline)	2X	3X	4X	5X	6X	7X	8X	9X	10X
Mean (GW)	30.6	30.5	30.5	30.5	30.3	30.1	29.9	29.5	29.0	28.3
Median (GW)	29.5	29.6	29.6	29.7	29.7	29.6	29.5	29.4	29.2	28.8
Min (GW)	16.1	17.5	18.9	15.4	13.9	10.3	5.5	3.0	2.2	1.9
Std (GW)	5.8	5.6	5.5	5.3	5.3	5.3	5.5	5.8	6.4	7.0

Table 6 summarizes the impact of gradually increasing levels of wind generation on the hypothetical amount of headroom provision. The mean and median of the distributions remain relatively constant throughout. Conversely, the distribution of headroom provision becomes progressively more negatively skewed (along with a wider dispersion in terms of standard deviation) as the amount of wind generation increases. This is another way of seeing how the low load hours, with already lower headroom provision originally, would tend to see more severe headroom declines in the presence of greater renewables. Another implication is that the hypothetical decline in headroom levels is gradual, rather than abrupt, in accordance with the renewables increase. So the share of energy from intermittent renewables can still grow before the status quo mechanism leads to headroom provision at a level significantly below current levels.

In summary, even though renewables and other alternative resources have the technical ability to provide frequency response, doing so may not yet be economical. Just as it is for traditional generators, leaving power output headroom as backup for under-frequency events constitutes an opportunity cost for renewable and alternative resources. PJM explains that "[m]arket changes may be needed… to incent non-synchronous generators to provide primary frequency response because, to do so, they need to operate below their maximum output" (PJM Interconnection 2017). The Sandia report similarly raises the possibility that "[i]n the future, market mechanisms may be needed to ensure sufficient provision of primary frequency control reserve" (Ellison, et al. 2012).

Mechanism Design for Frequency Response Headroom Provision

Different resources face divergent costs in providing frequency responsive headroom. The previous section shows how curtailing zero marginal cost renewables for headroom generally incurs the largest economic costs to the grid system. In addition, curtailing production has materially significant financial implications for the renewable energy project. A generating resource usually faces no additional costs to provide at least some amount of headroom if it is not fully utilized according to economic dispatch, but costs are incurred (both by generators and the system) if lower marginal cost generators are curtailed to allow a higher cost generator to come online and provide headroom.

Load demand choosing to participate in providing headroom would also face divergent opportunity costs. The "value of lost load" is an indicator estimating the "economic consequences of power blackouts and the monetary evaluation of uninterruptedness of power supply," which can differ according to factors including the region and industry as well as time and duration (Schroder and Kuckshinrichs 2015). For example, a value of lost load study using an input-output model, dividing the German economy into 51 sectors and including inter-sectoral linkage effects, estimates that an outage affecting the pulp and paper sector causes economic damages of 0.82 Euros per kWh, compared to 10.70 Euros per kWh for the printing and publishing sector; overall, sectors like coal, railway, and metals face lower costs from interrupted electrical supply (Praktinknjo 2016). Power blackouts are clearly harmful to society in general, but different parts of society are harmed by varying levels.

Therefore, as long as the proper frequency response equipment is installed, a load consumer has the potential to provide primary frequency response at lower opportunity costs when compared to generation resources, particularly renewables. For companies facing relatively lower values of lost load, "[l]oad shedding measures, meaning selective (voluntary) interruptions of costumers with the lowest Value of Lost Load, could prove to be more cost-efficient than the construction and operation of storage or generation capacities for only a few hours in a year" (Praktinknjo 2016). Whereas a wind or solar generation resource must curtail output during all normal operation hours and thereby incur opportunity costs even when an under-frequency event is not occurring, a "load that is set up to contribute frequency response will operate at its economically optimal level at all times other than the seconds or minutes when system frequency falls below the predetermined set points" (Union of Concerned Scientists 2016).

An energy storage system can be thought of as a hybrid of generation plus load demand characteristics, and therefore its costs to provide headroom might vary throughout the day as its role in the grid system may change dynamically.

When costs for the same service are heterogeneous between different suppliers, markets can play a role to ensure the most efficient resource allocation. From the earliest days of economics, Ricardo's theory of comparative advantage showed that gains from trade arise when each firm specializes in the product that it can produce at the lowest marginal cost. The analogy to frequency response is that between providing energy generation versus frequency response headroom, variable renewables with zero marginal cost are the most effective at the former and so ought to specialize in generating energy. Conversely, non-generating resources (like storage at certain times and frequency-responsive demand) may be the most effective at providing primary frequency response. Each load consumer firm is best informed about how much opportunity cost a particular application faces when its electricity supply is interrupted and whether potential frequency response service payments are enough to offset that cost; a market structure would enable this information to be incorporated with the potential of lowering overall system costs. According to Ela et al. (2012), the "new trend of different market participants being able to supply different forms of frequency-responsive reserve at different costs, may reveal a frequency-responsive reserve market that is a more efficient alternative to unit-specific requirements." This economic principle is also well-understood in environmental economics, where "it is widely recognized that abatement-cost heterogeneity is a fundamental determinant of the potential cost-savings with market-based policy instruments" (Newell and Stavins 2003).

In addition to efficiently allocating resources within a single time period, markets can help achieve dynamic efficiency in the longer term when cost structures change, as well as direct investments into research. Competition means resource providers will seek to innovate and minimize costs of a particular service, leading to the least-cost solution to provide a desired level of service. As demonstrated in the study by the California Independent System Operator (CAISO) on solar energy's ability to provide essential reliability services, ongoing research and development is continually expanding the technical capability of power electronics; flexible markets can provide a price signal that incentives technological innovation.

To be sure, the previous sections show that a requirement for frequency response headroom provision, if mandated today, would not be strictly binding because economic dispatch seems to be providing sufficient headroom as a byproduct. Equivalently, if a market for primary frequency response headroom were to be introduced today in North American electricity systems, the market price would be zero. The real challenge arises not now, but in the future, when such a requirement might become binding and the market price would become positive; however, grid operators do not currently have a mechanism to implement this price.

As part of the FERC proceedings leading up to its new frequency response rule, industry stakeholders have revealed distinct perspectives on whether establishing a primary frequency response market within independent system operator (ISO) or regional transmission organization (RTO) regions makes sense. Organizations that more purely represent the generation side—which would enjoy increased ancillary service payments if primary frequency response were to be explicitly compensated—appear to be more likely to favor market-based solutions. The Electric Power Supply Association (EPSA), for example, represents independent power suppliers in competitive wholesale markets and is much more explicit in its support of a market-based solution to frequency response provision. EPSA's comment supports procuring frequency response via a competitive mechanism: "in well-functioning wholesale markets primary frequency response should be a capability-based service with defined attributes to attract those resources which can provide the service most competitively" (EPSA et al. 2017). Similarly, the American Petroleum Institute (API), which represents the oil and gas industry, including natural gas generators, recommends that "FERC should direct the RTO/ISO to develop a market-based or cost-based approach to ensure the adequacy of essential reliability services, such as primary frequency response" (American Petroleum Institute 2017). The API more strongly favors the market approach because a "market based response is more consistent with the Commission's goal of promoting competitive market principles" (American Petroleum Institute 2017).

Alternative energy companies tend to favor a market approach as well. For example, SolarCity's comment supports market-based procurement of frequency response and notes that an uncompensated frequency response provision requirement "would create additional barriers to the participation of distributed energy resources by increasing costs to distributed resources that cannot easily provide that service and by eliminating a value stream for distributed energy resources" (SolarCity 2016). The American Wind Energy Association strongly supports a market-based solution, because markets "are ideally suited for procuring PFR at the lowest possible cost to consumers, given that not all resources need to provide PFR at any point in time, and the fact that different power system resources face widely divergent costs for providing PFR, with individual resources also facing varying costs at different points in time" (American Wind Energy Association 2016). The Energy Storage Association also supports competitive pricing for market procurement of frequency response, while pointing out that alternative resources like energy storage "can provide frequency response performance superior to that of generators; in particular, energy storage technologies provide instantaneous response and ramping performance critical for cost-effective and more efficient frequency response service" (Energy Storage Association 2016).

Finally, advocacy groups also generally support a market solution for frequency response provision. The group comment by public interest organizations, including the Sierra Club and Natural Resources Defense Council, says they "strongly prefer market-based procurement of primary frequency response to the alternative of requiring all new generators to provide this service" (Sierra Club et al. 2016). The R Street Institute also commented on the FERC docket in support of frequency response markets based on the principled support of free markets in general, stating that a uniform provision requirement is a "'one-size-fits-all' prescription more reminiscent of 'command-and-control' policy than market-based policies that foster competition, drive innovation and meet reliability requirements at least-cost" (R Street Institute 2017).

Grid Operator Perspectives on a Market for Frequency Response Headroom

In its initial comments to FERC's notice of inquiry, the consortium of five independent system operators (herein referred to as the RTO/ISO Group) said they do not believe a separate compensation mechanism is necessary. Specifically, the grid operators view the cost of frequency response capability and provision as simply another "cost of reliable generator operation (claiming it is similar to, for

example, maintenance, staffing, metering, software, and communications) that should be priced into generator operation" (ISO New England et al. 2016).[17] Furthermore, the RTO/ISO Group raised the issue of administrative costs required to "engage in debating the 'right' level of compensation for primary frequency response"—holding stakeholder processes and using staff time for market design represent real opportunity costs for solving other important issues such as PJM's proceedings about price formation reform. In a separate comment, the Midcontinent Independent System Operator concurs that "there is no need for a market for frequency response because there is sufficient frequency response in all Interconnections" (Midcontinent Independent System Operator, Inc. 2016).

More fundamentally, for the ISOs with the crucial responsibility of ensuring system reliability at every instant, establishing "uniform capability requirements" provides the "system operator with a known amount of response, which facilitates the operator's ability to manage dynamic conditions in the most efficient and effective way possible" (ISO New England et al. 2016). As mentioned earlier, primary frequency response is a concept that is viewed as a property of system robustness, so the RTO/ISO Group states "[p]rimary frequency response should be an inherent characteristic of resources" and "be provided broadly across each Interconnection" so that this essential service is not concentrated to just one grid region (ISO New England et al. 2016). Since various analyses such as NERC's report indicate that interconnections currently have sufficient frequency response as summarized in Table 2, the RTO/ISOs in this group are effectively asking: why fix something that is not broken, and in the process possibly break it? For example, based on the hypothetical PJM headroom estimate earlier, PJM's natural headroom arising from economic dispatch would not drastically degrade until roughly after a five-fold increase in wind generation as summarized in Table 5.

For regions with higher levels of existing and projected renewables deployment, grid operators appear to be more sensitive to the issue of primary frequency response headroom provision. The California ISO is undergoing a stakeholder initiative to explicitly "evaluate a market mechanism to ensure it has sufficient primary frequency response performance over the long-term" (California Independent System Operator 2016a).[18] In particular, the stakeholder initiative examines how the market design could best "[p]roduce price signals to incent capability or provision" of frequency response (California Independent System Operator 2016b). This inquiry into market-based signals for primary frequency response in California is reasonable considering the state's law targeting 50% renewable energy penetration by the end of 2030.

The Texas ISO has already attempted to implement a market-based solution for frequency response. ERCOT's proposal was to unbundle existing ancillary service products into more specific attributes to be procured, one of which is "Primary Frequency Response" (Electric Reliability Council of Texas 2016). According to a cost-benefit analysis by the Brattle Group incorporating both "day-ahead energy opportunity costs" and "real-time option value foregone," this unbundling proposed by ERCOT would have led to more "efficient procurement" that "reduces the quantities of ancillary services needed" and leads to estimated annual savings of about $22 million (Newell, et al. 2015). Incorporated into this net benefits number are implementation costs, which ERCOT estimated to entail an initial budgetary cost of

[17] In practice, "maintenance, staffing, metering, software, and communications" would most likely consist of fixed costs, which is different from the highly divergent marginal costs of providing frequency response headroom illustrated in the earlier example. To even start generating any amount of energy, the funding for "maintenance, staffing, metering, software, and communications" is already economically sunk. Contrastingly, every additional hour of leaving headroom for frequency response provision represents an ongoing, variable opportunity cost.

[18] In comparison, PJM's Primary Frequency Response Senior Task Force, created on May 25, 2017, is tasked more generally with evaluating the adequacy of primary frequency response and relevant documents, as well as "discuss any potential compensation mechanisms associated with providing primary frequency response capability." Based on its published documents, this Senior Task Force does not appear to be discussing market mechanisms for frequency response provision.

$12 to $15 million and an annual operations and maintenance budget cost of $250 to $270 thousand; these implementation costs would cover ERCOT software system revisions as well as increased staffing to support the operational impacts from the proposed market design changes (ERCOT 2014).

Despite being initially proposed and supported by staff members, the Electric Reliability Council of Texas (ERCOT) proposal on "Future Ancillary Services" was rejected in 2016 with an overwhelming vote of 23 versus 3 during a May stakeholder meeting (Kleckner 2016). Overall, stakeholders did not perceive a pressing urgency for an ancillary services overhaul, because frequency response quality has recently been stabilizing despite the rapid growth of wind and solar resources in Texas (Texas Coalition for Affordable Power 2016). Furthermore, stakeholders disliked the potentially large implementation costs that might grow beyond ERCOT's initial estimates and would have to be absorbed by load customers, as well as operational risks inherent in adjusting mission-critical dispatch software (Frazier 2016). Whereas allowing specialization can theoretically improve economic efficiency in the presence of cost heterogeneity, stakeholders also raised concerns that splitting ancillary services into many smaller segments would harm market liquidity for bilateral ancillary service contracts and potentially raise hedging costs (Frazier 2016). Overall, market participants' comments suggest a greater willingness to tackle "more modest incremental changes to Ancillary Services to plan for a future resource mix (Frazier 2016).

The experience in ERCOT can offer lessons about tradeoffs in efforts to introduce new market mechanisms for primary frequency response. Incremental changes might be easier to garner stakeholder buy-in, because the implementation costs would be smaller and more credibly quantified. But gradual improvements today can create policy path dependence that might make it harder to communicate value propositions or urgency in the future. Changes to critical software and procedures entail risks of implementation mistakes, leading to possible disruption of system operations. Such risk appears unjustified when current frequency response quality and quantity are adequate. On the other hand, it would not be prudent to wait until headroom provision is actually degraded to start considering system changes. To help navigate these tradeoffs, grid operators and policymakers may need more detailed studies about the impact of increasing renewable energy on frequency responsive headroom provision (rather than only frequency response capability). This report has provided a high-level numerical analysis of this very question focusing on only PJM and wind energy; more detailed modeling for other specific grid regions will be valuable to derive practical insights and enable foresighted action.

Conclusion

The current electric system involves externalities not being fully priced into market outcomes. Negative externalities include the full costs of climate and environmental damages from emissions, whereas positive externalities include the full benefits of essential reliability services. In particular, primary frequency response—one of the essential reliability services—has conventionally been provided "for free" as a bundled byproduct of traditional energy generation. A market-based approach to procuring primary frequency response headroom has advantages because of cost heterogeneity. Furthermore, some resources (i.e. load demand and energy storage) can provide frequency response reliability services that are completely independent of net energy generation.

In this context, RTO/ISOs have the difficult task of incorporating multiple considerations when ensuring that the future grid with growing renewables usage will have enough primary frequency response, at reasonable cost. From a theoretical perspective, a market approach can be the most economically efficient, but its main drawback is its complicated nature that may introduce operational risks and requires lengthy stakeholder processes in order to properly determine all the market design rules and parameters. However, time is an available resource right now: power grid interconnections are not in a frequency response crisis currently. The fact that current frequency response levels are stable should not

be used as an argument against exploring and implementing changes in response to reasonably expected outcomes like increased penetration of renewables (especially in RTO/ISOs where state policy requires increasing levels of renewable energy penetration). ERCOT's market-based proposal was perceived as being premature to implement. Yet, the effort represents a methodical, stakeholder-inclusive opportunity to carefully consider design options, an opportunity enabled by available time and lack of urgency.

At the end of the day, the optimization question is whether the potential improvements in economic effectiveness and system robustness (by expanding the base of participating resources to load demand and energy storage) outweigh the introduction of operational complexities that could undermine the crucial responsibility of ensuring the grid's reliability at all times. Nevertheless, if market and policy forces continue to promote greater penetration of renewable energy and loss of traditional generators, the way in which primary frequency response is obtained may be forced to evolve.

Appendix: Estimating Hypothetical Headroom in PJM

The data sources published PJM are as follows.

- "Daily Energy Market Offer Data": http://www.pjm.com/markets-and-operations/energy/real-time/historical-bid-data/unit-bid.aspx
- "Metered Load Data": http://www.pjm.com/markets-and-operations/ops-analysis/historical-load-data.aspx
- "Wind Generation": http://www.pjm.com/markets-and-operations/ops-analysis.aspx

Simplifying assumptions:
Ignore losses, transmission system constraints, and generator level constraints such as minimum run times. For simplicity ignore the impact of transmission system tied solar generation.

1. Every day, each generator submits day-ahead bids, specifying the marginal costs in $ / MWh for each MW segment of its individual supply curve. Generators also specify their economic maximum ("ecomax") values, which is the maximum output level that can be scheduled as part of economic dispatch. So even if a generator is fully committed in terms of reaching the economic maximum level, the economic dispatch method would still entail some remaining headroom if the economic maximum level is less than the "emergency maximum" level (which is assumed to equal the maximum MW bid curve point).
2. Sort each generator bid curve segment by the marginal cost to construct the daily supply curve. Note that a unit's bid curve segments do not have to be contiguous. A portion of a generator 1 might clear, then a portion of another generator 2 might clear, and generator 1's next marginal portion might clear afterwards.
3. Omit for now the generators with only negative price bids, as these are likely renewable generators that would not have headroom available.
4. Calculate available headroom as the sum of remaining differences between each generator's max power output (assumed to be the maximum bid curve point) and the cleared amount of generation (capped by each generator's stated economic maximum level). In the dataset the economic maximum values come from the real-time market, so the maximum of economic maximum values is used.
5. Find the set of generators that clear for each hour to achieve the historical hourly load demand, and calculate the hourly values for the implied headroom based on the above method.
6. Repeat step 5 for the net load at varying levels of wind penetration where net load = load – (wind multiplier)*(wind generation in hour).

Bibliography

American Petroleum Institute. 2017. *Comments of the American Petroleum Institute.* FERC Docket No. RM 16-6-000.

American Wind Energy Association. 2016. *Comments of the American Wind Energy Association.* FERC Docket No. RM 16-6-000.

California Independent System Operator. 2016a. *Comments of the California Independent System Operator Corporation.* FERC Docket No. RM 16-6-000.

—. 2016b. *Frequency Response Phase 2 – Issue Paper.* December 15. Accessed November 12, 2017. http://www.caiso.com/Documents/IssuePaper_FrequencyResponsePhase2.pdf.

Cheszes, Jonathan. 2012. *Impact of Curtailment on Wind Economics.* March 20. Accessed November 11, 2017. http://www.renewableenergyworld.com/articles/2012/03/impact-of-curtailment-on-wind-economics.html.

Delille, Gauthier, Bruno Francois, and Gilles Malarange. 2012. "Dynamic Frequency Control Support by Energy Storage to Reduce the Impact of Wind and Solar Generation on Isolated Power System's Inertia." *IEEE Transactions on Sustainable Energy.*

Ela, Erik. 2013. *Variable Renewable Generation Can Provide Balancing Control to the Electric Power System.* National Renewable Energy Laboratory.

Ela, Erik, Aidan Tuohy, Michael Milligan, Brendan Kirby, and Daniel Brooks. 2012. "Alternative Approached for Incentivizing the Frequency Responsive Reserve Ancillary Service." *The Electricity Journal* (Elsevier) 88-102.

Electric Power Research Institute. 2017. *Comments Regarding the Request for Supplemental Comments on Provision and Compensation of Primary Frequency Response Relating to Energy Storage and Small Generators from the Electric Power Research Institute.* FERC Docket No. RM 16-6-000.

Electric Reliability Council of Texas. 2016. *Future Ancillary Services: Preparing to maintain reliability on a changing grid.* April. Accessed November 12, 2017. http://www.ercot.com/content/wcm/lists/89476/FAS_TwoPager_April2016_FINAL.pdf.

Ellison, James F., Leigh S. Tesfatsion, Verne W. Loose, and Raymond H. Byrne. 2012. *Project Report: A Survey of Operating Reserve Markets in U.S. ISO/RTO-managed Electric Energy Regions.* Sandia National Laboratories.

Energy Storage Association. 2016. *Comments of the Energy Storage Association.* FERC Docket No. RM 16-6-000.

EnerNOC. 2017. *Earn Revenue and Protect Your Equipment in the Responsive Reserve Service Program.* April 26. Accessed November 10, 2017. https://www.enernoc.com/resources/datasheets-brochures/faq-ercots-responsive-reserve-service-program.

EPSA et al. 2017. *Comments of the Electric Power Supply Association, Independent Power Producers of New York, Inc., The New England Power Generators Association, Inc., The PJM Power Providers Group and the Western Power Trading Forum.* FERC Docket No. RM 16-6-000.

ERCOT. 2014. *ERCOT Impact Analysis Report: NPRR667.* November 18. Accessed May 18, 2018. http://www.ercot.com/mktrules/issues/NPRR667#keydocs.

Eto, Joseph, John Undrill, Ciaran Roberts, Peter Mackin, and Jeffrey Ellis. 2018. "Frequency Control Requirements for Reliable Interconnection Frequency Response." Energy Analysis and Environmental Impacts Division, Lawrence Berkeley National Laboratory.

Federal Energy Regulatory Commission. 2018. *Final Rule: Essential Reliability Services and the Evolving Bulk-Power System—Primary Frequency Response.* FERC Docket No. RM 16-6-000.

Frazier, Amanda J. 2016. *Position Statement on NPRR Appeal of Decision: NPRR667.* May 24. Accessed May 18, 2018. http://www.ercot.com/mktrules/issues/NPRR667#keydocs.

Gevorgian, Vahan, and Yingchen Zhang. 2016. *Wind Generation Participation in Power System Frequency Response.* National Renewable Energy Laboratory.

Greenwood, D.M., K.Y. Lim, C. Patsios, P.F. Lyons, Y.S. Lim, and P.C. Taylor. 2017. "Frequency response services designed for energy storage." *Applied Energy.*

Ingleson, James, and Eric Allen. 2010. "Tracking the Eastern Interconnection Frequency Governing Characteristic." *IEEE Power and Energy Society* 1-6.

ISO New England et al. 2016. *Joint Comments of ISO New England Inc., New York Independent System Operator, Inc., PJM Interconnection, L.L.C., Southwest Power Pool, Inc. and Independent Electricity System Operator.* FERC Docket No. RM 16-6-000.

Kleckner, Tom. 2016. *ERCOT Stakeholders Reject Ancillary Service Revisions.* May 30. Accessed October 10, 2017. https://www.rtoinsider.com/ercot-ancillary-service-revisions-27118/.

Loutan, Clyde, Peter Klauer, Sirajul Chowdhury, Stephen Hall, Mahesh Morjaria, Vladimir Chadliev, Nick Milam, Christopher Milan, and Vahan Gevorgian. 2017. *Demonstration of Essential Reliability Services by a 300-MW Solar Photovoltaic Power Plant.* National Renewable Energy Laboratory.

Midcontinent Independent System Operator, Inc. 2016. *Comments of the Midcontinent Independent System Operator, Inc.* FERC Docket No. RM 16-6-000.

Monitoring Analytics. 2018. "2017 State of the Market Report for PJM: Section 5 Capacity." 244.

National Grid. 2015. *Frequency Control by Demand Management.* May 5. Accessed November 9, 2017. https://www.nationalgrid.com/sites/default/files/documents/FCDM%20v1.1.pdf.

NERC. 2014. *Essential Reliability Services Task Force: A Concept Paper on Essential Reliability Services that Characterizes Bulk Power System Reliability.* Atlanta: North American Electric Reliability Corporation.

NERC. 2017. *State of Reliability.* North American Electric Reliability Corporation.

Newell, Richard G., and Robert N. Stavins. 2003. "Cost Heterogeneity and the Potential Savings from Market-Based Policies." *Journal of Regulatory Economics* 43-59.

Newell, Samuel A., Rebecca Carroll, Pablo Ruiz, and Will Gorman. 2015. *Cost-Benefit Analysis of ERCOT's Future Ancillary Services (FAS) Proposal.* The Brattle Group.

PJM Interconnection. 2017. "PJM's Evolving Resource Mix and System Reliability."

PJM. 2017. *RPM 101: Overview of Reliability Pricing Model.* April 18. Accessed May 19, 2018. https://www.pjm.com/-/media/training/nerc-certifications/markets-exam-materials/rpm/rpm-101-overview-of-reliability-pricing-model.ashx?la=en.

PJM State & Member Training. 2014. *Reserves Scheduling, Reporting and Loading.* PJM Interconnection.

Praktinknjo, Aaron. 2016. "The Value of Lost Load for Sectoral Load Shedding Measures: The German Case with 51 Sectors." *Energies* (MDPI) 9 (2): 116.

R Street Institute. 2017. *Comments of the R Street Institute.* FERC Docket No. RM 16-6-000.

Schroder, Thomas, and Wilhelm Kuckshinrichs. 2015. "Value of Lost Load: An Efficient Economic Indicator for Power Supply Security? A Literature Review ." *Frontiers in Energy Research* 3: 55.

Sierra Club et al. 2016. *Comments of Public Interest Organizations.* FERC Docket No. RM 16-6-000.

SolarCity. 2016. *Comment of SolarCity Corporation.* FERC Docket No. RM 16-6-000.

Texas Coalition for Affordable Power. 2016. *ERCOT Stakeholders Say No to Major Redesign Effort.* June 2. Accessed November 11, 2017. http://tcaptx.com/policy-and-reform/ercot-stakeholders-say-no-to-major-redesign-effort.

Union of Concerned Scientists. 2016. *Comments of the Union of Concerned Scientists.* FERC Docket No. RM 16-6-000.

WECC Control Work Group. 1998. *WECC Tutorial on Speed Governors.* Western Electricity Coordinating Council.