Diversity of Reliability Attributes

A Key Component of the Modern Grid

PREPARED FOR

American Petroleum Institute

PREPARED BY

Ira Shavel

Michael Kline

Roger Lueken

Pablo Ruiz

May 17, 2017

THE Brattle GROUP

This report was prepared for the American Petroleum Institute. All results and any errors are the responsibility of the authors and do not represent the opinion of The Brattle Group or its clients.

Acknowledgement: We acknowledge the valuable contributions of many individuals to this report and to the underlying analysis, including members of The Brattle Group for peer review.

Table of Contents

Executive Summary

The United States ("U.S.") electric industry is currently in the midst of a transformation driven by technological innovation, changes in the industry's cost structure, and environmental concerns. Throughout the latter half of the twentieth century, large coal, nuclear, and hydroelectric generators provided the bulk of the country's electricity, while natural gas and oil generators operated only during peak demand hours. Variable energy resources, such as wind and solar photovoltaic ("PV") generators, provided only a small portion of the nation's electricity. Most generators could change output only very slowly, but because virtually all capacity was dispatchable and variations in load were both relatively small and highly predictable, the system operator could readily deal with changes in load.

Over the last decade, reliance on non-dispatchable[1] generation from wind and solar facilities has grown rapidly. At the same time, falling natural gas prices and more stringent environmental regulations have led natural gas generation to replace output from coal, and, to some extent, retired nuclear resources. Because their output varies based on wind speed and solar insolation, wind and solar generators are sometimes described as "intermittent" or "variable energy" resources. Large variations in output from such resources happen with regularity. At high penetration levels even relatively predictable output variations, such as solar output falling at night, can strain the grid operator's ability to manage the system.

State-level environmental policies, net-energy metering, falling capital costs, federal tax incentives,[2] and improving technology make it likely that variable energy generators will constitute an even greater share of U.S. electrical generation in the future. To ensure reliability as variable energy resources' share of total generation grows, the system requirements will change. Grid operators will need access to power plants, storage, and demand response resources that have a diverse set of reliability attributes that can meet these requirements. Some of these

---

[1] Independent System Operators ("ISOs") can curtail wind and solar generation when needed, providing some control over their output. However, variable energy resources are not dispatchable in the usual sense (beyond curtailment) since their output largely is a function of constantly changing weather conditions outside the control of plant operators.

[2] Some tax incentives are phasing out over the next few years, but overall tax incentives remain an important factor.

reliability attributes are new to markets. For example, the California Independent System Operation ("CAISO") is exploring the creation of a market for primary frequency response to address North American Electric Reliability Corporation ("NERC") Standard BAL-003, which defines the amount of frequency response balancing authorities need to maintain frequency within acceptable limits.[3] The products CAISO is considering would be designed to address frequency within the NERC defined measurement period of 20 to 52 seconds.[4] Other reliability attributes, such as the ability to provide reserve capacity, have been traded for years. However, the changing make-up of generation may require even mature markets to rethink the way reliability attributes are priced by markets.

Going forward, all jurisdictions should consider two key principles when determining their needs for reliability attributes. First, variable generation resources can create additional reliability needs for the system. Second, resources with reliability attributes that meet these needs should be appropriately valued. With these two principles in mind, we identify three important issues for market designers and system planners to consider going forward. First, the marginal capacity value of variable generation resources can decrease as penetration increases. Second, obtaining frequency response will be increasingly important as increased reliance on variable generation resources decreases system inertia and increases frequency volatility. Third, fast ramping resources will be critical to integrating variable generation resources that increase the variability of net load.

In the remainder of this paper we identify and describe the attributes that contribute to grid reliability and discuss their importance in the context of a changing grid.[5] We review the ways existing markets compensate resources and provide high-level recommendations for what can be done to improve current market design going forward. We also include a discussion of how ensuring an appropriate diversity of reliability attributes may be different for vertically integrated utilities than for load serving entities operating in deregulated electricity markets.

---

3 CAISO Issue Paper, Frequency Response Phase 2, December 15, 2016.

4 NERC Standard BAL-003.

5 The authors note that cost and environmental attributes may also affect market design. This paper focuses solely on reliability attributes and on the appropriate principles for compensating resources that provide those reliability attributes.

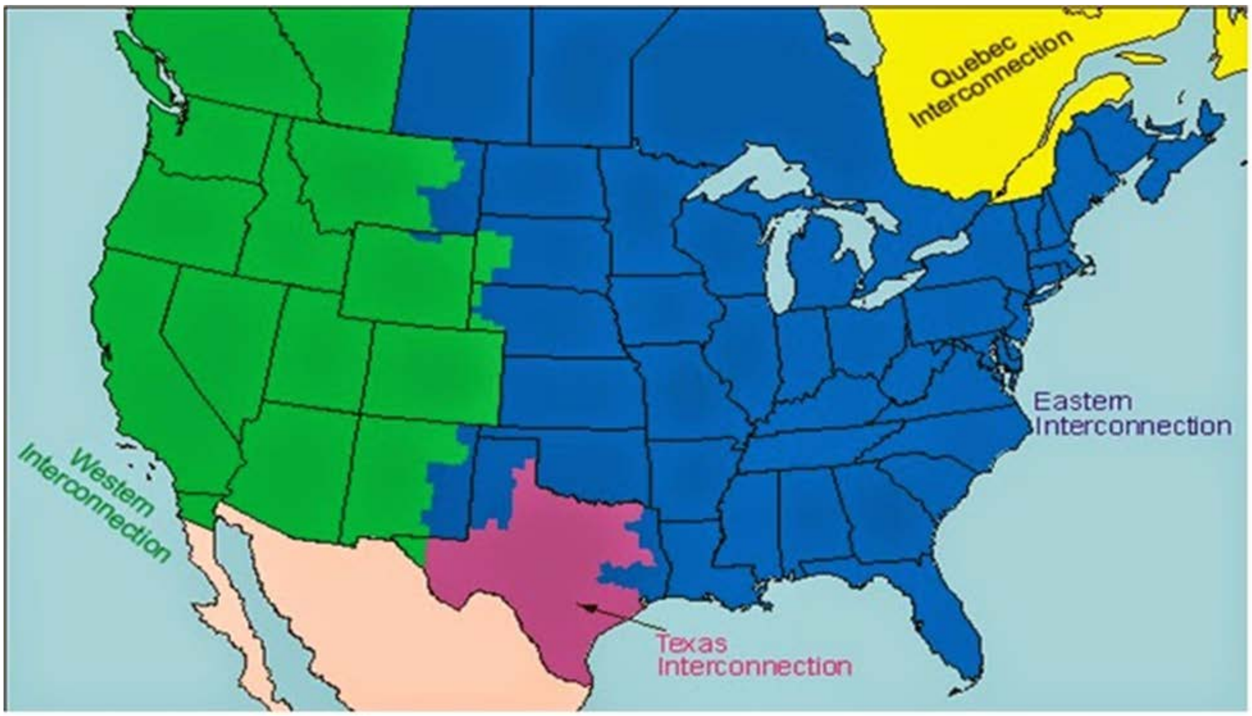
Finally, we conclude by identifying the most important challenges that the changing power system will create for system reliability.

I. Background

A. North American Interconnections

The North American transmission network is made up of four separate grids or alternating current ("AC") interconnections: the Eastern Interconnection, the Western Interconnection, the Texas Interconnection, and Hydro Quebec as depicted in Figure 1. The four interconnections are electrically independent of each other and are only connected together through a handful of low-capacity High Voltage Direct Current ties that allow for relatively small amounts of scheduled power to transfer between the grids. Balancing Authorities ("BAs") within each interconnection balance load (customer demand) and generation (see Figure 2).

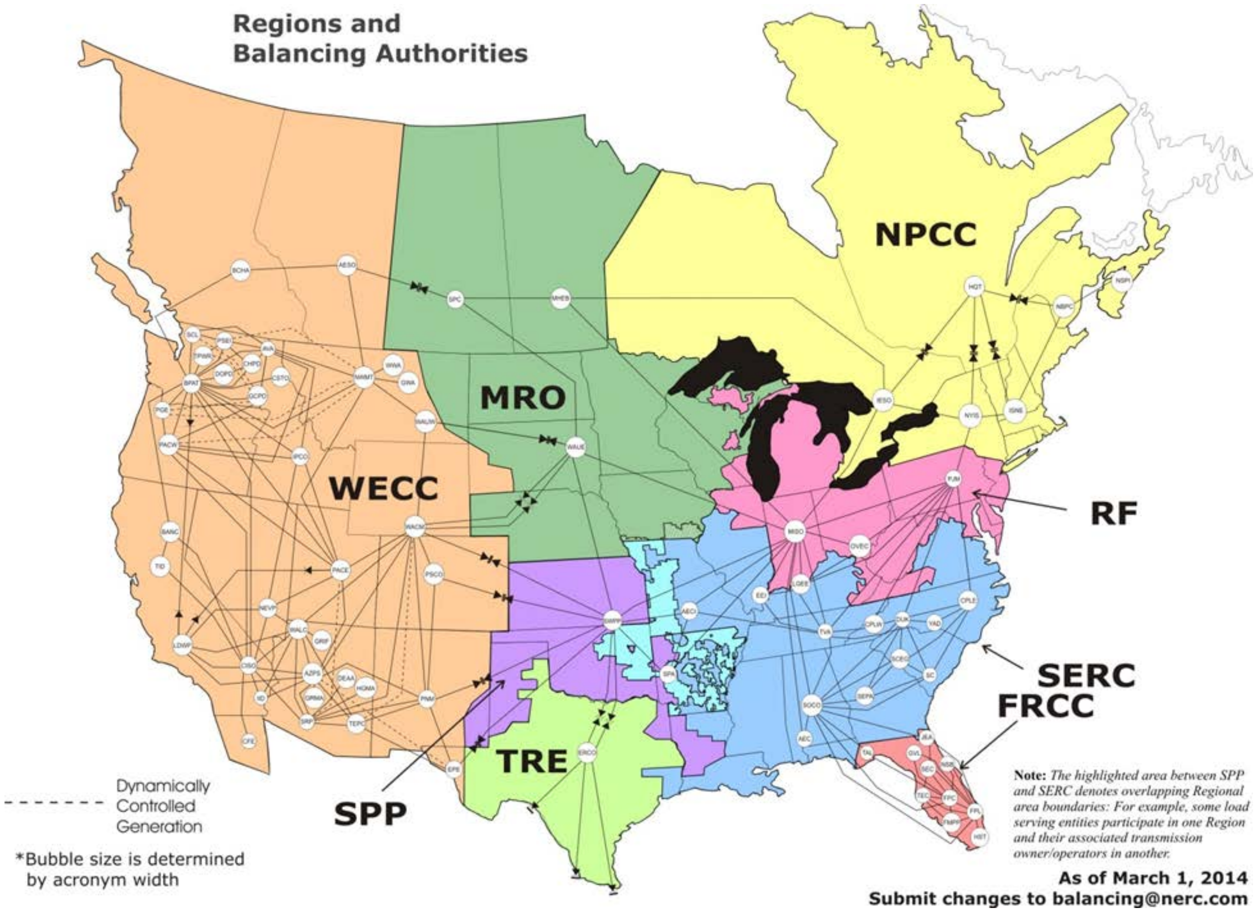
Figure 1
North American Interconnections



Source: North American Electric Reliability Corporation, "Balancing and Frequency Control," January 26, 2011.

There are about 75 BAs in North America; each is connected to neighboring BAs via transmission lines. The BAs are coordinated by 16 Reliability Coordinators ("RCs"). The BAs operate the local systems for which they are responsible, while the RCs are responsible for wide-area coordination.

Figure 2
North American Regions and Balancing Authorities



Source: North American Electric Reliability Corporation website. *http://www.nerc.com/AboutNERC/keyplayers/Pages/default.aspx*

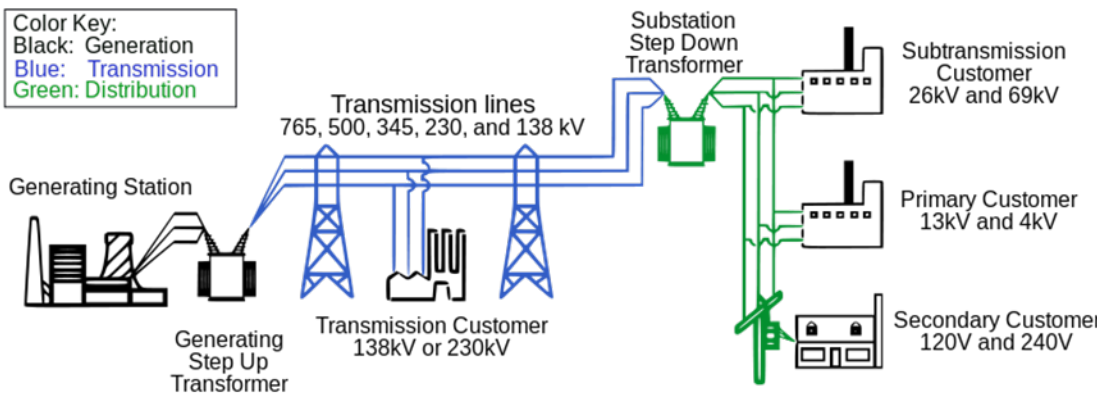
Power system operations focus on three goals, which are: 1) maintain an instantaneous balance of real power between supply and demand across the transmission network; 2) ensure power flows on the transmission system remain within safe limits under a wide range of conditions; and 3) maintain voltage within specified tolerances across the transmission system. Achieving these goals is very complex. Power system operations became more complex over time as the system grew from approximately 130 relatively small "control areas" in the U.S. that roughly corresponded to utilities in 2002 to eight Regional Entities, and seven Regional Transmission Organizations ("RTOs"),[6] that today provide approximately 65% of U.S. generation supply. The

6 RTOs and ISOs differ slightly in that RTOs meet additional requirements and have greater responsibility for the transmission network. However, the distinction is not crucial to this overview of market structure, and much of the industry uses the two terms interchangeably.

largest RTO, the PJM Interconnection ("PJM"), covers most of 13 states plus the District of Columbia and is responsible for the power supply of about 20% of the nation's population.

The power system is primarily an alternating current system. Electricity flows over a very complex network of transmission lines and transformers from generators to customers. Figure 3 provides a simple illustration of AC power system fundamentals.

Figure 3
Illustration of AC Power System Fundamentals



Over decades, power system operators developed rules-of-thumb to meet the three goals. The solutions differed from region to region but had common high-level elements:

1. Established a level of installed generating capacity above projected peak load to ensure a level of predicted reliability (usually 1 day loss of load in 10 years[7]), a metric known as "resource adequacy."[8]

2. Developed a set of ancillary services that the generation system must provide to deal with variations in customer demand and unexpected power plant outages at time scales from seconds to hours.

3. Developed a set of ancillary services that the generation and transmission system must provide to insure voltage stability on the transmission system.

---

7 NERC Rule BAL-502-RFC-02

8 ERCOT does not have a planning reserve margin requirement and relies on market signals to provide new capacity when needed. The NERC reliability planning council TRE has established a target reserve margin but it is not incorporated in the market rules.

4. Developed a set of rules for monitoring the transmission lines and transformers for overloads and potential overloads to ensure that the transmission system can continue to operate in the face of a transmission line failure or a generator outage.

These rules of thumb worked well when power system generation was almost entirely dispatchable and customer demand variations were relatively predictable and small in aggregate.[9] In addition, the power systems of the past were almost all based on generators with large rotating mass that had significant *inertia* that restored the supply-demand balance almost instantaneously after a disturbance. With the addition of significant wind resources about a decade ago and the addition of solar PV resources in the past few years, the nature of the power system began to change. At the same time, market conditions shifted and regulations came into effect that caused some of the baseload coal and nuclear plants that had provided inertia to retire.

Utility scale wind and solar plants are bringing new challenges to the power system due to the variability of their output. These resources cannot be "dispatched"[10] when needed and their output varies with wind and solar patterns with some degree of uncertainty. The result is that "net load" – the difference between customer demand and the output of variable resources – has much higher variance in systems with significant wind and solar penetration. At the customer level, rooftop PV further increases the variability of net load. These developments have resulted in RTOs rethinking their ancillary service needs.

---

[9] Dispatch refers to the ability of power plant or system control operators to vary a resource's consumption or generation of electricity to achieve balance between overall electricity supply and system demand.

[10] Modern wind machines can be dispatched down but cannot provide more power than available from the wind when needed.

B. Bulk Power Adequacy

Bulk power adequacy refers to having sufficient generating capacity to meet instantaneous demand at all times "… taking into account scheduled and reasonably expected unscheduled outages of system elements."[11] NERC considers the bulk power system to be generation and transmission elements that are connected and operate at 100 KV and higher voltages. Prior to the advent of RTOs/ISOs, regulated utilities planned to meet a very high level of reliability, usually the "one day in ten year" loss-of-load expectation standard described in the previous section. The planning process was fairly simple with the utilities often facing few internal transmission constraints and generators having a large degree of certainty and control over their output. As markets developed these same requirements were adopted by the RTOs/ISOs.

States that are not in RTOs/ISOs today mandate planning reserve margins consistent with the one-in-ten standard. These standards have worked very well over the years. Bulk system-level outages that affect customers are very rare. The outages that customers experience are almost all due to the distribution system. These are the lower-voltage wires and transformers that connect the bulk system to customers.

Most of the RTOs/ISOs have some form of capacity market, with the exception of the Electric Reliability Council of Texas ("ERCOT") and the Southwest Power Pool ("SPP"). A capacity market enables those with available generation to sell the attribute of firm electric power supply to buyers who need assured access to electric power at peak times. All of these markets are designed to maintain NERC reliability requirements. The approach taken by each RTO is different and to some extent is related to the underlying region's generation sector structure. PJM, the New York Independent System Operator ("NYISO"), and the New England Independent System Operator ("ISO-NE") mostly cover states that have deregulated wholesale markets. Those RTOs require that most generators offer into the capacity market. The Midcontinent Independent System Operator ("MISO"), which covers many regulated states, relies more heavily on utility-owned generation and bilateral transactions for capacity with a parallel, smaller market to balance the residual supply and demand. All of the capacity markets have local requirements that recognize transmission constraints. Each RTO's/ISO's reliability standards are derived from a NERC standard described in the next section.

---

11 http://www.nerc.com/files/glossary_of_terms.pdf

The one in ten reliability standard is almost always translated into a *reserve margin*, which is a measure of how much extra capacity is needed over and above peak load to maintain the required level of reliability. Reserve margins are usually in the range of 15% to 20%. One major challenge that arose during the last decade has been how to place variable energy resources into a reserve margin context. Variable resources, such as wind and solar, cannot dependably provide generation when needed during system peak conditions. For example, wind tends to blow hardest overnight and in winter months. On a hot summer day winds are often light in most places resulting in low wind generator output. One of the challenges facing both markets and regulated utilities is how to calculate an accurate capacity value for wind and solar. A widely used method is the effective load carrying capability ("ELCC"). ELCC is a probabilistic measure of the contribution of variable generating units to meeting load at the time of system peak. Wind units generally have an ELCC of between 10% and 30% of their nameplate capacities. Thus, 100 MW of installed wind capacity is equivalent to approximately 10 to 30 MW of a dispatchable plant in terms of contributing to reserve margins.

As penetration of variable resources increases, it will be increasingly important that the concept of ELCC include the marginal value of renewables towards meeting peak net load. Peak net load is the system's highest demand for generation from dispatchable resources. At high levels of renewable penetration additional renewable capacity may have substantially less, or even no, marginal value as a capacity resource. As a simple example, in a system with a peak load that occurs during daylight hours, solar PV will have value as a capacity resource. However, at a high enough level of solar PV penetration, the peak hour of net load could move to a nighttime hour. Because solar PV does not generate during the night, additional solar PV MW will have no additional capacity value.

C. NERC Adequacy Standards

NERC has published a set of mandatory standards that covers all aspects of the bulk power system from generation adequacy to cyber security.[12] One of these, BAL-502-RFC-02, establishes resource adequacy requirements to meet a "one day in ten year" loss-of-load expectation standard. This standard is translated into a planning reserve margin. The idea is to have sufficient capacity online to meet demand at the time of system peak load. This is the first line of defense

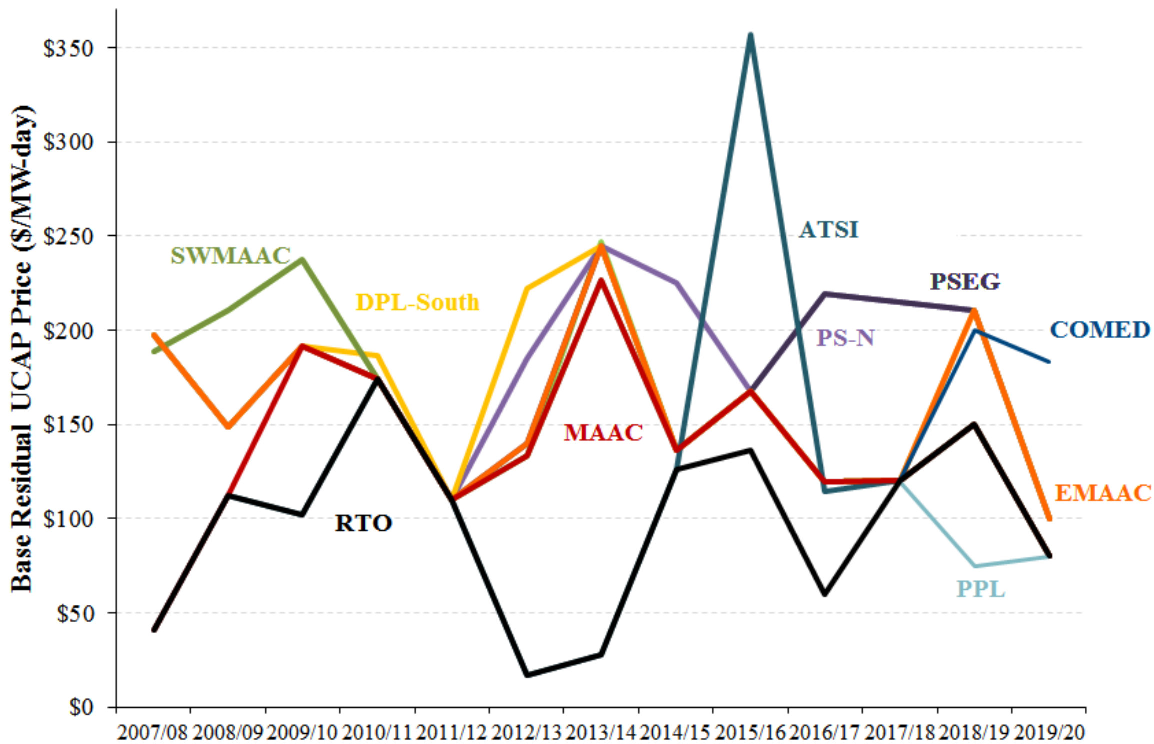
[12] http://www.nerc.com/pa/Stand/Reliability%20Standards%20Complete%20Set/RSCompleteSet.pdf

and is a static requirement based on the estimated capacity that is likely to be available at system peak. The one day in ten year standard had already been used for decades. BAL-50-RFC-02 made the standard an official requirement and specified how to calculate it.

Each RTO or utility establishes a planning reserve margin to achieve the resource adequacy standard. Even ERCOT, which has no requirement to maintain any specific reserve margin, has determined a "target reserve margin." All of the U.S. RTOs except ERCOT and SPP have market mechanisms that compensate market participants for providing capacity. In most instances, RTO capacity requirements have a locational element, recognizing the broad geography of these markets and the internal transmission constraints that can limit the deliverability of remote capacity.

For example, PJM has 27 load delivery areas ("LDAs") in its Reliability Pricing Model ("RPM"). The LDAs are used for evaluating transmission constraints. Auctions for capacity are conducted three years forward and the different LDAs can separate at each auction. When this happens, capacity prices differ across PJM. The RPM prices change with each annual auction as do the separate prices for LDAs that are transmission constrained. Figure 4 shows historical PJM RPM prices in different LDAs.

Figure 4
PJM RPM Prices Over Time[13]



Source: Energy Velocity

13 In delivery years 2018/19 and 2019/20 graph reflects prices for the Base Capacity product.

II. Requirements and Attributes

The power system faces a number of requirements that may change over time. The system operator needs to:

1. Meet bulk demand;
2. Follow load or net load;
3. Maintain voltages; and
4. Maintain frequency stability.

Generators, storage, and demand response (i.e., resources) have attributes that enable the system to meet each of these requirements. As the power system changes, the attributes necessary to meet these requirements change with it. Key resource attributes include:

1. Generation capability;
2. Dispatchability;
3. Security of fuel supply
4. Start times and ramp rates;
5. Inertia and frequency response capability;
6. Reactive power capability;
7. Minimum load level;
8. Black start capability;
9. Storage capability;
10. Proximity to load.

This section discusses these attributes and explains how they fulfill different requirements for the system.

A. Generation

No attribute is more fundamental to system requirements than the ability to generate electrical energy. While resources that lack this attribute may provide valuable services to the grid, the system cannot meet any of its requirements without resources that can generate electricity. Energy efficiency, and to some extent demand response, reduce the need for resources that have this attribute, but ultimately the system must have resources that generate electricity to serve customer demand when it does appear.

B. Dispatchability

Dispatchable resources have the ability to change their output or consumption levels in response to an order by the system operator. Dispatchable resources fall into two main categories – dispatchable generation and dispatchable load. Dispatchable generation resources can respond to orders from the system operator to increase or decrease the amount of electricity they send to the grid. Similarly, dispatchable load resources can respond to orders from the system operator to increase or decrease the amount of electricity they withdraw from the grid. Some resources such as storage can both inject and withdraw power from the grid.

Virtually all resources are dispatchable to some degree. However, some resource types have greater capabilities than others. Various factors affect a resource's dispatchability. Thermal units and pondage hydro can be partially or fully derated by outages, reducing their dispatchability. Thus, units with lower outage rates are more dispatchable. For similar reasons, seasonal runtime limits can reduce dispatchability in nonattainment regions. Access to a reliable fuel supply also affects the dispatchability of generation resources. Onsite storage of fuel (generally coal, uranium, distillate oil at gas plants, or dammed water) provides the most reliable fuel supply. Firm contracts for fuel supply (generally natural gas and coal) also provide a reliable source of energy. Variable resources have no control over their fuel supply, which limits their dispatchability.

Most variable energy resources can respond to an order to reduce generation through curtailment or increase their generation when already intentionally curtailed and not on outage. However, the maximum output for such generators at any given point in time is a function of factors (e.g., wind speed, solar insolation, or hydrology) outside the control of the system operator. By contrast, fossil-fueled units can respond to orders to increase generation up to their maximum seasonal output, so long as they are not experiencing an outage.

The system requires dispatchable resources to ensure that electrical output continuously matches electric demand. Without dispatchable resources the system operator would not be able to meet the requirements to: a) meet peak demand; or b) to follow load.

C. SECURITY OF FUEL SUPPLY

Security of fuel supply is an attribute that describes the dependability of a resource's energy inputs, or fuel. Generation resources primarily rely on uranium, water, coal, natural gas, oil, wind, and solar energy for fuel supply. Demand response and storage resources do not have fuel supply in the traditional sense, but their ability to supply power does depend on the availability of their inputs. Demand response resources can only respond to instructions from the system operator when consuming electrical energy. Thus, the "fuel supply" for demand response resources is their demand for electricity. Storage resources use stored electricity as a "fuel supply."

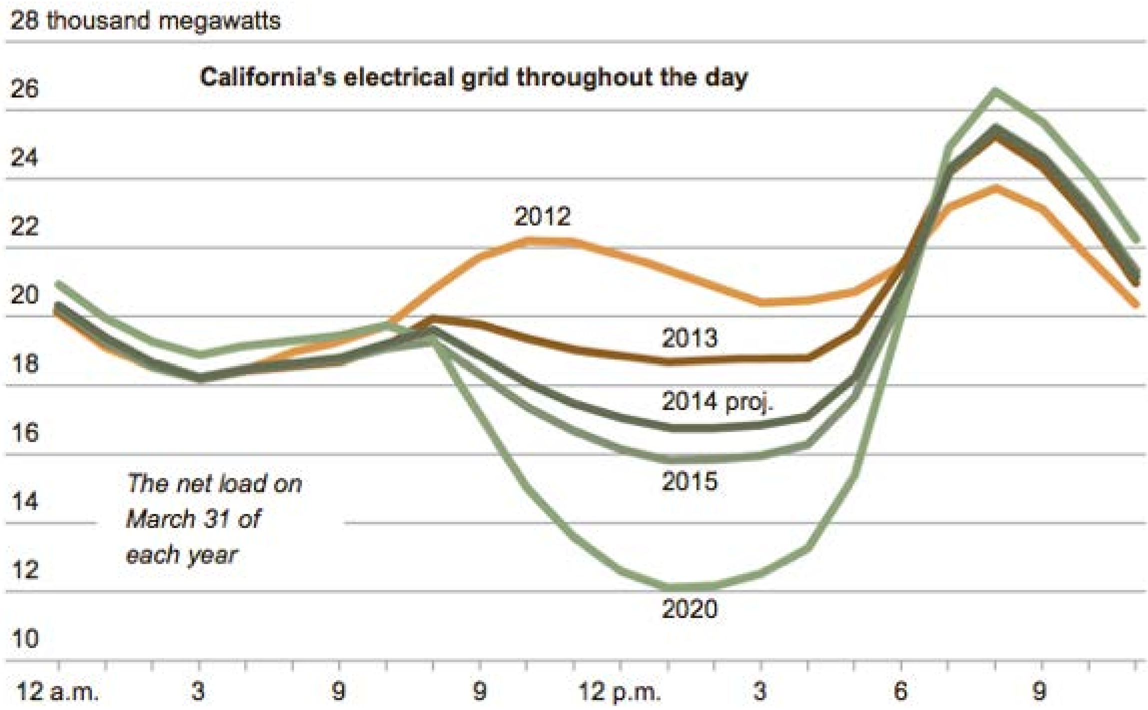
Resources with very secure fuel supplies rarely experience fuel supply interruptions, while resources with unsecure fuel supplies frequently experience fuel supply interruptions. Resources with very secure fuel supplies are especially important to meeting the system requirement to meet bulk power demand.

D. START TIMES AND RAMP RATES

Closely related to dispatchability, start times and ramp rates determine the speed at which resources can respond to system operators' orders to increase and decrease electricity delivered to the grid. A resource's start time describes the length of time needed to begin delivering energy to the grid when it is not already delivering energy to the grid. In other words, it is the length of time needed to "turn on" the resource. A resource's ramp rate describes the speed at which it can change output levels once it is already delivering electricity to the grid.

The system requires resources that can respond to dispatch orders from the system operator in order to follow load. As net load becomes more variable, the need for resources that can respond quickly will increase. The much discussed California "duck curve," shown in Figure 5, illustrates how higher solar PV penetration has dramatically increased, and is expected to continue increasing, the variability of net load. The dramatic spike occurs because solar PV generation naturally falls when the sun sets. In order to meet this spike, CAISO will need access to resources that it can ramp quickly. This will increase demand for resources with short start time and high ramp rate attributes.

Figure 5
Duck Curve



Source: CAISO

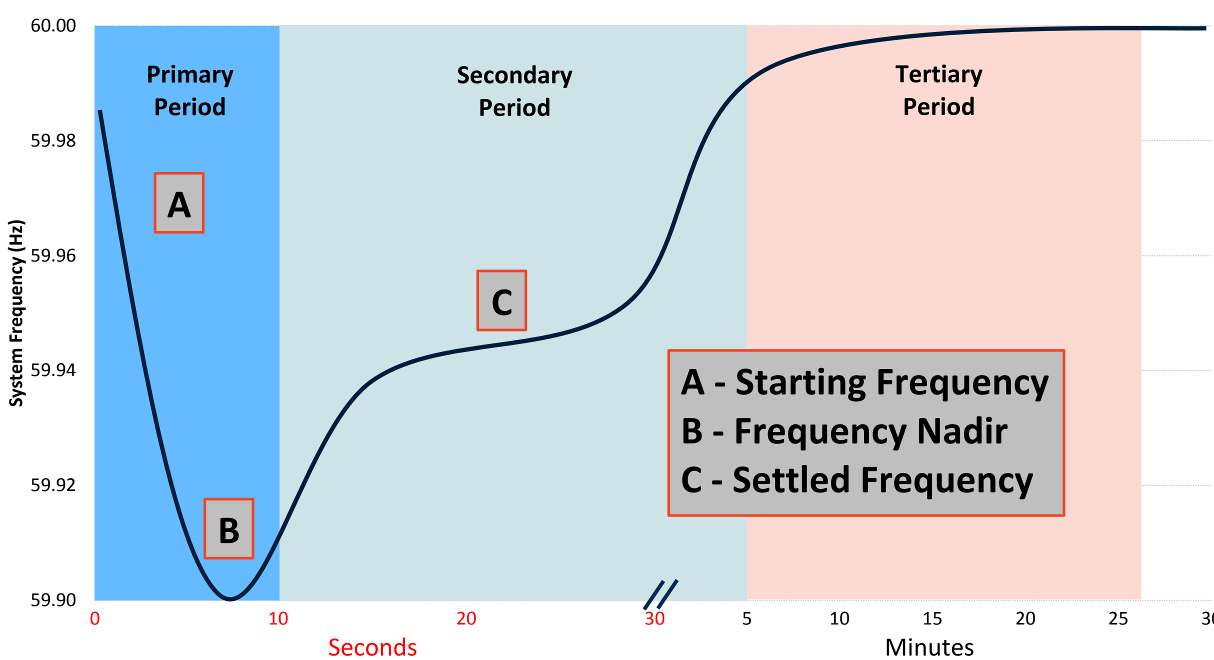
E. Inertia and Frequency Response

Inertia and frequency response are attributes of resources that help the system meet the requirement to maintain frequency stability. The North American power grid is an alternating current ("AC") system that is designed to operate at a constant frequency of 60 hertz ("Hz"). If the grid's frequency deviates too far from 60 Hz, then mechanical failures can occur. To avoid these failures, when large frequency deviations occur automatic safeguards result in load shedding and/or generator shutdowns. For this reason, the system's requirement for stable frequency is critical to reliability.

When demand exceeds supply on the grid, the frequency falls as rotating generators slow down; when supply exceeds demand on the grid, the frequency rises as rotating generators speed up. The heavy rotating turbines have large rotational inertia. When changes to demand occur, leading to transient supply-demand imbalances, the inertia of the turbines resists changes in speed and therefore opposes frequency changes. This helps maintain a nearly constant 60 Hz frequency on the grid by giving the system operator time to adjust generation (or load) to correct the frequency deviation from 60 Hz.

In addition to inertial response, primary, secondary and tertiary frequency response services play a necessary role in maintaining frequency stability by quickly injecting or withdrawing energy. Primary frequency response occurs more quickly than secondary frequency response which, in turn, occurs more rapidly than tertiary frequency response. Turbine governors on synchronous generators and frequency responsive load provide primary frequency response. Primary frequency response occurs automatically over the course of up-to 15 seconds. Secondary frequency response, also called regulation, occurs over a period of 10 seconds to several minutes. The system operator orders regulation resources to inject or withdraw energy from the system as needed using an Automatic Generation Control ("AGC") signal. This shifts the response from primary frequency regulation providers to secondary frequency regulation providers to ensure that the primary frequency regulation resources are ready to respond in case of a subsequent event. Finally, tertiary frequency response occurs over a period of 5 to 30 minutes. This response is also controlled by the system operator and, in the organized markets, it is organized through the real-time energy markets through a change in the generation dispatch. Figure 6 illustrates the timing of different categories of frequency response.

Figure 6
Illustrative Example of Frequency Response



Historically, resources have not been compensated for providing inertia, even though inertia has value because it helps maintain frequency stability. Inverter-connected wind and solar PV installations do not provide rotational inertia (with the proper equipment, wind farms can provide some synthetic inertia and there has been discussion that other inverter-connected devices may also provide synthetic inertia in the future).[14] As the proportion of generation provided by wind and solar has risen, FERC has raised concerns about the decline in total system inertia, leading to larger and more rapid variations in frequency.[15] Similarly, NERC has raised the concern that, "[w]ind, solar, and other variable energy resources that are an increasingly greater share of the BPS [Bulk Power System] provide a significantly lower level of essential reliability services than conventional generation."[16]

While some regions, such as California, are currently developing market products to address problems caused by lower system inertia, decreases in inertia associated with higher levels of renewables may be offset by increases to inertia from other parts of the system. For example, ERCOT found that under a high renewables scenario, inertia increased in most hours relative to a recent historical scenario because coal generation was displaced by not only renewable generation but also by combined-cycle generation, which has nearly twice the inertia per MW as coal.[17]

Resources have been compensated for providing frequency response services, but the growth of variable renewable generation has led to changes in the markets for those services. The variable nature of wind and solar PV has contributed to greater variability in net load, further increasing the need for frequency response. NERC Standard BAL-003 defines the amount of frequency response that balancing authorities need to maintain frequency within acceptable limits.

---

14 "The kinetic energy stored in rotational parts of wind turbines can be extracted through a control strategy referred to as "synthetic inertia". The control system detects the frequency deviation and adjusts the power flow into the grid based on this. In this way the turbine contributes to the system as if it would have inertia just like conventional units; hence the term 'synthetic inertia'."

The Utilization of Synthetic Inertia from Wind Farms and its Impact on Existing Speed Governors and System Performance; Mohammad Seydi, Math Bollen, STRI, January 2013, pages 6-7.

15 Essential Reliability Services and the Evolving Bulk-Power System—Primary Frequency Response, Docket No. RM16-6-000, February 18, 2016.

16 State of Reliability 2015, May 2015, page 16.

17 Cost-Benefit Analysis of ERCOT's Future Ancillary Services (FAS) Proposal, http://www.ercot.com/content/wcm/key_documents_lists/30517/667NPRR_12a_Cost_Benefit_Analysis_122115.pdf pages iv and 9.

Increased renewable penetration will lead to greater need for resources with inertia and frequency response capability.

F. Reactive Power

The ability to provide reactive power is an attribute necessary for meeting the system's requirement to maintain voltage within certain limits. Increasing reactive power supply leads to local voltage increases; decreasing reactive power supply decreases local voltages. Voltage control in an AC power system is important for the proper operation of electrical equipment, which is designed to operate within certain voltage limits. Voltage control is also needed to enable power flow over the transmission system and to reduce transmission losses. Under sustained and pronounced voltage reductions, generators automatically disconnect from the system to avoid equipment damage. This can lead to cascading generator shutdowns and widespread blackouts.

G. Minimum Load Level

A resource's minimum load level describes the lowest level of electrical output the resource can continuously send to the grid. For example, a generator with a minimum load level of 200 MW and a capacity of 800 MW can continuously generate as little as 200 MW of electricity or as much as 800 MW of electricity. The generator cannot, however, continuously generate only 100 MW of electricity. Dispatchable resources with lower minimum load levels (as a percentage of maximum potential output) are better able to help the system meet its requirement to follow load because the system operator has greater flexibility to dispatch the resource in response to changes in net load. This flexibility gains value as net load variability increases.

H. Black Start Capability

Most generating units need electricity from an external source to start after they have shut down. After a single power plant experiences an outage under normal conditions, the generator can simply draw electricity from the grid. However, to restart after a widespread power outage, generators need resources with black start capability. Black start capability is the ability of a power plant to restart without relying on the transmission network to deliver power.

The black start process generically involves using small generators or storage devices to provide the electricity that can provide sufficient energy to restart a power plant. The restarted plant can, in turn, energize transmission lines enabling plants without black start capability to restart. The black start attribute is critical to meeting the system requirement to serve bulk power demand in the aftermath of an area-wide outage.

I. Storage Capability

Resources with the attribute of storing electricity help the system meet multiple requirements including meeting bulk demand, following load or net load, and maintaining frequency stability, but not all resources with the ability to store electricity contribute to meeting all of the requirements. To provide capacity and contribute to the requirement to meet bulk demand, storage must be able to provide output for several hours. To contribute to the requirement to follow load, the resource must be able to provide output for at least several minutes. To contribute to the requirement to maintain frequency stability, the resource does not need to store energy for a significant length of time, but it does need to be able to respond rapidly to operator instructions. For the resource to provide primary frequency response, it needs to be able to respond automatically to changes in the frequency.

J. Proximity To Load

The ability to site resources close to load is an attribute that helps the system meet bulk demand and maintain voltages. Resources that are close to load that also have the ability to generate reduce transmission losses and transmission congestion. This helps the system meet bulk demand because the system needs to generate less electricity in aggregate and the transmission system does not need to be as extensive. Proximity is also important for maintaining voltage because reactive power is more effective at controlling voltage when it is located close to the reactive power load.

Early power systems had generators that were quite close to load, often within metropolitan areas. Today, some of those early generating stations in major cities remain in service. Often, those stations have been modernized and upgraded with new, usually gas-fired generators. The Mystic power plant in the Boston area was first developed in 1957 and has been modernized as a gas plant in recent years.[18] These sites remain valuable in part because of their proximity to major metropolitan areas.

---

18 http://www.prnewswire.com/news-releases/new-exelon-power-plant-providing-clean-energy-to-boston-area-71274792.html

As the power system expanded, larger power plants, usually coal-fired and nuclear, were built further away from load with the power being brought to growing load centers by an expanded transmission system. However, modern natural gas plants, storage facilities and rooftop solar are often built close to load, sometimes in major urban areas. Recent gas plant additions in metropolitan areas include a combined cycle at Astoria in New York City, and the Mystic units in the Boston area. In addition, gas peaking plants (gas turbines or reciprocating engines like those under development in Denton, Texas) can be added near load centers because of their small physical and environmental footprints.

K. ATTRIBUTES AND TECHNOLOGIES

Table 1 shows the relative advantages that different technologies have in providing the attributes needed for system reliability.

Table 1
Reliability Attributes and Technology

	Natural Gas - CC/CT/RICE/ Aeroderivate	Coal	Nuclear	Wind	Solar	Pondage Hydro	Run of River Hydro	Demand Response	Storage
Generation	●	●	●	●	●	●	●	N/A	N/A
Dispatchability	●	●	◍	○	○	●	○	●	●
Security of Fuel Supply	●	●	●	○	◍	●	◍	◍	◍
Start Times	●	○	○	N/A	N/A	●	N/A	●	●
Ramp Rates	●	◍	○	N/A	N/A	●	N/A	●	●
Inertia	●	●	●	◍	○	●	●	○	○
Frequency Response	●	◍	○	○	○	●	○	○	●
Reactive Power	●	●	●	◍	◍	●	◍	N/A	N/A
Minimum Load Level	●	◍	○	N/A	N/A	●	N/A	●	●
Black Start Capability	●	N/A	N/A	○	○	●	●	N/A	◍
Storage Capability	N/A	N/A	N/A	N/A	N/A	◍	N/A	N/A	●
Proximity to Load	●	◍	◍	◍	●	○	○	●	●

● Relatively Advantaged
◍ Neutral
○ Relatively Disadvantaged

Other than demand response and storage, all of the technologies listed have the generation attribute. All of the technologies have the dispatchability attribute, but the variable generation resources (wind, solar and run of river hydro) have a disadvantage relative to other technologies. The system operator can dispatch variable generation resources down, but only at times when the resources are already generating. If wind, insolation, or hydrology conditions prevent generation in the first place, the system operator cannot dispatch the resources down. The system operator's ability to dispatch variable generation resources up is even more limited. In addition to the conditions required to dispatch down, the resources must also be generating less than their maximum potential output at the time. Because variable resource generators have very low variable costs (if any), it rarely makes economic sense to operate the resources in a fashion that would make upwards dispatch viable. Nuclear capacity is less dispatchable than some resources because, in most cases, it is designed to run at its full output level at all times (except when on outage).

Thermal resources and pondage hydro generally have the highest degree of fuel security. Nuclear plants and coal plants maintain local supplies on-site. Nuclear plants store fuel in the reactor core and coal generation resources store coal in "piles". While natural gas is rarely stored in large quantities on-site, some natural gas-fired plants have the capacity to burn distillate oil stored in tanks on-site in the event of a natural gas supply interruption. Pondage hydro stores water (i.e. its fuel supply) in large reservoirs.

Fuel must be delivered to nuclear, coal, and natural gas plants. Nuclear plants store as much as two years of fuel in their reactors. The large storage capacity makes it unlikely that delivery issues will interrupt fuel supply. Coal plants receive their fuel by rail, truck and/or barge. Because most coal plants also store relatively large amounts of coal in their piles, delivery is unlikely to interrupt fuel supply. Natural gas plants receive their fuel by pipeline. Pipeline delivery arrangements to generators can be interruptible or firm. Natural gas pipeline supply can be interrupted due to lack of capacity when demand is very high, but firm supplies have a very low probability of interruption.[19] Pondage hydro must have stored water, but because reservoirs are typically very large, only prolonged drought conditions are likely to affect the security of the fuel supply. Moreover, the restrictions on hydro due to weather usually only limit energy over a long

---

[19] Firm supply delivered through the natural gas system is highly reliable with New England being a notable exception due to pipeline capacity constraints. Please refer to the Eastern Interconnection Planning Collaborative Gas-Electric Target #3 Final Draft Report for further details, http://nebula.wsimg.com/4f9c07a87edd4a873d447e16208e2b6e?AccessKeyId=E28DFA42F06A3AC21303&disposition=0&alloworigin=1.

Additional work on gas-electric system resilience is underway at NERC and across various RTOs/ISOs to better understand and measure the resilience of the integrated gas and electric systems.

period of time; these restrictions do not limit the amount of capacity available to the system operator at any given time (e.g. during a scarcity event).

Run-of-river hydro, demand response, and storage all have a fairly high degree of fuel security, but it is can be less than dispatchable generation resources during scarcity events. While run-of-river hydro almost always has some level of generation, it is impossible to know the exact amount that will be available at any given time. Demand response also has a fairly secure fuel supply. However, there is no guarantee that a demand response resource will be drawing electricity from the system (for example, air conditioners rarely run during the winter). Finally, while electricity from the grid is almost always available to storage, storage cannot simultaneously deliver electrical energy to the grid and draw from it. Because storage can only "store" a relatively limited amount of energy, its fuel security is lower than that of dispatchable generation resources.

Variable generation wind and solar PV resources have lower fuel security. They cannot store fuel (i.e. they cannot store wind or solar energy) and the supply of fuel is frequently interrupted (i.e. the wind speed or insolation falls). As a result, wind and solar PV resources frequently have their fuel supplies interrupted. However, because solar energy is almost always available to some degree during the day, solar PV has a more secure fuel supply than wind.

Newer natural gas combined cyclers ("CCs") and combustion turbines ("CTs"), reciprocating internal combustion engines ("RICE units"), aeroderivatives, pondage hydro, demand response, and storage have relatively short start times and fast ramp rates. RICE units, aeroderivatives, batteries, and demand response are particularly fast. Older natural gas CCs and CTs generally start and ramp more slowly, but are still generally quicker than coal plants. Some coal units have been designed to ramp quickly, but even those units have slow start times. Nuclear plants ramp very slowly and are difficult to start or stop. Because variable resources are largely not controllable, the concept of start times and ramp rates (as presented in this report) do not apply to them.

Traditional turbine-based generators provide inertia naturally, by design. Wind can provide some inertia and additional "synthetic inertia" by using appropriate control functions in its inverter. Inverter-connected solar and batteries could theoretically provide synthetic inertia, but without any rotating mass they would need to rely on stored energy, such as that stored in their capacitors. (It should be noted that both pumped storage and compressed air storage can provide significant inertia because they generate electricity with turbines). Demand response associated with motor load can provide inertia, though most demand response cannot.

Modern natural gas units, pondage hydro, and storage have an advantage in providing frequency response because in many cases they can provide primary, secondary and tertiary frequency response. While coal can provide primary frequency response, it has a more limited ability to provide secondary and tertiary frequency response because of its slow ramp rate. Because nuclear and variable generators usually operate near their maximum output levels to maximize economic efficiency, they have an even more limited ability to provide secondary and tertiary frequency response under normal circumstances. Demand response could theoretically provide frequency response, but – in practice – the system operator's ability to call demand response is generally limited, reducing its usefulness for providing frequency response.

All generators can provide reactive power, but variable generation resources have less of an advantage than other generators. Reactive power experiences high loss levels and variable generators are usually not close enough to load centers to meet reactive power demand. Moreover, because utility scale wind and solar are often located in areas that are remote from load, they need to generate additional reactive power demand to support local transmission and distribution. This limits the usefulness of variable generation resources for providing reactive power to the system. Demand response and storage do not generate reactive power.

Some modern natural gas units (particularly RICE units) and pondage hydro generally have relatively low minimum load levels. Demand response and storage resources, while not generators, essentially have very low minimum load levels (0 MW for batteries). Coal units often have higher minimum load levels than natural gas plants (measured as the percent of total capability represented by minimum load level) and nuclear units generally operate near their maximum output levels. Because variable generation resources' output is dependent on outside system conditions, the concept of a minimum load level (as defined in this report) does not apply to them.

Natural gas CCs, CTs, RICE units, aeroderivatives, and hydro facilities can all provide black start services. Storage can provide black start, but it would need to remain partially charged at all times to do so reliably. Otherwise, there would be a risk that the storage would be fully discharged during an outage event. Wind and solar can provide black start, but would not be able to do so unless wind or insolation conditions permitted. Coal and nuclear units cannot start without an outside power source (though a plant may have that power source onsite). Because it cannot deliver electricity to the system, demand response cannot provide black start.

For the purposes of this report, we have defined storage as the ability to store electricity provided to the grid for use, as electricity, at a later time. Many types of resources can store energy onsite or through contracts. However, as we have defined the attribute, only storage and pondage hydro can store electricity for use at another time. Storage directly converts electricity from the grid into another form of stored energy, while pondage hydro indirectly converts grid electricity into stored energy by allowing other resources to serve demand so that it can reduce its output and store water to generate at a later time.

Natural gas (particularly aeroderivatives and RICE units), storage, demand response, and solar (rooftop solar) are relatively easy to site near load. Historically, some coal and nuclear plants have been sited near large industrial facilities and load centers, but more recent installations have been sited a considerable distance from major load centers. Wind can theoretically be sited near load, but the best wind resources are generally remote from the places people live. Similarly, the best hydro resources are generally not close to load.

III. Markets for Attributes

A. Ancillary Service Markets

Ancillary services are essential services the power system needs to provide grid reliability. They are critical services that allow the system operator to meet all four reliability requirements. Ancillary services refer to those services that are necessary to support the transmission of energy from resources to loads while maintaining reliable operation of the transmission system in accordance with good utility practice.[20] Table 2 below summarizes the different categories of ancillary services.

Table 2
Summary of Different Ancillary Service Products

Ancillary Service	Response Time	Description
Regulation	Seconds	Capacity that responds to RTO regulation signals, increasing or decreasing generation to manage short-term imbalances of supply and demand.
Spinning Reserves	Within 10 minutes	Capacity that is online but unloaded and that can respond within 10 minutes in response to a contingency.
Non-Spinning Reserves	Within 10 minutes	Capacity that may be offline (not synchronized with the grid) and can respond within 10 minutes.
Black Start	n/a	Resources that can start up without assistance from a power system.
Frequency Response	Seconds	Generation ensuring the grid frequency stays within a specific range of the nominal frequency.
Reactive Power	n/a	Generation used to compensate for voltage drops within transmission system.

Most ancillary services (except reactive power and black start) require the resource have the capability to respond rapidly to orders from the system operator. This means a resource must

[20] See Federal Energy Regulation Commission (2016), https://www.ferc.gov/market-oversight/guide/glossary.asp, accessed on February 2, 2017

either a) operate below its full output level (in case the system operator orders it to increase output) or b) allow the system operator to order it to reduce output in a case where it may have a positive energy margin. For these reasons, providing an ancillary service generally reduces the amount of electrical energy a resource can deliver to the grid.

Despite their importance, policy and market design had not focused on ancillary services until relatively recently. After the provision of energy, generation resources historically had the capability to provide more ancillary services than systems required. Because of this surplus, in recent years providing ancillary services created only modest incremental costs for the system. However, with higher renewable penetration, ancillary service markets are becoming increasingly important. Renewables increase the uncertainty and variability in net load and make ramps larger, thereby increasing the ancillary service requirements. In addition, higher renewable penetration depresses energy market prices. This reduces margins earned by resources in the energy market and increases the need to compensate resources for the ancillary services they provide.

Responding to the increased need for ancillary services, the U.S. ancillary services markets have been undergoing changes. At the FERC level, Order 784 requires that markets for frequency regulation take into account the speed and accuracy of regulation resources. Traditionally, generators were only rewarded for the amount of regulation services they provided. No additional compensation was offered for providing a more rapid response time or greater accuracy following a regulation signal. For example, battery storage technologies can respond to system changes in a much faster way than traditional generators. Speed and accuracy are important metrics that impact resources' abilities to provide both frequency response and other fast ramping services.

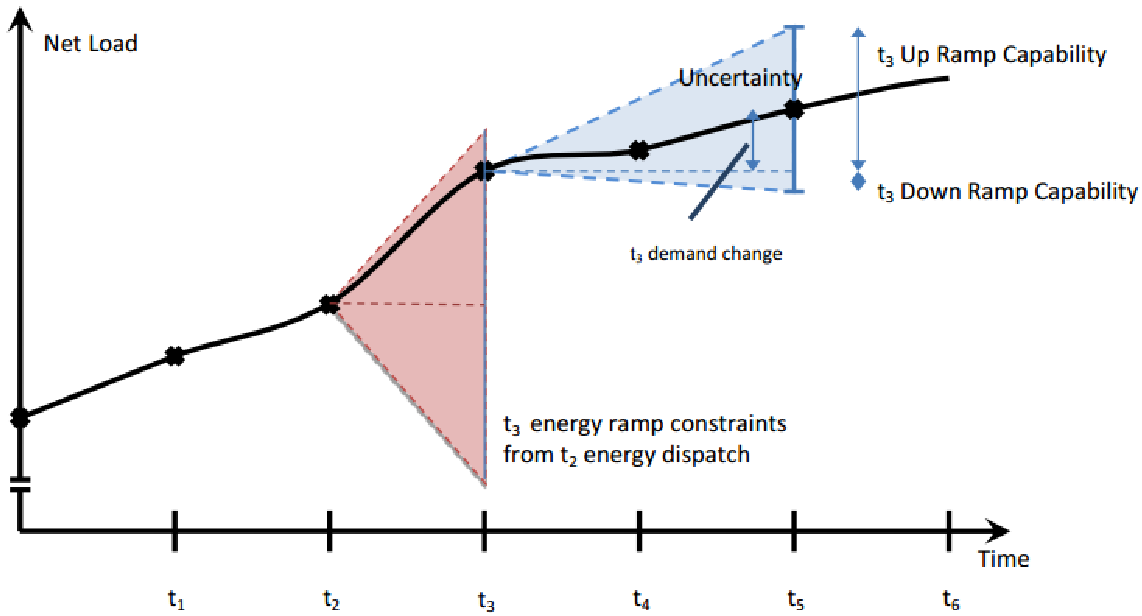
Some ISOs also implemented other reforms in ancillary service markets in order to better reflect the need of the system for flexibility. Both MISO and CAISO have established new ancillary services to manage the challenge of rising variability and uncertainty in net load due to increasing levels of renewables.[21] These products will help dispatchable resources respond to uncertainty and variability in non-dispatchable resources (including most load and renewables). At a high level, the ramp capability products allow the real-time dispatch algorithm to dispatch resources in a way that reduces the likelihood of future scarcity events. The products are designed to strike a balance between the higher operating costs required to provide additional ramp capability and the avoided costs of prevented scarcity events.

Figure 7 illustrates generically how these ramping products work. The red area illustrates the aggregated ramping capabilities of all online resources between periods t_2 and t_3, imposed as a constraint on generation during period t_3. The blue area illustrates the new ramp capability constraints associated with dispatch for time interval t_3. Prior to including ramp capability products, MISO did not consider any requirements beyond t_3 in dispatch decisions made for period t_3, and CAISO accounted for a deterministic forecast of future load. With ramp capability, the dispatch in period t_3 positions resources such that the range of potential load requirements at t_5 can be met by the available resources.

---

21 See https://www.misoenergy.org/Library/Repository/Communication%20Material/Key%20Presentations%20and%20Whitepapers/Ramp%20Capability%20for%20Load%20Following%20in%20MISO%20Markets%20White%20Paper.pdf and https://www.caiso.com/Documents/RevisedDraftFinalProposal-FlexibleRampingProduct-2015.pdf

Figure 7
Illustration of Ramp Product



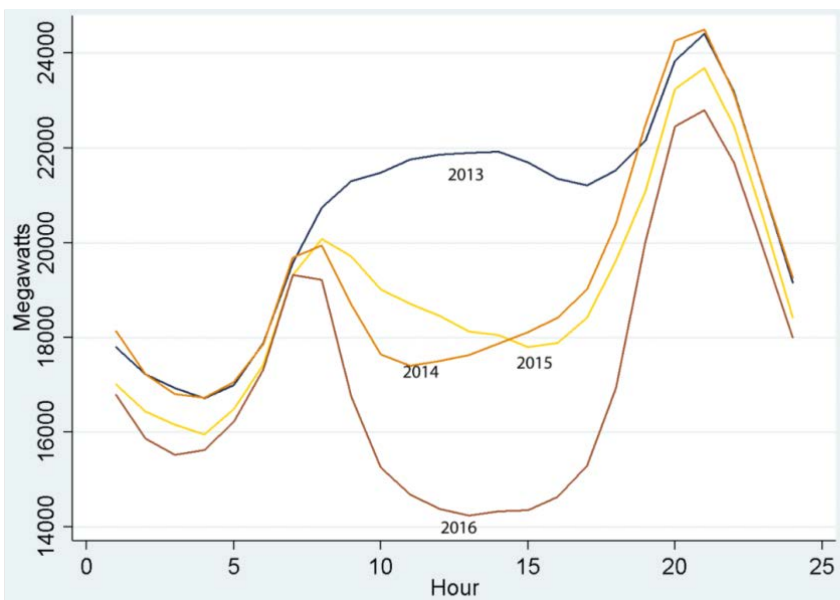
Source: MISO

Both MISO and CAISO established separate up ramp and down ramp products. These products have symmetric features but the auctions clear for separate quantities and prices. The quantities procured of each resource are based on the expected future change in net energy demand and uncertainty in forecast net load. Both MISO and CAISO include ramp constraints in their co-optimization of existing energy and ancillary service products.

B. CALIFORNIA FLEXIBLE RESOURCE ADEQUACY PRODUCT

Growth in renewables has changed what attributes are valuable for supporting reliability. In light of concerns that the California Renewable Portfolio Standard ("RPS") requirements would increase the need for flexible resources that can ramp quickly, on November 6, 2014 CAISO implemented the Flexible Resource Adequacy ("Flexible RA") product. This product supplements the Flexible Ramping Product by ensuring the market will attract and retain sufficient flexible capacity to achieve the California Public Utility Commission's ("CPUC's") desired level of reliability while integrating California's increasingly high penetration of renewable resources. As shown in Figure 8, growth in solar PV has increased the size of the evening ramp. Specifically, the program ensures enough flexible capacity exists to meet the largest three-hour system ramp each month. Flexible RA resources are procured via bilateral contracts on a multi-year forward basis.

Figure 8
California Hourly Net Load
March 28 – April 3, 2013 – 2016



Source:

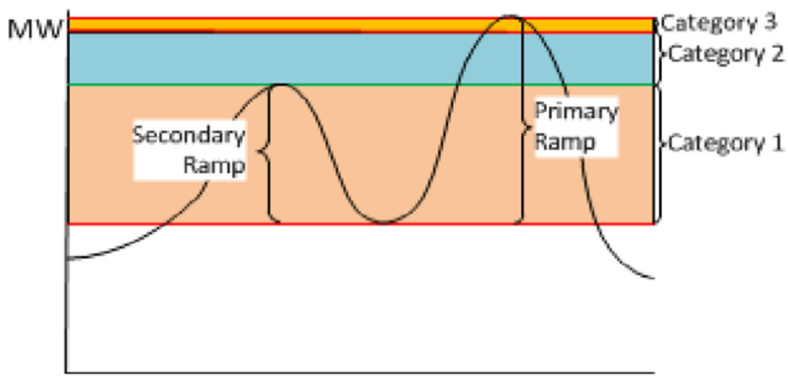
Meredith Fowlie, *The Duck has Landed,* https://energyathaas.wordpress.com/2016/05/02/the-duck-has-landed/

CAISO annually assesses system flexibility needs, accounting for load forecasts, the quantity of renewable resources under its RPS, and renewable generation profiles. CAISO then determines the maximum forecasted 3-hour net-load ramp for each month, calculated as the quantity of

MWs that resources must be ramped or demand must be curtailed across a 3-hour period. CAISO divides flexibility requirements into three product categories:

- **Category 1 (Base Flexibility):** Requirement set by the magnitude of the largest 3- hour secondary net-load ramp.
- **Category 2 (Peak Flexibility):** Requirement set by the difference between 95 percent of the maximum 3-hour net-load ramp and the largest 3-hour secondary net-load ramp.
- **Category 3 (Super-Peak Flexibility):** Requirement set by five percent of the maximum 3-hour net-load ramp of the month.

Figure 9
California Flexible RA Product Categories



Source: CPUC

Once CAISO has determined the total quantity of flexible capacity to procure, it then allocates this requirement across load serving entities ("LSEs"), which must contract for the capacity. Each LSE's obligation is calculated based on its historical contribution to the ISO's largest monthly 3-hour net-load ramp.

Each resource's Effective Flexible Capacity ("EFC") is calculated as the maximum change in net output over a three-hour period. More flexible resources, such as storage, can provide greater flexibility per MW of capacity than inflexible resources with long startup times or high minimum generation levels. The tariff provides specific guidelines for calculating EFC.

C. ERCOT Future of Ancillary Service Reforms

In 2013, ERCOT proposed reforms to the ancillary service markets, collectively referred to as the "Future Ancillary Services" ("FAS").[22] At that time, the design of ERCOT's ancillary services framework was largely unchanged from when it was first established in the 1990s. Although the FAS reforms were ultimately rejected by stakeholders, they serve as a useful example of the types of ancillary service reforms that systems may undertake going forward as deployment of wind and solar resources increase.

The proposed reforms reflect the changing need for ancillary services within ERCOT as inverter-based wind and solar generation displace traditional generation. This transition could result in lower system inertia, thereby increasing reliability risks posed by sudden power imbalances that cause frequency to decay more rapidly than in a system with higher inertia. With less inertia, more reserves are needed to maintain frequency. ERCOT has the additional challenge of not being synchronously connected to neighboring systems, meaning ERCOT itself must manage short-term deviations between load and supply.

ERCOT's proposed FAS reforms were intended to more effectively procure the resources needed to ensure system reliability, based not on the capabilities of specific technologies, but on the fundamental needs of the system. The proposed FAS reforms were multi-faceted, including: (1) enabling a broader range of resources to help meet system needs; (2) more finely tuning requirements to system conditions; and (3) using a procurement approach that better recognizes the relative effectiveness of different ancillary services and resources under different conditions.

The proposed FAS redesign had six main products, ordered below from fastest- to slowest-responding:

---

22 Following description based on two sources:

- ERCOT (2013). ERCOT Concept Paper: Future Ancillary Services in ERCOT. Posted at https://www.ferc.gov/CalendarFiles/20140421084800-ERCOT-ConceptPaper.pdf
- Newell et al. (2015). Cost-Benefit Analysis of ERCOT's Future Ancillary Services Proposal. Prepared for ERCOT. December 21, 2015. Posted at: http://www.ercot.com/content/wcm/key_documents_lists/30517/667NPRR_12a_Cost_Benefit_Analysis_122115.pdf

- **Synchronous Inertia Response Service ("SIR"):** Targeted procurements of inertia under FAS would support system reliability following a disturbance. Inertial response helps to slow the decay in frequency, providing more time for slower-responding resources to respond. Resources are not typically compensated for the inertia they provide; FAS would have provided payment for these resources. SIR would respond immediately to a contingency.
- **Fast Frequency Response Service ("FFR"):** ERCOT would have procured fast-responding resources to slow the decay in frequency following a disturbance, providing more time for primary frequency response ("PFR") to deploy. FFR was designed to accommodate all fast-responding resources, including energy storage. FFR would fully respond to a contingency within 0.5 seconds.
- **Primary Frequency Response Service ("PFR"):** Similar to traditional frequency response products, PFR was intended to re-set frequency closer to defined limits following a contingency. PFR would be provided by generators with governor control or load. PFR would fully respond to a contingency within 16 seconds.
- **Regulating Reserves (Reg):** The FAS proposal included no major changes to the existing Regulation Up and Regulation Down products. These products are intended to balance short-term deviations from the net-load forecast due to unforeseen changes in renewable generation and load. Regulating reserves would follow the ERCOT regulation signal at all times; performance incentives would reward resources that respond more accurately.
- **Contingency Reserve Service ("CRS"):** FAS would have procured fewer contingency reserves, which are slower responding resources intended to help restore the frequency to defined limits. After a contingency, CRS would replace deployed PFR and FRR. CRS would fully respond to a contingency within 10 minutes.
- **Supplemental Reserves ("SR"):** The proposal may have procured SR as a placeholder to fill any additional needs that arise, but ERCOT did not identify any need for SR under anticipated system conditions.

The FAS reforms offered both reliability and economic benefits. Reliability benefits would have included more rapid response to contingencies through SIR and FFR products; ensuring sufficient reserves were available at all times, even immediately after a contingency; and providing incentives for resources to improve performance over time. Economic benefits would have resulted from ERCOT procuring resources to meet a more finely tuned set of requirements. A Brattle analysis found a ten year net present value of the reforms of over $120 million, with a benefit cost ratio of approximately 10. The scenario analyzed had only a modest increase in variable renewables from what was then projected. Although stakeholders rejected the entire slate of reforms, components of the reforms are still under discussion.

In another Brattle study conducted for the Texas Clean Energy Coalition,[23] the authors found that under high wind penetration, a new type of ancillary service would be required to cope with the net load uncertainty caused by wind. Under two scenarios for 2032, one with 26% of ERCOT's energy coming from wind and the other with 43% of ERCOT's energy coming from wind, a new type of ancillary service was needed for system reliability. It was dubbed "Inter-hour commitment option" or ICO. ICO ensures that operators have recourse in the event of net load under-forecast. Resources that supply this service are those that can be brought online within four hours, generally CTs, internal combustion engines ("ICs"), and CCs. The combination of needed non-spin and ICO commitment requirements are a function of renewables and net load forecast uncertainty. With high wind penetration, the study found that up to 9,700 MW of ICO were required for reliable operation of the ERCOT system.

D. CAPACITY MARKETS

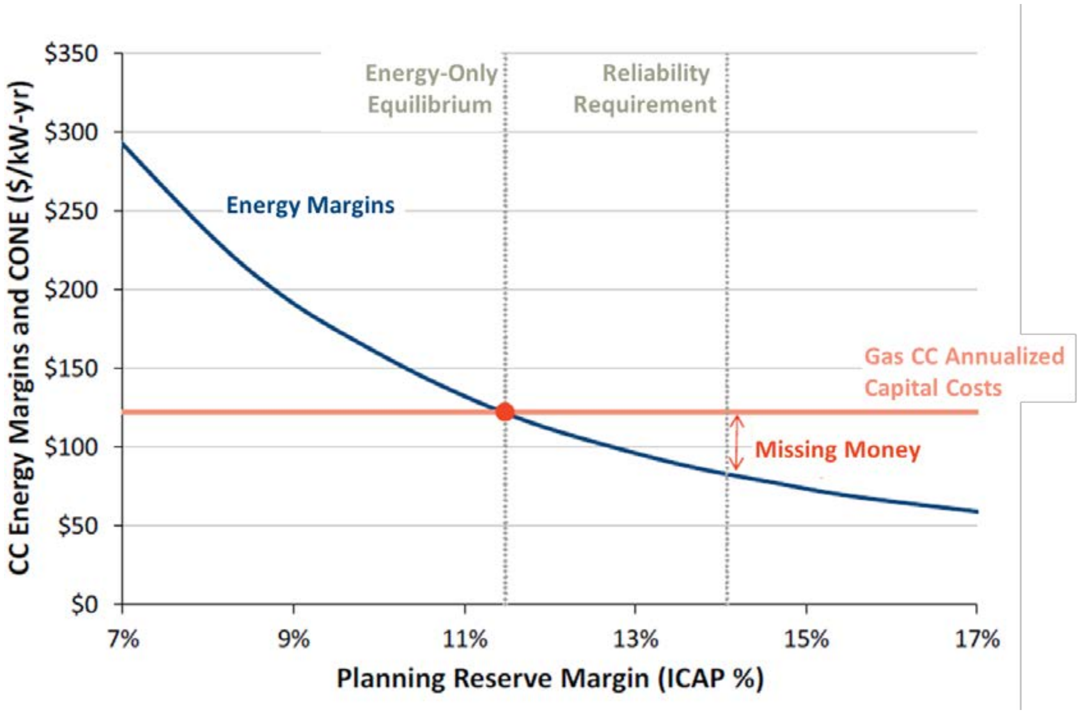
Capacity markets are an administrative construct designed to competitively procure sufficient capacity to achieve mandated resource adequacy reliability standards in competitive wholesale markets. NERC-mandated levels of reliability exceed the energy-only equilibrium level that marginal cost based wholesale energy markets alone are likely to attract and retain.[24] As reserve margins rise, the frequency of high priced hours falls and generators earn smaller energy margins. As illustrated in Figure 10, the so-called "missing money" problem reflects the idea that at target levels of reliability, net revenues from the energy market will be insufficient to cover a resource's total going forward costs. This challenge is unique to restructured markets; suppliers in regulated regions earn regulated rates of return on their invested capital, or are under contract with a regulated utility.

---

23 http://www.ercot.com/content/meetings/lts/keydocs/2014/0113/5._ERCOT_01_13_14_shavel.pdf

24 The ERCOT energy only market relies on high prices during scarcity hours to attract and retain capacity without a capacity market. During scarcity hours, market prices may reach levels substantially higher than the marginal cost of the most expensive resource in the system.

Figure 10
Illustration of Energy Margins vs. Planning Reserve Margin



Notes:
Illustrative figure, adapted from non-public analyses. Planning reserve margin measured in % installed capacity (ICAP).

PJM established the first capacity market in 1999. Other RTOs followed in the early 2000s as a way to attract and retain sufficient supply to meet mandated reliability standards. Most RTOs have some form of competitive market for capacity, although the design details vary across markets (only ERCOT operates without a mandated reliability standard). These markets can be the primary source of capacity procured (PJM, ISO-NE, and NYISO) or they can be backstop markets in an otherwise bilateral or self-supply environment (MISO and CAISO).[25]

---

[25] SPP filed a tariff revision at the FERC on March 3, 2017 to implement a mandatory Resource Adequacy Requirement. The proposal establishes penalties for entities that serve load in SPP that fail to have adequate capacity based on a SPP-wide 12% reserve margin. The proposal envisions bilateral capacity trading and bases penalties on the Cost of New Entry ("CONE") for a gas combustion turbine. The penalty for non-compliance increases as the SPP-wide reserve margin falls. See https://www.spp.org/documents/48681/2017-03-03_tariff%20revisions%20to%20implement%20resource%20adequacy%20requirement_er17-1098-000.pdf

Capacity markets are administered by the RTO. The RTO first identifies how much capacity (MW) is needed to achieve the mandated reliability standards, using probabilistic modeling that accounts for uncertainty in projected peak load and generator availability. As described earlier, the most common resource adequacy standard is a loss of load expectation ("LOLE") of one day per ten years, or the so-called 1-in-10 LOLE standard. Once the target level of reliability is determined, the RTO constructs a demand curve to procure that capacity. The demand curve's shape is based on administrative parameters, including a price cap and slope. This shape is set such that a generic new entrant, typically the most common new plant type (generally a gas CC or CT), would earn enough revenue to enter the market if supply were at or below the target quantity. The RTO approximates the generic new entrant's CONE net of energy and ancillary service revenues, or Net CONE, when setting the demand curve. Over time, RTOs have moved from vertical to downward sloping demand curves to reduce year-to-year capacity price volatility.

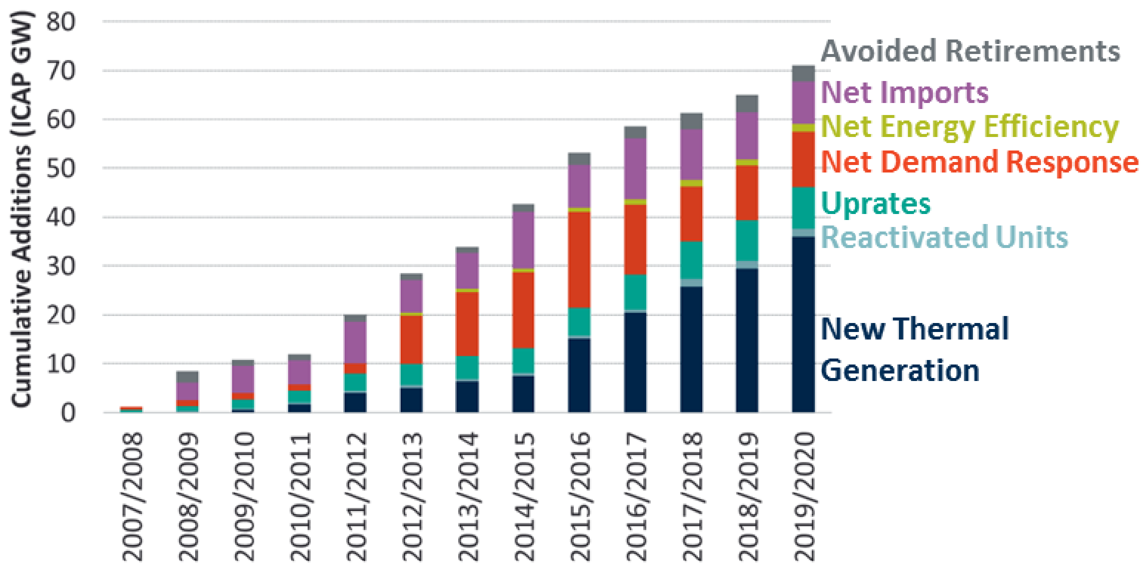
Resources submit competitive offers into the market reflecting their going-forward fixed costs, net of revenues earned on energy and ancillary service markets ($/kW-yr or equivalent). The market clears the lowest cost offers until supply intersects demand. All cleared resources within an RTO zone receive the same price. Prices may separate between zones because of transmission limitations. All resources are derated to an unforced capacity ("UCAP") value that reflects their likely output during peak events, accounting for outages and ELCC deratings. For example, 100 MW nameplate of wind may have 15 MW UCAP value, whereas 100 MW nameplate of gas CTs may have 95 MW UCAP. The goal is to ensure the RTO is procuring a consistent product, irrespective of the type of supplier providing it.

Other design details vary between markets. PJM and ISO-NE procure capacity under one year contracts, whereas NYISO holds separate 6-month summer and winter auctions. PJM and ISO-NE hold base capacity auctions three years forward of the delivery year, whereas NYISO and MISO do not have forward auctions.

Table 3
Summary of U.S. Capacity Constructs

Market	Procurement Method	Forward Period (years)	Delivery Period (years)	Demand Curve
California	Bilateral	Prompt	1	n/a
MISO	Bilateral + Mandatory Auction	Prompt	1	Vertical
NYISO	Bilateral + Voluntary & Mandatory Auctions	Prompt	1	Sloped
PJM	Bilateral + Mandatory Auctions	3	1	Sloped
ISO-NE	Bilateral + Mandatory Auctions	3	1 (7-yr lock-in for new resources)	Sloped
SPP	Bilateral	Prompt	1	n/a

Figure 11
PJM Capacity Additions, 2007/08 to 2019/20



Sources and Notes:

PJM 2007/08 to 2019/20 Base Residual Auction Results. Net imports includes reductions in exports from the 2007/2008 auction.

All jurisdictions continue to refine capacity market designs. PJM's recent "Capacity Performance" reforms serve as one example. Prior to Capacity Performance, PJM paid all resources that receive a capacity supply obligation ("CSO") based on their derated UCAP value. This payment was made regardless of whether the resource was available when needed, such as during a scarcity event. This became problematic in the January 2014 cold snap, referred to as the Polar Vortex, when extreme cold simultaneously forced 22% of PJM's supply out of service during periods of high load caused by the extremely cold weather. Natural gas supply disruptions, mechanical failures, and other factors all contributed to the outages.[26] PJM's Capacity Performance reforms sharpen the incentives for resources with CSOs to be available when needed by penalizing those resources that underperform relative to their UCAP rating and rewarding those resources that over-perform.

26 http://www.pjm.com/~/media/library/reports-notices/weather-related/20140509-analysis-of-operational-events-and-market-impacts-during-the-jan-2014-cold-weather-events.ashx

Appropriately determining the capacity value for variable generation resources is another issue that markets must continuously work to address. As the penetration of wind and solar PV increases, at least two factors will increasingly affect the relative value of variable generation resources versus dispatchable resources. First, the output of wind and solar PV generators is somewhat correlated within most geographic regions. As penetration of each resource type increases, the potential magnitude of an unexpected drop in output during peak hours also increases, reducing the capacity value of incremental wind and solar resources. Second, to maintain reliability the system operator must meet peak load net of wind and solar generation. As wind and solar penetration increases, peak net load hours may shift to hours with less wind and solar output. As a result, at high penetrations each additional MW of wind and solar has less of an impact on reducing peak net load than the previous MW. At low levels of penetration this has little impact, but at higher levels the impact from peak shifting can be significant.

Multiple studies have examined how the capacity value of variable generation resources changes at different penetration levels. MISO has concluded that the ELCC for wind capacity falls as wind capacity increases.[27] A study conducted for Arizona Public Service ("APS") found that under its base case assumptions for solar PV expansion, the marginal ELCC of solar PV would fall from 34.1% of capacity in 2015 to only 5.3% in 2025.[28] Another study, published in 2015, found that while the capacity value of hypothetical solar resources might exceed 40% of nameplate capacity at penetration levels under 2%, the value falls below 20% of nameplate capacity at penetration levels of 15%. The same study found that while the capacity value of hypothetical wind resources is under 20% of nameplate capacity at low penetration levels, the value falls to 10% of nameplate capacity at penetration levels of 15%.[29] These studies demonstrate that market design needs to ensure that compensation to resources reflect their real capacity values and take into account both of the factors described above.

---

27 Planning Year 2016-2017 Wind Capacity Credit, page 7. https://www.misoenergy.org/Library/Repository/Report/2016%20Wind%20Capacity%20Report.pdf

28 2013 Updated Solar PV Value Report, SAIC, page 2-7.

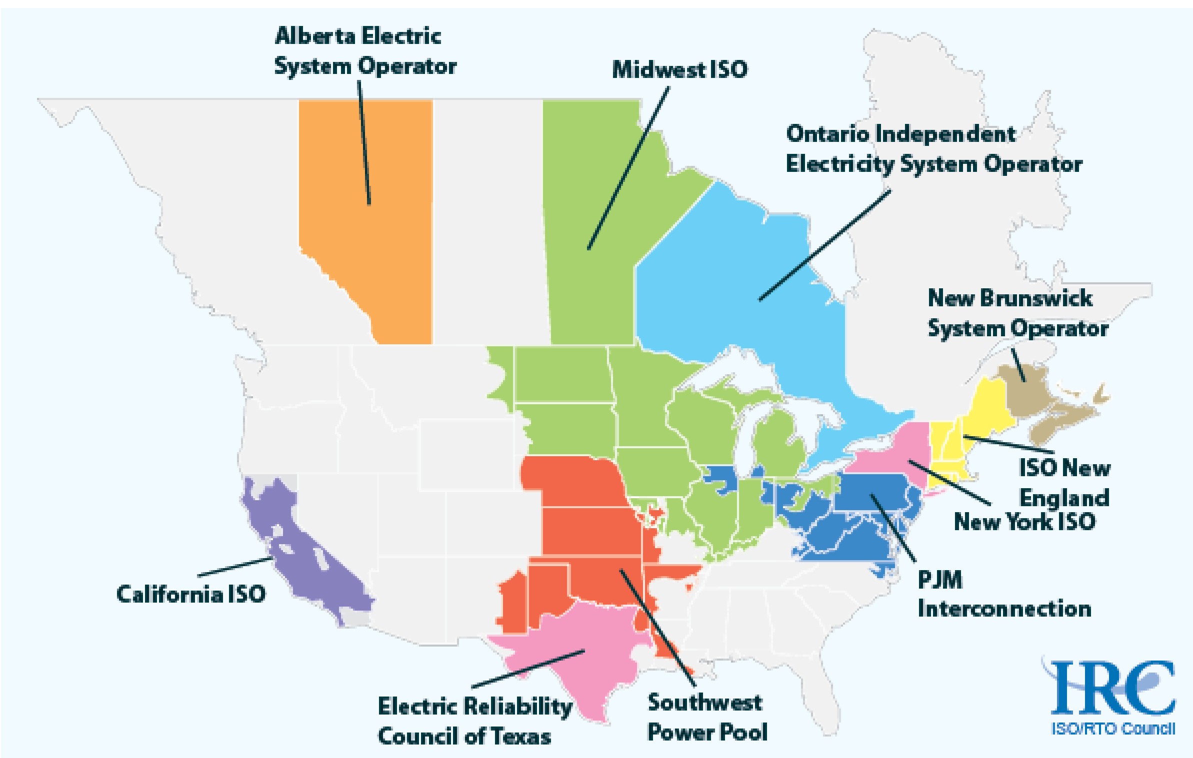
29 Endogenous Assessment of the Capacity Value of Solar PV in Generation Investment Planning Studies, Francisco D. Munoz, Member, IEEE, and Andrew D. Mills, page 8.

E. ENERGY MARKETS

Wholesale energy markets are centralized RTO markets for the electricity commodity in its most basic form. For most system resources, energy markets and capacity markets (described in Section III.D) provide the vast majority of revenue and earnings. Energy markets help the system meet bulk power demand on an instantaneous basis, but in most cases they are not intended to contribute to reliability (ERCOT is the one U.S. exception). Energy markets primarily help the system operator meet the requirement to meet bulk demand, and follow load or net load. Energy markets provide some of the compensation for the generation, dispatchability, start time/ramp rate, and minimum load level attributes described in Section II.

Figure 12 shows the North American RTOs and ISOs.

Figure 12
Map of North America Wholesale Energy Markets



Providing precise details of the energy market rules of each of the seven U.S. RTOs exceeds the scope of this paper, but all share similar rules designed to serve load from the least-cost suppliers available. Dispatch decisions in wholesale energy markets are made by the ISO, which also calculates the market compensation for generators, through an auction process. Load serving entities submit expected load schedules approximately one day in advance. Supply resources also submit bids one day in advance. The bids generally involve multiple components such as start-up costs, no–load costs, and variable cost components. In most ISOs, these bids are tied to an estimate of the generator's marginal costs, but in ERCOT some resources may bid up to the offer cap (currently $9,000 per MWh).[30]

Based on the load schedule and the energy bids, the ISO conducts a Day Ahead auction that determines the ISO's dispatch instructions to generators. The ISO uses complex optimization software to determine the least-cost way to serve the forecast load in each hour. Prices are a function of the most expensive bid (i.e., the marginal bid) to clear the auction along with the marginal costs of congestion and losses caused by each generator. Because marginal losses and congestion vary by location, the price each resource receives also varies by location. Location specific energy prices are called Locational Marginal Prices ("LMPs"), Locational Based Marginal Prices ("LBMPs"), or nodal prices. The Day Ahead auctions are held for each hour of the next day.[31]

Shortly before the actual delivery of energy, the ISO conducts another balancing or Real Time auction. The purpose of this auction is to correct for any real world deviations from the forecast load, the forecast output from variable generators, and generator availability. The process for determining prices in a Real Time auction is similar to the Day Ahead auction. The Real Time auction results in adjustments to the dispatch instructions the ISO made based on the Day Ahead auction results. Real-time markets operate over shorter time intervals (generally five minutes) with new dispatch instructions and prices generated in each interval.

In energy markets with cost-based bids, the most expensive resource dispatched in every interval covers its variable costs but nothing more. Because resources have significant fixed costs,

---

30 https://www.ferc.gov/CalendarFiles/20160629114652-3%20-%20FERC2016_Scarcity%20Pricing_ERCOT_Resmi%20Surendran.pdf

31 The ISO must also determine which units to commit. This is a complicated process and, because of forecast errors, committed units do not always recover their costs. ISOs generally have a "make-whole" mechanism for providing outside of market payments to committed resources that do not recover their variable costs through the market.

marginal resources do not earn enough revenue in the energy markets to justify remaining online. The additional revenue needed to cover costs comes from the capacity market.[32]

---

[32] For most system resources, particularly generators, ancillary service revenues are not a significant source of revenue.

IV.Diversity of Reliability Attributes in Regulated States

The physical operating requirements required by regulated utilities are not fundamentally different from RTOs. In both cases, reliability standards are set by NERC. However, in regulated regions the state regulator directly approves the resources that utilities develop. Regulators therefore have direct control over how the diversity of attributes is implemented. Most regulated utilities have an Integrated Resource Planning ("IRP") process that identifies resources to be added. Once the regulator approves an IRP, utilities identify and acquire the actual resources via a competitive Request for Proposal ("RFP") process or self-build, overseen by state regulators.

In contrast, deregulated regions rely on market competition to provide some attributes related to reliability. RTOs specify competitive products (e.g. capacity, regulation, etc.) and quantify how much of each product is needed. Market participants then competitively offer to provide each attribute. RTOs procure some reliability attributes (e.g. black start, voltage control, and sometimes capacity) via bilateral contracts, not market competition.

Many regulated utilities explicitly account for diversity of reliability attributes in their IRP process. A recent Brattle review of the IRPs of eight regulated utilities found that all IRPs explicitly stated attributes such as reliability and flexibility were priorities.[33]

Recently, several regulated utilities with high levels of variable renewables have undertaken efforts to improve their management of such resources. These utilities include Xcel Energy Colorado ("Xcel"), Westar Energy, and Puget Sound Energy. In many ways, these reforms are similar to those described above in restructured markets. But the challenges of renewable integration can be even greater for regulated utilities, which often have smaller thermal generation resource bases with which to balance renewables. Utilities need to develop new tariffs that appropriately allocate costs to the resources that impose them on the system, while compensating the resources that offset these costs.

---

33 Reviewed IRPs include Ameren, Arizona Public Service, Dominion, Florida Power & Light, Long Island Power Authority, PacifiCorp, Tennessee Valley Authority, and Xcel Colorado.

Xcel

In 2014 FERC approved ancillary service tariff provisions filed by Xcel.[34] Xcel's proposed tariff revisions were in response to rapid growth in variable energy resources ("VERs") in their system; in 2014, Xcel had 2,251 MW of VERs and only 5,000 MW of thermal generation. The proposed changes were two-fold. First, Xcel proposed to allocate the costs of regulation and frequency response services to transmission customers, VERs, and non-VERs in a manner that accounted for their relative contribution to costs or offsetting benefits. Previously, such costs were borne only by native load customers, to the extent they were recovered at all. For example, any regulation and frequency response costs above the established rate due to the addition of intermittent VERs were not recovered. Xcel's proposed rates were $0.18/kW-yr for load, $0.23/kW-yr for non-VER generation, and $1.92/kW-yr for VER generation.

Xcel's second proposed change was to add a new type of reserves called "Flex Reserve Service". This product helps manage sustained, downward wind ramps that can occur due to a loss of wind speed. Such down-ramps can occur over tens of minutes or even a few hours. Xcel calculated 411 MW of Flex Reserve Service would be required. As with regulation and frequency response, the proposal called for costs to be allocated to those transmission customers that create the need for the service.

Westar

Westar is a Kansas public utility located within the footprint of SPP. As a balancing area within SPP, Westar is responsible for maintaining the balance between load and generation in its balancing area. Historically, Westar charged all transmission customers for regulation and frequency response when their generation was used to serve customers in the control area. This charge was calculated by multiplying a regulation requirement percentage of 1.35% by the amount of transmission service and the cost to provide regulation and frequency response. [35]

---

[34] 149 FERC ¶ 61,208. Order Conditionally Accepting and Suspending Proposed Tariff Revisions, Subject to Refund, and Establishing Hearing and Settlement Judge Procedures. December 5, 2014

[35] 137 FERC ¶ 61,142. Order Granting Rehearing in Part, Denying Rehearing in Part, Instituting Section 206 Proceeding, and Establishing Refund Effective Date. November 17, 2011.

This arrangement allowed Westar to recover costs associated with serving load within the control area, but did not cover the cost of providing regulation or frequency response to generators within the control area that exported to load outside of the Westar control area. Even though these resources were selling outside of the control area, Westar still needed to procure sufficient regulation and frequency response to cover their generation. It also only accounted for the need for these services due to the variability and intermittency of load, and did not account for the increased need due to variable renewable resources.

In November 2011, FERC approved modifications to the Westar tariff. The approved changes allow Westar to charge for and provide regulation and frequency response to generators that export outside of Westar's balancing area. Westar was also approved to assign higher regulation obligations on variable generation due to their contribution to regulation and frequency response requirements. Westar will refile every 3 years to modify the requirements for dispatchable and variable generation to account for changes in technology and improved management experience.

Puget Sound

Puget Sound Energy is a balancing authority in Washington with responsibility for maintaining regulation and frequency reserve within its control area. As with Westar, Puget Sound proposed updates to their tariff to reflect the cost of procuring sufficient regulation and frequency response to integrate variable generation resources.[36] FERC approved changes to the Puget Sound tariff requiring variable generation resources to purchase regulation capacity equal to 16.77% of their transmission reservation. The regulation obligation for load and exporting dispatchable generation remained unchanged at 2%. These obligations reflected the regulation burden each resource created. Specifically, the obligations were approximated based on the 95% confidence interval of the expected differences between actual and scheduled MW for load, wind, and dispatchable generation. FERC also approved increasing the capacity rate for regulation and frequency response from $5.50/kW-mo to $12.39/kW-mo, reflecting increased costs of the pool of resources that provide regulation.

---

[36] 137 FERC ¶ 61,063. Order Accepting and Suspending Proposed Tariff Revisions, Subject to Refund, and Establishing Hearing and Settlement Judge Procedures. October 20, 2011.

V. Conclusion

Historically, the U.S. has relied largely on dispatchable generation resources to meet its electricity needs. However, with advances in technology that have driven down costs, non-dispatchable variable energy resources now generate a significant amount of energy. For a variety of reasons related to both policy and economics, the shift towards variable energy resources will likely continue. This change to the system will require the restructured markets and the regulated states to rethink the way they value the different reliability attributes of resources. To a large degree, market designers and planners at vertically integrated utilities are aware of these issues and have begun to take action. However, as the penetration of variable energy resources increases, more work will be needed in rethinking the traditional way resources' contributions to reliability are valued. In both markets and regulated states, ensuring reliability will depend on rules that recognize two important reliability principles. First, integrating variable generation resources can increase the need for resources with the reliability attributes discussed in this paper. Second, reliability attributes should be valued in an economically efficient way.

Going forward system operators in both restructured markets and vertically integrated states will face three issues that are particularly important for reliability. First, the marginal capacity value of variable generation resources with correlated output tends to decrease as penetration increases. To minimize costs and maintain reliability for consumers, it is important that variable generation resources receive compensation for the value of the capacity resources they provide. The potential decline in capacity value must be carefully accounted for or the system may not be able to meet demand during system peak load conditions. Second, increased reliance on variable generation resources will most likely decrease inertia and increase frequency volatility. This will create a greater need for primary, secondary, and tertiary frequency response products. Resources with the ability to provide these services will be necessary to prevent the cascading blackouts that can occur when frequency deviates too much from 60 Hz. Third, increased reliance on variable generation resources will increase the variability of net load. This will create a greater need for resources with fast start times and quick ramping capabilities. These resources will be necessary to ensure that load can be served during a rapid change in net load.

Changes to the composition of the U.S. generating fleet are creating new challenges for maintaining reliability. While addressing these three issues will be particularly important to the electricity grid, all of the attributes identified in this report are important for reliability. The mechanisms for valuing these reliability attributes will differ between restructured markets and regulated states, and the relative value of different attributes will also vary between different regions. However, in all regions and jurisdictions, ensuring economically appropriate compensation for the attributes identified in this report will be important to maintaining system reliability.

CAMBRIDGE
NEW YORK
SAN FRANCISCO
WASHINGTON
TORONTO
LONDON
MADRID
ROME
SYDNEY

THE Brattle GROUP