Evolution of Wholesale Electricity Market Design with Increasing Levels of Renewable Generation

E. Ela,[1] M. Milligan,[1] A. Bloom,[1] A. Botterud,[2] A. Townsend,[1] and T. Levin[2]

[1] *National Renewable Energy Laboratory*
[2] *Argonne National Laboratory*

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

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

Technical Report
NREL/TP-5D00-61765
September 2014

Contract No. DE-AC36-08GO28308

Evolution of Wholesale Electricity Market Design with Increasing Levels of Renewable Generation

E. Ela,[1] M. Milligan,[1] A. Bloom,[1] A. Botterud,[2] A. Townsend,[1] and T. Levin[2]

[1] *National Renewable Energy Laboratory*
[2] *Argonne National Laboratory*

Prepared under Task No. WE14.9B01

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

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

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Technical Report
NREL/TP-5D00-61765
September 2014

Contract No. DE-AC36-08GO28308

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

Available electronically at http://www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

Acknowledgments

The authors would like to thank the U.S. Department of Energy Wind and Water Power Technologies Office—in particular Charlton Clark and Jose Zayas—for its support in this research. The team would also like to thank a group of reviewers from the National Renewable Energy Laboratory (NREL) and elsewhere for their valuable contributions to the report. In particular, the team thanks Mark Ahlstrom, Alex Papalexopoulos, Charlie Smith, Eduardo Ibanez, and Barbara O'Neill. The team finally wishes to thank Katie Wensuc and Mary Lukkonen for editorial review.

List of Acronyms

AC	alternating current
ASCP	ancillary service clearing price
CAISO	California Independent System Operator
CONE	cost of new entry
DAM	day-ahead market
DC	direct current
ELCC	effective load-carrying capability
ERCOT	Electric Reliability Council of Texas
EUE	expected unserved energy
FERC	Federal Energy Regulatory Commission
FOR	forced outage rate
FTR	financial transmission rights
ICAP	installed capacity
ISO	independent system operator
ISO-NE	Independent System Operator of New England
LMP	locational marginal pricing
LOLE	loss-of-load expectancy
LOLH	loss-of-load hours
LOLP	loss-of-load probability
LSE	load-serving entity
MISO	Midcontinent Independent System Operator
NERC	North American Electric Reliability Corporation
NGCC	natural gas combined cycle
NGCT	natural gas combustion turbine
NREL	National Renewable Energy Laboratory
NYISO	New York Independent System Operator
ORDC	operating reserve demand curve
PFC	primary frequency control
PRM	planning reserve margin
PV	photovoltaic
RTM	real-time market
RTO	regional transmission organization
RUC	reliability unit commitment
SCED	security-constrained economic dispatch
SCUC	security-constrained unit commitment
SPP	Southwest Power Pool
TEPPC	Transmission Expansion Planning Policy Committee (of the Western Electricity Coordinating Council)
UCAP	unforced capacity
VG	variable generation
VOLL	value of lost load
WWSIS-2	Western Wind and Solar Integration Study Phase 2

Executive Summary

Variable generation (VG) such as wind and photovoltaic solar power has increased substantially in recent years. VG has at least four unique characteristics compared to the traditional technologies that supply energy in the wholesale electricity markets. It increases the variability of net load (load minus VG) because its available power changes through time because of the changing source (i.e., wind speed or solar irradiance). VG also increases the uncertainty of the net load because the available power can be only partially predicted at all time horizons. Although it has significant fixed capital costs, VG has near-zero or zero variable production costs because of the free source of fuel. When production-based subsidies exist, this variable cost can be negative. Finally, VG has unique diurnal and seasonal patterns that make it primarily an energy resource. This is because periods of high-energy output may not correspond to times of high demand (or risk of insufficient generation) when considering the power system's resource adequacy requirements.

These characteristics create unique challenges in planning and operating the power system, and they can also influence the performance and outcomes from electricity markets. At the same time, electricity market design must be robust enough to allow the market/system operator to make the most efficient use of the system, given the many constraints; thus, the market characteristics must be reflective of the physical nature of the power system. This report provides a comprehensive review of wholesale electricity markets and how the introduction of VG has impacted these markets. The report then focuses on two particular issues related to market design: revenue sufficiency for long-term reliability and incentivizing flexibility in short-term operations. Throughout the report, the authors provide an overview of current design and some designs that have been proposed by industry or researchers. Although certain market characteristics described in this report may point to advantages of specific market design elements, we avoid making specific recommendations in this report.

In the United States, wholesale electricity markets consisting of market operators—referred to as an independent system operator (ISO) or regional transmission organization (RTO)—now account for two-thirds of the nation's electricity consumption. Figure ES-1 shows the ISO/RTO markets in all of North America. In the United States, this includes the California Independent System Operator (CAISO), Electric Reliability Council of Texas (ERCOT), Southwest Power Pool (SPP), Midcontinent Independent System Operator (MISO), PJM Interconnection, New York Independent System Operator (NYISO), and Independent System Operator of New England (ISO-NE). Although there are many variations among the rules of these regions, there has been a large degree of convergence in the general principles of market design and their applications to existing markets. The common market design framework reflects a pool-based market in which there exists a two-settlement system for day-ahead market (DAM) and balancing/real-time markets (RTM), with co-optimized energy and ancillary services, locational marginal pricing (LMP) for energy, and financial transmission rights (FTR) markets for financial hedging. Energy is sold in day-ahead markets and balanced in 5-min real-time markets with LMPs for every generator bus on the system.

At present, some ancillary service markets are in place in all of these ISOs, including those for spinning contingency reserve, non-spinning contingency reserve, and regulating reserve. These ancillary services are used to support power system reliability and perform the necessary services that the energy market cannot provide. FTRs are cleared in forward markets and are instruments that hedge against locational differences in energy prices. Market mitigation procedures exist to limit gaming and market power. These procedures protect against market participant bids that do not reflect true variable costs and have a significant impact on market outcomes and competition.

Finally, in some regions there exist forward capacity markets. The purpose of these markets is to ensure long-term reliability so that enough capacity is maintained and built to meet future resource adequacy needs. Capacity markets are in place to ensure that resources that are needed for long-term reliability can recover their total costs of building and operating large generating facilities.

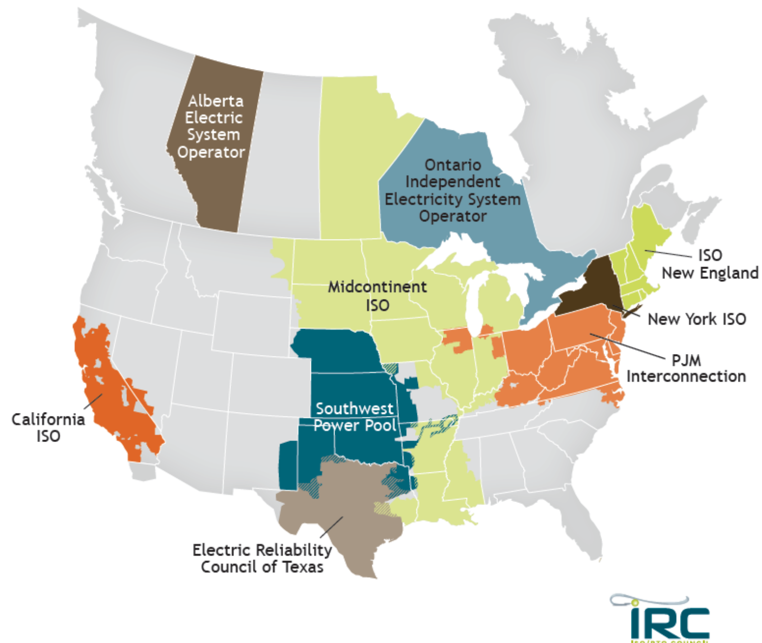

Figure ES-1. ISO/RTO market regions in North America

Each of the different market products discussed above includes auctions conducted by the market operator so that market participants can buy and sell services at various prices. Many of these markets are complex in the way they combine the physics of electric delivery with the principles of economics. Each element of a given electricity market design is linked to a facet of the reliability needs of the system, along with measures that allow for economically efficient operation and competition, while limiting market power. There are some differences in the way that quantities and prices are determined among regions. The rules that govern the operation of these markets are always evolving as new technologies enter the market. Throughout this report, we describe the differences of scheduling quantities and prices and how the mechanisms that are in place and are changing may be allowing for the integration of VG in an efficient and reliable manner.

Many of the tools necessary to manage the challenges of VG already exist, but the market structures may not be properly positioned to incentivize their efficient use or additional investment. Variability and uncertainty are not new challenges for system operators. Various operational practices and tools have been developed to manage different sources of variability and uncertainty, including the impacts of varying load and uncertain generation availability (i.e., from forced outages). Some of the tools used to reliably and efficiently manage these historical challenges include operating reserve requirements, network transmission service, and frequent system redispatch. Newer tools are also being proposed, including advanced scheduling models, operational VG forecasting, balancing authority area cooperation, intelligent operating reserve requirements, and new or changing ancillary service markets.

When new and existing tools are effectively and well designed, and in the absence of unintended consequences of market design elements, the impacts of VG can be mitigated; however, it is important for the market to incentivize these market players so that they are willing and able to provide the various services in an economically efficient manner. Initial evidence suggests that high penetrations of VG will require increased levels of flexibility to manage increased net load variability and uncertainty. It is unclear whether or not the current market designs have the right incentives to provide this flexibility when the system flexibility need is increased with rising penetrations of VG.

Other tools have been developed over the years to determine resource adequacy and revenue sufficiency. Numerous metrics have been developed to determine the level of capacity needed to meet future peak load projections considering the outage probabilities of capacity resources on the system. These metrics have been improved to also include the unique characteristics of VG. In addition, numerous market mechanisms have been introduced over the years to allow for the recovery of sufficient revenue for generators to recover total costs. It is still unclear whether these tools are supporting and will continue to support these issues of resource adequacy and revenue sufficiency in future systems with increasing penetrations of VG.

With increasing penetrations of VG, we have identified a number of potential impacts to the wholesale markets. These impacts are not necessarily negative per se; they may be reflective of a changing resource mix and completely appropriate. However, because these markets were initially designed without the notion that large penetrations of VG would be participating, the impacts should continue to be monitored to determine if the existing designs are still effective. If the original market designs lead to inefficiency, increased market power or reduced competition, or reliability degradation, modifications to these market designs may be required.

In the energy markets, a number of changes to the market schedules and prices have been and may continue to be seen with increased penetrations of VG:

- VG can reduce average LMPs over time because of its low, zero, or negative bid-based costs. VG can also cause more occurrences of zero or negative LMP periods because of its bid costs. These outcomes can affect the revenue stream of other resources that may depend on revenue from the energy market. Because it has the lowest operating cost, VG nearly always enters at the bottom of the bid stack. This means that VG can also reduce the energy schedules of other resources and will have a similar impact on energy market revenue streams.
- VG can cause the LMPs to be more volatile from one time period to another because of its increased variability. VG can also cause greater disparity between DAM and RTM

LMPs because of its increased uncertainty. Both can cause greater uncertainty to all market participants on price outcomes.

- Finally, with increased variability and uncertainty, VG can increase the need for flexibility in the system. If there is no way to incentivize for this flexibility when needed, potential reliability issues or costly out-of-market actions can occur.

VG can also impact ancillary service markets:

- Typically, operating reserve is used to better manage variability and uncertainty. When system variability and uncertainty are increased with increasing levels of VG, more operating reserves may be needed. VG can increase the requirement for normal balancing reserve, such as regulating reserve, which can increase the demand and therefore the ancillary service clearing prices for those services.
- Further, because of the variability and uncertainty of VG impacts on the operating reserve requirements, it is possible that the reserve needs may differ through time and even change between the DAM and RTM. This can cause greater uncertainty from the market participants in ancillary service demands and prices.
- VG can also displace synchronous, frequency-responsive power plants, and when not equipped with technology to provide a comparable response, VG can cause the need for supplemental actions or market designs to ensure that sufficient frequency response is made available.
- VG may also cause a higher probability of scarcity events when its variability and uncertainty lead to ancillary service requirements being unmet by the resource mix because of some combination of commitment, transmission, or ramp limit constraints. This can cause more price spikes, which may cause higher costs to consumers and higher revenue to generators.
- Finally, when VG increases the need for flexibility, it is important that flexibility is incentivized in the ancillary service markets, particularly if it is not incentivized in the energy markets.

This report does not focus heavily on the impact that VG has on FTR markets; however, we do identify a few impacts that researchers and market designers should continue to evaluate.

- The increased variability of VG can cause greater variation on power flows, causing FTR holders to be uncertain of expected congestion patterns.
- The increased uncertainty of VG can also cause greater deviation of power flows between DAM and RTM. Because FTR revenues are typically based on DAM congestion, there could be a greater divergence between FTR revenues and actual congestion patterns.

Finally, the increasing penetration of VG can influence forward capacity markets that are now in place in some U.S. market regions.

- The reduction in LMP and energy schedules from conventional resources will result in reduced revenues in the energy market. If these resources are still required to be available for short periods of time and to meet long-term reliability requirements, more resources will become capacity-based resources rather than energy-based resources. They may have to rely on forward capacity markets or revenues other than energy market revenues to earn the level of revenue needed to remain in the market. This is the revenue-sufficiency question.

- The variability and uncertainty of VG may require more flexible resources to maintain resource adequacy—inflexible resources may not be enough. Incumbent, inflexible forms of generation may be induced to increase their ability to provide flexibility, and market designs should provide appropriate incentives that may encourage retrofits if they are found to be cost-effective.
- Finally, must-offer price rules, which are designed to limit market power, may increase risk that a resource built to satisfy a state renewable portfolio standard will not clear the capacity market at the applicable minimum offer floor.

This summary of impacts leads us to the two key issues explored in this report. First, we examine the revenue-sufficiency question posed above: do energy plus ancillary service markets provide the revenue required to cover all costs? Insufficient revenue may lead to an unreliable system when those resources choose to leave the market. Reliability issues can also arise from miscalculations of the long-term reliability need, both for capacity and other attributes. Proper market designs should allow for a sufficient level of resources needed for long-term reliability to recover their total costs and remain in the market.

For the second issue, we explore whether sufficient structures are in place in current market designs to ensure that the resources necessary to balance the variability and uncertainty are available and used efficiently. Improper utilization of existing flexibility, or unwillingness of resources to provide flexibility, can lead to efficiency or reliability degradation. It can lead to insufficient flexibility available to the market operator to meet the changing net load, resulting in increased energy imbalance. It can also lead to higher costs when more expensive flexibility is used out of market instead of economic flexibility that is not offered into the market. Proper market designs to incentivize flexibility are critical to meeting these challenges in an efficient manner. These two issues become the focus of this report.

Revenue Sufficiency and Long-Term Reliability

To meet long-term resource adequacy needs, system planners use a variety of different metrics to understand how much capacity is required and how each resource on the system can contribute to meeting that requirement. Although it is used in many of the market areas within the forward capacity market, we show how the planning reserve margin, used in isolation from reliability-based metrics, loses a lot of the information required to make efficient decisions about resource adequacy. We also show how the planning reserve margin can be more difficult to use when resources such as VG are a large part of the resource mix, with availability based on weather patterns rather than forced outage probabilities. System metrics, such as loss of load expectation for system needs and effective load-carrying capability for individual contributions, especially if done on an hourly rather than daily basis, can provide a better estimate for meeting long-term reliability needs.

The effective load-carrying capability metric measures how much more peak load can be added reliably with the introduction of new capacity. This metric can show how each resource contributes to long-term resource adequacy and therefore provides information about how many and which resources the market should allow for recovery of capital costs. Wind and solar generation are unique in their dependence on weather for capacity availability and diurnal and seasonal patterns that dictate their effective load-carrying capability as opposed to forced outage rates. Many areas have adopted time-based approximation methods to calculate the capacity value of VG using a certain time period (month, hours, and number of years for calculation) and different calculation methods (median generation value or average capacity factor), as shown in Table ES-1.

Table ES-1. Methods for Determining the Capacity Value of Wind and Solar in RTO/ISOs

RTO	Season	Months	Time	Term	Method
ISO-NE	Summer	June–September	1–6 p.m.	5-y rolling	Medium net generation
ISO-NE	Winter	October–May	5–7 p.m.	5-y rolling	Medium net generation
ISO-NE	All	Default based on summer and winter wind speed data and ISO oversight			
NYISO—Existing	Summer	June–August	2–6 p.m.	Previous year	Average capacity factor
NYISO—Existing	Winter	December–February	4–8 p.m.	Previous year	Average capacity factor
NYISO—New Onshore Resources	Summer	10% default capacity credit			
NYISO—New Onshore Resources	Winter	30% default capacity credit			
NYISO—New Offshore Resources	All	38% default capacity credit			
PJM	All	June–August	2–6 p.m.	3-y rolling	Average capacity factor
PJM—New Wind	All	13% default capacity credit			

The much-debated missing-money issue in power systems has been described in much of the literature. It predates the current large-scale expansion of VG and is said to be caused by flaws with retail consumers shielded from wholesale price swings, the existence of price caps and inadequate scarcity pricing, and, because of the interconnected nature of the power system, the inability for different customers to purchase different levels of reliability. With increased penetrations of VG—which are likely to offer energy at near-zero, zero, and negative-bid costs—both the average energy prices and the cleared energy levels of existing generating plants are likely to be reduced. This can cause the overall revenue for these existing generators to be reduced, and depending on other revenue streams and changing reliability needs, it can exacerbate the missing-money issue by making revenue sufficiency more difficult to achieve.

The differences in long-term reliability needs introduced by VG are not only because of the calculation methods of VG capacity availability. VG also increases the need for special attributes of new and existing resources other than only capacity. The resources in existence in the market may need to have qualities relating to the increased flexibility needed to efficiently integrate VG into the bulk power system. This question differs from the second question introduced earlier regarding incentivizing flexibility in system operations. This question focuses on the availability of flexible attributes to meet long-term reliability needs—the question of how much and what type of generation to build or in which to invest. The earlier question focused on whether and how market mechanisms could incentivize providing flexibility in the short-term. These are separate yet related questions—incentivizing the development of sufficient flexible resources is a prerequisite to utilizing that flexible generating asset in market operations. Figure ES-2 shows the maximum ramp magnitudes and durations to expect on an example system—one way to determine the flexibility need. Many new methods, some very similar to those developed previously for capacity valuation (i.e., loss of load and effective load-carrying capability), are being developed to better understand the long-term flexibility need. If flexible resources are never built, it does not matter what incentives are introduced in the short-term market—the operator would clearly not have access to flexible resources that are not built. Flexibility, similar to provision of energy, can have both fixed capital and variable operating costs. Careful evaluation of short- and long-term markets should take place to determine whether incentives are in place for flexibility to be part of new installed capacity, for flexibility to be part of existing resources through modifications, and, as discussed later, for flexibility to be offered and provided in system operations when needed.

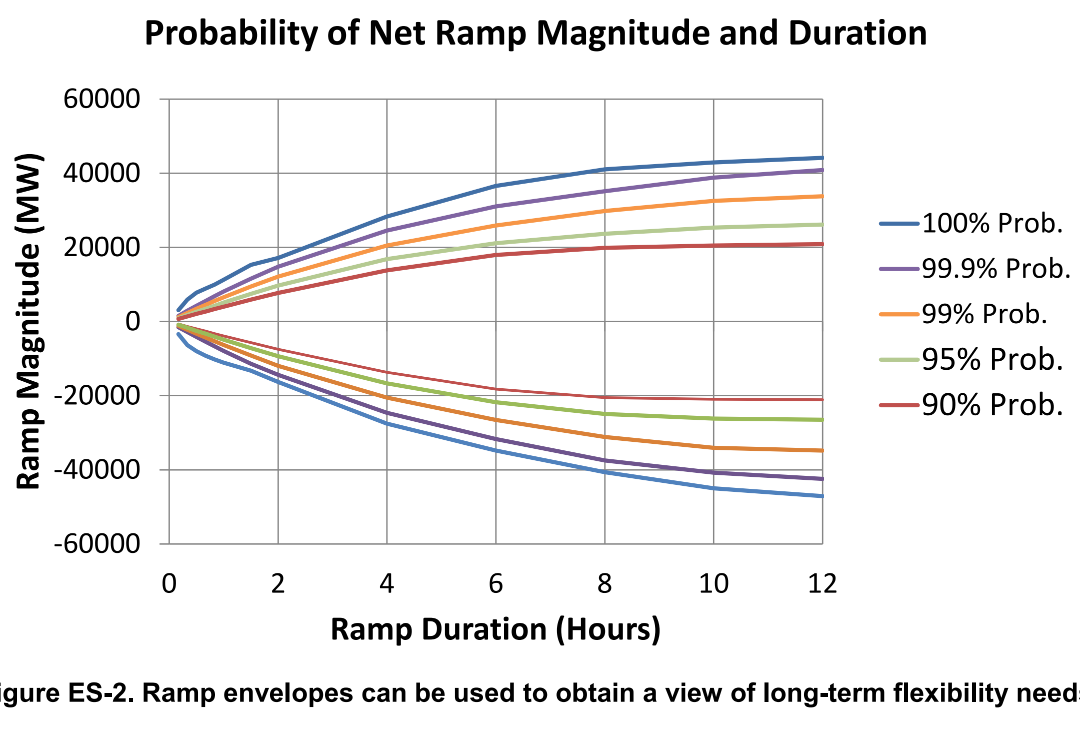

Figure ES-2. Ramp envelopes can be used to obtain a view of long-term flexibility needs

The two primary existing market mechanisms for providing revenue sufficiency are scarcity pricing and forward capacity markets. Low variable-cost resources are often able to recover some of their fixed capital costs in the energy market when the marginal price is set by resources with higher variable costs; however, for the highest variable-cost resources, the price would rarely exceed their marginal cost, thus reducing or eliminating their opportunity to recover their fixed costs in the energy market.

Scarcity pricing—through the value of lost load pricing, ancillary service scarcity pricing, and emergency demand-response pricing—allows the energy price to rise to levels much higher than the variable cost of the highest variable-cost resources when the system is capacity constrained. These prices can range from thousands to tens of thousands of dollars per megawatt-hour when the system cannot meet the demand. In reality, the prices are typically based on administratively set ancillary service scarcity prices. When the system is not able to meet the ancillary service requirements, the price will be set to an administratively set high level. Because energy and ancillary services are co-optimized, this price will also be reflected in the energy market. These prices, though difficult to predict in terms of investments into capital, can provide some needed revenue for capacity resources to recover fixed capital costs. Prices based on demand response bids can provide a similar effect.

Forward capacity markets have been established in NYISO, ISO-NE, and PJM. These markets have been put in place to establish more certain revenue streams for investors in new capacity that will be needed at a future date. These markets determine capacity prices depending on the current supply of capacity and the ISO's predicted capacity need. The capacity markets vary with respect to the time horizon for which they are securing capacity, the resource qualification criteria (e.g., how demand response and other nontraditional technologies can participate), and the slope of the demand curve. Price caps are typically used to limit prices to the cost of new entry, usually set to the annualized capital cost of a peaking plant. Price floors may also be in place to avoid price collapse and potential exercise of market power from purchasers of capacity. The forward capacity markets also usually have zonal markets, such that new capacity is incentivized to locate and be deliverable in areas that need that capacity the most.

Some recent changes have been made to the design of electricity markets to address some of the issues with resource adequacy and revenue sufficiency. In ERCOT, which has an energy-only market, recent changes have been made to the way in which scarcity prices are calculated. First, the offer cap that limits the bid price of generation resources is gradually being increased up to $9,000/MWh. Second, an operating reserve demand curve is implemented in the real-time market, such that the price of operating reserve will depend on the hourly probability of lost load, which depends on the amount of operating reserve available. This value is added directly to the energy price as a premium, so that the additional revenue may help recover fixed costs.

In CAISO, a new proposal will require that load-serving entities have sufficient flexible capacity to meet forecasted system needs in addition to their resource capacity adequacy requirements. Any resource that contributes to meeting the flexibility requirement must offer a flexible resource during daytime periods in the short-term energy markets. The flexibility requirement is proposed to be based on the maximum 3-h net load ramp that is projected to occur in a month (see Figure ES-3). This would be the first U.S. market area to include flexibility attributes in its resource adequacy planning requirements.

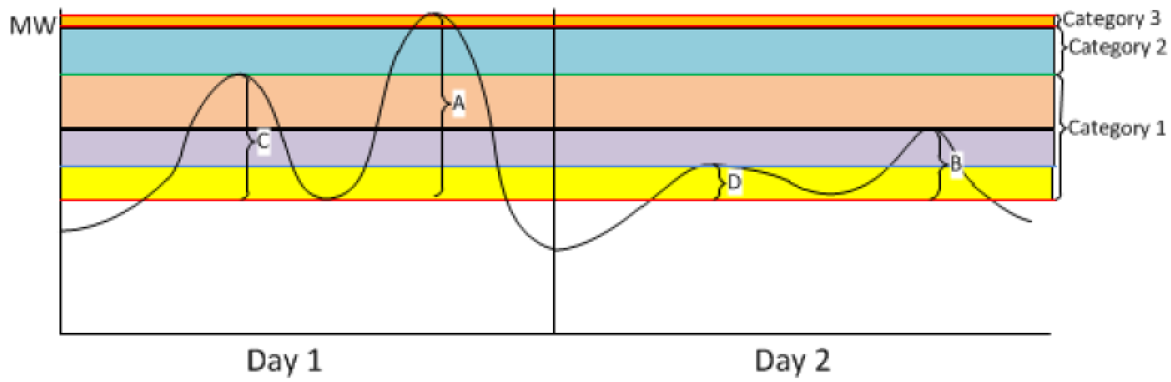

Figure ES-3. Flexible capacity requirements in CAISO based on 3-h net load ramps

Incentivizing Flexibility in System Operations

Defining flexibility has been a challenge that a number of industry members and researchers have attempted to address in recent years. With increased variability and uncertainty of VG, the resources on the system will have to be more flexible to adjust output, so that power output ranges, power ramp rates, and energy duration sustainability are sufficient to meet the needs of balancing supply with demand at various operational timescales. For clarity, we provide a definition of power system flexibility in system operations as *the ability of a resource, whether any component or collection of components of the power system, to respond to the known and unknown changes of power system conditions at various operational timescales.* Similar definitions have been proposed by other authors. Different types of resources excel at different forms of flexibility, and they also have different cost impacts when providing flexibility. When providing substantial flexibility to the system when the system requires that flexibility to maintain the generation and load balance, and the flexibility increases costs of fuel, wear-and-tear, or other factors, it is important that the market structure incentivizes resources to provide this flexibility.

VG output varies through time at different timescales. This causes the aggregate net load to change more and potentially faster as well. In turn, the resources used to balance the system need to change output more often and at potentially faster rates. In addition, VG output cannot be predicted perfectly in advance. This may require at least some of the balancing that occurs to come from resources that can quickly respond when the output of VG is finally apparent. In addition, VG is location constrained and nonsynchronous to the electrical frequency. These characteristics can also cause challenges and potentially increase the need for these types of flexibility from other resources that may be displaced by VG. Most of these flexibility needs can be met by existing market processes, tools, and methods, such as 5-min scheduling and ancillary service markets. However, sometimes new scheduling strategies or improved methods may more efficiently extract the flexibility that is needed from the system resources.

Several mechanisms are in place in the existing markets to induce resources to offer flexible bids to the market operator, therefore allowing the market operator to adjust resource output to meet the changing needs of the system. Some examples include efficient centralized scheduling and pricing, 5-min settlements, ancillary service markets, make-whole payments, and day-ahead profit guarantees. Centralized scheduling leads to more efficient prices that should, in theory, provide greater profit to resources who participate compared to those who self-schedule their output. This is shown in Table ES-2. The table shows that profit is maximized when the market operator schedules the resource, and profit is lost when it self-schedules itself in any manner.

Table ES-2. Revenue, Costs, Profits, and Profit Lost for Various Self-Scheduling Techniques

Scenario	Total Revenue	Total Cost	Total Profit (Total Revenue Minus Total Cost)	Profit Lost
Scenario 1 Market	$195,668	$147,313	$48,355	N/A
Scenario 2 Min Self	$167,326	$120,120	$47,206	$1,148
Scenario 3 Max Self	$220,567	$173,594	$46,972	$1,382
Scenario 4 Mid Self	$175,517	$128,079	$47,438	$916
Scenario 5 Lag LMP	$190,452	$142,909	$47,542	$812

Five-minute settlements, which are related to but not identical to 5-min scheduling, occur if the resources actually are paid based on the 5-min schedules and 5-min prices rather than based on hourly averages of the 5-min schedules and prices. This is done only in a few of the markets today. Table ES-3 shows how hourly average settlements may deter resources from providing 5-min flexibility. With 5-min settlements, following the market schedule is the most profitable ($4,313 is the highest in second column from right). With hourly settlements, performing output different than the market, and therefore inhibiting the flexibility that the market needs, is more profitable than following the market schedules ($2,443 is less than the other two methods of chasing the hourly average, as shown in the right-most column).

Table ES-3. Revenue, Costs, and Profits of Different Scenarios with 5-Min Settlements Versus Average Hourly Settlements

Scenario	Total Revenue		Total Cost	Total Profit (Total Revenue Minus Total Cost)	
	5-Min Settlements	Hourly Average Settlements		5-Min Settlements	Hourly Average Settlements
Scenario 1 (Market)	$15,208	$13,338	$10,895	$4,313	$2,443
Scenario 2 (Moving Hourly Average)	$17,479	$17,441	$13,373	$4,106	$4,068
Scenario 3 (Perfect Knowledge of Hourly Average)	$17,134	$17,134	$13,053	$4,081	$4,081

There are two other ways in which side payments can be made to suppliers to incentivize them to participate and offer their flexibility to the market: the first is the make-whole payment, which provides additional revenue that compensates for when bid-in costs are higher than revenues such that the supplier is guaranteed not to lose money in the short-term when participating and following the directions of the market operator; the second is the day-ahead profit guarantee, which maintain profits earned in the DAM in cases in which the generation would lose profits by performing actions that benefit the system in the RTM.

Newer market mechanisms have also been proposed in many markets to incentivize flexibility in system operations. Markets have started to allow for nontraditional resources, such as demand response, energy storage, and VG itself to provide flexibility, and they benefit from doing so. New ancillary service market designs, such as pay-for-performance regulation and primary frequency response markets, provide additional incentives to support reliability for resources that may not have had those incentives in the past. Finally, explicit products for flexible ramping provision, as has been proposed in CAISO and MISO, may be able to provide specific incentives that make resources more inclined to provide greater flexibility when the flexibility is needed in system operations. CAISO has already partially implemented a piece of this product with its introduction of the flexible ramping constraint in 2011. Table ES-4 shows some statistics of the payments made for providing this flexible ramping capability during 2012. More research is required to evaluate the potential value of flexible ramping products, along with the other products and design changes in development. It is important that these products incentivize necessary capabilities when they are necessary and that any revenues for providing a type of service are not double-counted through multiple market products.

Table ES-4. Statistics for the First Year of Flexible Ramping Constraint in CAISO

Month	Total Payments to Generators ($M)	Intervals Constraint Was Binding (%)	Intervals with Procurement Shortfall (%)	Average Shadow Price When Binding ($/MW-h)
January	$2.45	17%	1.0%	$38.44
February	$1.46	8%	1.3%	$77.37
March	$1.90	12%	1.0%	$42.75
April	$3.37	22%	1.5%	$39.86
May	$4.11	23%	6%	$79.48
June	$1.49	13%	2.3%	$77.37
July	$1.01	8%	1.4%	$42.75
August	$0.77	7%	1.2%	$39.86
September	$1.03	13%	0.8%	$79.48
October	$0.9	9%	1.0%	$39.19
November	$0.23	4%	0.5%	$53.34
December	$1.09	9%	1.6%	$61.84

Conclusions

VG has several unique characteristics that can have an impact on power system operations and the outcomes of wholesale electricity markets. The impacts of VG may be both positive and negative. Two important questions asked in this report are (1) whether or not resources needed for long-term reliability are recovering fixed capital and variable operating costs, and (2) whether resources are being incentivized to provide flexibility when flexibility is needed. We review several historical approaches taken by market areas in the United States to manage these issues and describe how these approaches have evolved in recent years due, in at least some part, to increased VG levels. A review of different market mechanisms was undertaken to gain insight into how the current markets function and how they contribute to managing the various challenges of VG. We have found that these questions are also now being addressed by many regions, although the approaches taken differ significantly among regions. Whether or not consistency across markets should be an objective, it is important that further research is carried out to ensure that evolving market designs are providing the incentives they set out to provide and can achieve optimal efficiency and reliability.

In our discussion of capacity market mechanisms, we showed the difficulty in using peak periods as a proxy for times of system risk and the difficulty—if not the impossibility—of the ability of a capacity market to capture the salient aspects of resource adequacy and reliability. The regions in the United States that have capacity markets have alternative ways of mapping the reliability target to the capacity acquisition target, though there are significant differences in market timing and questions about the ability to capture the long-term investment process to ensure adequacy.

In our discussion of flexibility incentive mechanisms, we showed that existing designs could help ensure that resources that are flexible have the incentive to offer that flexibility. However, newer designs might be needed to get the right quantity of flexibility being offered and provided to the market. We also showed that flexibility means many things, whether it is the speed of power adjustments, the range of power that can be provided, or just the fact that it can autonomously respond to a frequency deviation. We have found that the mechanisms to maintain reliability, both in the long term and the short term, are inseparable to the market design. This is especially true because of the way the United States wholesale electricity market designs have been established. System operators, planners, and reliability organizations and regulators cannot consider the means of meeting reliability without recognizing the electricity market design. Likewise, market designers cannot create designs without recognizing the reliability needs of the system. Markets that do not reward needed and valuable characteristics will not incentivize—or dis-incentivize—those characteristics; in the absence of alternatives, this will result in a lack of needed characteristics. This can compromise economic efficiency and/or reliability. The links between the market design, reliability, and system planning and operation must not be broken. Numerous additional questions about market design evolution should be researched analytically, and results should be investigated analytically and disseminated to the electric power community. Electricity markets, which for the most part were not designed with the notion that large penetrations of VG would be part of the supply mix, should continue to promote competition, incentivize services from market participants that reduce costs to the consumer, and maintain a reliable, secure electric system while also gaining the environmental benefits that renewable energy offers.

Table of Contents

List of Figures

List of Tables

1 Introduction

The levels of installed variable generation (VG) capacity, such as wind and photovoltaic solar energy, have increased significantly in recent years. Many states in the United States have incentives or requirements that will provide for a further increase in VG in the coming years. VG characteristics are unique compared to those of traditional thermal and hydropower plants that have been providing the majority of the electric power produced worldwide during the past 100 years. VG increases the variability of the net load (load minus VG) because its available power changes with time due to the changing source (e.g., wind speeds or solar irradiance). VG also increases the uncertainty of the net load because the available power cannot be perfectly predicted at all time horizons. In addition, VG has near-zero, zero, or negative variable production costs because of the free source of its fuel and the relevant production-based subsidies given but with high fixed capital costs. This means that VG is operated at its maximum available capacity limit during the majority of times, is the lowest-variable-cost resource, and would operate below this level only when the marginal cost to serve load in its location is zero or negative (i.e., the VG is the marginal resource). Finally, VG is primarily an energy resource that has limited capacity value relative to its rated capacity because periods of high energy output may not correspond to times of high demand or risk of insufficient generation. Together, these characteristics create challenges to reliably operating the power system at least cost and planning for the expansion of the power system to meet the changing needs of the future. The special characteristics of electricity markets, especially those in the United States, and their close ties to the physical power system's operational and planning needs creates unique challenges in the design of wholesale electricity markets. The operators and planners that have goals toward maintaining reliability, in the short and long term, cannot consider mechanisms to meeting reliability without some recognition of the electricity market design. Likewise, market designers cannot create new designs without recognition of the system reliability needs. This is an ongoing theme seen throughout this report.

Variability and uncertainty are not new challenges for system operators. A series of operational practices and tools have been developed during the last century to manage sources of variability and uncertainty. Some of the tools used to reliably and efficiently manage these historical challenges include the carrying of operating reserve, network transmission service, and frequent system redispatch. In regions with restructured power markets, the revenue to provide these services is derived from energy and ancillary service markets, and in some cases forward capacity markets. These markets are designed to provide an opportunity for resources to recover their variable and fixed capital costs, incentivize resources to contribute to a reliable and secure system, and provide indifference for resources on their operational directives. As the net variability and uncertainty of the system increases with higher penetrations of VG, new and existing market products to manage these challenges may need to be explored. Initial evidence suggests that high penetrations of VG will require increased levels of flexibility to manage increased net variability and uncertainty. It is unclear if the current market designs have the right incentives to induce the provision of this flexibility. At the same time, higher penetrations of VG may cause energy prices, the main method by which most electricity suppliers earn sufficient revenue, to decrease (Maggio 2012). This can cause the resources that are needed for capacity and flexibility to earn less revenue to recover their variable and capital costs.

It has been shown that the market regions can facilitate the integration of VG (Milligan, Kirby, Gramlich, and Goggin 2009). Market regions are typically large in geographic area, which reduces the aggregate variability and uncertainty of load and VG and provides a larger source of resources able to accommodate variability and uncertainty. Regional transmission organizations (RTOs) and independent system operators (ISOs) typically have transparent markets for services needed to reliably integrate VG and have fast energy markets that can meet the changing needs more efficiently. Finally, transparent pricing in all the different market forms (including energy and various ancillary service markets) can also incentivize market participants to install the needed capabilities and, because of location-specific pricing, build in the right locations. However, most of these markets were designed without significant penetrations of VG in mind. Significant improvements have been made to accommodate this unique resource in the various forms of electricity market design, but several areas of market design may still need further development (Smith et al. 2010). VG has a variety of unique characteristics that are of importance to the topic of wholesale electricity market designs. This paper focuses on four important characteristics:

1. The maximum available power from VG at any given moment depends on weather conditions—for example, wind power output depends on the current wind speed, and solar power output depends on solar insolation. Therefore, the maximum available power will vary with time. Additionally, a wide geographic spread of VG will make it likely that full rated output from all plants will not be concurrently achieved. This property of VG coupled with the effective inelasticity of demand can increase volatility in energy prices. The variability in output from VG also creates a need for other resources to provide flexibility to adjust output to meet the changing net demand and follow the volatile pricing patterns. The variability also may change the needs for ancillary service products and modify the design of ancillary service markets.
2. The maximum available power from VG cannot be predicted with perfect accuracy, which means that market participants cannot predict what the market outcome and prices will be with certainty. Forecast uncertainty also adds to the need for other resources to have flexibility to adjust when errors in net load forecast are realized, and this affects the needs for ancillary services and potentially the design of ancillary service markets.
3. VG has essentially zero variable costs and even negative variable costs when production-based subsidies are considered. This means that they enter at the bottom of the offer stack, reducing the production of marginal resources, thereby reducing energy prices. They can also set the price when they are the marginal resource, such as during times of low load, high VG output, and transmission constrained periods, making the price of energy zero or negative.[1]
4. The long-term availability of VG has seasonal and diurnal characteristics. Because of these characteristics, some VG may not provide significant load-carrying capability relative to their operating capacity. Because the load-carrying capability of VG is based on weather drivers rather than outage probabilities, adjustments to the procedures used to determine its contribution to resource adequacy are needed. In addition, systems with

[1] Note that negative prices occur in systems without VG due to transmission constraints, unit constraints that keep units online, and self-scheduled resources.

high levels of VG will require other capacity on the system that can contribute to resource adequacy, but that will likely generate at relatively lower capacity factors.

VG often has other unique characteristics as well, including being non-synchronous and constrained by location to be located where its fuel source is most prevalent. This paper addresses two related questions that become apparent from the characteristics shown above. The first question is whether resources needed to ensure a reliable system in the long run have sufficient opportunity to recover their variable and fixed costs to remain in the market. With zero variable costs from VG reducing prices and increasing the occurrence of zero or negatively priced periods, while at the same time reducing the energy that other suppliers can sell in the energy market, the suppliers that are still needed to provide capacity and flexibility to meet the long-term reliability requirements may earn less revenue in the energy market. We will explore how RTO/ISO markets in the United States are designed to ensure that suppliers who are needed for long-term reliability are given sufficient opportunity to remain in the market, proposals of how this may be evolving, and discussion about how this may affect electricity market designs in the future.

The second question is whether existing market designs provide adequate incentives for resources to offer their flexibility into the market to meet the increased levels of variability and uncertainty introduced by VG in the short-term operational timeframe. Increased variability and uncertainty can increase the need for flexibility, and market designs will have to make sure the flexible suppliers are incentivized to provide their flexibility when needed. We will explore how markets in the United States currently incentivize resources to offer flexibility, discuss how these designs are evolving, and discuss how this need will affect electricity market designs in the future.

At present, several RTOs/ISOs are investigating ancillary service products to increase access to flexibility. Some in the industry argue that energy prices during certain periods with significant ramps may not be sufficient to incentivize units to offer their flexibility into the market or may incentivize inflexible resources incorrectly. In the long term, it is not clear whether market signals in today's real-time markets (RTMs) will provide sufficient incentive for investors to develop new sources of flexibility, which may include some combination of flexible generation technology, demand response, or storage. Thus, the issue of flexibility has two dimensions: the long-term signals to ensure that sufficient flexibility is built and the short-term mechanism for ensuring that the flexible resources offer that flexibility to the market.

This paper is structured as follows. Section 2 gives an overview of current wholesale electricity market designs in the United States and how VG may be impacting the outcomes of various markets and market products. Section 3 discusses whether resources needed for long-term adequacy are earning sufficient revenue to stay in the market and how some regions are addressing this issue. Section 4 discusses whether flexibility is being incentivized in system operations to meet the increased needs from VG and new and evolving designs from several markets addressing this issue. Section 5 then gives insights and examples of further research needed to address these issues and concludes the paper.

2 Overview of Wholesale Electricity Market Designs

The goal of all electricity systems, whether they are operated by regulated monopolies or centrally administered by an RTO/ISO, is to ensure the reliable delivery of electricity at the lowest cost to consumers. These goals are rooted in a long history of regulatory principles that influence the entry of new market participants, set prices, prescribe the quality and condition of entry, and obligate a utility to provide service. The rationale for this regulation emerges from the physical constraints of the electric grid. This paper is not intended to explain the intricacies of the grid and electric utility regulation, but a brief review is important to understanding the challenges of electricity markets.

Three fundamental components comprise the wholesale electricity supply: generation, transmission, and coordination services. Each of these has a financial and physical component that must accommodate for the lack of a consumer response inherent to electricity markets while ensuring the constant balance of generation and load. Because of the extreme cost associated with failure of the power system, the physical requirements of the system must be ensured, even though market and operational inefficiencies are introduced to do so. Assuring this reliability requires procuring adequate generation, transmission, and coordination services. In short, resource adequacy—i.e., having enough available capacity in the system—is required to reliably meet load at all times. This includes adequate transmission capacity which is also required to ensure energy can be delivered to where it is needed. Because electricity demand is relatively inelastic, variable in time, and uncertain in quantity, both generation and transmission must be constantly coordinated to meet load in a reliable manner.

To gain the system requirements necessary to support the security and reliability of the electric grid, adequate market policies must be crafted that address the financial implications of these requirements. Ideally, these policies will provide sufficient opportunity for generators to recover both fixed and variable costs if they contribute to resource adequacy; promote the construction and upkeep of a viable transmission network; and incentivize generators to coordinate scheduling of resources to meet the variable and uncertain load while maintaining the reliability of the transmission network. Simultaneously, these policies must avoid incentivizing an overbuilt system or overcompensating inefficient units.

Historically the electricity industry has been operated as a natural monopoly, regulated by a combination of state commissions and Federal oversight for some aspects of interstate trade. The legal justification for electric utility regulation in the United States can be traced through British common law and a series of Supreme Court cases.[2] Generally speaking, these cases have found that utilities, such as those in the electricity and natural gas industries, provide services that are in the "public interest" and are necessary for the common welfare of the people. The economic justification for regulation has been focused on the inherently noncompetitive nature of the market. The market can be uncompetitive for a variety of reasons, including: (1) technology that allows a limited number of companies to provide adequate capacity to supply all demand; (2) the unique position of a principle buyer; and (3) conditions in the market that do not produce competitive results (Brown 1993). Because of these characteristics, regulators and policy makers

[2] Munn v. Illinois (94 U.S. 113, 1887), Smyth v. Ames (169 U.S. 466), and Federal Power Commission v. Hope Natural Gas Company (320 U.S. 591).

have adopted certain regulatory frameworks to meet the essential needs of society and to ensure that utilities are capable of earning a fair return on their investments. However, the potential benefits of competitive generation instigated the restructuring of the electricity markets in the late 1990s. Currently, more than two thirds of the electricity consumption in the United States is purchased within restructured electricity markets. The restructured markets have been designed to support the financial constraints of generation, transmission, and coordination that are necessary to secure a stable and reliable physical power system while addressing the problems of inefficient pricing, investment risk, and market power. The RTO/ISOs shown in Figure 2-1 evolved to administer these energy, ancillary service, financial transmission rights (FTRs), and in some instances forward capacity markets that have formed in response to these challenges.

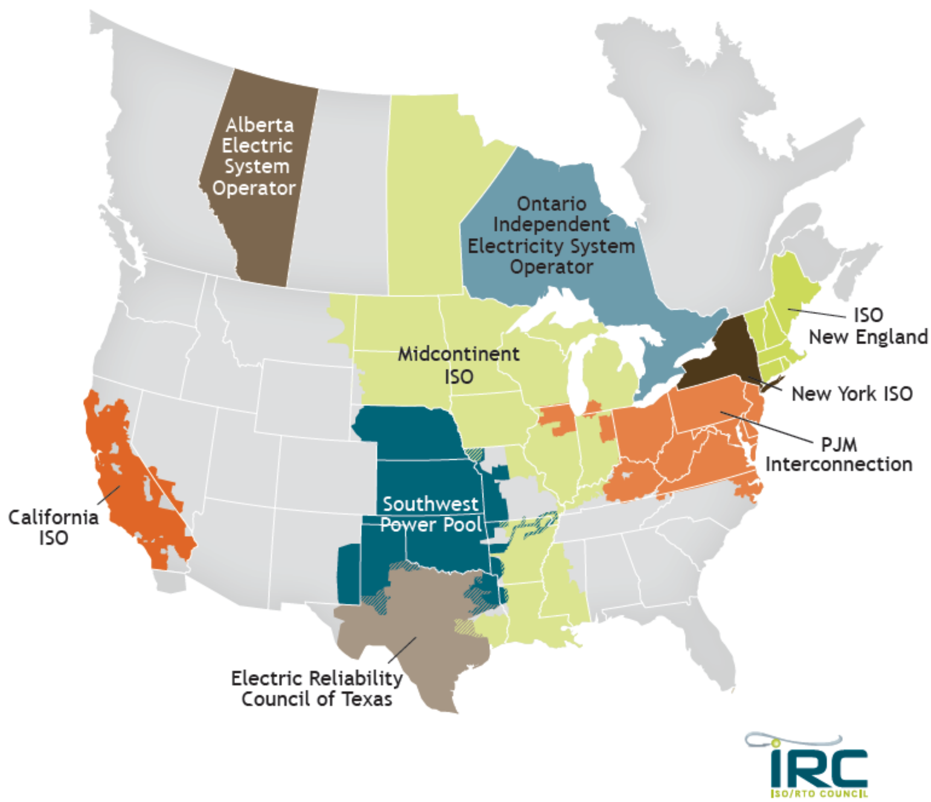

Figure 2-1. RTO/ISO regions in North America. ***Image from www.iso-rto.org***

In the United States, RTO/ISO administered markets have evolved in similar directions to a large extent following the principles proposed in the standard market design (Hogan 1998). This design reflects a pool-based market in which there exists a two-settlement system for day-ahead and balancing markets (e.g., RTM), with co-optimized energy and ancillary services, locational marginal pricing (LMP) for energy, and FTR markets in place for financial hedging. Energy is sold in forward (e.g., day-ahead, hourly) markets and balanced in 5-min RTMs with LMPs. Locational energy markets in the United States are cleared once a day for hourly trading intervals for day-ahead markets and every 5-min for real-time markets. At present, ancillary service markets are in place in all markets, including those for spinning contingency reserve, nonspinning contingency reserve, and regulating reserve. The ancillary service markets operate in a similar manner to energy markets and are cleared using the same model, with day-ahead and real-time prices and schedules for the capacity reservation of the ancillary service. FTRs are cleared in forward markets and are an instrument put in place to hedge against locational differences in energy prices.

Each RTO/ISO also has a process for procuring sufficient resources to meet the peak load requirements. In PJM, New York Independent System Operator (NYISO), and the Independent System Operator of New England (ISO-NE), mandatory capacity markets have been designed to incentivize investment in installed capacity and to allow peaking units to recover fixed costs. At present, the Electric Reliability Council of Texas (ERCOT), California Independent System Operator (CAISO), Midcontinent Independent System Operator (MISO), and SPP do not have mandatory capacity markets available and utilize various administrative processes and spot prices to provide fixed-cost recovery for resources. We describe each of these markets in further detail below.

For context, Table 2-1 shows the annual revenue received in the various markets that were part of ISO-NE's wholesale electricity market in 2011 and 2012 (Likover 2013). The majority of revenue was being sold in the energy markets, specifically the day-ahead energy market, with the capacity market revenues producing the next most revenue. Uplift payments (out of market), ancillary service markets, and FTR all followed, with approximately two orders of magnitude less in revenue received. Although the share of revenue volume will vary from year to year and from market to market, this chart gives a good perspective on each of the markets and how they contribute to total wholesale electricity payments.

Table 2-1. Annual Revenues of the Various Markets Within the ISO-NE Wholesale Electricity Market

Market	2011 Revenue	2012 Revenue
Day-Ahead Energy Market	$6.5B	$4.9B
Forward Capacity Market	$1.3B	$1.2B
Real-Time Energy Market	$300M	$300M
Uplift (Make-Whole Payments)	$74M	$87M
Ancillary Service Market	$36M	$46M
FTR Market	$23M	$16M

2.1 Energy Markets

Energy is bought and sold in most of the RTOs/ISOs through a two-settlement system. A forward market sells energy to load-serving entities (LSEs) and buys from sellers in advance of the time when the energy is produced and consumed. This is typically through the day-ahead market (DAM). The DAM clears to meet bid-in load demand for the entire day, one day in advance. Schedules and prices are calculated from the market-clearing engine, and this price-quantity pair is settled for all market participants regardless of their actual performance. The DAM is important because it provides a hedge against price volatility in the real-time markets caused by load forecast errors, generator outages, or other imbalances. The DAM also allows for make-whole payments when resources do not recover their costs, and it provides price incentives in advance toward reliable operation when resources may need ample notification time to be able to start their generating resources (Stoft 2002).

To reflect changes that may occur between the day-ahead market and real-time operations, a second market clearing is used by RTOs/ISOs to re-dispatch resources and commit new resources to meet system requirements. This is generally referred to as the RTM. Variability and uncertainty is present throughout the power system including changes in weather that can cause unexpected deviations in load and variable resource output, and forced outages that can take resources and network facilities offline unexpectedly. The RTM is in place to set prices and schedules to match the imbalances caused by such events. It reflects the actual operation of the resources participating in the market. Many markets also have intermediate scheduling procedures on the hour ahead or a few hours ahead to facilitate this process in advance of real time when the differing conditions from the DAM are apparent. These markets typically have advisory prices and schedules, but they may have binding commitment directions.

In both day-ahead and real-time markets, suppliers will offer energy bids as a price and quantity pair. In U.S. markets, there is further complexity in supplier offers, which are designed as three-part bids. Due to the non-convexity of costs of many generating resources, the generators submit a bid for (1) incremental energy, (2) no-load cost—i.e., a cost just to be online, or at its minimum generation level, and (3) a cost of starting up the generating unit and synchronizing it to the grid. The generators also submit to the ISO their unit constraints, including how fast they can ramp, how long they must stay online if committed, and other constraints. The market operator will select the least-cost set of suppliers to meet the demand based on these three-part bids and generating unit constraints while also obeying many of the physical power system constraints.

It is important that the average prices of the day-ahead and real-time markets converge, so that market participants should not have a strong preference to be in either market. Virtual trading, or convergence bidding, is used in most RTO/ISO markets to ensure that the prices of the DAM and RTM converge to the same price on average (ISO-NE 2004). Virtual traders will sell or buy energy in the DAM and buy or sell it back in the RTM. They have no requirement to have physical assets to supply or consume energy. By taking advantage in either market when there is a premium in one, they will drive down the difference in prices between these markets. This design feature of the market recognizes the natural tendency of traders to arbitrage across different markets. In the absence of virtual trading, there is potential for a premium in one market that can lead to uncompetitive and inefficient behavior.

In addition to the day-ahead market process, a subsequent process is used, generally referred to as the reliability unit commitment (RUC) process. The day before the operating day, an initial SCUC will solve to meet the bid-in load with bid-in generation and create the schedules and prices for the DAM. These bid-in quantities, in particular the bid-in load or bid-in VG capacity, may or may not be close to reality. To ensure the system has sufficient capacity available, a subsequent SCUC will be solved to meet the RTO/ISO forecasted load. The exact practices vary by region, but generally the RUC will only commit additional resources and will not decommit any resources needed in the DAM. For example, while most markets solve the RUC subsequent to the DAM, the NYISO solves the DAM and RUC iteratively, so that resources committed by the RUC can affect the DAM prices and schedules (NYISO 2009). Most ISOs are now also using the RTO/ISO forecasted VG as part of the RUC process as well.

2.2 Ancillary Service Markets

To support transparent trading of energy in power systems, all transmission providers are required to procure ancillary services. The six ancillary services required in the United States were defined by the Federal Energy Regulatory Commission's (FERC) landmark rule, Order No. 888 on Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities (FERC 1997). This rule was written in response to the restructuring of the utility industry. The functional unbundling of these services was deemed necessary by FERC to ensure that transmission access could be provided in an open and transparent manner. Although the requirement to procure and provide these services is consistent across all wholesale markets, the method of acquisition varies greatly. In non-RTO/ISO regions, these services are obtained and paid for according to a series of FERC-approved rate schedules. In the RTOs/ISOs, most of these services are procured in a competitive manner that is co-optimized with energy markets. Below is a list of these ancillary services and what type of market design the RTOs/ISOs use to procure these services (Ela et al. 2012c):

- **Scheduling, system control, and dispatch**: Provided by the RTOs/ISOs as they schedule and control the resources on their system; not applicable to ancillary service markets.
- **Reactive supply and voltage control from generation service**: Generally supplied as a static cost-based service without any dynamically changing prices.
- **Regulation and frequency response service**: Regulation is typically supplied and priced by dynamic markets in RTOs/ISOs; it is used to correct area control error. However, frequency response, defined as the local droop response of governors autonomously responding to frequency, is generally not included in any dynamic markets nor is it given cost-based rates.
- **Energy imbalance service**: Typically the service of the RTM correcting the imbalance from the forward/DAM markets and therefore a component of the real-time energy markets.
- **Operating reserve—synchronized reserve service**: Typically supplied and priced by dynamically priced markets in RTO/ISO regions.
- **Operating reserve—supplemental reserve service**: Typically supplied and priced by dynamically priced markets in RTO/ISO regions.

In addition to the ancillary services listed above, new forms of ancillary services are being discussed, in part because of the emergence of VG. In previous literature, we categorized all active power control services that are ancillary to energy scheduling, also defined as operating reserve, as shown in Figure 2-2 (Ela, Milligan, and Kirby 2011). The operating reserve types that have existing dynamically priced markets—including synchronized reserve (contingency reserve—secondary), supplemental reserve (contingency reserve—tertiary), and regulation (regulating reserve)—are bought and sold in day-ahead and real-time markets in a similar manner to energy markets. In fact, the U.S. markets that have ancillary service markets currently co-optimize energy and operating reserve when clearing DAM and RTM markets. This means that the markets are cleared simultaneously so that costs and requirements of both markets are considered when clearing the entire market.

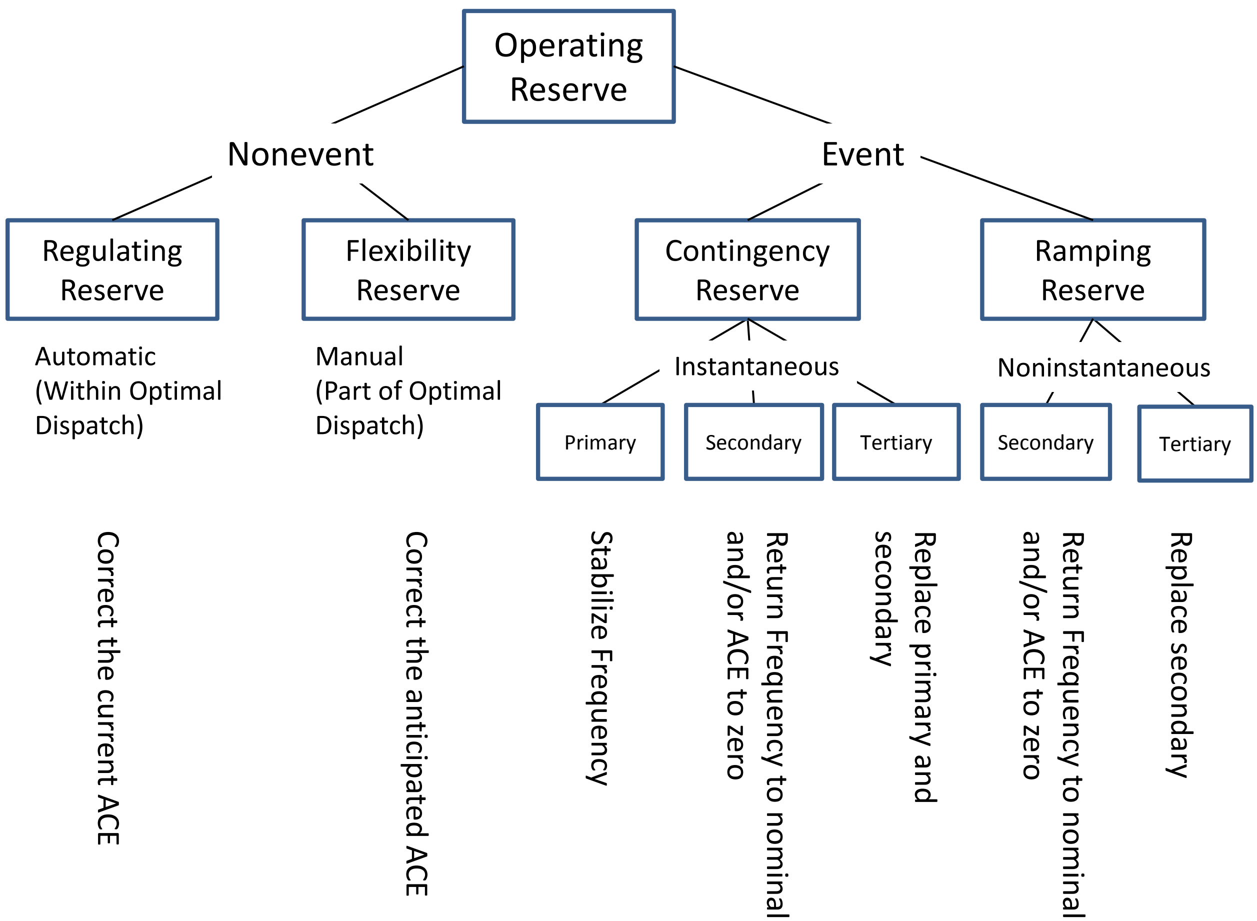

Figure 2-2. Operating reserve types and their uses, adapted from Ela, Milligan, and Kirby (2011)

Other ancillary services are not sold through dynamic markets. For example, reactive supply and voltage control are needed services both during steady state and disturbances. During the initial stages of wholesale market design in the United States, researchers evaluated the potential for reactive power markets (Hogan 1996). There are at least two issues for the implementation of reactive power markets. First, because of the network characteristics, reactive power does not travel far; therefore, the reactive power supplier must be very close to the reactive power need. This is contrary to active power, limits the competitiveness for this service, and can create market power issues. Reactive power prices also require the solution of full AC power flow for the market compared to the simplified linear DC power flow model used in most current markets. The use of AC power flow within the market can be complex and computationally intensive. The complexity may not meet the needs of market participants that require the market results to be solved and published in a timely manner. Therefore, no market in the United States has a reactive power market that solves for dynamic prices. However, most do have mechanisms in place to recoup lost opportunity costs and other costs to reactive power suppliers (CAISO 2009).

Other services are much more long term and are cost based. For example, black-start service is needed from generators for system restoration following blackout events. These resources must be capable of starting without outside power supply, able to maintain frequency and voltage under varying load, and able to maintain rated output for a significant period of time (e.g., 16 hours) (PJM Interconnection 2013; Saraf et al. 2009). Many markets will request black-start service proposals and will then have cost-based recovery mechanisms in place for these resources. Other services such as primary frequency response and inertial response currently lack markets or cost-based recovery mechanisms in many markets, which was detailed in Ela et al. 2012a. Flexibility reserve, for additional ramping requirements to meet increasing levels of variability and uncertainty, have historically not been an ancillary service market either, but, as discussed in Section 4, they are garnering more interest in some markets.

2.3 Pricing Energy and Ancillary Services

Prices for energy and ancillary services are calculated in similar ways throughout all of the restructured regions in the United States. We refer to these prices as LMP and ancillary service clearing prices (ASCPs) for energy and ancillary services, respectively. In U.S. markets, these prices are based on the marginal pricing concept, in which the prices are equal to the bid-based marginal cost to provide each service. Market participant bids are meant to reflect true variable costs, and the marginal pricing design theoretically drives resources to bid their true variable costs.

The energy and ancillary service schedules and prices are determined from the optimization solutions of the market-clearing engine, the security-constrained unit commitment (SCUC) and/or security-constrained economic dispatch (SCED). Typically, the SCUC is used in the DAM, and the SCED is used in the RTM. The LMP and ASCP are calculated from the dual solution of the market-clearing engine, often referred to as shadow prices. The LMP is mathematically represented as the partial derivative of the total costs with respect to the derivative in load demand. It includes components for energy, transmission congestion, and transmission electrical losses (Shahidehpour, Yammin, and Li 2002). For both LMP and ASCP, constrained systems will experience more variations than unconstrained systems. When transmission congestion is present, it causes more expensive resources to be needed on one side of the constraint, because the cheaper units are constrained by the transmission limits. This causes the price to be higher where the expensive unit is needed and lower at the location of the cheaper unit, based on the marginal costs of providing an increment of energy at each location. Typically, the market-clearing engine evaluates how power flows on large systems using a DC power flow that approximates the true AC power flow and provides a good estimate of congestion on the network, which in turn influences both LMP and ACSP. Finally, prices will also be higher at locations that are closer to the load, even without transmission congestion, because the energy injected by generation closer to the load will have fewer transmission losses as it is transported to the load. Therefore, 1 MW is worth more when injected closer to the load than 1 MW located farther away where more energy will be lost on the transmission lines. The level of locational detail will differ between energy and each ancillary service, with LMPs typically varying at the nodal level and ASCPs typically varying zonally or even not at all.

The calculation of ASCP for active power operating reserve is in some ways very similar to the calculation of LMP. We are discussing only the ancillary services related to active power operating reserve, so we ignore voltage control and reactive power support. The ASCP is mathematically represented as the dual value (i.e., shadow price) of the associated operating reserve requirement constraint. It is equal to the total cost increase of the system if an incremental amount of operating reserve is required. The costs associated with operating reserve are the combination of bid-in costs and lost opportunity costs. As discussed, in all United States markets that have ancillary service markets, energy and ancillary services are co-optimized. Lost opportunity costs are the lost opportunity of a resource to make profit in another market, typically the energy market, and are part of the calculated ASCP when energy and ancillary services are co-optimized in the market (Hirst and Kirby 1997). This lost opportunity cost is the LMP minus the resource's bid-in energy cost. Some markets allow market participants to also provide bid-in ancillary service costs, which may be due to the added costs such as wear and tear from providing ancillary services, efficiency reductions, and risk margins. Ancillary service markets will also typically follow a pricing hierarchy (Oren 2001). The hierarchy will price higher quality reserve services that share the same capacity to be greater than or equal to the lower quality service. This is because some ancillary services are more critical than others, and the incentives provide transparency to market participants on which service they should provide. ASCP may also have locational differences when deliverability issues arise. For example, when the transmission system is constrained, the contingency reserve capacity must be located on the receiving end of the constraint, so that when the outage happens on this side, the reserve can be delivered without overloading the transmission system. A different marginal cost reserve provider on each side of the constraint will result in different ASCPs in each location. Hence, ASCP zones typically reflect the major transmission bottlenecks in the system.

Most ASCP payments go to market participants for the provision of capacity to provide ancillary services. The payments usually are not modified based on how the market participant performs the ancillary service, or if the unit was even asked to respond, as long as its performance is satisfactory and the capacity reservation is held (although if deployed, the resource will be paid for the energy deployed with additional energy payments). Recently, there has been motivation to incentivize market participants based on how they performed. FERC Order 755 directs a pay-for-performance scheme for regulating reserve. Resources that provide greater movement and accuracy when providing regulating reserve are compensated more. This is an advantage for participants that can provide regulating reserve faster or more accurately.

Suppliers will be paid the DAM LMP at the DAM energy schedule and the DAM ASCP at the DAM ancillary service schedule. When asked to provide energy or ancillary services differently from the DAM schedule in the RTM, the suppliers will be paid the RTM LMP and RTM ASCP for the difference between the RTM- and DAM-scheduled energy and ancillary services, respectively. In both markets, load pays the LMP and generation is paid the LMP at their corresponding locations.[3] The prices in RTM can change because of changing load, changing VG output, change in committed resource, or change in network topology (i.e., due to transmission outage). The change in RTM prices should incentivize suppliers to adjust schedules accordingly. The introduction of virtual trading (i.e., convergence bidding) should result in the

[3] In many markets, generators are paid the nodal LMP, and loads pay the zonal (load-weighted) LMP.

average prices between DAM and RTM to converge, thereby not leading to suppliers, or consumers, to prefer one market than another.

Another important factor to the pricing of energy and ancillary service prices is the administratively-set scarcity prices. For example, all ancillary service markets in the U.S. markets have demand curves for when ancillary service requirements cannot be met or are very expensive to meet (Isemonger 2009). When this occurs, the ancillary service prices will be set to this administratively set level, which often depends on how much the system is scarce on meeting the ancillary service requirement. These prices can vary from as little as $80/MW-h to well over $1,000/MW-h. Because the ancillary service markets are co-optimized with the energy markets, these prices will flow to the energy market as well, causing price spikes that can induce additional incentives and assist in revenue adequacy for peaking units that require higher energy prices to recover their fixed capital costs.

These pricing methods are designed to incentivize resources to offer their true costs for energy and true capabilities for ancillary services. The RTO/ISO is responsible for solving an optimization problem to minimize the total costs to meet the energy and ancillary service demands while also meeting numerous generation and reliability constraints. This schedule should place each market participant in a position to make the most amount of profit given the prices generated by the market-clearing engine. However, because of issues such as non-convex costs, commitment constraints, and out-of-market reliability rules, it is possible for the RTO/ISO to direct a market participant to provide energy and ancillary services that cause that market participant to lose money. When this happens, the RTO/ISO provides a make-whole payment to ensure that the market participant does not receive a negative profit. After actual power data is measured, resources are paid this make-whole payment in addition to the scheduled payments. Sometimes penalties are in place for market participants that stray too far from their directed energy or ancillary service schedules. These vary depending on the market region but give further incentives to ensure reliable operation.

2.4 Financial Transmission Rights Markets

FTR markets, also called transmission congestion contracts and financial congestion rights, are markets designed to hedge the volatility in locational differences of energy pricing (Lyons et al. 2000).

When the transmission system is congested, the load at the receiving side of the constraint would typically pay more for energy than the generators supplying energy at the sending side. This difference is allocated to the FTR holders between the two locations. These FTRs are not part of system operation, because they are purely financial and do not affect the objective of the system operator to dispatch the supply at least cost. Bilateral agreements between supply and demand at different locations can avoid the volatility of pricing between their locations with the purchase of FTRs.

Before FTRs, physical transmission rights existed. Physical transmission rights are an exclusive right to transport energy between two locations on the transmission network. The owner of physical transmission rights has exclusive rights to the capacity on the transmission lines between two locations. The owner can also sell the rights to use those lines to others, allowing the owner to earn additional revenue when demand for transmission capacity is high. Thus, a major concern about physical transmission rights is the conflict between the physical transmission rights owner and the system dispatcher, because the system dispatcher aims to minimize the cost of operating the transmission system by dispatching the least-cost resources. A second concern about physical transmission rights is that they allow for the exercise of market power. The ability to control capacity on specific transmission paths can allow the physical transmission rights owner to influence LMP, because they can deny transmission capacity to cheaper resources, causing uncompetitive prices in those locations. This led to the need to utilize FTRs in the restructured regions that use LMP, although physical transmission rights are still in place in areas of the United States that are not in restructured markets, like most of the Western United States except California.

Market participants can obtain FTRs through an RTO-specific allocation process and auctions. Initial FTR allocations are based on historical usage and entities that fund the construction of new facilities. FTRs are typically auctioned at annual, seasonal, and monthly periods. They can also be traded bilaterally. Each auction can include new potential buyers and sellers of FTRs, and it will include a market-clearing engine similar to the one used in the energy market, in which the objective is to minimize the cost of all FTR bids while incorporating the network security constraints. The pricing that results from the FTR auction is performed in a very similar manner to the prices of energy, where in this case the marginal cost of transmission is paid to the seller and taken from the new buyer. Many other characteristics can be included in the FTR market (Sarkar et al. 2008). FTR options are rights in which the owner earns only the locational difference in energy prices if that difference is in their favor. Some markets will have FTRs that are different for on-peak and off-peak periods to signify the differences in transmission flows between these periods. Other areas also have multi-round auctions, in which each round will sell only a portion of the available transmission capacity to FTR purchasers. This is said to make the FTR market more flexible and competitive and allows for the market participants to adjust the bids each round after learning the results from the previous round.

The revenues that FTR holders receive when they own the rights are typically through the congestion costs that occur in the DAM rather than the RTM. The more the prices differ between the DAM and RTM, the more that FTRs may not reflect the true cost of congestion. The congestion patterns are well understood in most markets, although on-peak and off-peak times and transmission outages can certainly affect the outcomes differently than anticipated during the auction periods. Also, at the onset of FTRs it was thought that they could promote future investment in new transmission, but there is a lot of argument about whether FTRs provide sufficient incentives for transmission investment (Oren et al. 1995). How these markets may evolve in the future is still very unclear, as is the impact that higher penetrations of VG have on them. However, the scope of this report has only marginal relevance to FTR markets and so we provide little focus on this market product.

2.5 Capacity Markets

Capacity markets are motivated by the desire to employ a market mechanism to ensure that new generation is developed on time to meet resource adequacy targets and help these resources recover their capital costs. Power plants are large, capital-intensive resources that take considerable time to permit and build. The decision to build a power plant must be made well before the plant is needed. Some RTO/ISO regions rely on high and volatile energy prices that are sometimes constrained by administratively-set scarcity prices or price caps. Other RTO/ISOs operate explicit capacity markets to ensure that sufficient generation will be available to meet the expected load. In vertically integrated systems, resource adequacy assessments are carried out by the utility, and any needed additional capacity could be acquired internally or via contract, subject to regulatory oversight. The costs for procuring that capacity are typically subject to rate-making proceedings with state public utility commissions.

This section briefly focuses on mandatory capacity markets that are intended to address long-term reliability needs and ensure that resources have adequate opportunity to recover their variable and fixed costs over time. Capacity markets are often backstop mechanisms that evaluate potential capacity shortfalls after considering bilateral contracts or other power purchase agreements. In ISO-NE, NYISO, and PJM, mandatory capacity markets are characterized by (1) an obligation for load-serving entities to have sufficient capacity to reliably serve load; (2) a methodology to determine a capacity reserve margin and future capacity needs both for subregions within the RTO/ISO and the entire RTO/ISO; (3) a process for soliciting qualified supply and demand resources to meet future capacity needs; (4) a benchmark to judge the cost of new capacity; (5) a methodology or approach for creating a demand curve; and (6) a process to select resources and determine a capacity price (Rose 2011).

The methods for calculating capacity prices in each of the RTO/ISOs are based on the market design choices of each region. In general, regions with capacity markets find that the capacity prices tend to be limited to the capital cost of a new gas-fired plant that can be sited and built within three years (FERC 2012). As shown in Figure 2-3, prices generated by mandatory capacity markets have been considerably volatile (FERC 2013). These results are driven by a variety of market considerations that vary from one region to another. In Section 3, major components of these markets are discussed further.

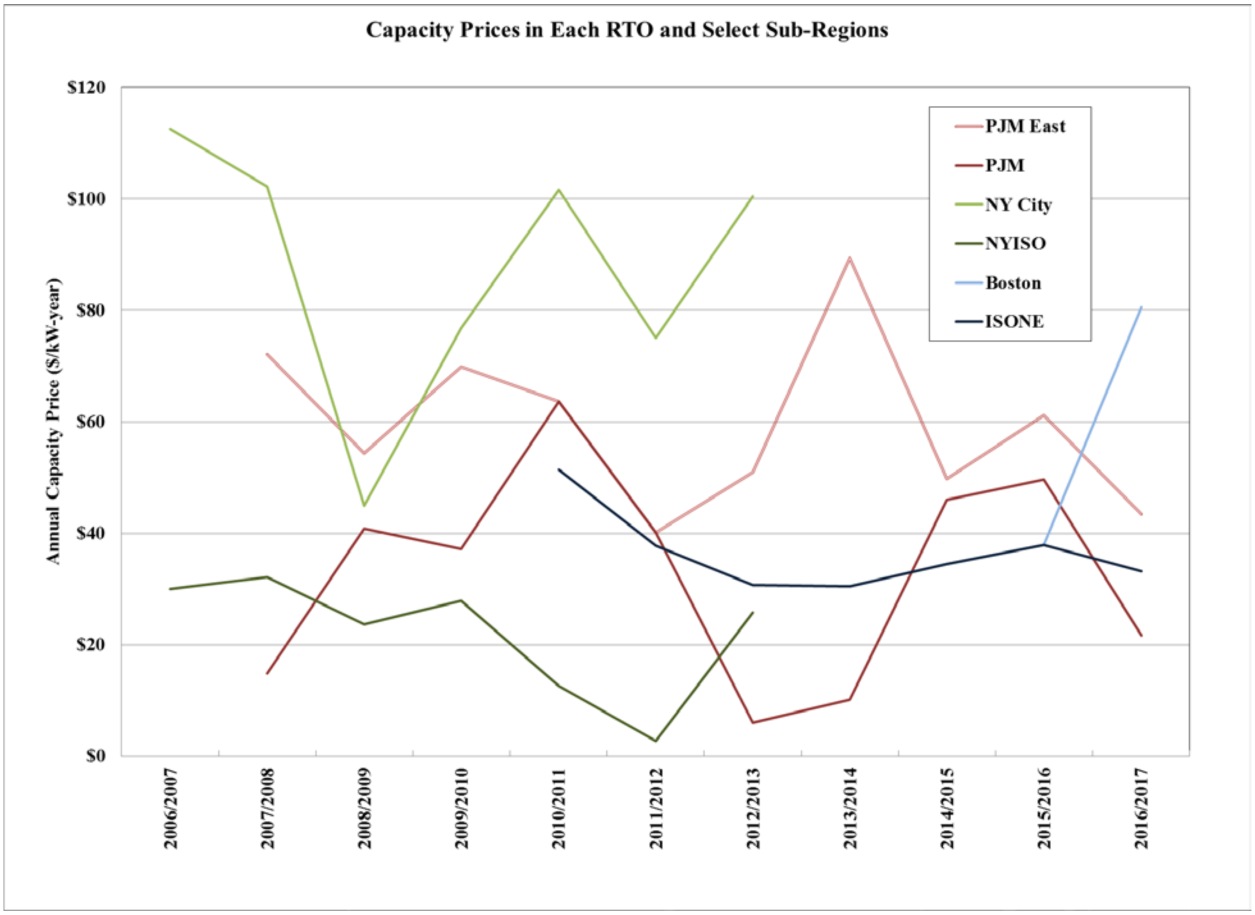

Figure 2-3. Capacity price in RTO/ISOs (FERC 2013)

The demand for capacity is based on an administrative process that determines the total amount of capacity necessary to meet peak load requirements. NYISO, PJM, and ISO-NE all use a downward-sloping demand curve for capacity rather than a fixed target. The downward-sloping demand curve is constructed to reflect the marginal value of capacity to load, and it serves to reduce the potential exercise of market power in capacity auctions. Although the specific demand curve parameters vary between the markets, the main principles are illustrated in Figure 2-4. The curve is constructed around a target for new capacity at which the price is set equal to the cost of new entry (CONE). CONE is typically set equal to the annualized capital cost of a new peaking plant (e.g., a combustion turbine), and it may be adjusted for the expected revenue from the energy market (i.e., net CONE). Administered price caps are common and are designed to protect against potential market power and provide a backstop mechanism in case insufficient bids are received from the market.

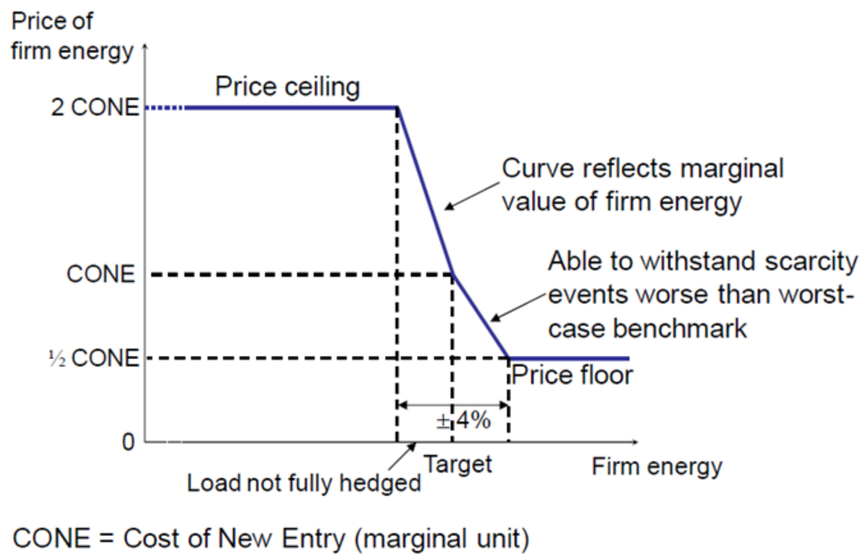

Figure 2-4. Illustration of demand curve for capacity. ***Image from Crampton and Ockenfels 2012***

Resources participating in the capacity markets must verify their capabilities to determine the total capacity they can bid into the market. Each of the mandatory capacity markets has a process for qualifying as a capacity resource. Generally speaking, resources interested in participating in capacity markets must verify their operating capability in MW for a specified time period, usually the winter or summer peak. Each organized market has different capacity qualification rules for existing resources, new resources, external resources, demand response, and renewables. Many of the markets will require capacity market resources to offer their capacity in the day-ahead market. Current capacity markets typically do not require capacity resources to have specific attributes other than the provision of capacity during periods of peak demand.

The physical location of a resource is also important for capacity markets. Transmission limitations can limit the ability of a load to access a resource. Local capacity obligations are enforced in each of the markets to ensure that load-serving entities have adequate supply and transmission capacity to deliver energy to an area. The issue is most prevalent in regions with constrained export and import capabilities. Accurately identifying zones that have deliverability constraints is critical to developing efficient capacity markets.

2.6 The Impacts of VG on Wholesale Market Outcomes

The outcomes of each of the markets discussed above may be impacted by the introduction of high penetrations of VG. Possible impacts are briefly discussed in the list below.

- Energy markets
  - VG can reduce average LMPs because of its low variable costs.
  - VG can cause more occurrences of zero or negative LMP periods because of its variable cost and zero or negative bid-in costs.

- VG's increased variability can cause LMPs to be more volatile from one time period to another.
- VG's increased uncertainty can cause greater differences between DAM and RTM LMPs (although on average they are likely to remain converged as a result of virtual trading).
- VG can cause a greater need for flexible resources in the energy market, and the energy market may or may not provide sufficient incentive for this flexibility.

- Ancillary service markets
  - VG can increase the requirements for normal balancing reserve, such as regulating reserve, which can increase the ASCP for those services.
  - With higher balancing reserve demands and increased variability and uncertainty, administratively-set scarcity ASCP may be triggered more often, resulting in more frequent extreme price spikes.
  - VG can displace synchronous, frequency-responsive resources, and when not equipped with technology to provide a comparable response, it can cause the need for supplemental actions or market designs to ensure that sufficient frequency response and/or system inertia is available.
  - VG can cause the ancillary service requirements to change from one day to another and from DAM to RTM, if the requirements are based on correcting the variability and uncertainty of VG, which can cause uncertainty in ancillary service demands and changing demands for the same time periods between DAM and RTM, similar to load.
  - VG can cause a need for greater flexibility from the resources that correct for its variability and uncertainty. Certain forms of flexibility may or may not be built into the current ancillary service markets.
- FTR markets
  - VG's increased variability and uncertainty can cause greater variation on power flow, which causes FTR holders to be uncertain about expected congestion patterns.
  - VG's increased uncertainty can cause greater deviations of power flows between DAM and RTM. Because FTR revenues are typically based on the DAM, there could be greater divergence between FTR revenues and actual congestion patterns.
- Capacity markets
  - The reduction in LMP and energy schedules from conventional resources will result in reduced revenues in the energy market. If these resources are still required to be available for short periods of time, more resources become capacity-based rather than energy-based.
  - VG's variability and uncertainty can cause the need for different types of resources to be built and available. In other words, it might require the need to

plan and build more flexible resources to prepare for future needs and not to focus on the need for MW capacity alone.

- VG's variability and uncertainty can cause the need for existing resources to modify their flexible capability potential. Market designs may need to incentivize the existing resources to spend the capital on retrofits to increase the flexible capability that it can provide.
- Must-offer price rules, designed to limit the ability of buyers to suppress capacity prices by subsidizing relatively higher-cost new capacity to replace lower-cost existing capacity, may increase risk that a resource built to satisfy a state renewable portfolio standard will not clear the capacity market at the applicable minimum offer floor.

These topics are not necessarily concerns about the current market design, but rather theories on how increased penetrations of VG may impact the market outcomes when compared to today's outcomes. From this exploratory list, we focus on two specific topics. First, we explore if the current market design ensures that resources required for long-term adequacy earn sufficient revenue to cover both variable and capital costs. Second, we discuss if there are sufficient structures in place in the current market designs to ensure that the resources necessary to balance the variability and uncertainty of the system in real-time are available and used efficiently. These two topics are the focus of the rest of this report.

3 Long-Term Resource Adequacy, Long-Term Flexibility Requirements, and Revenue Sufficiency

In this section, we focus on two related topics: (1) resource adequacy, including newer methods of determining adequacy metrics, and (2) revenue sufficiency and how existing and evolving market designs may enable resources to retrieve sufficient revenue to ensure long-term resource adequacy. The focus here is on the investment time horizon and the installation of sufficient generation capability. Operational issues, which are closely related, are addressed in Section 4. The topics of resource adequacy and revenue sufficiency—the process of determining the quantity and acquiring that quantity of capacity that will be needed at some future date and ensuring that those resources that offer the capacity receive sufficient revenue to recover their costs—are the focus here. Resource adequacy is generally based on one or more metrics that quantify the long-term reliability of the generation supply (and possibly demand-response resources) and its ability to meet load. When sufficient capacity is acquired—whether through a market, payment, or other incentive, or through a direct regulatory process—resources must have an opportunity to earn sufficient revenue to remain in the market. Without revenue sufficiency to recover both variable and fixed costs, it is likely that resources would retire from the market, potentially compromising long-term reliability. Also, as more VG is brought online, there is a need for greater flexibility attributes from new and existing market participants and technologies. This brings a new dimension to the traditional resource adequacy question of whether there is enough capacity and adds the question of whether there is sufficient flexibility within that capacity.

Several alternative approaches are available to solve the problems that we identify in this section. Rather than recommending any single approach, we present an overview of these alternatives. First, we discuss the issues that make ensuring long-term resource adequacy and revenue sufficiency in electricity markets challenging. Then we discuss the current mechanisms for ensuring resource adequacy and the importance of revenue sufficiency. Next, we discuss how the increased penetrations of VG, with its diurnal and seasonal availability patterns, high capital costs and low variable costs, and its increased variability and uncertainty can change the methods and needs of resource adequacy and revenue sufficiency. We present the historical designs that U.S. wholesale electricity markets have used to address these issues. Finally, we present a review of the most recent market design changes to address resource adequacy and revenue sufficiency with a focus on how they are evolving to meet the needs due to increased VG.

3.1 Challenges to Ensuring Long-Term Reliability in Electricity Markets

In a perfect world, reliability would be bought and sold in a competitive market. This market would feature consumer choice, as exercised by individual demand curves that would be aggregated to the market level, and supply curves that would result in economically efficient market equilibrium.[4] In this perfect world, there is no market power (among buyers or sellers), no free riders, and consumers are free to buy as much (or as little) reliability as they desire at the

[4] An outcome is economically efficient if it (a) provides a mix of goods and services—including energy and reliability—that society values more highly than any other achievable combination and (b) cannot be provided by other means at a lower total cost.

market price. However, as is well known, most retail consumers purchase electricity via administered prices that are most often a characterization of average total cost plus an administered profit rate. Most consumers are thus insulated from price swings that are, or would be, a function of the relatively volatile cost and marginal wholesale prices of electricity at the bulk system level. In addition, electrical neighbors who wish to purchase different levels of electric reliability must instead purchase the same amount because there is currently no way to differentiate reliability among customers on the same feeder. These issues—that consumers are insulated from actual time-sensitive prices and consumers cannot choose the level of their individual electric reliability—are the two primary demand-side flaws that impact the way that electricity markets are designed.

To be economically efficient—by providing the level of product desired by society at the lowest cost—markets must have functioning supply and demand functions. Market equilibrium is achieved when the plans of buyers coincide with the plans of sellers. In a free market, buyers can choose whether or not to purchase the product at the market price. Price is related to cost, although this relationship may be complex. However, the fundamental principle is that if cost, and therefore price, were to increase, at least some consumers would withdraw from the market.

Well-functioning markets also allow for producers to differentiate among customers—i.e., a customer who is not willing to pay a given price does not receive the product. In the case of electric reliability, it is usually not technically or economically possible to differentiate levels of reliability to customers (especially residential customers) who may be willing to pay more for reliable service—or, conversely, customers who may be willing to receive a lower level of reliability in return for lower prices.

Thus, resource adequacy is not based on a true market outcome. Instead, in most cases it is based on a long-term reliability standard defined by policy. Instead of prices that ration electricity usage, a somewhat-arbitrary reliability standard is introduced along with administered pricing rules that have the effect of muting most forms of price-response from most consumers. There is yet another complication: in some regions, there is only an approximation of a reliability standard, known as a planning reserve margin (PRM). The PRM is usually defined as the percentage by which installed capacity exceeds annual peak demand. In some cases, there is an adjustment to the PRM or to installed capacity that allows for some consideration of the reliability target. This is discussed in more detail in the next section.

In general, markets require the ability of participants to define the product, its quantity, and price. Price is determined by the intersection of the demand and supply curves of buyers and sellers. Reliability targets in the power system exist because the two demand flaws (Stoft 2002) of the market for electricity prevent markets from functioning effectively. Thus, bridging the gap between reliability and electricity markets is challenging. Fundamentally, there are several requirements, including the following:

- A method for choosing and assessing the resource adequacy target
- Determining whether the resource adequacy target has been, or will be, achieved
- Determining the contribution of each entity toward meeting the resource adequacy target

- Utilizing the right time horizon for meeting resource adequacy targets (e.g., 1 year ahead, 3 years ahead, etc.)

When the resource adequacy target is achieved, will the energy and ancillary service markets result in revenue sufficiency? If not, implement measures to ensure the long-term viability of resources that are needed to achieve reliability or other objectives.

In systems with significant levels of VG, additional questions may need to be addressed:

- How do the resource characteristics of VG influence reliability calculations and resource adequacy targets?
- Does the capacity have the right flexibility attributes to effectively handle the increased variability and uncertainty characteristics of VG in grid operations?
- How does VG itself contribute toward the required capacity adequacy requirements?

3.2 Achieving Long-Term Resource Adequacy and Revenue Sufficiency

As described above, because of the limited price elasticity of demand and the inability of suppliers to curtail load based on consumers' reliability preferences, and also because of the length of time involved to build new supply resources, the combined wholesale and retail electricity markets will not function like other commodities markets. Resource adequacy must be considered to ensure that the electricity supply is sufficient to serve load that appears as an inelastic demand. Determining whether resource adequacy targets are achieved is a probabilistic problem. Well-known methods exist and are based on loss-of-load probability (LOLP) and related metrics, which are briefly described in the accompanying text box: Resource Adequacy Metrics. PRM, the ratio of installed capacity to peak demand, does not directly address resource adequacy. This is why the existing capacity markets in the United States perform some type of mapping to a reliability-based metric. In nonrestructured areas, there is a mixture of whether and how PRM and reliability-based metrics are used. In this section, we show why the PRM, by itself, is an inadequate tool for measuring reliability and why a probabilistic reliability-based metric is a more rigorous reliability target. This is why some markets derive a PRM from a probabilistic assessment. LOLE, or a related reliability-based metric, is essential to ensuring that the long-term supply is adequate, and it also gives a meaningful way of determining whether resources are needed for long-term reliability and whether they should be incentivized to remain in the market.

Because of the reasons described above, including the large capital costs for generation, effective inelasticity of demand, and regulatory price caps in the wholesale markets, the level of revenue and profit determines whether a resource remains in the market. A resource that is needed to maintain a reliable and secure power system must earn enough revenue to recover both its variable and fixed costs. This concept of revenue sufficiency is the second focus of this section.

Resource Adequacy

Several metrics describe the reliability of generators. Although these metrics are related, they convey different information. The choice of which metric to use depends on the type of modeling performed and the judgment of the practitioner.

EFOR – equivalent forced outage rate. EFOR = (FOH + EFDH) / (SH + FOH), where SH = service hours, FOH = full forced-outage hours, and EFDH = equivalent derated outage hours. This rate describes the unit's failure rate regardless of whether it is needed by the system.

EFORd – equivalent forced outage rate demand; the probability that the unit will fail (partially or completely) when needed. EFORd = (FOH * f + EFDH * p) / (SH + FOH * f), where SH, FOH, EFDH are defined as above, f = full f factor, and p = partial factor.

XEFORd – same as EFORd, except XEFORd excludes the impacts of outages caused by events outside management control.

Resource adequacy is a function of the demand forecast and aggregate generator availability, which is in turn a function of total capacity that is not experiencing a forced outage, as described by one of the metrics above. Several system-wide metrics can be used to describe overall resource adequacy:

LOLP – loss of load probability; the probability that there will be insufficient generation at some point in time to meet load. Depending on differences in how LOLP is calculated and interpreted, it may also describe the probability of not meeting demand plus operating reserve, or the probability that neighboring emergency imports may be needed. By definition, all probabilities lie in the range of 0–1.

LOLE – loss of load expectation. This is generally LOLP multiplied by a unit of time. Common uses of LOLE are measured in days/year of outages. Thus, 1 day/10 years, a common target for resource adequacy, is a LOLE, not a LOLP. LOLE is often, but not exclusively, referred to in units of days/year or days/10 years. LOLE counts only the number of shortfalls and does not differentiate the size of the shortfall. A common target for LOLE that has been used for many years is 1 day/10 years.

LOLH – loss of load hours, which is a form of LOLE

EUE – expected unserved energy. This metric does not count the number of occurrences of inadequacy, but it does capture the energy component of inadequacy.

ELCC – effective load-carrying capability. This metric describes the contribution that a given generator or group of generators makes toward resource adequacy. In general, thermal generation will have an ELCC value, measured in MW, that is relatively close to the rated capacity. The ELCC can be approximated by the unforced capacity of the unit, calculated as UCAP = (1 – EFORd) x Capacity. Variable generation ELCC will generally be somewhat or significantly lower than nameplate capacity and is a function primarily of fuel availability (wind or solar irradiance).

3.2.1 Resource Adequacy Calculations

A common reliability target in the United States is a LOLE of 1 day in 10 years. This means that generation supply is sufficient nearly all the time. Alternative targets can be adopted if desired, and the choice of the reliability target is largely one of policy. In modern interconnected systems, it is likely that the LOLE of any one particular area is overstated because if there is insufficient generation within a given balancing authority area, emergency imports may be available from neighboring systems as long as transmission capacity is available. In any case, using reliability metrics provides information regarding how often and/or how much of a generation shortfall might exist.

Reliability models are used by system planners to calculate the LOLE or similar metric for existing or future system configurations. If the calculated LOLE value is higher than the target, alternative resources can be added until the actual LOLE matches the target LOLE. Traditionally, LOLE was calculated using a single data point per day, chosen from the peak hour of the day (Astrape Consulting 2013). To calculate the daily LOLE, each generating unit's capacity and forced outage rate (FOR—the probability the unit would be in an unplanned outage state) are used in a mathematical convolution with forecast demand values (Billinton and Allan1996). This approach explicitly considers each of these data points to the contribution of a generator to meeting load on a statistically-expected basis. For example, a 100-MW unit with an FOR of 0.10 would have a higher statistically expected output than another 100-MW unit with a 0.20 FOR. Thus, the convolution of multiple units with differing capacities and FORs forms the basis of the reliability calculation and related metrics. Using this simple example, the first generating unit would be considered to have 90 MW of unforced capacity (UCAP) [100 MW x (1 – FOR)].

Additional metrics may be used instead of LOLE. A relatively simple extension is to apply the basic LOLE convolution algorithm to hourly data instead of to daily data. Interest is increasing in evaluating the performance of this hourly metric, expected loss-of-load hours (LOLH), with the increasing levels of VG that behave differently than more conventional thermal generation.

The relationship between LOLH and LOLE (measured in days) is not straightforward. Unless otherwise specified, LOLE will be taken to mean a LOLE measured in days. The traditional LOLE calculation accounted for whether an outage might occur in a given day, given the LOLE. Because no hourly data were used in the calculation, there was no way to know, or to calculate, how many hours within the day an outage might occur. Generally, the calculation assumed that if an outage occurred in a given day, there could be anywhere from one hour to many hours of outage. The LOLE days did not have that information. Conversely, LOLH explicitly calculates the number of hours in a year in which there may be insufficient generation supply. For further discussion on the differences between LOLE days and LOLH. From this discussion, we can conclude that an LOLE of 1 d/10 y is not the same as an LOLH of 2.4 h/y.[5]

Another, simpler approach to measuring resource adequacy is the use of the PRM, often expressed as a percentage of capacity above the forecasted peak demand. Because the PRM

[5] For example, SPP uses 2.4 h/y, which results in a lower reserve margin than 0.1 LOLE (see Astrape Consulting 2013).

includes only data regarding capacity and ignores data regarding forced outages, it is easy to see that PRM is not fundamentally a reliability metric—it cannot distinguish between two systems with the same installed capacity and peak loads but with different FORs.

Milligan and Porter (2008) provide an example in a parametric study of a system based on the CAISO system. Increasing the FOR on a subset of the thermal units and simultaneously increasing the number of new 100-MW units to maintain a 0.1 d/y reliability target, they found that the PRM required to maintain reliability increased from approximately 15% to nearly 24%, as illustrated in Figure 3-1.

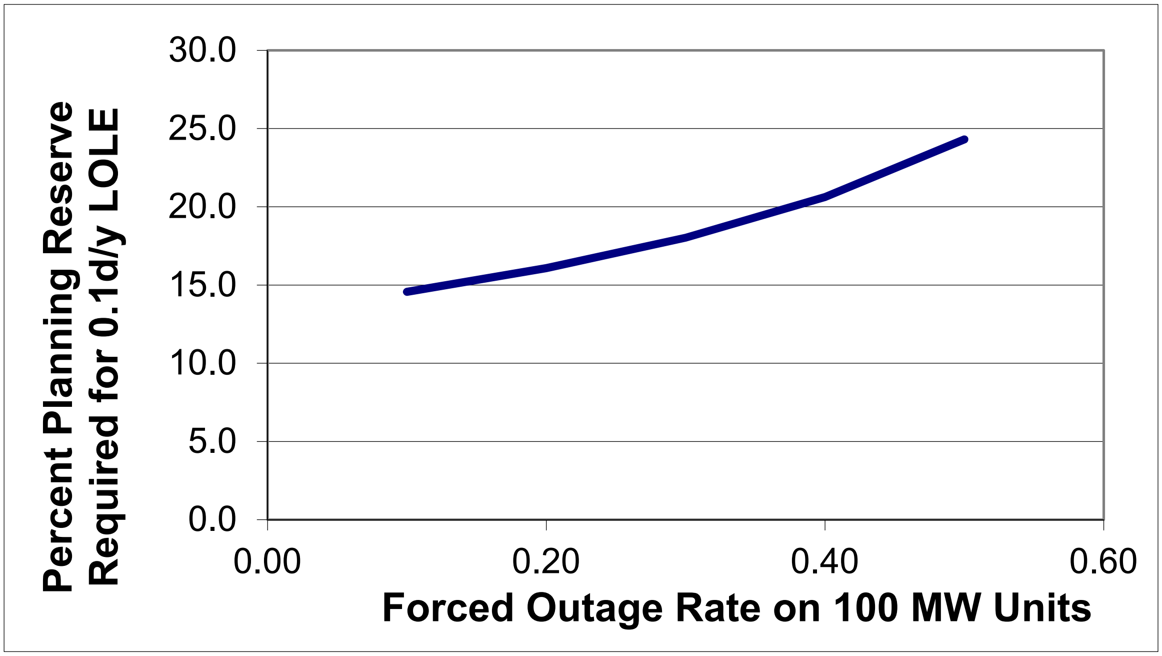

Figure 3-1. PRM to achieve 0.1 d/y LOLE is a function of FORs (Milligan and Porter 2008)

The implications of this analysis seems clear: when establishing resource adequacy targets, the PRM metric may not prove to be very useful, especially in systems with large penetrations of VG (or any other generation that has a relatively low ratio of capacity value to installed capacity). Although it is possible to calculate the PRM that would result in a 0.1 d/y LOLE target, which is done by some markets, the value of the PRM metric is still questionable because it can no longer be used to compare different systems, nor does it provide consistent information regarding resource adequacy. Common values for a PRM on historical systems range from approximately 13% to 18% (North American Electric Reliability Corporation, or NERC, 2013).

The contribution of any resource, or group of resources, to resource adequacy can be calculated using the effective load-carrying capability (ELCC) method, which is built on one of the more fundamental reliability metrics, such as LOLE, LOLH, or expected unserved energy (EUE). This approach can be applied to conventional generation, and it can also be applied to VG. Details can be found in Keane et. al 2011 and NERC 2011. The ELCC represents the additional load that can be supplied by the resource being evaluated, holding long-term reliability constant. Historically, this set of calculations has been computationally demanding, so alternative approaches have been developed to rate individual generators. These simplified methods, however, should be benchmarked against the full reliability approach to ensure that they are reasonable (Keane et al. 2011, Milligan and Porter 2008, NERC 2011, Rogers and Porter 2012).

For example, a 200-MW gas unit with an FOR of 0.10 would have an ELCC of approximately 180 MW. A 200-MW wind power plant with a capacity factor of 35% might have an ELCC of 30 MW, or 15% of its installed capacity. Note that this example points out a fundamental difference in the ratio of capacity value, as measured by ELCC, to installed capacity when resource types are compared. We discuss the implications of this in more detail later in this section.

The ELCC calculation is graphically illustrated in Figure 3-2. The example uses a target of 1 d/10 y LOLE. The left curve shows the relationship between the level of peak load that can be served and the LOLE. At the target of 1 d/10 y, a 10-GW load can be served, and as the curve shows, a lower load will have a higher reliability level and a higher load would have a lower reliability level. When a new generator is added to this system, the reliability curve shifts to the right, and the distance of this shift depends on a combination of system and generator attributes. The example diagram shows that the additional load that can be served while maintaining the 1 d/10 y level of reliability is 150 MW; thus, the new generator has a 150-MW capacity value.

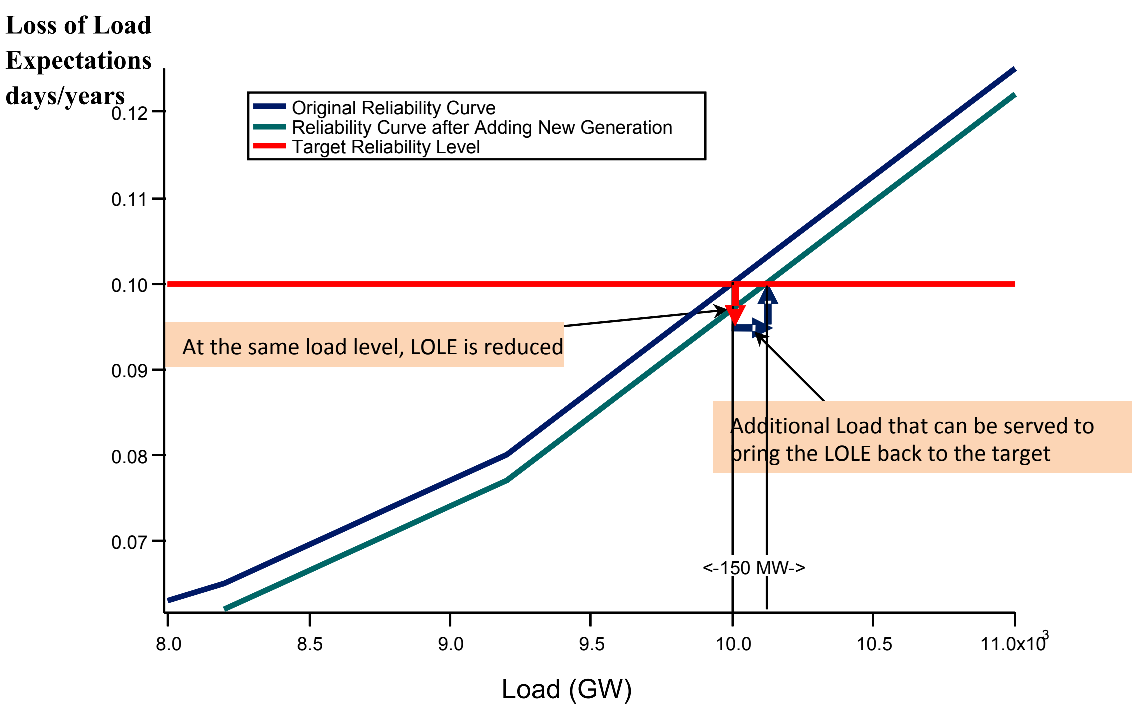

Figure 3-2. ELCC is the horizontal difference between the reliability curves evaluated at the target reliability level. *Illustration from NERC*

The usual mathematical formulation for LOLE is based on the daily or hourly estimates of LOLP.[6] The LOLE is the sum of these probabilities, converted to the appropriate timescale. The discussion that follows is based on the calculation of the LOLE in terms of days/year; however, the same procedure can be applied to the LOLH calculation with minimal changes. The annual daily LOLE can be calculated as:

$$LOLE = \sum_{i=1}^{N} P\left[C_i < L_i\right] \tag{1}$$

where *P()* denotes the probability function, *N* is the number of days in the year, C_i represents the available capacity in day *i*, and L_i is the daily peak load. To calculate the additional reliability that results from adding VG, we can write *LOLE'* for the *LOLE* after the new capacity is added to the system as:

$$LOLE' = \sum_{i=1}^{N} P\left[(C_i + g_i) < L_i\right] \tag{2}$$

where g_i is the power output from the generator under evaluation during hour *i*. The ELCC of the generator is the additional system load that can be supplied at a specified level of risk (LOLP or LOLE).

$$\sum_{i=1}^{N} P(C_i < L_i) = \sum_{i=1}^{N} P\left[(C_i + g_i) < (L_i + \Delta C_i)\right] \tag{3}$$

Calculating the ELCC of any generator amounts to finding the values ΔC_i that satisfy equation (3). This equation states that the increase in capacity that results from adding a new generator can support ΔC_i more MW of load at the same reliability level as the original load could be supplied (with C_i MW of capacity). To determine the annual ELCC, we simply find the value ΔC_p, where *p* is the hour of the year in which the system peak occurs after obtaining the values for ΔC_i that satisfies equation (3). Because LOLE is an increasing function of load, given a constant capacity, we can see from equation (3) that increasing values of ΔC_i are associated with declining values of *LOLE*. Unfortunately, it is not possible to analytically solve equation (3) for ΔC_p. The solution for ΔC_p involves running the model iteratively for various test values of ΔC_p until the equality in equation (3) is achieved to the desired accuracy. Historically, this calculation was considered to be computationally expensive, and many simplified approaches were developed as shortcuts (Garver 1966). However, modern computers can easily manage the computations in a short amount of time.

---

[6] This discussion is based on daily LOLE but can be easily adapted to LOLH.

In modern interconnected power systems, it is likely that during emergency events, such as generation outages, neighboring systems can provide emergency capacity provided there is an unconstrained transmission path and operational procedures in place that allow this response. This may fundamentally alter the LOLE calculation, and this issue was recognized by the Integration of Variable Generation Task force convened by the North American Electric Reliability Corporation (NERC). The task force recommends full transparency in reliability assessments with regard to the way interconnected systems are modeled (NERC 2011). Examples of the impact that transmission plays in resource adequacy have been shown in recent research (EnerNex Corporation 2011; Ibanez and Milligan 2012). Transmission can enhance long-term reliability, and in some cases it can reduce the need for generating capacity (Ibanez and Milligan 2012).

Capacity contributions of any generator will be subject to interannual variations, although the properties of this variability will differ among technologies. As an example, a thermal plant with an ELCC of 90% to 95% of its installed capacity, can experience a forced outage event during high-LOLE peak periods, and could conceivably contribute nothing toward meeting load in that year. Similarly, wind and solar generation is a function of the weather and thus may vary from one year to another around the long-term value. More details on LOLE and ELCC for systems with VG can be found in NERC (2011).

Resource adequacy is measured in terms of a reliability metric, as discussed above. However, there is no market (or market characteristic) for reliability. Because two otherwise identical plants with different FORs have different ELCC values, utilizing installed capacity as the metric for measuring resource adequacy will result in at least a small, but possibly a relatively large, divergence from the goal of resource adequacy. Conversely, if a target level for installed capacity is utilized, no information about FORs are incorporated in the market. The simplified methods for calculating wind capacity value only include reliability information insofar as the input data represents the time periods of high risk—high LOLP—and that data time series is sufficiently long.

One way to improve this approach is to utilize UCAP. This is done by each of the existing capacity markets in the United States—PJM, MISO, and NYISO—although the mechanisms are different (The Brattle Group and Astrape Consulting 2013). The various capacity auctions are conducted so as to account for UCAP, even though the market may directly address installed capacity. This means that all capacity acquisitions via the market account for the units' contribution to system reliability and do not simply account for installed capacity. The specific UCAP value will vary based on resource type and by region. This allows two plants of the same size but with different FORs to be differentiated. We discuss the potential treatments of VG in Section 3.3.1.

Calculating resource adequacy for a future time period is complex. Resource Adequacy targets are predicated on demand forecasts for future time horizons, and the likelihood of over- or underestimating demand as these time horizons increase. The contribution of different types of resources—including demand response, for which characteristics may not be well-known today—can be challenging to ascertain. The future is always uncertain. Is it more appropriate to develop a range of targets using multiple scenarios? Can they be combined to form a stochastic expected value, or evaluated subject to a probabilistic hedge? Are there other strategies such as minimizing maximum regret?

3.2.2 Revenue Sufficiency

The missing-money problem (Stoft 2002) is the concern that even robust energy markets may not provide sufficient revenue for at least some generators to earn sufficient revenue to pay for both variable cost and fixed costs. It predates large amounts of VG. It occurs fundamentally because of the market failures described in Section 3.1 and the concern that insufficient revenue will result in insufficient installed capacity to serve the load, especially during peak periods, because there is insufficient incentive for generators to build new capacity or even maintain existing capacity. In particular, limited demand-side participation may lead to inadequate scarcity pricing that suppress revenues from the energy market.

Other technical characteristics of electricity markets will impede them from functioning at or near the level of an ideal market. The two demand-side flaws of (1) the inability for buyers to respond to price and (2) the inability for suppliers to supply different levels of reliability imply that there will be some degree of market failure unless a clever approach can be discovered to overcome these obstacles.[7]

Therefore, the market for electricity does not possess the necessary characteristics to perform as a perfectly competitive market; however, markets for specific electricity products, such as bulk energy and ancillary services, can still perform well if there are many buyers and sellers, along with limited congestion and limited market power (i.e., increasing the level of competition). These markets still suffer because the consumer usually cannot respond to price. The system operator is then constrained to provide energy up to an administered level of reliability. When reliability is compromised, or if the operator has concerns that it might be, the threat of penalties or actual reliability events can result in paying high prices for electricity that may exceed the value of the energy from some consumers' point of view. Additionally, it is also well-known that outages are very costly, therefore a strong argument can often be made that the system operator must incur high purchase costs to avoid outages. Thus, the value of lost load (VOLL) has been estimated as high as $77,000/MWh, although there is a wide variation in estimates for different consumer groups and regions (London Economics 2013).

In the early periods of market design, industry and researchers have made numerous arguments that energy-only markets cannot incentivize appropriate investment. Crampton and Stoft (2008) succinctly state several of the issues of energy-only markets and conclude that a forward capacity market is required to ensure reliability:

> "The misconception…is the notion that a cleverly designed 'energy-only' market can induce optimal adequacy, or something close to it, even while the market has insufficient demand elasticity…In an ideal market, with sufficient demand elasticity, the market always clears. This means there can be no adequacy problem because involuntary load shedding occurs only when the market fails to clear and demand exceeds supply…energy prices do what every economics text says they do, they determine the efficient (not reliable) level of capacity. … The concept of an energy-only market solving the reliability problem without selling a reliability

[7] To some extent, the advent of the smart grid may enable more price response among end users and possibly also the ability for system operators to curtail consumers based on reliability preferences; however, such a scenario is still a long way from reality in most areas, and the two fundamental problems remain.

> product is logically impossible. It suggests that 'the market' can do something 'fairly well' when logic shows that It cannot do it at all." —Crampton and Stoft (2008).

Long-term reliability needs and short-term economic costs also contribute to revenue sufficiency concerns. When the variable cost of resources unexpectedly goes down (e.g., reduced fuel costs compared to those originally anticipated), the revenues made could be much less in certain years than they are in others. The expected load may be less than anticipated as well. Although this might mean that some of these resources are not needed for reliability, because of the reduced load, they may still be needed in the near future, and building this capacity is a long-term investment. At the same time, long-term markets for electricity are typically few and not very liquid, making it difficult to hedge price and quantity risks for investors.

3.3 Increasing Penetrations of VG Impacts on Long-Term Resource Adequacy and Revenue Sufficiency

The introduction of VG can have an impact on resource adequacy and revenue sufficiency in many ways. First, determining the contribution of VG toward resource adequacy is very different than the method used for conventional generation. Although the FORs of an entire collection of wind turbines or PV cells is very rare and not likely to contribute significantly to the resources' unavailability, the availability of the fuel source of VG can be quite variable. Changing weather patterns drive how VG can contribute to meeting long-term reliability needs; it is not is caused by the random forced outages that occur. Second, VG increases the amount of variability and uncertainty on the system, which can require an increased need for flexibility (see Section 4). Although certain changes to short-term energy and ancillary service markets may be needed to ensure that the flexibility that is available is provided, this may not guarantee that sufficient flexibility is built or available in the first place. This could lead to the need for new ways in which resource adequacy evaluation is performed. Finally, VG has total costs that are almost entirely fixed capital costs rather than variable operating costs. This can bring down the energy prices further, while potentially increasing (or keeping constant) the total variable and fixed costs in the power system. This could lead to further reliance on markets or incentives other than the energy market to ensure that the resources needed for long-term reliability can recover both variable and fixed capital costs.

3.3.1 Calculating the Capacity Value of VG

For VG, the recommended approach begins with the use of time-synchronized VG data with load. This will implicitly capture the underlying weather that drives load, solar generation, and wind generation. If data from different years are used for load and VG, a situation could be easily envisioned in which the load on a given day is based on hot, sunny weather that induces significant air-conditioning loads, whereas wind data is based on a cloudy, stormy day. Many other similar examples can result in a mismatch between the implicit weather driver of load and the VG resource. This mismatch would result in an implausible data foundation for the convolution algorithm. The process underlying the calculation is essentially the same as that described above for conventional power plants; the exception is that hourly VG production data (real or simulated, depending on data availability and the specific study requirements) replaces the use of the generator capacity and FOR. Details on the method can be found in Keane et al. (2011). The ELCC of wind power plants typically ranges from approximately 5% to 40% (Keane et al. 2011).

Several of the RTOs in the United States use simplified methods to calculate the capacity value for wind power. Generally, these methods have been adopted because of their simplicity and transparency, and they define a peak time period and calculate the capacity factor during that period. For example, PJM calculates the wind capacity factor for the hours ending 3:00 p.m. to 6:00pm, June through August for the most recent 3-y period. For wind power plants with at least 3 years of operational data, actual data is used for the calculation. For new wind power plants, a default value of 13% is used initially, which is replaced as operating data become available. Rogers and Porter (2012) surveyed RTOs and utilities' methods for calculating wind capacity value during the period from September 2010 to February 2012.

The RTO/ISOs with capacity markets evaluate capacity values for wind resources as described in Table 3-1. The table shows the time periods from which the capacity value was calculated, including months, time of day, and the number of evaluation years. The table also shows the method of calculation (median generation value or average capacity factor).

Table 3-1. Methods for Determining the Capacity Value of Wind and Solar in RTO/ISOs (Rogers and Porter 2012)

RTO	Season	Months	Time	Term	Method
ISO-NE	Summer	June–September	1–6 p.m.	5-y rolling	Medium net generation
ISO-NE	Winter	October–May	5–7 p.m.	5-y rolling	Medium net generation
ISO-NE	All	Default based on summer and winter wind speed data and ISO oversight			
NYISO—Existing	Summer	June–August	2–6 p.m.	Previous year	Average capacity factor
NYISO—Existing	Winter	December–February	4–8 p.m.	Previous year	Average capacity factor
NYISO—New Onshore Resources	Summer	10% default capacity credit			
NYISO—New Onshore Resources	Winter	30% default capacity credit			
NYISO—New Offshore Resources	All	38% default capacity credit			
PJM	All	June–August	2–6 p.m.	3-y rolling	Average capacity factor
PJM-New Wind	All	13% default capacity credit			

Milligan and Porter (2008) and Keane et al. (2011) recommend periodic analyses and refinements to ensure that the simplified approaches provide good estimates of contributions toward resource adequacy using a more probabilistic approach.

However, note that it is not possible to ensure that these simplified approaches can accurately capture the reliability aspect of resource adequacy. A simple method such as that used by PJM (or other similar approaches) may miss times of significant risk. As an example, Kirby et al. (2004) performed ELCC calculations on a 3-y data set from 2001 to 2003, supplied by CAISO. In 2001, the top 20 peak hours all occurred in July or August. In 2002, peak demand in July, September, and June ranked above all hours of August. In 2003, as in 2001, June did not appear in the top 20 peak hours. In the same study, one year experienced an unusually hot period in late September and early October. In the early autumn, some generation was taken out of service for scheduled maintenance as a result of prior planning, and hydro runoff was no longer providing as much energy and capacity as it was during the usual peak season from July to August. Some high LOLP hours thus occurred in the autumn, when load was relatively low compared to levels during the peak season, some generation was out on maintenance, and hydro was not contributing as much as it was during peak periods. Thus, the use of predefined peak windows may miss times of system risk when generating capacity is needed and therefore provide an incomplete picture of the state of reliability of the generation fleet.

Several possible approaches can overcome some of these obstacles, although they may also fall short of providing a true picture of resource adequacy. One approach is to use the top daily or hourly loads. For example, the top 2% of load hours, approximately 175 hours, could be evaluated post hoc, and the VG capacity factor could be calculated for that period.

One approach is to identify periods of time during which there is (or may be) high LOLE. These time periods are closely linked to LOLE or related metrics as well as the periods of time during which the capacity value of a resource is determined. Examples of this type of approach have been incorporated into the methods used by NYISO and PJM to determine the capacity value of wind energy, as described in Table 3-1. As described in Porter et al. (2014), these approaches calculate the capacity factor of the wind resource during the critical time periods and use that as a proxy for the capacity value. In some cases, such as in GE (2005), the capacity factor may match the more rigorous ELCC fairly closely, but this is not guaranteed. One example of inconsistent matches was shown in Shiu (2006).

In the study, the capacity value (ELCC) of wind and solar generation was evaluated for a 3-y period. Note that the study suffered from some data anomalies[8]; however, the results discussed below are most likely robust against the data concerns.

Table 3-2 illustrates the results of comparing ELCC for three wind power plants and one solar power plant with the capacity factors of these respective plants that were calculated over the peak period, defined as June to September, 12:00 p.m. to 6:00 p.m. The calculations were based

[8] Some solar plants, for example, had gas-assist capability. The public data from the study did not define how much energy/capacity came from gas and solar separately. Subsequent studies of solar capacity credit in the West have found lower values. However, for the purposes of our discussion, which is intended to show that results of different calculation methods for capacity credit may provide different answers, this data issue is not believed to be important.

on hourly wind, solar, and load data. The utilities and CAISO define this time period as the likely time when the system peak would occur. The table shows that in some cases and for some years, the ELCC matches the time-period factor method reasonably well. San Gorgonio in 2003 is a good match; whereas Northern California in 2004 is not. The implication of these results is that unless a comparison is made between ELCC (which is a reliability metric) and time-period capacity factor (which is not a reliability metric but is used to approximate it), the extent to which the approximation method matches the preferred metric of ELCC is not known. This means that the ability of a capacity market to capture reliability is not likely. This is not surprising, because there is no specific reliability information content in capacity factors in spite of the possibility that high-risk (high-LOLP) hours generally correspond to peak periods.

Table 3-2. ELCC Does Not Always Match Peak-Period Capacity Factors

	2002		2003		2004		3-Y Average	
Resource	ELCC (% of Rated Capacity)	Peak Capacity Factor	ELCC (% of Rated Capacity)	Peak Capacity Factor	ELCC (% of Rated Capacity)	Peak Capacity Factor	ELCC (% of Rated Capacity)	Peak Capacity Factor
Solar	88	97	83	93	79	94	83	95
Wind (Northern California)	24	19	25	20	30	24	26	21
Wind (San Gorgonio)	39	36	24	23	25	28	29	29
Wind (Tehachapi)	26	30	29	24	25	25	27	26

In fact, the assertion that LOLP and load are highly correlated turns out to be incorrect for at least some regions. Another part of the Shiu (2006) study examined the change in imports and scheduled hydro energy into the CAISO system during various peak and near-peak periods. During many of the highest peak demand periods, additional energy imports were brought in to help support the load. At the same time, to the extent that the hydro could be scheduled to support these peak periods, hydro generation was higher during many of the peak periods. Conversely, at some near-peak, and at many off-peak periods, imports and hydro were lower than at peak times. This means that during near-peak periods when there was less hydro and imports, LOLPs could sometimes be higher than during peak periods.

Figure 3-3 illustrates this condition for the 2002 study year (Kirby et al. 2004). The black curve is a load duration curve that shows the top 300 hours of load. In contrast to the monotonically decreasing load duration curve, the more erratic behavior of the trace that shows the ranking of load by LOLP (hourly) shows that there is only partial correlation between load and LOLP. This partial correlation is clearly not linear, but the downward trend of the LOLP ranking curve follows the trend of the load duration curve. A similarly erratic trace is shown when the curve is ranked by load, net of hydro and imports.

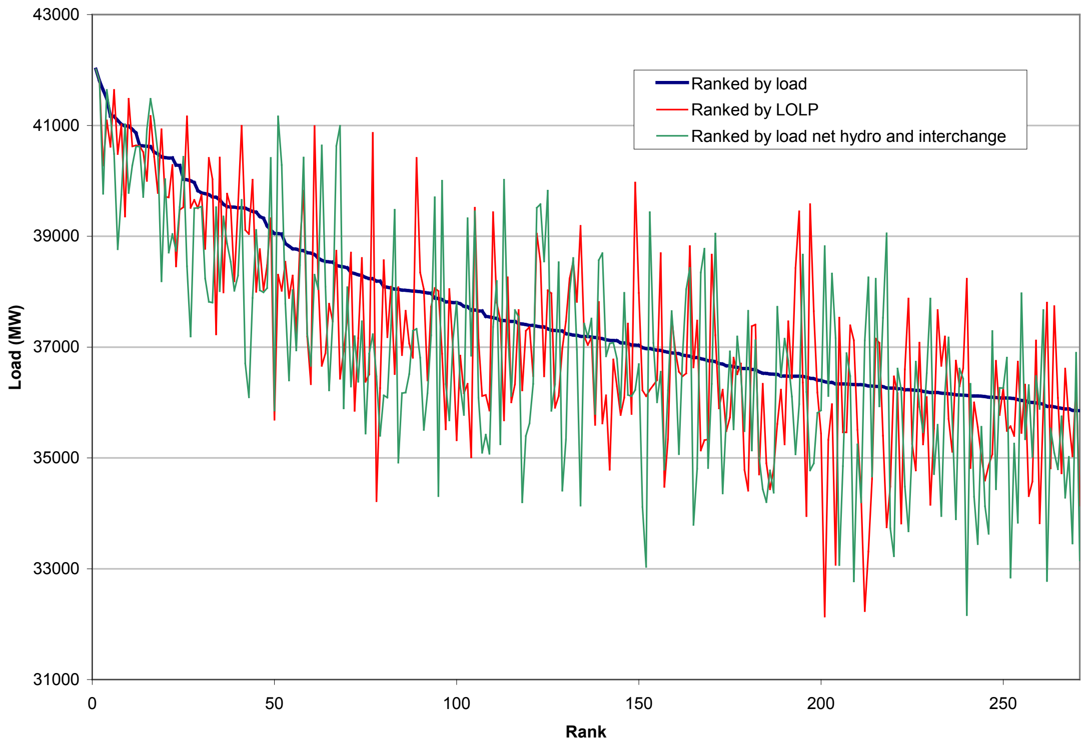

Figure 3-3. Relationship between rankings of load, LOLP (hourly), and load net of hydro and interchange for 2002. ***Image from Kirby et al. 2004***

Kirby et al.'s (2004) analysis provides evidence that there is only a partial correlation between demand and LOLP. A perfect correlation would imply monotonically decreasing curves for the LOLP-ranking and load duration curve. From this, we can conclude, that a capacity market that relies on peak periods instead of LOLP or similar metrics to calculate the capacity value of VG will have difficulty achieving reliability targets and that approximations of capacity value by the use of capacity factor calculated over peak periods will be unlikely to capture the reliability contribution of wind and solar generation to resource adequacy.

Another view of these results can be taken. Because peak periods generally are the highest risk times for loss-of-load events, system operators and markets will schedule additional generation, including imports, during those times. The impact of increasing imports is that LOLP will be reduced for the period of import. Thus, the system was operated in the way it was intended. Instead, however, some hourly LOLP values were higher during times that they may not have been expected to be significant.

Improvements can be made to the time-period capacity factor approximations. One approach is to perform a ranking of top loads, such as the top 5% or 10% of loads. Multiple years would provide a more robust indicator of possible future critical periods, but predicting the future based on the past is always somewhat problematic. When the sorting exercise is done, it would then be possible to utilize the days/times that are in the top of the ranking as the basis for a modified capacity factor calculation for the VG.

As shown, even a load-based ranking does not necessarily capture reliability. To utilize a simple approach that does not explicitly calculate ELCC, an LOLP ranking of days or hours could be performed, proceeding similarly to that discussed above. Rankings of LOLP could be done for 3 y or more, with the times noted, and VG capacity factors could be calculated over that period. One advantage to either of these ranking approaches is that it provides transparency to market participants with regard to the times that capacity availability can help achieve resource adequacy.

3.3.2 Incorporating Flexibility into Resource Adequacy Needs

As mentioned above, resource performance, in particular the flexibility attributes of a resource, may be significant enough that these attributes should play a role in long-term resource adequacy assessment, and potentially also in forward capacity markets. As experience with VG has increased, there has been a growing recognition that flexibility needs will change in the future, and how to plan for that flexibility has become increasingly relevant. The precise mechanism(s) that would ensure resource adequacy along with the required levels of flexibility are active areas of research.

As an example of one approach to extend the traditional resource adequacy techniques to flexibility analysis, Lannoye et al. (2010) adapted LOLP analysis, which is based on the changing levels of load and generation, using the speed of how rapidly resources could respond. The adaptation makes it possible to apply LOLP, ELCC, and related metrics to ramping. Figure 3-4 from Lannoye et al. (2010) shows how standard LOLP-related metrics map to ramping. This approach makes it possible to put ramping analysis in the context of reliability because a generator's ability to ramp will depend in part on whether it is on forced outage. Generators that have high FORs will have a lower effective ramping capability compared to a unit with a low FOR, all else equal. It is not clear what target value to use for the inadequate ramp resource probability (IRRP) because this is a new area of research. However, a good starting place is to perform this analysis on existing systems to provide a benchmark that may be useful in setting the target.

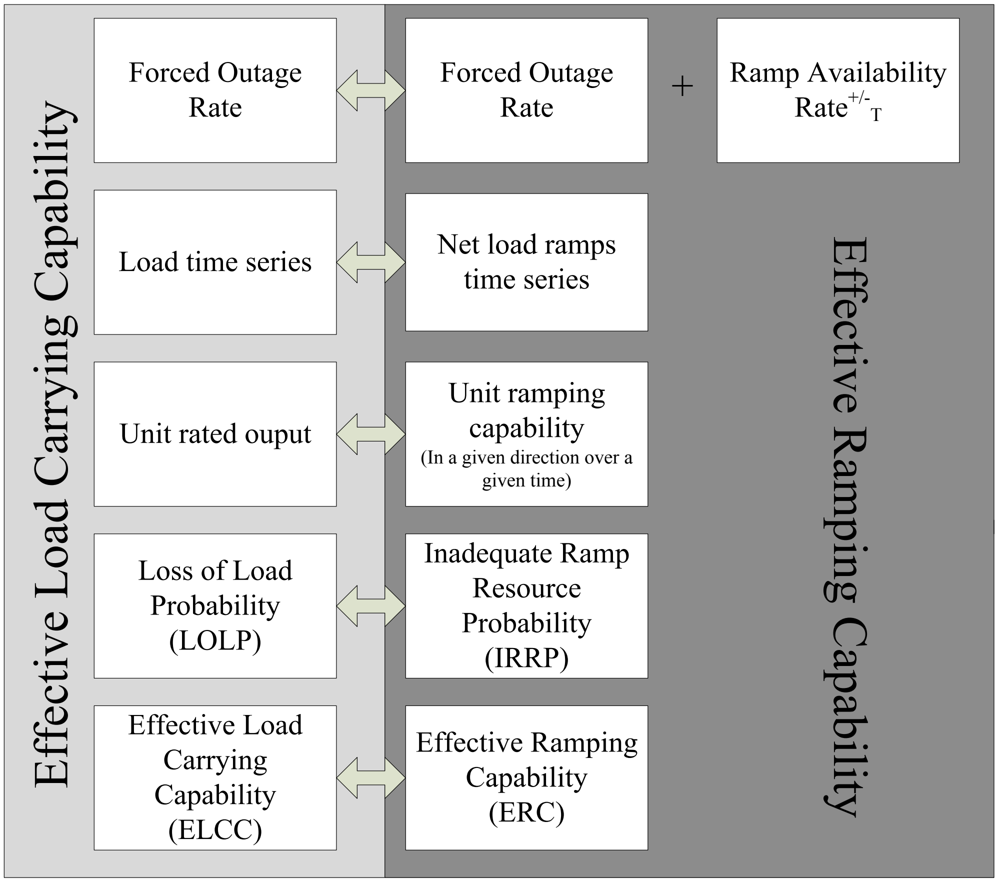

Figure 3-4. Proposed metrics to understand how effective load-carrying capability can similarly be adapted to understand the capability of resources to meet expected flexibility needs

In some parts of the electricity market, a mandate to provide a certain service is more useful than a market in which prices incentivize only some of those resources to provide it. For example, some regions require all resources to have synchronous inertia or a capability similar to inertia (Brisebois and Aubut 2011). Some reasons that might lead to this approach include the cost of such service, such as when it is extremely low, lower than the cost of administering a market to achieve this service, when the market is too complicated, or when there is low diversity in the costs to provide that service (i.e., when it is difficult to innovate and provide the service better or at a lower cost) (Ela et al. 2012a). If a mandate requires all resources to have a given level of flexibility, then much of the following discussion is not relevant. What may be relevant is how to determine how much flexibility is needed relative to required new capacity and somehow pro-rate that across units. For example, a requirement that all new generation must provide a ramp rate of at least 30 MW/min could be based on an assessment that determines that there will be sufficient ramping if all units have this ramping capability. However, it is likely that any mandate like this for all resources would be very inefficient because the cost to provide that flexibility would vary extensively between different plants/providers and the required flexibility needs would not be achieved at the least cost.

Ramping is only one part of flexibility. We define and discuss flexibility more thoroughly in Section 4, but for the discussion related to planning, we assume that key flexibility attributes can be grouped into two categories:

- Operating range—the difference between maximum and minimum stable output. A larger operating range suggests a more flexible unit than a small operating range.
- Rate of change from one state to another—including ramping, start-up, shut-down, etc. A high rate of change per minute or per hour denotes a more flexible unit than a low rate of change. Quick-start units that can ramp quickly are more flexible than slow-start units with low ramp rates.

To assess forward flexibility needs, whether for a market or not, some relatively simple approaches can be used as a starting point. One approach to assessing ramp magnitude and timing utilizes hourly (or sub-hourly, if available) data for demand, wind power, and solar power. Similarly to LOLE studies, these should be based on the same weather year so that the often-complex underlying weather impact on each of these variables is consistent. Recognizing that system balance is achieved when the sum of all demand equals the sum of all supply, making maximum use of the installed wind and solar energy means that the net load—demand less wind less solar—must be balanced by the remaining fleet.[9] It remains a simple exercise to calculate this net load. An example graph or up-ramp needs are shown in Figure 3-5. The y-axis shows each of the 52 weeks of the year, and the x-axis shows the time of day. Ramps in this example are average ramps, but the method can be adapted easily so that the graph shows maximum, minimum, or other ramp metrics, such as mean plus standard deviation, etc.

To read the graph, find a time and then examine the color/legend. For example, the morning hours of approximately 5 a.m. to 11 a.m. in Week 26 show large up-ramp needs. This can be compared to autumn morning up-ramp needs, which are not as high, and do not last as long. These plots can be compared so that alternative scenarios or data views can be analyzed. Some examples can be found in *Western Wind and Solar Integration Study Phase 1* (GE Energy 2013) and King, Kirby, Milligan, and Beuning (2011).

[9] For the discussion, we ignore the possibility of wind/solar curtailment. In reality, some limited curtailment or downward dispatch may help achieve economic and/or reliable system operation.

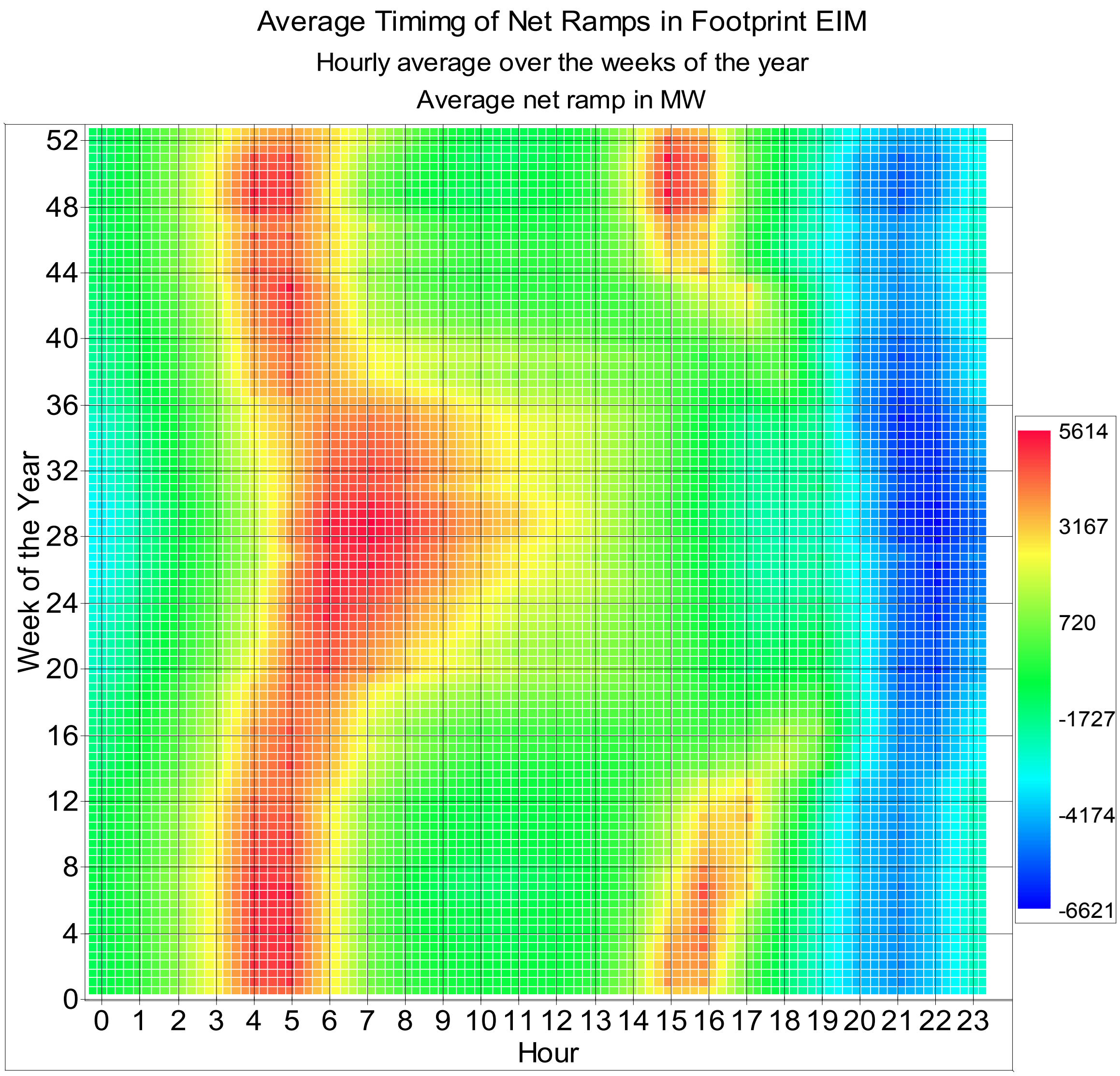

Figure 3-5. Estimates of ramp timing and magnitude. ***Image from Milligan et al. 2013***

Using the same data set as a starting point, various ramp envelopes can be generated and graphed, as illustrated in Figure 3-6. This example calculates envelopes of up to 12 hours, although there is no limit to the number of hours that can be considered in such a graph. Each envelope corresponds to a given percentile boundary or probability. For example, the blue "100% Prob." curve states that all ramps are bounded by this curve. The yellow 99% curve bounds 99% of all ramps.

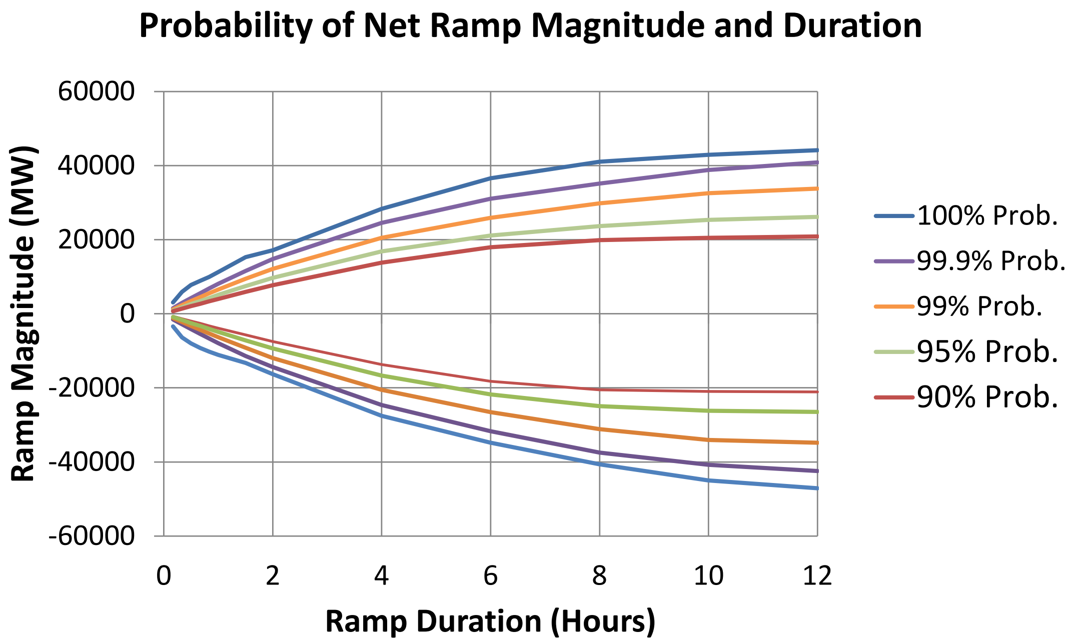

Figure 3-6. Ramp envelopes can be used to obtain a view of alternative percentage exceedance levels. ***Image from King, Kirby, Milligan, and Beuning 2011***

3.3.3 Revenue Sufficiency Challenges as a Result of Increased Amounts of Low-Variable-Cost Resources

Recent analyses of regions with high penetrations of wind generation indicate that the low marginal cost of wind generation may decrease LMPs and thus reduce revenue of all suppliers in the energy market (Maggio 2012). This may add to the revenue insufficiency problem for some resources, preventing them from recovering both variable and fixed costs.[10] This impact of wind power on market prices is not confined only to the United States but has been apparent in other countries as well (Milligan et al. 2012). In addition to the empirical evidence regarding lower LMPs resulting from VG, the *Western Wind and Solar Integration Study Phase II* (WWSIS-2) found that prices were suppressed under VG penetrations of approximately 33% (see Figure 3-7) (Lew et al. 2013). In addition to reducing electricity prices, VG displaces other resources via the merit-order effect, such that capacity factors for other generator types are also reduced. The question is what the combination of lower energy prices and lower capacity factors of existing plants implies for revenue sufficiency. It may be possible that even though the majority of prices are being depressed, the occasions of high price spikes may increase, helping to capture needed revenue. This may depend, however, on price caps, market mitigation procedures, and the price levels of administratively-set scarcity pricing. Finally, the existence and design of forward capacity markets can have a large impact on the level of revenue sufficiency.

---

[10] The potential impact of high penetrations of wind generation on revenue sufficiency is not unlike the recent drop in natural gas prices in the United States. The dramatic decrease in the cost of natural gas in 2008, coupled with low load growth, decreased LMPs in most U.S. energy markets.

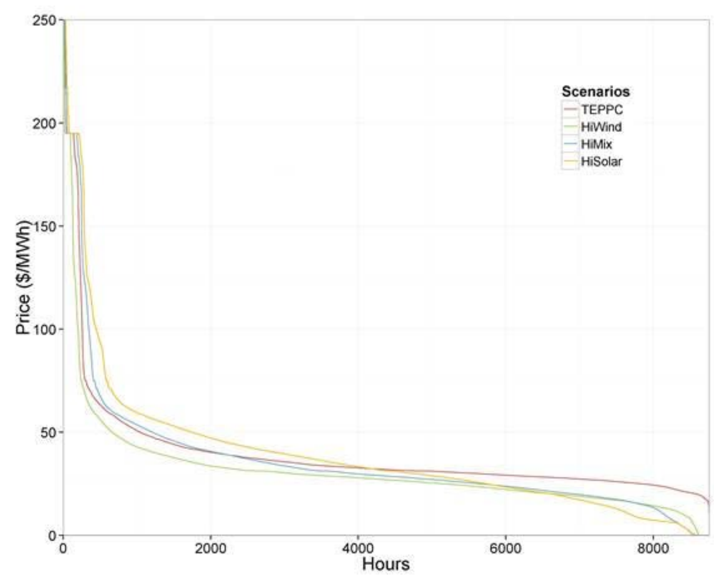

Figure 3-7. Price-duration curve from WWSIS-2 illustrating energy price suppression with higher VG penetration (Lew et al. 2013)

To demonstrate the reduction in operating profits as a result of price suppression and reduced capacity factors caused by widespread deployment of wind and solar generation, we examined archived results from WWSIS-2. WWSIS-2 used a unit-commitment and dispatch model to analyze how the system would operate under different penetration levels of wind and solar generation. For this analysis, revenues and costs for each generator type (nuclear, coal, gas combined cycle, gas steam boiler, and gas combustion turbine) from electricity sales were calculated from the model results. WWSIS-2 was not intended to study price formation, and the results included penalty prices that significantly impacted the conclusions from this analysis of revenues and operating profits. After controlling for the penalty prices by applying a price cap, the analysis found that capacity factors of certain generator categories decline and electricity prices are suppressed, leading to lower revenues and operating profits. This analysis emphasizes the importance of a model providing accurate price formation in addition to operations, especially in the context of high penalty prices, such as reserve shortage penalties. The full analysis is presented in Appendix A.

In a separate study, summarized in Appendix B, capacity adequacy and revenue sufficiency were analyzed using a generation expansion model that finds the optimal portfolio of thermal power plants for a given wind penetration level considering the variability in the wind resource and the increased need for operating reserve due to wind power forecast uncertainty. After the optimal expansion plan was determined for a given wind level, expansion variables were fixed, and the model was solved again to derive the prices for energy and spinning reserve that would result under the optimal expansion plan. A case study found that increasing wind penetration reduces energy prices while the prices for operating reserve increase. Moreover, scarcity pricing for operating reserve through reserve shortfall penalties significantly impacts the prices and profitability of thermal generators. This was the case regardless of the wind penetration level. Without scarcity pricing, no thermal units are profitable; however, scarcity pricing can ensure profitability for peaking units, also at high wind penetration levels. Capacity payments can also ensure profitability, but the payments required for baseload units to break even increase with the amount of wind power. The results indicate that baseload units are most likely to experience revenue sufficiency problems when wind penetration increases.

The study builds on a number of simplifying assumptions. For instance, expansion decisions were based on a system-wide least-cost objective function assumed to represent a fully competitive market. Also, spinning reserve up was the only reserve product considered in the analysis. Moreover, similar to the study of revenue sufficiency for the WWSIS-2 results, it was found that the level of reserve shortfalls and corresponding frequency of scarcity prices were very sensitive to minor changes in parameters inputs, as discussed in more detail in Appendix B. However, the study still provides some insights into what a future electricity market with high levels of renewable generation may hold in terms of prices and revenue sufficiency. A more detailed summary of assumptions and results are presented in Appendix B.

3.3.4 Revenue Sufficiency Challenges as a Result of Increased Flexibility Needs

With increasing penetrations of VG anticipated during the next several years, it has become clear that a system design perspective is needed to determine the best resource mix for the non-VG fleet. Markets that incentivize both the investment in and use of flexible generation should be designed from a perspective of how they can best achieve an economically optimal solution subject to various reliability constraints. What is needed is a clear market signal to investors that communicates how much flexibility is needed at some future date. Of course, there are many uncertainties around this question and about how the need for flexibility will change over the lifetime of a new generator. This section provides a discussion of potential market structures that can incent investment in new flexible resources. This can be broken into two areas of exploration:

- Whether the market design provides incentives for new resources entering the market to have the flexibility attributes that are needed
- Whether the market design provides incentives for existing resources to increase their ability to provide flexibility (e.g., through retrofits), subject to technical barriers and economic trade-offs

This discussion focuses on long-term investment in flexibility. Section 4 deals with flexibility market issues in the operational timescale.

There is no widespread agreement as to whether volatile energy prices will induce suppliers to invest in the needed level of flexibility, or whether an explicit market for flexibility is required. Another open question is whether existing capacity markets (whether modified to better capture resource adequacy aspects) should be conducted in tranches of differing flexibility needs or whether there should be separate, linked markets for capacity (resource adequacy) and flexibility. Adding to these complications is that the electric power system is not in a steady state, but is rather in a transition between a low-VG past and a potentially high-VG future. Thus, it is unlikely that current price signals can provide a good indication of the flexibility requirements 5 or 10 years from now. Therefore, investors in flexibility will need to determine the needed level of flexibility over the lifetime of potential new flexible technologies, such as demand response. This is a difficult determination to make because there is a large number of variables that can influence the need for flexibility that cannot easily be determined over the asset life of the flexible resource.

The question as to whether a given suite of flexibility attributes should be required from all new generators is an important one. Proponents of setting a requirement that applies to all new generation argue that this way there is no discrimination between different types of units, and that markets may result in unintended arbitrage that, coupled with market power, may needlessly increase costs. Proponents of the market view argue that specifying a requirement for all new market entrants will result in some types of technology that may be able to provide the required flexibility at a very high cost compared to others. This can result in getting the flexibility at a much higher cost compared to a least cost solution or getting more flexibility than is really needed, driving up costs further without apparent benefit.

3.4 Traditional Market Design Elements to Ensure Resource Adequacy and Revenue Sufficiency

Several potential approaches can be used to help address potential revenue sufficiency issues. Revenue sufficiency has been discussed since the initial stages of electricity market design (Stoft 2002, Hogan 2005). Market designs had traditional ways of meeting this issue from the inception of wholesale electricity markets. The extent to which existing market designs provide revenue to recover the fixed costs of enough resources to remain available in the market for long-term reliability is still an ongoing debate. Also, the way in which they incorporate the changing resource adequacy needs of systems with increasing VG penetrations is also somewhat unclear. Two distinct directions in terms of market designs for long-term resource adequacy have emerged. First, scarcity pricing, both through administrative prices as well as offered prices, may provide prices that go above the variable costs of the most expensive operating resources for short periods of time, when capacity is scarce, such that those prices can help recover the fixed costs of the peaking units. Second, forward capacity markets have been a part of several U.S. wholesale electricity markets for a long time. These markets look ahead to ensure that enough available capacity will be available to meet load in peak periods and aim to provide incentives for new capacity to be built in locations where it is most needed. Below, we provide a brief summary of these two designs that many of the U.S. wholesale markets have traditionally adopted to mitigate the issues of revenue sufficiency.

3.4.1 VOLL or Ancillary Service Scarcity Pricing

In markets that approach perfectly competitive markets, price volatility provides signals to both buyers and sellers. In the absence of market power and with the other attributes of perfectly competitive markets, prices would be free to fluctuate. The interplay between buyers and sellers in the market would result in an economic profit of zero in the long-run. In cases such as this, there is no concern regarding market power or a level of profit that is above and beyond what is needed to elicit the economically efficient level of supply.

The VOLL evaluates the potential cost of supply shortages where load must be involuntarily curtailed. It may be used to determine price caps in the energy market. The importance of the VOLL concept in electricity markets stems from the limited demand response, which may prevent end-users' preferences to be reflected in prices during times of scarcity. In typical power system operations, shortages will first result in insufficient reserve while load is maintained.

Therefore, ancillary service scarcity prices rather than VOLL are more frequently used to determine prices during scarcity. Stoft (2002) shows that the system operator can induce the same optimal capacity expansion as under VOLL pricing by setting an administratively determined price cap for operating reserve. Sometimes the administratively-set ancillary service scarcity prices use the VOLL in its calculation so that VOLL is still reflected in the resulting prices. Research studies have looked at many new ways of incorporating VOLL in the ancillary service pricing explicitly (Wang, Wang, and Wu 2005, Ortega-Vazques and Kirschen 2007), accounting for the probability of not being able to meet load (e.g., Hogan 2005, Zhou and Botterud 2014). The VOLL and associated administratively-set ancillary service scarcity prices are typically very high because of the large economic losses that are usually associated with outages, such as food spoilage, failure of expensive industrial processes, etc. For example, ERCOT VOLL estimates vary from $110/MWh to $6979/MWh for residential and commercial customers, respectively. However, estimated VOLL ranges even more widely elsewhere, up to $42,256 for MISO's small commercial/industrial consumers (London Economics 2013). In New Zealand, VOLL has been estimated as high as $77,687/MWh. In practice, a price cap based on VOLL exists only to protect the purchaser from market power, which could potentially cause prices to move even higher than VOLL pricing.

These administratively-set ancillary service scarcity prices will set the price for both operating reserve and energy when the system is short of reserve because it means there is an overall capacity shortage (or ramping capability shortage). The scarcity prices are set such that they are triggered only during rare instances when capacity (or ramping capability) is unavailable and therefore no supply resource is marginal to set the price. The number of times these prices are triggered combined with their high price are intended to provide the revenue needed to make the peaking units (those that have the highest variable cost, but still nonzero capital cost) recover their capital costs over the lifetime of the resource. Thus, these administratively-set scarcity prices, when triggered an appropriate number of times to reflect the reliability target of how often the reserve requirements should be scarce, would help resources recover sufficient revenue to recover their capital cost. In practice, operators still attempt to reduce the number of times scarcity pricing occurs, and it is extremely difficult to predict the number of occurrences. Hence, for investors in new generation capacity, it is difficult to secure financing based on revenues from volatile scarcity prices and it is being questioned whether these prices alone provide sufficient investment incentives.

In practice, the administratively-set scarcity prices throughout the United States vary significantly. The prices for various services have different meanings, and the ways in which these are triggered are also very different. Some regions have fixed stepwise curves, in which the greater the scarcity, the higher the price. Others have more dynamic prices depending on the system conditions. For example, Midwest Independent System Operator (MISO) bases its regulation reserve scarcity price on the monthly peaker proxy price, whereas it bases its spinning reserve scarcity price according to a formula that evaluates the VOLL and the probability of an outage for resources that are online. Other scarcity prices are very low, thereby not strictly in place to assist in capital cost recovery. For example, the ISO-NE spin scarcity price, at $50/MW-h, may be in place to reflect that a system operator sometimes prefers to be scarce by a small amount rather than make additional costly commitment decisions.

Table 3-3 shows a summary of some of the scarcity prices in the market areas in the United States.

Table 3-3. Administrative Ancillary Service Scarcity Prices for Selected U.S. Markets

Market	Regulation	Spin Reserve	Total Contingency	Other
NYISO[a]	• $400 if scarcity greater than or equal to 80 MW • $180 if scarcity between 25 and 80 MW • $80 if scarcity is less than 25 MW	• $500[b]	• $450[b]	30-min reserve[b] • $200 if scarcity greater than or equal to 400 MW • $100 if scarcity between 200 and 400 MW • $50 if scarcity is less than 200 MW
ISO-NE[c]	• N/A	• $50	• $850	30-min operating reserve[d] • $500
MISO[e]	• Monthly peaker proxy price[f]	• $98 if scarcity is greater than 10% of requirement • $65 if scarcity less than 10%	System-wide operating reserve[g] • Min:$1,100 • Max:VOLL - RegDC	• N/A
CAISO[hi]	Regulation up • $200 Regulation down • $700 if scarcity greater than 84 MW • $600 if between 32 and 84 MW • $500 if scarcity less than or equal to 32 MW	• $100	• $700 if scarcity greater than 210 MW • $600 if scarcity between 70 and 210 MW • $500 if scarcity less than or equal to 70 MW	• N/A

[a]NYISO Market Services Tarriff, Accessed May 2014. http://www.nyiso.com/public/markets_operations/documents/tariffviewer/index.jsp
[b]These numbers represent statewide scarcity prices. Locational scarcity prices exist for Eastern and/or Long Island Spinning Reserve as well, not shown here.
[c]ISO-NE Market Services Tariff, Section III – Market Rule 1. Accessed May 2014. http://www.iso-ne.com/regulatory/tariff/sect_3/

[d]These results represent the system-wide 30-min operating reserve. Local rules exist for this product as well, not shown here.
[e]MISO Business Practices Manual-002, Energy and Operating Reserve Markets, Accessed May 2014. https://www.misoenergy.org/Library/BusinessPracticesManuals/Pages/BusinessPracticesManuals.aspx.
[f]The monthly peaker proxy price is equal to the average cost per MW of committing and running a peaking unit for an hour. The price is updated on a montly basis.
[g]The scarcity price for operating reserve is determined based on the product of VOLL and the conditional probability that a resource contingency will occur. The minimum price of $1,100 is set based on the sum of the energy and reserve offer price caps ($1,000+100). The maximum price is set based on VOLL (~$3,500) minus the regulating reserve demand curve price.
[h]California ISO, Final Draft Proposal: Reserve Scarcity Pricing Design. Accessed May 2014. http://www.caiso.com/Documents/DraftFinalProposal-ScarcityPricing05-Oct-2009.pdf.
[i]The scarcity prices depend on the bid cap. The numbers presented here assume a $1000 bid cap. A lower bid cap would result in lowe scarcity prices.

In theory, VOLL pricing in the energy markets can be shown to support the optimal mix of generation capacity, because the generators recover their capital and operating costs from the resulting market-clearing prices as a result of the infrequent periods of very high prices at VOLL (e.g., Stoft 2002). The optimal level of reliability is also obtained if the price cap is set equal to the true VOLL. However, VOLL pricing gives rise to extreme price volatility, high investment risks, and the potential exercise of market power. Therefore, in practice, whether VOLL pricing will appropriately compensate peaking units to provide sufficient revenue to cover fixed costs when operating only during a limited number of hours is in question.

In some ways, the introduction of demand response for energy provision can provide a similar effect to scarcity pricing. During times when otherwise the system would be in scarcity conditions, demand response can be utilized to avoid the scarcity condition. When this demand response is used, the price can rise above the cost of the peaking units but below that of VOLL or even ancillary service scarcity price. For example, in both NYISO and PJM, the real-time LMP will be set to $500/MWh whenever emergency demand response is called upon (Walawalkar et al. 2010). Depending on frequency of occurrence of the emergency demand response and the level at which the price is set, the affect can similarly provide additional revenue for capital cost recovery.

Depending on how it is done, VOLL energy pricing or ancillary service scarcity pricing may directly link reliability to economics. For example, a price cap may be established so that it corresponds to a target LOLE, LOLH, or EUE. One example of this is the NEM in Australia, which establishes the price cap for energy at an EUE of 0.002%. Assuming that the VOLL is precisely calculated, this would appear to allow prices to rise up to the cap so that the reliability target is achieved. The price cap is applied to energy prices so it does not explicitly account for ancillary service pricing or reserve pricing, which could potentially be used as substitutes to VOLL pricing.

3.4.2 Forward Capacity Markets

In the United States, PJM, ISO-NE, and NYISO all have mandatory forward capacity markets, which establish the level of needed capacity and allow for fixed cost recovery in a transparent market environment.[11] The main characteristics of these markets are discussed in Section 2.5. The mechanisms to procure the capacity are somewhat different, but several common steps are undertaken.

1. *Establish a capacity target*. A forward capacity market is designed to procure capacity for a future period that is above and beyond what exists today. Existing resources and bilateral power purchase agreements can be used to count toward a load-serving entity's capacity obligation.
2. Determine the future capacity need for the period(s) in question.
3. Determine how existing resources will count toward the target. When this has been accomplished, the difference between the future capacity need in (2) and the existing resources in (3) is the new capacity that must be acquired via the forward capacity market.

A recent FERC white paper (FERC 2013) categorizes some of the key features of the capacity markets currently operating in the United States. The main design elements include:

1. Demand curves for capacity
2. Forward and commitment periods
3. Definition of the capacity product
4. Performance requirements
5. Market power mitigation (FERC 2013)

The overall objective of a forward capacity market is to provide a mechanism to send signals to investors regarding the need for new capacity at a future date. As noted above, the existing markets in the United States have different forward and commitment periods, and the shorter periods can provide a challenge because they may not be long enough to adequately stimulate the needed investment and construction time required for needed resources in the future.

The ultimate objective of resource adequacy, and the associated capacity markets, is to achieve some level of long-term reliability that is consistent with society's preferences. The existing capacity markets use differing measures of the product and different methods of "trueing up" or adjusting for the difference in what was acquired by the market compared to what was delivered in the key operating periods during which capacity resources are expected to perform. Thus, the market attempts to link resource adequacy/long-term reliability of supply to a market.

In general, the existing forward capacity markets have a somewhat general definition of capacity that is resource neutral, although this may not be strictly true in all cases. Demand response, for example, does not always qualify or is limited in how it is able to participate. However, several types of resources can participate in these markets, including VG. The methods for calculating the contribution of these resources may be different than those used for conventional generation,

[11] Mandatory capacity markets are not present in MISO, CAISO, ERCOT, or SPP. These regions generally use a combination of integrated resource planning and bilateral agreements to meet capacity needs. MISO, however, does operate a voluntary capacity markets.

as described in Table 3-1 and Section 3.3.1. Also capacity needs are differentiated by constrained transmission zones, which may result in different capacity prices even within the same overall market.

Two aspects of generator performance may be relevant for capacity markets. The first aspect relates to the capability that the unit can provide in terms of ramping, minimum up/down times, and generally flexible operations. It is likely that the increase in renewable generation will increase the need for flexibility in the system, as discussed in Section 3.3.2. Current markets do not explicitly recognize this issue, which is the subject of intense interest, research, and debate (although some concepts are being developed, see Section 3.5.2).

The second aspect of generator performance is how the unit responds as a capacity resource during the critical system times. Generally, this is the question of how the level of installed capacity relates to delivered capacity, and markets can handle this issue in different ways. The U.S. markets address this by using either installed capacity (ICAP) or UCAP—i.e., accounting for a resource's likelihood of forced outage at a certain time—in their capacity auctions (Pfeifenberger et al. 2013). This linking of the capacity market to a reliability metric is key, and it results from the absence of total price transparency from generation to end user. In general, system operators in the United States use a probabilistic assessment to assess how much capacity is needed to meet the required reliability standard (1 event in 10 y), but the specific approaches and software packages used differ (Pfeifenberger et al. 2013).

The details regarding the size and composition of the jurisdiction that establishes and operates the capacity market, along with the rules for how and what capacity counts toward the target, may have significant impacts. For example, consider an RTO market that includes multiple LSEs. Each LSE is allowed to acquire its own capacity via long-term power purchase agreements or other similar mechanisms. Thus, a long-term purchase of 100 MW would result in 100 MW, derated as per its ELCC, to count toward the capacity that is needed to serve the load obligation of the LSE. Consider the case that multiple LSEs secure long-term capacity in this manner, each acquiring capacity toward its own target. Generally, such LSEs will have noncoincident peak loads, and absent significant transmission constraints between them, they will be able to share contingency reserve under some circumstances. As shown by Ibanez and Milligan (2012), there is capacity value that can be acquired via aggregation: if a pooled area considers its combined adequacy needs, it will often be lower than the required capacity needed if each area undertakes its own individual assessment. Therefore, the assessments of resource adequacy, even when considered on a reliability basis, are affected by the level of aggregation in establishing the appropriate target, crediting LSEs for capacity acquired via power purchase or other mechanisms, and establishing a resource gap that must be provided via a capacity market. The ultimate level of reliable supply that is "in the ground" is what it is. However, whether there is a need for additional capacity to be secured via a market mechanism can be driven in part by the assessment level. This means that the market may incorrectly assess the quantity and price of new capacity unless a regional (such as within an RTO) target is developed first and then allocated to the appropriate load-serving entities.

Table 3-4 outlines general differences of the mandatory capacity markets including time horizon, resource qualification, slope of the demand curve, and type of capacity product (ICAP or UCAP). The forward period is defined as the length of time the auction takes place ahead of when the capacity resource is needed. The commitment period is the length of time that the capacity resource must provide capacity. The administratively-established demand curve, discussed further below, is used to determine the amount of capacity that is procured by the market. The auction product is either installed or unforced capacity.

Table 3-4. Design Features for U.S. Capacity Markets

Market	Longest Forward Period	Longest Commitment Period	Demand Curve	Auction Product
ISO-NE	3 y	5 y	Vertical with descending clock auction	ICAP
NYISO	30 d	6 mo	Downward sloping	UCAP
PJM	3 y	3 y	Downward sloping	UCAP

The table shows some of the key characteristics of these markets and illustrates that there is significant variation. For example, the NYISO capacity market considers periods of 30 d and required selected resources to continue providing its capacity service for 6 months. Conversely, both ISO-NE and PJM use a 3-y forward period and require selected resources to perform for 5 y and 3 y, respectively. A forward period of 30 d is clearly insufficient to incent the development of new resources that have not been built, thus the NYSIO mechanism secures operating capacity over a short forward horizon. A period of 3 y is sufficient to develop some forms of generation, such as natural gas combined-cycle or peaking units, along with wind or solar generation.

Thus, the longer forward periods aim to balance the time necessary to construct new resources with the risk of over-procurement. If the forward period is too short, a resource may incur significant costs before being able to participate in an auction to cover those costs. If the forward period is too long, there is a risk that an inefficient level of capacity will be procured. NYISO has elected to use a comparatively short forward period; whereas ISO-NE and PJM have longer periods and realignment auctions closer to the commitment period (FERC 2013).

The final two columns in the table illustrate the different approaches for setting the capacity target and valuing capacity additions as well as the type of capacity product being auctioned.

3.5 Evolving Market Design Elements to Ensure Resource Adequacy and Revenue Sufficiency

The impact of renewable energy on resource adequacy and future capacity needs is receiving increasing attention in U.S. electricity markets. In regions with existing capacity markets, there are discussions regarding the potential need to revise market designs to ensure that sufficient flexible capacity is procured in the capacity auctions (FERC 2013). Other developments include improved scarcity pricing through dynamic operating reserve demand curves (ORDCs) in ERCOT and the introduction of specific requirements for flexible capacity in the centralized resource adequacy program in CAISO. We discuss these two developments, which represent two different approaches to address different challenges, in more detail below.

3.5.1 Dynamic Demand Curves for Operating Reserve (ERCOT)

ERCOT is the only electricity market in the United States that relies on an energy-only market design to ensure capacity adequacy (as discussed, some markets have bilateral agreements for capacity even when no mandatory forward capacity market exists). Under the energy-only design, prices in the electricity market need to reach high levels during periods of scarcity to ensure that generators recover their fixed and variable costs, as discussed above. Therefore, scarcity pricing therefore becomes particularly important to obtain revenue sufficiency and reduce the missing-money problem, because there are no additional incentives for generation investment.

ERCOT has recently taken two steps to improve scarcity pricing. First, offer caps in the energy market are gradually being increased (e.g., to $5,000/MWh in June 2013, $7,000/MWh in June 2014, and $9,000/MWh in June 2015). Increasing offer caps mean that energy prices can increase to higher levels during periods of extreme scarcity. Second, a demand curve for operating reserve is being introduced in the real-time market to better reflect the marginal value of reliability from reserve on electricity market prices (Hogan 2012). The ORDC influences the prices for both reserve and energy because the two products are closely linked through opportunity costs.

The proposed ORDC is derived based on a probabilistic assessment of the LOLP for different reserve levels and an estimate of VOLL. Ideally, the ORDC should be used in a co-optimization of energy and reserve markets. However, because there is currently no co-optimization in the ERCOT real-time market, a post calculation that mimics co-optimization and derives prices for reserve and price adders for energy is used instead. The procedure involves the following main steps (Hogan 2013):

- Estimate probability distributions for LOLP based on historical differences between hour-ahead scheduled reserve and available reserve in real time. These estimates are done for 4 seasons and 6 time-of-day blocks, a total of 24 time segments. The resulting LOLP probability distributions, assumed to take the shape of a normal distribution, reflect the likelihood of forced outages, load forecasting errors, and wind power forecasting errors for the 24 different time segments.
- An adjusted LOLP distribution, LOLP', is derived by assuming that a certain minimum level of contingency reserve, X, is required to avoid load shedding. LOLP' is assumed equal to 1 whenever the available reserve is below X. Moreover, the LOLP' probability distribution is shifted to the right by X compared to the original LOLP distribution. Separate LOLP' curves are also derived for spinning and nonspinning reserve to reflect the different response times.
- The price on the demand curve for operating reserve for a given reserve level, *R*, equals LOLP'(R) multiplied by (VOLL minus LMP).
- Price adders for spinning and nonspinning reserve are derived based on the actual available reserve in real time.
- The LMP is increased by the calculated price adder for spinning reserve. ERCOT does not currently co-optimize energy and reserve in the real-time market, so this adder must be manually added to the energy price.

- An ancillary service market imbalance settlement ensures that resources are indifferent between energy and reserve in real time.

Figure 3-8 compares the proposed ORDCs for four different seasons to the current practice of using a fixed reserve requirement and scarcity price, which is equivalent to a vertical demand curve (in black). The downward-sloping demand curves will result in higher prices whenever the reserve margin is above the current requirement of 3,300 MW, but it may actually give lower prices when the reserve level drops below 3,300 MW. Simulation results for ERCOT (Hogan 2013) indicate that if the ORDC had been in place in 2011—i.e., a year with several extreme weather events—the average energy price would have increased in the range of $7/MWh to $26/MWh, depending on the VOLL parameter and the minimum contingency reserve, X. In contrast, in 2012 the average energy price increase would have been much more modest, in the range of $1/MWh to $4.5/MWh. Overall, any increases in the prices for energy and reserve would increase the incentives to invest in new system resources. Improved scarcity pricing could also provide improved operational incentives for both supply and demand resources. The ORDC concept does not prevent the introduction of additional incentives for capacity and flexibility. In fact, several other ISO/RTOs have simple ORDCs in place already, along with other capacity adequacy incentives. Current demand curves for ancillary services in other markets are versions typically not based on the same dynamic, rigid, probabilistic assessment as the one proposed for ERCOT (see 3.4.1 and Table 3-3). For instance, other ORDCs do not account for the actual probability of load shedding for that particular time. However, improved scarcity pricing through the ORDC and increased offer caps are currently the main new resource adequacy initiatives in ERCOT, which continues to rely on the energy-only market design to provide adequate investment incentives as wind energy penetration continues to rise.

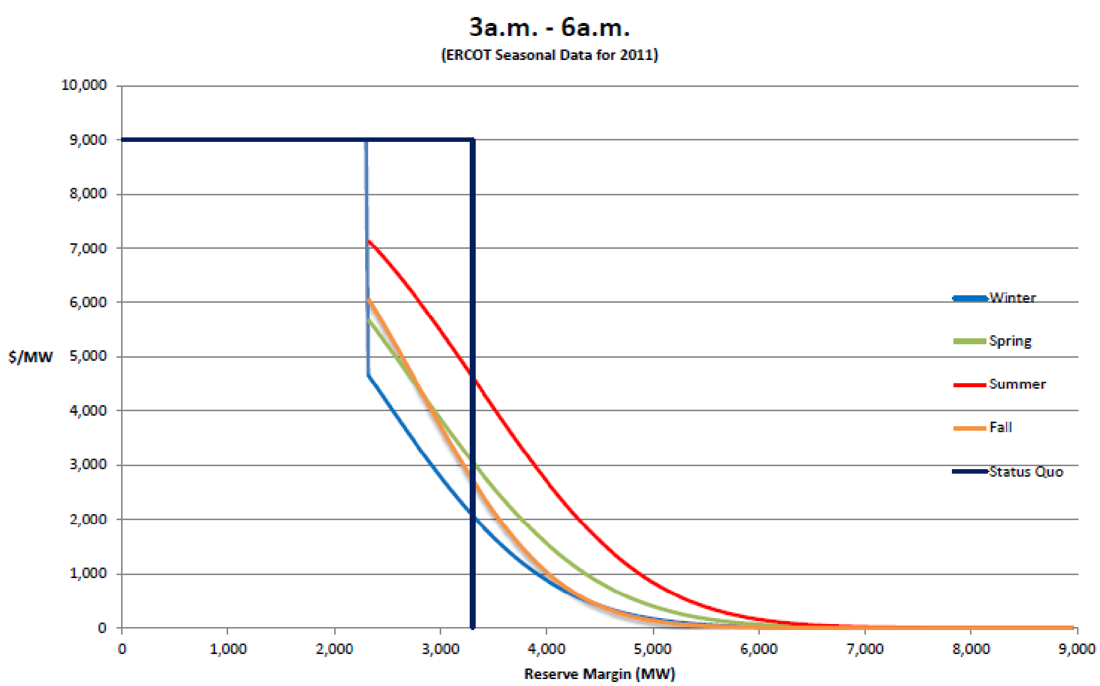

Figure 3-8. Illustration of ORDCs for different seasons in ERCOT for the time period from 3 a.m. to 6 a.m. based on data from 2011 (Anderson 2013)

An ORDC is not a new idea in the academic literature (e.g., Hogan 2005). However, the current ERCOT initiative represents the first rigorous implementation based on a first-principles probabilistic assessment of the reliability impacts and marginal value of reserve. Still, the ERCOT implementation could be enhanced in several ways. For instance, full co-optimization of energy and reserve in the real-time market would ensure more efficient operation and pricing. Moreover, the ORDC could also be considered in the day-ahead market to avoid inconsistencies and to achieve better price convergence between day-ahead and real-time markets. The ORDCs could be derived based on a dynamic assessment of relevant information in real time rather than using predefined LOLP distributions for different time segments. Zhou and Botterud (2014) propose to calculate ORDCs dynamically, accounting for the uncertainty in wind power through a probabilistic wind power forecast along with load forecasting uncertainty and the probability of forced outages from thermal generators. Hence, the expected forecast uncertainty, which is likely to change depending on the weather situation, would be factored into the ORDC. Zhou and Botterud (2014) conducted simulations of scheduling, dispatch, and market-clearing prices in co-optimized day-ahead and real-time markets with different reserve strategies. The results also indicate that the ORDC would lead to higher prices in many hours, whereas extreme price spikes are less likely to occur because of the downward-sloping demand curve (Figure 3-9).

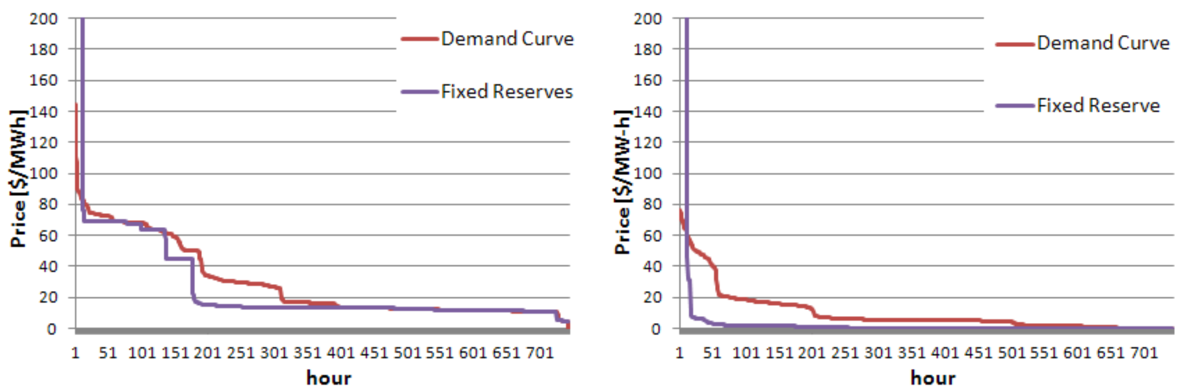

Figure 3-9. Simulated price duration curves in a day-ahead market for energy (left) and spinning reserve (right) with an ORDC and a fixed reserve requirement for one month (Zhou and Botterud 2013)

3.5.2 Forward Flexible Capacity Requirements (CAISO)

CAISO is proposing a different approach to ensure resource adequacy in the long term. Since 2005, a resource adequacy program has been in place in which load-serving entities are required to meet a local PRM. Certain deliverability criteria are in place for resources to qualify and count toward the PRM, which must be documented at monthly and yearly levels. Moreover, qualified resources must make themselves available to the system operator. There is no centralized capacity market, but market participants can use bilateral trading to ensure that sufficient capacity is available to meet reserve margins (CAISO 2013a). The bilateral resource adequacy market provides generators with a potential source of income to help ensure revenue sufficiency.

Until recently, the focus of the resource adequacy program has been to ensure that sufficient capacity was available to meet peak load. However, several developments in recent years have prompted an ongoing revision and extension of the resource adequacy program. In particular, California's renewable portfolio standard, which requires the state to meet 33% of its load with renewable resources by 2020, has raised concerns about whether or not there will be sufficient system flexibility to efficiently operate the system with the rapid increase in renewable energy. Therefore, changes to the resource adequacy program are being introduced so that LSEs will be required to not only procure sufficient capacity to meet forecasted peak load but also to meet additional flexibility requirements with their capacity. In short, the current proposal, scheduled to be introduced in 2015 if approved, requires that LSEs in aggregate have sufficient flexible capacity available to meet forecasted system needs. The new flexible capacity initiative includes six measures (CAISO 2014).

1. Flexibility requirement determination for the system for the upcoming year. This is based on net load forecasts (Figure 3-10) considering the most current RPS contracts. In the figure, Category 1 is base flexibility, Category 2 is peak flexibility, and Category 3 is super-peak flexibility. "A" refers to the maximum 3-h net load ramp in a month, and "C" refers to the largest secondary 3-h net load ramp for the month (CAISO 2014).
2. An allocation methodology that translates the system flexibility requirements to the individual LSEs based on their historical contributions to net load ramps.
3. Flexible capacity showings in which LSEs are required to demonstrate adequate flexible capacity procurement. Flexible capacity can come from several sources (generation, storage, demand).
4. Assessment of the capacity showings in which the ISO uses a flexibility counting methodology to ensure that LSEs' flexible capacity requirements have been met.
5. Must-offer obligations—i.e., flexible capacity must provide economic bids to the ISO's day-ahead and real-time markets from 5 a.m. through 10 p.m.
6. Backstop procurement, which allows the ISO to procure flexible capacity on a one-year forward basis if there is a deficiency in the aggregate supply of flexible capacity to the system.

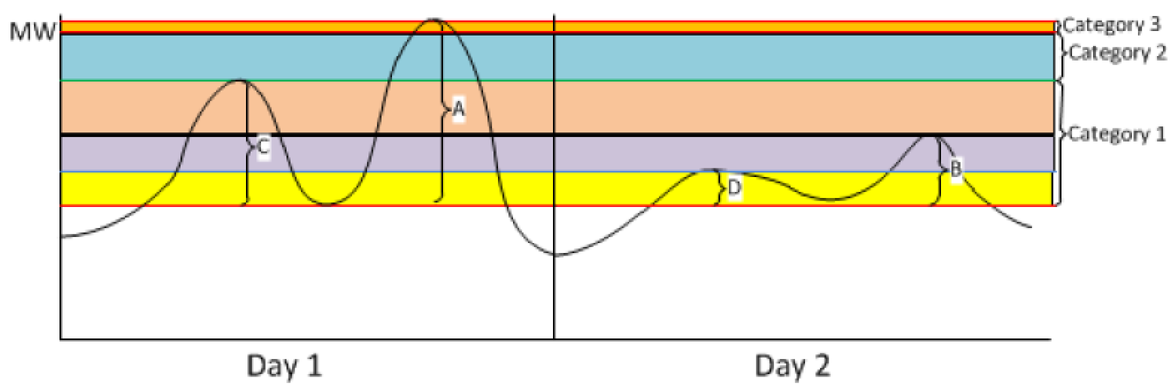

Figure 3-10. Capacity requirements for different flexibility categories as a function of forecasted net load ramps (CAISO 2014)

Overall, the approach taken by CAISO relies on administrative and centralized planning procedures to ensure both the amount of future capacity and level of flexibility in the future resource mix. This is in contrast to the approach in ERCOT, which relies on price incentives in the short-term markets for energy and reserve to ensure capacity adequacy and revenue sufficiency.

3.6 Summary

At this point, it should be clear that designing a market for the reliability of supply—resource adequacy—is a very difficult proposition. The large-scale expansion of VG adds further complexities to resource adequacy assessments and market design to ensure revenue sufficiency. Whether current market designs provide the incentives that are needed to ensure adequacy in the long run or if new approaches are needed is still an open debate.

If electricity market designs rely on the energy-only approach, the key challenge is to ensure that the short-term prices for energy and reserve provide sufficient incentives for investments in a resource mix with sufficient capacity and flexibility. Appropriate scarcity pricing has been and still is the main solution that is required to ensure that resources recover both capital and operating costs in an energy-only market. This will require clever regulatory market interventions that allow prices to rise during scarcity conditions without creating opportunities for market manipulation. Renewable resources may lead to higher frequency of low and negative energy prices and increase the variability in market prices, increasing the reliance on scarcity pricing under the energy-only designs. At the same time, prices for operating reserve may increase because of higher reserve requirements associated with the increased variability and uncertainty of renewables. However, it is an open question whether future prices for energy and reserve will adequately compensate ramping capabilities to meet system flexibility needs as well as the required need for available capacity. Analyzing revenue sufficiency is difficult because it depends on having an accurate model of price formation in a future system, which is in itself a challenging problem that deserves more research. Finally, stochastic programming approaches have been proposed to more efficiently handle the challenges of renewables in electricity market operations, but setting prices for energy and reserve under stochastic scheduling is not straightforward.

If the solution is to use an additional incentive mechanism for resource adequacy, such as a capacity market, several other questions need to be addressed:

- What reliability metric should be used at the system level: LOLE, LOLH, EUE, or other?
- What target reliability level should be used? Is the traditional target of 1 d/10 y the right target?
- Is a separate metric required to ensure sufficient flexibility, in addition to capacity?
- How many years should be used in rolling reliability and capacity assessments?
- Do time-period capacity factor approximation methods sufficiently capture the link between performance attributes and resource adequacy?
- Does the combination of metric and reliability target exhibit consistency across generator types and their contribution to resource adequacy?
- What type of performance assessment should be required to ensure and incentivize resources to be available when needed?
- How should capacity market auctions be designed to minimize the potential exercise of market power?

In this section, we have argued that it may be useful to consider ELCC itself as a candidate metric for a capacity/resource adequacy market, because it would better capture a resource's contribution to reliability than current metrics such as unforced capacity. However, there are several challenges to implementing such a probabilistic ELCC metric. As illustrated, ELCC is highly nonlinear and potentially sensitive to several other influences. Clearly, more work would be needed to develop an ELCC-based auction for resource adequacy. As an alternative, more rigorous mapping between ELCC and UCAP or a similar metric may result in achieving the goal of retaining some reliability information in the market, yet perhaps overcome some of the concerns regarding the nonlinearity and other issues, such as sensitivity to ordering, that may exist.

A variety of market designs are considered and being introduced in different ISO/RTO markets to address the issues of long-term flexibility needs, resource adequacy, and revenue sufficiency. We believe that the industry will go through several iterations before consensus emerges on the specific best practices on these complex topics of resource adequacy and revenue sufficiency in electricity markets with large-scale penetrations of renewable resources.

Several other open research questions remain, many of which exist independently of the future structure of long-term resource adequacy and flexibility markets.

- What is the behavior of VG ELCC over multiple years, and what is the appropriate number of years of data to use in a resource adequacy study to achieve statistically-expected results and/or behavior and quantification of statistical tails?
- What metric, or family of metrics, can best describe future flexibility needs? What is an appropriate choice for a target level of flexibility, especially given that there is some uncertainty surrounding the levels of VG that will be experienced during the next several decades?
- What are the desired properties of these flexibility metrics, and how do they perform when considering multiple technologies? Are the metrics robust enough to provide consistent evaluations of new, possibly unknown, technologies that may emerge?
- How can multiple flexibility metrics be established within an incentive mechanism? Will it incentivize both new and existing resources to have flexibility capabilities?
- Is there a right combination of scarcity pricing and forward capacity markets that can take the best attributes of both concepts?
- Is it important to have standard resource adequacy and flexibility definitions and markets across all regions? If there are differences in product definitions and/or assessment algorithms, will that create gaming from entities that could potentially sell into multiple markets?

As these markets undergo changes and potentially new markets evolve to address long-term issues described here, it will also be critical to identify and correct any unintended consequences that undermine one or more markets and the way they interact. And, as always, the potential for market power must be assessed and managed.

4 Incentivizing Flexibility in System Operations

The existing wholesale electricity market designs are unique in their complex relationships between economics and the physics of electricity. These markets aim to incentivize resources to provide a variety of services, including energy and various ancillary services. However, it is unclear as to how much the existing markets may be incentivizing flexibility from the market participants in an efficient manner. Questions that should be asked are (1) whether the market designs are incentivizing new resources entering the market to have the needed flexibility capabilities, (2) whether the market designs are incentivizing existing resources to upgrade their technology to offer additional flexibility capabilities if more flexibility capabilities are needed, and (3) whether the market designs are incentivizing resources that have flexible capabilities to offer those capabilities to the short-term energy and/or ancillary services market when flexibility is needed most. The first two considerations focus more on long-term capacity adequacy, revenues, and incentive needs and are discussed in Section 3. This section focuses on the third question and how power system flexibility is incentivized during short-term system operations. First, we describe some definitions and examples of flexibility. Then we discuss how the introduction of VG might be making the topic of flexibility incentives more of a pressing issue. Next, we cover a number of historical, recent, and then proposed market design principles and elements that may affect how flexibility is incentivized in wholesale electricity markets. Many of the existing market design elements provided incentives whether or not a unit could be flexible and offer its flexibility to the market operator. However, some of the more recent changes are being designed to more explicitly incentivize an increasing quantity of flexibility, in many ways because to the increasing variability and uncertainty that is brought to the system by VG. Finally, we discuss ongoing issues and remaining questions and provide a summary of this complex topic.

4.1 Defining Flexibility

Flexibility as a term is gaining in popularity in the electric power and energy industry. Although the general meaning is understood, some clarification of the definition is important. In Lannoye et al. (2012), flexibility is defined as the ability of a system to deploy its resources to respond to changes in net load. In Ma et al. (2013), flexibility describes the ability of a power system to cope with variability and uncertainty in both generation and demand while maintaining a satisfactory level of reliability at a reasonable cost over different time horizons. Both of these definitions focus on the system, but they can be easily disaggregated to individual resources by replacing "ability of a system" with "ability of a resource." Other definitions also exist (see Text Box). A more general definition can be converged from all of these as *the ability of a resource, whether any component or collection of components of the power system, to respond to the known and unknown changes of power system conditions at various operational timescales.* For the purposes of this section, the changes are those that occur with the active power of individual or aggregate generation, demand, or network elements at multiple timescales. System and market operators should have an objective to utilize the system flexibility to meet the reliability requirements in the most cost-efficient manner possible. A lack of flexibility can lead to imbalance of generation and load, overloading of transmission elements, and other potential reliability issues. Improper utilization of existing flexibility, or unwillingness of resources to provide existing flexibility, can also lead to the same reliability issues, or to higher costs when more expensive flexibility is required when economic flexibility is unwilling to be offered.

Defining Flexibility

A number of authors have defined power system flexibility in a number of different ways. A few examples are shown below. The definition used in this report is not very different from those proposed in past work. We make an effort to make the definition general to all system conditions and both individual contributions and system contributions and needs.

This Report
The ability of a resource, whether any component or collection of components of the power system, to respond to the known and unknown changes of power system conditions at various operational timescales

Lannoye et al.
The ability of a system to deploy its resources to respond to changes in net load

Ma et al.
The ability of a power system to cope with variability and uncertainty in both generation and demand while maintaining a satisfactory level of reliability at a reasonable cost over different time horizons

Bouffard and Ortega-Vazquez
The potential for capacity to be deployed within a certain time frame

Ulbig et al.
The technical capabilities and related constraints of individual power system units to modulate power and energy in-feed into and out-feed out of the grid, respectively

International Energy Agency (2014)
(1) The extent to which a power system can adapt the pattern of electricity generation and consumption to balance supply and demand in a cost-effective manner
(2) The extent to which generation or demand can be increased or reduced over a timescale ranging from a few minutes to several hours in response to variability, expected or otherwise

A number of characteristics have been emphasized as qualities that are needed for increased flexibility. We focus on active power flexibility of generating units, but this can be applicable to other types of resources that can control active power as well as reactive power. The main characteristics of flexibility will generally fall into three categories: absolute capacity range, speed of power output change, and duration of energy levels. Resources that have a large range of absolute output levels between their minimum and maximum capacity levels can be classified as more flexible because they have greater ability to adjust to changing power system conditions. Resources with greater ramp rates can also be classified as more flexible because they are able to adjust faster to changes in power system conditions of varying speeds. Last, resources that are able to hold energy levels for longer periods of time can be classified as more flexible because they are able to better meet power system conditions that sustain for significant periods.

Many of the existing generating technologies have numerous limitations that may affect absolute power range, speed of power output change, and energy level durations in different ways. For example, a number of constraints will limit thermal plants on how and when they can be committed on or off. This includes minimum on times, minimum off times, maximum starts per day limits, and other commitment constraints. Hydro plants may have rough zones at certain power limits when they cannot provide power without incurring damage (Rux 1993). Combined-cycle plants also have constraints on how they can be configured and how configurations can be transitioned (Lu and Shahidehpour 2004). These constraints can affect the absolute power range. Resources that can be easily turned on or off provide increased flexibility because the absolute power range can be taken between zero and maximum capacity. Similar constraints can limit the speed of power output change. Start-up times and shut-down times can limit how fast resources can provide power when needed and provide nothing when not needed. All of these types of "discrete" constraints limit the flexibility and add further complexity when evaluating the flexibility of individual resources and of the system. Further constraints can limit how long resources can provide energy. Examples are hydro resources with limited inflow and pumped storage hydro and other storage resources that have limited energy because of insufficient charging from other time periods. VG has considerable flexibility, subject to its maximum available power and sustainment times. VG's absolute power range is large when wind speeds are high and it has minimum generation levels of zero. It can also rapidly change its power output response with the assistance of power electronic controls.

Additional characteristics demonstrate a resource's flexibility as well. Resources that have automatic control capability, are frequency responsive, and can contribute to restoration can be classified as more flexible. Reactive power range and ability to ride through voltage and frequency fault conditions are also an important flexibility parameter. These qualities are important to ensuring overall power system reliability. The more of these flexible qualities a resource has, the more valuable it is to the power system.

It is possible that some resources can provide different levels of flexibility, and that increased flexibility may increase resource costs. For example, a resource may be able to reduce its minimum generation level, but that may in fact cause an inefficient heat rate. Similarly, a resource may have emergency ramp rates it can provide at faster rates, but it may cause more wear and tear to the unit, leading to higher operations and maintenance costs. In these instances, it is important that the extra flexibility that has increased costs is requested only if necessary, and that incentives are present for those periods so that resources can at least recover costs. These topics are crucial and the focus of this section.

4.2 The Impact of Increasing Penetrations of VG On the Need for Flexibility

System flexibility is required to meet the known and unknown changes in power system conditions, referred to as variability and uncertainty, respectively. In power systems, the load has variability characteristics, both diurnally, and at shorter timescales. The load is also uncertain, because it cannot be perfectly predicted at all times and over all time horizons. Conventional generation and transmission elements also have uncertainty, because they can fail without any certainty of when and if that may happen. Over time, the industry identified various procedures to accommodate this variability and uncertainty, including operating reserve (see Section 2.2) and security-constrained scheduling models (e.g., n-1 preventative procedures for transmission outages).

VG adds variability and uncertainty to the existing amounts at multiple timescales (Ela and O'Malley 2012d). The maximum available power changes based on the changing weather driver, such as wind speed and solar irradiance, and cause variability. A number of studies have quantified this variability (Wan 2005, Wan 2011, and Mills et al. 2009). Although the variability of a single turbine or PV cell may be quite high, the variability of an entire wind power plant or large PV array is relatively reduced. Further reduction comes from multiple plants in a balancing authority area because of geographic diversity. The variability also differs depending on the timescale. For example, 1-s per unit variability is typically less than 1-min per unit variability, which is typically less than 10-min per unit variability.

Correspondingly, the maximum available power cannot be predicted perfectly, causing uncertainty. These characteristics have been quantified in numerous studies (Zhang et al. 2013, Hodge et al. 2012). Intuitively, the farther ahead the horizon of the forecast, the more difficult it is to predict. Wind and solar power uncertainty have very different characteristics because a general pattern is much easier to predict for solar power than it is for wind power. Some studies have now also shown how variability and uncertainty at these different timescales can affect both reliability and efficiency in different ways (Ela and O'Malley 2012d, Ela et al. 2013b).

The variability and uncertainty of VG can be met by different operational strategies depending on the timescale and time horizon. For instance, short-term (e.g., minute-to-minute) variability might be met by regulation reserve, longer-term (e.g., tens of minutes to hours) variability might be met by flexibility reserve, and long-horizon uncertainty might be met by improved forecasting. It is possible that some of the existing procedures for accommodating the variability and uncertainty of the past can be used to meet the increasing variability and uncertainty of VG; however, it may be that new procedures and tools can do so in a more reliable and efficient manner. Some examples of these evolving strategies include shorter scheduling intervals (Milligan, Kirby and Beuning 2010) to meet increased variability and stochastic or robust unit commitment and dispatch solutions for addressing uncertainty (Bouffard and Galiana 2008, Meibom et al. 2011, Wang et al. 2011, Bertsimas et al. 2013). In addition, incorporating increased look-ahead horizons in the scheduling model and intelligent operating reserve requirements can help meet increased variability and uncertainty if done efficiently (Price and Rothleder 2011). Finally, and for all the above strategies, increased variability and uncertainty, with all else unchanged, will require increased flexibility requested of the resources. If there are increased costs or decreased revenue to the resources, additional incentives should be provided with the request for increased flexibility. The new strategies may be more complex and may require changes to how energy and ancillary services are currently priced.

Besides variability and uncertainty, a few other traits that VG carry may increase the need for flexibility on power systems. VG is not a synchronous machine, nor is it inherently responsive to the frequency of the grid. Although this should not explicitly increase the need for frequency response in itself, increasing penetrations of VG without these characteristics can displace resources that do. This may produce a need for incentives to ensure that this capability is available. For example, primary frequency response and synchronous inertia currently are not part of the ancillary services market, but they may be necessary in the future as penetrations of nonsynchronous VG increase (Ela et al. 2012a). It is also possible that VG can create the controls to provide these services itself (Ela et al. 2014), but that may also depend on the incentives present and whether any revenues from providing this service will justify VG installing these capabilities.

Other traits include location constraints because VG will be located in areas where there is the highest power production potential. These areas may be located far from load centers. This may increase the need for localized flexibility that has to be able to accommodate the variability and uncertainty of VG without overloading the elements of the transmission network.

All of the above characteristics of VG may require increased needs for flexibility on the power system. As we have discussed throughout this paper, most of these characteristics and the need for flexibility is not necessarily new. The past needs of flexibility on the power system may have been met by the current system operating procedures and wholesale market designs. It may be possible, however, that increasing penetrations of VG may push the needs to the point at which the issue of specific flexibility incentives can no longer be ignored. This is a difficult discussion with varying opinions throughout the industry.

Throughout this section, we will describe some of the existing wholesale market designs and some recent and proposed changes from the U.S. market operators that may have been implemented and proposed due to increased VG penetration that all attempt to address incentivizing flexibility in system operations. It is important to understand these characteristics, especially those of variability and uncertainty at all operational timescales, as well as asynchronism and non-proximity to load, to understand how different market designs may incentivize the need for increased flexibility in these different manners.

4.3 Traditional Market Design Elements That Impact Flexibility Incentives

A few mechanisms that provide some incentives for market participants to provide flexibility have been in existence since the inception of the U.S. wholesale energy markets or shortly thereafter. The extent to which they provide sufficient flexibility is an ongoing debate. It is not obvious whether some of these elements incentivize flexibility or not. The first step in obtaining flexibility from market participants is to have a mechanism that allows the market operator to commit the resource and dispatch the resource's output when it is needed. A portion of the generating fleet—sometimes as much as 50% to 70% of the energy (CAISO 2010b and Monitoring Analytics, LLC, 2010)—operates through bilateral contracts outside of the pool-based electricity markets. From the market operator perspective, these suppliers, along with others who are not responding to price signals, are self-scheduled resources offering absolutely no flexibility for the market operator to utilize. Although the market operators still receive the energy from the resources to meet the expected demand, they do not have any flexibility from these self-scheduled units to meet ancillary service demands or changing energy demands. The more self-scheduled resources exist in the system, the less flexibility the system operator has access to, holding all else constant. Therefore, in addition to having the ability to quantify needed levels of flexibility, it is important to incentivize suppliers to offer their flexibility into the market. Without this feature, significant levels of physical flexibility may be unavailable to the market operator. Suppliers should be incentivized to allow the market operator to dispatch its output to meet the changing energy and ancillary service demand, while still operating within design parameters.

We focus on five specific examples of traditional market mechanisms in place that in some manner could incentivize some form of flexibility, or at least provide an incentive so that suppliers may offer their flexibility to the market operator. These include centralized scheduling and efficient dispatch, frequent scheduling and frequent settlement intervals, existing ancillary service markets, make-whole payment guarantees, and day-ahead profit guarantees.

4.3.1 Centralized Scheduling and Efficient Dispatch by the Market Operator

Current markets have mechanisms that incentivize resources to offer their operating capabilities to be dispatched by the RTO/ISO. This allows for the market operator to ensure that the resource is flexible, but it does not guarantee how flexible the resource is. The first mechanism is somewhat obvious but often overlooked. When suppliers participate in the pool market, the market operator will operate them at their most efficient operating point based on their offered bid-cost curve. The market operator minimizes the bid-production costs from all these bids to meet the energy and ancillary service demands subject to power system security and unit constraints. The LMP and ASCP, as discussed earlier, are calculated based on the marginal bid-based cost to provide energy or ancillary services. Therefore, the cost of supplying energy for a unit participating in the market and allowing for the market operator to commit and dispatch the supplier's output should theoretically not be greater than the resultant price (reasons that costs can be higher than the price are discussed later in Section 4.3.4). When the price increases, the market operator gives the supplier a position that reflects that it is efficient to increase its output, and this allows the supplier to earn more revenue. When the price decreases, the market operator gives the supplier a position to reduce output, because it may be that the current output is no longer efficient when receiving the reduced price.

Under good electricity market design, the supplier output level should always reflect the changing prices and should avoid operating at levels that cost more to produce energy than the price they receive. Self-scheduled resources provide the market operator with the scheduled output before the market clears, and this schedule is fixed regardless of the price. During periods of high prices, the self-scheduled resource could miss out on additional profit. During low prices, the self-scheduled resource could lose money when the cost to supply energy is greater than the energy payments they receive. When substantial bilateral contracts are self-scheduled into the market, there may come a point at which the flexibility that is available to the market operator is insufficient, inducing a need for other mechanisms to obtain this flexibility. For example, a very high proportion of self-scheduled resources may drive the need for more expensive sources of flexibility, such as additional flexible generating capacity or storage. In cases such as this, the system may possess more flexibility than is needed; however, much of this flexibility may be stranded. It is important to note that the levels of *physical* flexibility may be sufficient; however, some of this may not be contractually available.

To illustrate the potential impacts of self-scheduling, we show a simple example. Table 4-1 shows a bid-in cost curve for a thermal generating unit which is taken from real bid-cost data. This bid-in cost curve reflects representative costs of thermal plants based on a convex, monotonically increasing incremental heat rate. We ignore no-load costs in this example for simplicity. The incremental cost in column 1 is the cost bid for the specific capacity represented in column 2. Therefore, the first 286 MW in this example will always cost (35*286) = $10,010.

Table 4-1. Hypothetical Thermal Plant, Piecewise Linear Cost Curve

Incremental Cost ($/MWh)	Energy/ Capacity Segment (MWh)
35.00	Up to 286.0
47.25	286.1–295.0
47.60	295.1–304.0
47.95	304.1–313.0
48.30	313.1–322.0
48.65	322.1–331.0
49.00	331.1–340.0
49.35	340.1–349.0
49.70	349.1–358.0
50.75	358.1–376.0
52.50	376.1–377.0

The cost data in this table forms the basis of how this resource would bid into the market. We next turn to the relationship between this cost data and LMPs and an examination of how various self-scheduling strategies compare to how the unit would be dispatched in the absence of self-scheduling.

Table 4-2, column 2, shows a 12-hour period of LMPs. Scenario 1 (Market) allows the market operator to efficiently dispatch the resource every hour. For simplicity, ramp rates and other constraints that may cause inefficiencies are ignored.[12] In nearly every time period, the unit's output changes as a function of the LMP. This is in contrast to each of the self-scheduling scenarios shown in Scenarios 2 through Scenario 5. In Scenario 2 (Min Self), the supplier simply schedules itself at its minimum capacity level for all hours. In Scenario 3 (Max Self), the supplier schedules itself at its maximum capacity. In Scenario 4 (Mid Self), the supplier schedules itself at a level in between its minimum and maximum capacity. Finally, in Scenario 5 (Lag LMP), the supplier uses the LMP from the previous hour to predict where it should schedule itself for the following hour.

[12] This is a simplistic example; for a more rigorous analysis of this topic, these constraints should be included.

Table 4-2. Twelve-Hour Example for Allowing the Market Operator to Efficiently Dispatch the Output of a Resource (Scenario 1) Versus Various Self-Scheduling Techniques (Scenarios 2–5)

Hour	LMP ($/MWh)	Scenario 1 Market	Scenario 2 Min Self	Scenario 3 Max Self	Scenario 4 Mid Self	Scenario 5 Lag LMP
1	$45.41	286	286	377	300	286
2	$49.65	349	286	377	300	286
3	$52.27	377	286	377	300	349
4	$51.37	376	286	377	300	377
5	$48.32	322	286	377	300	376
6	$46.45	286	286	377	300	322
7	$46.35	286	286	377	300	286
8	$50.97	376	286	377	300	286
9	$49.44	349	286	377	300	376
10	$44.70	286	286	377	300	349
11	$48.51	322	286	377	300	286
12	$51.13	376	286	377	300	322

Table 4-3 shows the revenue, cost, and profit results for all five scenarios. The profits from each of the self-scheduling cases are compared to Scenario 1: Market. Thus, the right-most column shows how much profit the supplier loses by not offering its flexibility into the market. Note that each of the self-scheduling cases results in lost profits compared to the market case. Even with intelligence in the self-scheduling strategy (Scenario 5), it would still lose out on $812 during a 12-hour period. Although the lost profit is small relative to the total profit, it will be highly dependent on the cost curve and prices during different time periods. For example, if the cost for the first segment of Table 4-1 (up to 286 MW) were $47 rather than $35, the profits lost would be the same as Table 4-3, but the total profits would be an order of magnitude less, making the relative profit loss much more significant.

Table 4-3. Revenue, Cost, Profit, and Profit Lost for Various Self-Scheduling Techniques

Scenario	Total Revenue	Total Cost	Total Profit (Total Revenue Minus Total Cost)	Profit Lost
Scenario 1 Market	$195,668	$147,313	$48,355	N/A
Scenario 2 Min Self	$167,326	$120,120	$47,206	$1,148
Scenario 3 Max Self	$220,567	$173,594	$46,972	$1,382
Scenario 4 Mid Self	$175,517	$128,079	$47,438	$916
Scenario 5 Lag LMP	$190,452	$142,909	$47,542	$812

4.3.2 Five-Minute Scheduling and Five-Minute Settlements

The real-time market scheduling intervals of all the regions that operate wholesale markets in the United States are shorter in time resolution than scheduling intervals of utilities prior to restructuring. In fact, all market regions in the United States now schedule the real-time market and real-time output of resources that offer their flexibility at a 5-min interval, updated every 5 minutes. This allows for better pricing of actual conditions on a more granular scale and provides incentives for resources that can follow the prices. Because ramp constraints are used in the market-clearing model to constrain the ability of a supplier to sell energy into the market when they do not have the flexibility to follow prices, the selection of supply into the energy market should be based on actual capability when ramp constraints, provided by the resources are based on physical ramp rates. For example, in the hypothetical example in Table 4-2, Hour 7 to Hour 8, the supplier changes its output from 286 MW to 376 MW. Although it may be possible that this change could be made during 60 minutes, it may not be possible for most thermal plants to execute this ramp during a 5-min period. If the supplier had a ramp rate of 5-MW/min, it could reach only 311 MW in the next 5-min interval, resulting in $99 of lost profit (although ignoring the fact that it is now MW in 5 min rather than MWh). In this way, the 5-min dispatch provides an incentive for flexibility in response to quickly changing prices.

Most of the 5-min energy markets that are currently in place in the United States do a good job of extracting flexibility without resorting to a separate market for a specific product for ramping capability, as illustrated by the example above. Units that bid into the 5-min energy market are obligated to ramp to their set point by the time the market period begins. Because these set points are calculated so frequently, many units ramp a substantial portion of the time. This allows for the most economic provision of energy given the constraints on the transmission system, and units can take advantage of price volatility when they can ramp faster. However, in some instances, it may be that ramp constraints can give the opposite effect. A resource that is ramp constrained will not set the LMP, because a faster, more expensive unit would have to be used to make up for the slower unit's ramp limitation. Thus, the more expensive unit will set the price, while at the same time giving a higher revenue opportunity for the slower unit.

Milligan and Kirby (2010) provide a simple illustration of the issue depicting the real-time market. In this example, a single time period market assumption is used without any look-ahead function in the dispatch interval, so that the future expectations cannot affect the current energy pricing. The example is a simplistic power system with only two generators: a baseload unit that has a marginal cost of $10/MWh and a peaking unit that has a marginal cost of $90/MWh, as illustrated in Figure 4-1. During a steep ramp that is beyond the speed that the baseload unit can respond, but within its capacity range, the peaking unit is dispatched to cover the ramp and meet the load. After the baseload unit catches up, the peaking unit is shut off. However, the peaking unit sets the energy price at $90/MWh during the time it is used to meet the load during the ramp period. Although this is not a problem per se, the baseload unit also collects $90/MWh during this period, which does not provide the baseload unit any incentive to become more flexible, and in fact it may provide a disincentive. This topic should be studied further to see what consequences may occur. For example, a multi-period dispatch looks ahead to ensure that units can meet upcoming ramping requirements. This can change the market prices and revenues for market participants, even if pricing is set only for the current interval. In fact, many U.S. market operators have or are developing proposals to move toward multi-period dispatch when solving the real-time market (O'Neill et al. 2011).

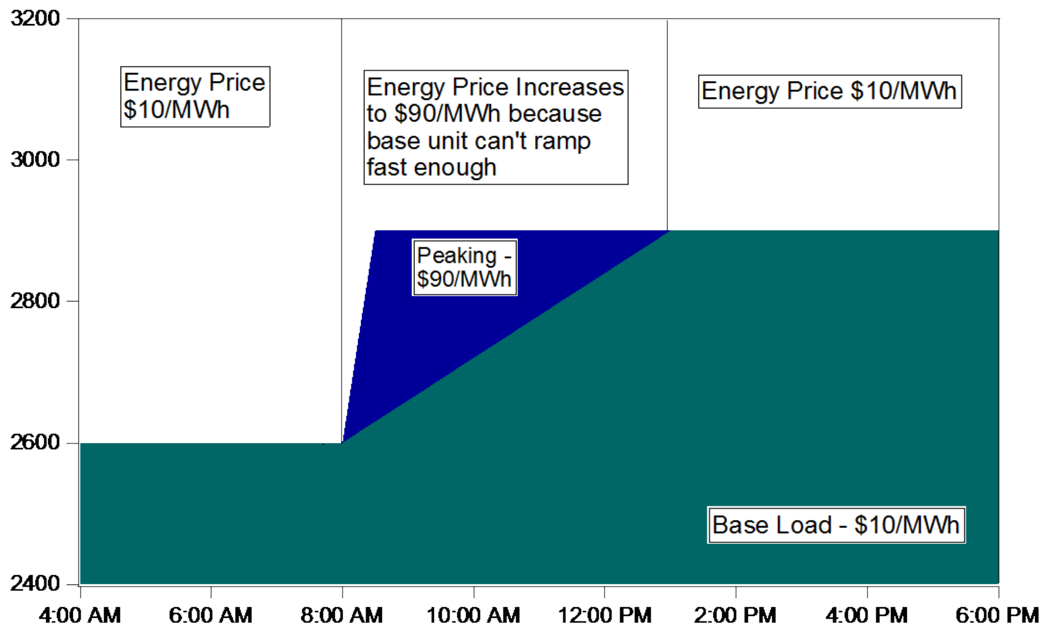

Figure 4-1. Simple example of ramp-limited resources and resulting prices

Although all of the U.S. markets have 5-min real-time energy markets that dispatch and price energy at 5-min intervals, not all of these markets settle at this granularity (Hirst 2001). Many settle the real-time markets based on the average hourly price of all intervals within that hour. Some areas, including NYISO and SPP, however, do settle at 5-min intervals. SPP states that 5-min settlement incents the submission of ramp capability by resources precisely because the capability to move quickly is rewarded by an LMP commensurate with the 5-min instructions.

SPP further explains that without this settlement feature, resources may be disinclined to offer all of their ramp capability, perceiving that they are not being fully compensated for the actions required.[13]

To provide an illustration of how the settlement period can have an impact on incentives for flexible operations, we develop a simple example. Table 4-4 shows 12 5-min LMPs, from real LMP data. Column 3 shows the average LMP for the hour, which is calculated based on the cumulative average LMP from the beginning of the hour to the current time period. We use the same incremental costs for the supplier as shown in the earlier example in Table 4-1. In Scenario 1 (Market), the supplier follows a schedule, as in Section 4.3.1, based on the most efficient output level that the market operator computes and directs each 5-min period. In Scenario 2 (Moving HourlyAverage), the supplier follows an output that is based on the current hourly average LMP (from column 3), because it gets updated throughout the hour. We ignore any impacts from uninstructed deviation penalties throughout this example. Finally, Scenario 3 (Perfect Knowledge) shows a hypothetical example of what the most efficient output would be from the supplier if it had perfect knowledge of the final average hourly price ($49.10).

[13] See http://elibrary.ferc.gov/idmws/common/OpenNat.asp?fileID=12905230.

Table 4-4. Incentive Differences When Settlements Are Done on an Hourly Average Versus 5 Min

Interval	LMP ($/MWh)	Current Hourly Average	Scenario 1 Market	Scenario 2 Moving Hourly Average	Scenario 3 Perfect Knowledge of Hourly Average
0:05	$73.68	$73.68	377	377	340
0:10	$41.87	$57.78	286	377	340
0:25	$43.48	$53.01	286	377	340
0:20	$44.17	$50.80	286	376	340
0:25	$45.75	$49.79	286	358	340
0:30	$46.69	$49.27	286	340	340
0:35	$46.73	$48.91	286	331	340
0:40	$45.91	$48.54	286	322	340
0:45	$61.25	$49.95	377	358	340
0:50	$47.88	$49.74	304	358	340
0:55	$47.88	$49.57	304	349	340
1:00	$43.85	$49.10	286	340	340

An examination of the different dispatches in the table shows that the maximum flexibility is achieved in Scenario 1 (Market). When maximizing profit based on the anticipated or predicted hourly settlement (Scenario 2 and Scenario 3), the unit provides the incorrect level of or no flexibility. It is a simple extrapolation of this scenario that would illustrate similar behavior if the assumption of perfect foresight is relaxed and the unit bids to another hourly average price level. Because price changes in each of the 5-min periods, this is an indication that the system needs a varying level of output; otherwise the price would have remained constant throughout the hour. Next, we turn to an examination of the profits earned in these scenarios.

Table 4-5 presents the total revenue, total cost, and total profit for each of these scenarios with both 5-min settlement and hourly average settlement procedures. If settlements are based on the hourly average, as they are in many markets today, the supplier will earn more profit by producing output differently than the dispatch schedule that was given by the market operator. This would result in output levels that are not the most efficient and could potentially result in reliability issues. In the hourly settlement case, the profit almost doubles when the supplier follows the hourly average price compared to the market schedules ($4,068 compared to $2,443). This shows that even though 5-min scheduling is present in almost every U.S. energy market, it is important that the settlement interval length follow the same interval length as the scheduling to incentivize suppliers to provide the flexibility that is needed by the market operator. On the other hand, numerous uninstructed deviation penalties and ex-post pricing rules may also incentivize the supplier to follow the efficient schedules when hourly average prices are used for settlements. There could also be other reasons that require hourly settlements, such as data retention and storage, as well as a desire to limit market complexity. However, from this simplified example, it appears likely that settlements that match the interval length are the most efficient for extracting the desired flexibility needed from market participants.

Table 4-5. Revenue, Costs, and Profits of Different Scenarios With 5-Min Settlements Versus Average Hourly Settlements

Scenario	Total Revenue		Total Cost	Total Profit (Total Revenue Minus Total Cost)	
	5-Min Settlements	Hourly Average Settlements		5-Min Settlements	Hourly Average Settlements
Scenario 1 (Market)	$15,208	$13,338	$10,895	$4,313	$2,443
Scenario 2 (Moving Hourly Average)	$17,479	$17,441	$13,373	$4,106	$4,068
Scenario 3 (Perfect Knowledge of Hourly Average)	$17,134	$17,134	$13,053	$4,081	$4,081

4.3.3 Existing Ancillary Service Markets

As discussed in Section 2.2, a number of ancillary service markets exist in both day-ahead and real-time electricity markets. These services are generally active power capacity that is held as operating reserve and used for various reasons and at various timescales. These markets are usually co-optimized with the energy market so that the market operator is able to efficiently schedule suppliers for both energy and ancillary services and price both services accordingly. In co-optimized energy and ancillary service markets, the ancillary service prices will be set based on the bid-in cost to provide the ancillary service as well as the lost opportunity cost to provide energy or a separate ancillary service. This will cause the marginal suppliers to provide energy or ancillary services and give further flexibility to the market operator.

Because the existing ancillary service markets were discussed in detail in Section 2.2 and Section 2.3, we refer the reader back to those sections for more information on the current ancillary service markets. From Figure 2-2 and the bullet points describing the ancillary service markets in Section 2.2, regulating reserve (or regulation), secondary reserve—contingency (or spinning reserve), and tertiary reserve—contingency (nonspinning, 30-min reserve, or replacement reserve) are the three ancillary services that are common among all U.S. markets. An important factor is that all of these ancillary services limit the amount of capacity that can be sold to the market based on the response speed of the supplier. For example, regulating reserve typically limits the capacity by how much the resource can provide in 5 mins. Secondary contingency reserve (spin) is typically limited by 10 min worth of ramp response. The faster the resource can adjust its output, the more it can sell into these ancillary service markets.

For example, if two resources have 50 MW of capacity available, with one having a 1-MW/min ramp rate and another having a 2-MW/min ramp rate, they would be able to provide 5 MW and 10 MW of regulation reserve, respectively. If the price of regulation were 10$/MW-h, the second resource would receive $50 more revenue than the first, even though they had the same capacity available. This is an obvious incentive for resources to improve flexibility by way of faster response rates.

4.3.4 Make-Whole Payment Guarantee

The mechanism of scheduling and pricing suppliers offering into the market described in Section 4.3.1 will theoretically place each market participant in a position to maximize profit, subject to various market and technical constraints. However, because of issues such as non-convex costs, commitment constraints, and out-of-market reliability rules, it is possible for the market operator to direct the flexible supplier to provide an energy and ancillary service quantity and for the market participant to lose money when following this direction. For this reason, additional business rules have been designed as part of the U.S. electricity market design to further incentivize suppliers to offer into the market and allow the market operator to commit and dispatch the supplier's output when market prices alone may not provide sufficient revenue to cover all operating costs. One of these is the make-whole payment, also called bid production cost guarantee (NYISO), revenue sufficiency guarantee (MISO), and operating reserve credit (PJM.[14] This payment ensures that suppliers that offer flexibility into the market are guaranteed to be made whole to their offer cost when that bid clears the market. If the revenue that the supplier makes based on the market prices (LMP and ASCP) is less than the supplier's bid cost, the supplier is made whole, with the market operator paying the supplier the difference as a side payment. The offer cost will typically include a three-part offer, including no-load (or minimum generation) costs, start-up costs, and incremental energy costs. It may also include the costs that the supplier has bid in to supply ancillary services. The make-whole payment will make it so the total profit is at least zero. A simplified form of the make-whole payment is shown in the equations below. This is typically netted for all hours of a single day, and it is typically performed separately for both day-ahead and real-time markets to incentivize participating as a flexible resource in both markets.

$$
\begin{aligned}
TotalCost \\
&= NoLoadCost + StartupCost + IncrementalCost \\
&\quad * EnergySchedule + AncillaryServiceCost \\
&\quad * AncillaryServiceSchedule
\end{aligned}
$$

$$TotalRevenue = EnergySchedule * LMP + AncillaryServiceSchedule * ASCP$$

$$If\ TotalRevenue < TotalCost$$

[14] PJM, PJM Manual 28: Operating Agreement Accounting, Revision: 65, Effective Date 4/24/2014, accessed May 2014; NYISO Market Services Tarriff, accessed May 2014; MISO Business Practices Manual-005, Market Settlements, accessed May 2014.

$$MakeWholePayment = TotalCost - TotalRevenue$$

$$Else$$

$$MakeWholePayment = 0$$

If a supplier is the marginal resource and sets the LMP, it will earn enough revenue to recover its incremental energy cost. However, assuming it is marginal for its entire period being online, it will not earn enough revenue to recover its no-load or start-up cost because these costs are generally not part of the LMP (see Section 4.4.4 for an exception). This would cause a disincentive for offering flexibility into the market. Similarly, a unit may be needed for voltage or stability constraints, which are typically not part of the market model constraints. To offer the flexibility to maintain reliability, the market operator will guarantee that a flexible supplier recovers all operating costs associated with supplying energy and ancillary services.

Self-scheduled resources would not receive this guarantee, because they are not giving a bid-cost to the market nor are they offering the flexibility for the market operator to commit and dispatch the supplier's output. Therefore, self-scheduled resources will have no guarantee that they will be made whole to their costs, and, depending on pricing outcomes, they can make less revenue than it costs them to be committed and supply energy, leading to negative profits.

4.3.5 Day-Ahead Profit Assurance

Another settlement mechanism in place today to incentivize suppliers to participate in the market and allow for the market operator to commit and dispatch the supplier's output is the day-ahead margin assurance payment (DAMAP). This mechanism is in existence today in a number of ISOs. It ensures that when energy schedules are reduced in the real-time market from their day-ahead energy schedules, this will not adversely affect the profit margin the suppliers made in the day-ahead market. The purpose of the DAMAP is to provide an incentive for the market participants to be flexible in offering into the real-time market and to be used by the market operator when conditions in the real-time market have changed without being negatively affected. If the real-time market adjusted the supplier output such that it would receive more profit by not operating as the market operator suggests and operating as it was scheduled in the day-ahead market, reliability could be adversely affected. The DAMAP will incentivize resources to offer their flexibility in the real-time market by guaranteeing the profit it received regardless of real-time outcomes. The DAMAP calculation is described below.

$$\begin{aligned} DAMAP &= Max\{0, (DayAheadEnergySchedule - RealTimeEnergyOutput) \\ &* (DayAheadPrice - RealTimePrice)\} \end{aligned}$$

An example of a unit receiving a DAMAP payment after providing flexibility in both the day-ahead and real-time markets is shown in Table 4-6. The right-most column also shows the combined effect from day-ahead and real-time in brackets.

Table 4-6. Example of Unit Receiving Day-Ahead Margin Assurance Payment (DAMAP) After Being Reduced in the Real-Time Market

	Day-Ahead	Real-Time (Combined)
Cost	$50/MWh	$50/MWh
LMP	$60/MWh	$55/MWh
Schedule	200 MWh	-50 MWh (150 MWh)
Revenue	$12,000	$-2,750 ($9,250)
Cost	$10,000	$7,500
Profit	$2,000	($1,750)
DAMAP	$250	

The supplier was asked to reduce output in the real-time market. In doing so, it would lose the profit made initially in the day-ahead market because its total profit goes from $2,000 to $1,750 after the real-time market. In this case, the market operator wants the resource to reduce output to maintain reliability and increase efficiency. The DAMAP of $250 ($1,750 + $250 equaling the initial $2,000) is paid to make up for the lost profit and ensure that the supplier will have an incentive to continue to provide its flexibility to the market.

This calculation will ensure that suppliers that offer the flexibility to be adjusted in real time when conditions require a reduction in output are not financially harmed from the position the market operator scheduled them at in the day-ahead market. The DAMAP applies to day-ahead energy and ancillary service markets. Self-scheduled resources that do not offer flexibility are not guaranteed this payment when output changes in real time. This results in a further incentive to offer flexibility to the market rather than self-scheduling.

4.4 Emerging Market Design Elements That Impact Flexibility Incentives

Efficient operation through centralized dispatch, frequent scheduling and short settlement intervals, ancillary service markets, make-whole payments, and day-ahead profit assurance payments are traditional ways that suppliers have been incentivized to offer flexibility to the market operator and allow it to commit and schedule the supplier’s output. However, it is unclear whether these existing design elements provide sufficient incentives to ensure an adequate level of system flexibility and that the available flexibility is accessible to the system operator when needed. To some extent, new market designs have recently been implemented to incentivize increased flexibility on the system when that flexibility is needed. We discuss a number of more recent market design changes that may have some influence on incentivizing further flexibility from suppliers.

4.4.1 Flexibility from Nontraditional Resources

The suppliers that have traditionally provided flexibility in the energy and ancillary service markets are thermal and hydro power plants. These resources are able to adjust output at various response speeds with absolute power ranges typically in the range of 50% of total capacity. The wholesale electricity markets were initially designed with these technologies, thermal plants in particular, and their characteristics in mind. Given new characteristics of emerging technologies, adjusting market rules that were designed with other technologies in mind may be required. Some recent changes in the wholesale markets have been made to accommodate such resources as energy storage, demand response, and VG itself. In a number of ways, these emerging technologies can provide flexibility as well as or in some ways superior to the traditional thermal and hydro resources.

One of the most significant market rule changes has been made for further adoption of demand-response resources as suppliers of energy and ancillary services. In 2011, FERC issued Order 745 (FERC 2011a) on Demand Response Compensation in Organized Wholesale Energy Markets. The order directed the wholesale market operators to pay demand-side resources that curtailed their load when directed by the energy LMP, as long as a net-benefits test was used that showed providing the demand response reduced costs per unit to consumers. In addition, many of the market operators have also implemented ways in which demand can participate in ancillary service markets. In ERCOT, nearly half of the contingency reserve needed is supplied by demand-response resources that curtail when system frequency reaches some level below nominal frequency (Huang et al. 2009). Other markets have limitations on how much ancillary services can be provided by DR (ISO/RTO Council 2014). These limitations do have justification in terms of both reliability and economics, but at the same time the ability to utilize demand response introduces a great new source of flexibility that was not historically available. The participation of demand response on wholesale markets is however, subject to significant uncertainty. In 2014, the US Court of Appeals vacated FERC Order No. 745, finding that the FERC overstepped its authority by encroaching on states' jurisdiction of the retail electricity market. The court also noted substantive errors with the FERC's compensation rules. The impacts of this ruling have yet to be realized and may result in substantial changes in how this form of flexibility can participate in wholesale energy markets.[15]

Energy storage is another resource that has tremendous flexibility. Energy storage can effectively double its absolute power range because it can act as a supplier as well as a demand. Most energy storage resources also have superior response rates and much less limiting commitment constraints than thermal units. However, they do have limitations on the amount of time they are able to sustain energy levels. This energy limitation varies depending on the type of energy storage, with some, such as large pumped storage hydro, able to sustain for multiple hours to days; whereas other storage devices, such as flywheels, although extremely fast-responding, can sustain maximum power for only 10 to 30 min. Some of the market operators have been adjusting market rules to allow for extraction of the tremendous flexibility from energy storage. For example, PJM now has full optimization of pumped storage resources in day-ahead markets (O'Neill, Dautel, and Krall 2011). Other markets have adjusted automatic generation control

[15] See http://www.cadc.uscourts.gov/internet/opinions.nsf/DE531DBFA7DE1ABE85257CE1004F4C53/$file/11-1486-1494281.pdf.

algorithms to allow for extracting the extremely fast speed and flexibility of limited energy storage resources while keeping track of its energy discharge level (Allen et al. 2009). Some issues still exist in the markets on furthering the ability to treat storage in a way that incents storage to provide its full flexibility potential (Ela et al. 2013a).

Finally, recent changes have enabled market operators to extract flexibility from VG. Previously, all VG were treated very similar to negative load, in that their outputs were considered fixed and other resources needed to adjust output for the variability and uncertainty that occurred. At first, this practice seemed intuitive. VG has zero or very low variable costs, meaning that it is most cost efficient to use as much power from VG as available. However, due to transmission constraints or minimum generation constraints of thermal units, it was found that there are some instances in which VG could curtail its output, do so quickly, and help balance the system and maintain reliability and security limits. The NYISO proposed this with its wind resource management program. Subsequently, PJM adjusted market rules to allow for negative prices and allow for wind curtailment for economics and reliability. MISO then implemented its Dispatchable Intermittent Resources program, which allows for the economic dispatch of wind energy. Although the rules governing each implementation are slightly different, they all allow for wind to be dispatchable in real-time markets to manage transmission congestion and meet load efficiently. This provided a great new source of flexibility.

To explain why these programs are beneficial, we show an example from Ela and Edelson (2012b). Figure 4-2 shows a simplistic three-bus system, with a cheap generator at Bus 1 (G10), an expensive generator at Bus 3 (G3), and a wind power plant at Bus 2. The 250-MW load (L3) is located at Bus 3, the reactances of all lines are equal (X12=X13=X23), and there is a transmission limit on the branch from Bus 2 to Bus 3 of 100 MW (L23). We perform two market solutions, first with wind as a non-dispatchable price taker. The second—similar to the programs in NYISO, PJM, and MISO—allows for wind to be dispatchable in the market at an offer price of $0/MWh.

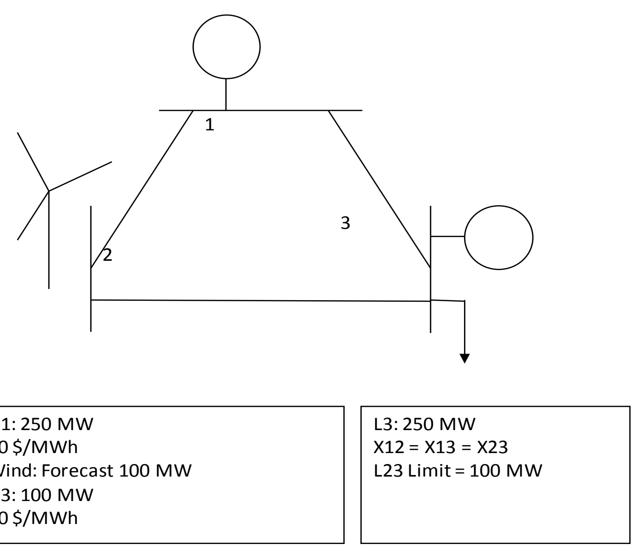

Figure 4-2. Three-bus example to explain the benefits of wind on dispatch

Table 4-7 shows the production, production costs, and LMPs with wind power fixed, and Table 4-8 shows the same results with wind as a flexible producer. Rows 2 through 4 add 1 MW of load to each bus as a way to approximate what the LMP is (because LMP is the marginal cost of meeting an increment of load at each location). With wind fixed, G3 is required because of the transmission constraint that limits G1's output to 100 MW. The wind power plant receives a negative LMP, and the load must pay the expensive LMP based on G3's marginal cost. A different solution results with a market that allows for wind to be flexible in the market (Table 4-8). The wind generator reduces its output and no longer has a negative price, the production costs are reduced by more than 40%, and the price the load pays is cut by 60%. This shows that enabling VG to be a flexible resource has great benefits for improving efficiency and increases the flexibility pool.

Table 4-7. Production, Production Costs, and LMPs with Wind as a Price Taker

	Wind MW	G1 MW		G1 Cost		G3 MW		G3 Cost		Total Cost	LMP
Base Case	100	100	*	$10/ MWh	+	50	*	$50/ MWh	=	$3,500	
Add 1 MW to Bus 1	100	101	*	$10/ MWh	+	50	*	$50/ MWh	=	$3,510	$10
Add 1 MW to Bus 2	100	102	*	$10/ MWh	+	49	*	$50/ MWh	=	$3,470	$-30
Add 1 MW to Bus 3	100	100	*	$10/ MWh	+	51	*	$50/ MWh	=	$3,550	$50

Table 4-8. Production, Production Costs, and LMPs With Wind on Dispatch

	Wind MW	G1 MW		G1 Cost		G3 MW		G3 Cost		Total Cost	LMP at Bus
Base Case	50	200	*	$10/ MWh	+	0	*	$50/ MWh	=	$2,000	
Add 1 MW to Bus 1	50	201	*	$10/ MWh	+	0	*	$50/ MWh	=	$2,010	$10
Add 1 MW to Bus 2	51	200	*	$10/ MWh	+	0	*	$50/ MWh	=	$2,000	$0
Add 1 MW to Bus 3	49	202	*	$10/ MWh	+	0	*	$50/ MWh	=	$2,020	$20

In addition to VG providing flexibility in the energy market and assisting in congestion management, recent discussions have also explored the ability of VG to offer its flexibility in the ancillary service markets. Wind power can provide various forms of active power control using a combination of mechanical pitch and torque control and power electronics control (Ela et al. 2014). In many ways, it can provide a desired response faster than thermal plants are able to. Research has looked at the ability of wind power to participate in regulating reserve markets (Kirby, Milligan, and Ela 2010; Liang, Grijalva, and Harley 2011; Tuohy et al. 2012). These works have shown that in certain instances wind power can earn revenue and reduce total costs to consumers by providing regulating reserve. Further research is evaluating the effect that wind output forecast errors may have on its provision of regulating reserve. It is possible that wind, and potentially other VG such as solar, can play a role in providing a number of ancillary services and further add to the flexibility pool.

4.4.2 Evolving Regulating Reserve Markets (Order 755)

Some recent changes have been made to the ancillary service markets to change the ways in which resources are incentivized. The most significant changes have been made to the regulating reserve markets. In late 2011, Order 755 was issued by FERC on Frequency Regulation Compensation in the Organized Wholesale Power Markets (FERC 2011b). The order directed market operators that are part of the organized wholesale markets to include market-based payments for regulating reserve performance, lost opportunity costs for all regulating reserve capacity prices, and incentives and rules for accuracy. The order did not require any standardization between markets and also made no changes to the net energy payments that came as a result of the energy from regulating. At present, all markets except ISO-NE have implemented the changes for Order 755. In addition, ERCOT, though not FERC jurisdictional, has initiated a pilot program on fast regulation response service, which is in many ways analogous to the implementations made to meet Order 755 in the other markets.

Although many markets already included lost opportunity costs in regulating reserve markets, the order enforced this. Historically, a few areas had paid only the lost opportunity cost to the suppliers that incurred these costs. With the order, it was decided that the lost opportunity cost is part of the marginal cost of providing regulating reserve capacity and should therefore be a part of the price paid to all regulating reserve suppliers. The order also stated that the market operators are responsible for assigning the lost opportunity costs, but that intertemporal opportunity costs (i.e., by providing regulating reserve in the current hour, a supplier may lose out on energy profits in future hours) must be verifiable and can be included in a supplier's regulating reserve offer.

Historically, ancillary service markets are paid only for the capacity that suppliers held to provide the ancillary service and not the actual utilization of the capacity for the ancillary service (Rebours 2007). This order adjusts the payments for regulating reserve so that the resource is paid based on how much it was asked to control during each market interval as well as how accurate it followed its automatic generation control signal. The performance price must be market based rather than administratively set, and the performance is based on the absolute amount of movement that a supplier performs in a market interval. Suppliers that are asked to move up and down at a higher frequency would therefore be paid more for performance than those being asked to move more slowly. In addition, the closer in accuracy the supplier followed the automatic generation control signal, the more value it would receive as well. Exactly how the accuracy was measured would vary in each market, but the order required that the accuracy is based on how well a resource follows the control signal and not how well it follows area control error, and that all resources' accuracy is measured by the same means. This design would then incentivize resources that are more flexible and can provide regulating reserve faster and more accurate by providing greater payment than that made to slower and less accurate regulation resources.

Although the order had its objectives toward incentivizing suppliers that provided regulating reserve to be faster and more accurate, there was not a consensus on the benefits of the order. Many commenters on the order believed that the faster response would not have any significant reliability benefit and would only raise costs to consumers. Proponents of the order suggested that the introduction of performance payment would reduce the regulating reserve capacity prices. Other proponents also argued that the use of faster ramping resources would improve efficiency of meeting regulating reserve requirements and thereby reduce the capacity requirement of regulating reserve. This was also shown in other studies that analyzed the impact of faster responding resources, such as Makarov et al. (2008). Table 4-9 shows ancillary service prices for a time period when Order 755 was implemented in NYISO (June 26, 2013) and then the prices for the same time period during the previous year without Order 755. Although there could be many other reasons this occurs rather than Order 755, prices for all ancillary services increased with the new design, not necessarily supporting the efficiency improvement. Although the argument of whether the implementation of Order 755 improves efficiency as well as the argument of how much it improves reliability should continue to be evaluated, it is clear that it does make the regulating reserve market better suited to incentivize response speed as well as response accuracy, giving a great push toward improved flexibility incentives.

Table 4-9. Ancillary Service Prices of the NYISO During a Period With and Without Regulation Performance Payment

Before 755 (June 26–Oct. 22, 2012)	Spin ($/MW-h)	Nonspin ($/MW-h)	30-min. ($/MW-h)	Regulation Capacity	Regulation Mileage ($/MW)
Average price	$4.00	$1.80	$0.08	$6.44	N/A
Average intervals at 0 price	84.5%	98.0%	99.9%	0.4%	N/A
After 755 (June 26–Oct. 22, 2013)	**Spin ($/MW-h)**	**Nonspin ($/MW-h)**	**30-min. ($/MW-h)**	**Regulation Capacity**	**Regulation Mileage ($/MW)**
Average price	$5.82	$3.26	$1.70	$10.59	$0.23 ($2.30)[16]
Average intervals at 0 price	86.4%	98.1%	99.4%	1.4%	10.2%

[16] NYISO uses a multiplier of 10 for regulation mileage, in that a typical unit will perform mileage equal to 10 times its capacity; therefore, multiplying the regulation mileage price by 10 can give a more relative comparison.

4.4.3 Ancillary Service Markets for Primary Frequency Control

A common argument to Order 755 discussed in the previous subsection was that it considered only secondary frequency control and not primary frequency control (PFC). The arguments stated that FERC cannot look at only secondary frequency control because of the interrelationship between primary and secondary frequency control. In some ways, the argument was that the faster a supplier can follow the automatic generation control signal, the more it could earn, until the automatic generation control signal is too fast and the supplier follows frequency, in which case the supplier gets paid nothing.

PFC is the response typically from synchronous generator turbine governors that responds proportionally to frequency deviations. The aggregate PFC will arrest frequency decline and bring it to a new steady-state level. Synchronous inertia service, which may or may not be included in the definition of PFC, is typically defined as the immediate injection of active power through the stored kinetic energy of the rotating mass of synchronous machines. This response will slow down the rate of change of the frequency decline. Both PFC and synchronous inertia are crucial services needed to maintain a reliable and secure system and avoid under-frequency load-shedding, machine damage, and potential blackouts.

In the United States, there is no reliability requirement for a balancing authority area or market area to have sufficient synchronous inertia or PFC. A recent draft standard, BAL-003-1, would require a minimum amount of PFC that balancing authority areas must have available at all times. Also, currently there are no incentives in place for individual resources to provide either service. Some studies have shown that the frequency response in the United States, especially in the Eastern Interconnection, has been declining during the past 20 years or more (Ingleson and Allen 2010). Some reasons for this include high governor deadbands, generators operating in modes that do not offer frequency-responsive reserve, governors that are not enabled, a reduced percentage of direct drive motor load, and others (Schulz 1999; Virmani 1999). Without any controls or changes to meet the needs, increased penetrations of nonsynchronous VG could further degrade frequency response. However, some have claimed that the wholesale electricity market design, lack of incentives, and even the presence of disincentives to provide the service are among the major causes of the decline (IEEE Task Force 2007).

A potential disincentive was discovered in Ela et al. (2012a). Many market operators have financial penalties in place when suppliers produce outputs different than those that were scheduled. Suppliers providing PFC would automatically adjust output when frequency deviates from its nominal level (60 Hz in the United States) without any control room operator intervention. Few of these markets would use system frequency in their market settlements rules. Therefore, as in the example in Ela et al. (2012a), in a market with a 3% tolerance band, a supplier with a 5% droop curve would be automatically penalized when frequency deviates more than 90 mHz. Meanwhile, this resource is doing exactly what is required to maintain a reliable and secure power system.

Numerous ancillary service market designs for PFC and synchronous inertia have been proposed in the literature, including that from Ela et al. 2014. As VG increases, displacing resources that typically offer PFC and synchronous inertia, the need for incentivizing this service can become more apparent. It can also incentivize resources that would not typically provide these services, such as VG, to install the capability and offer into that market. Recently, in its ancillary service market redesign initiative, ERCOT was the first market to mention its intentions to implement a PFC ancillary service market (ERCOT 2013). After BAL-003-1 was passed, and with the increasing need to incentivize resources to be more flexible and provide these services, it is likely that this trend will continue in the future.

4.4.4 Convex Hull Pricing

Some differences do occur in the way that each ISO prices energy and ancillary services. The marginal pricing theory for energy and ancillary service markets is based on continuous, convex, monotonically increasing variable costs. As a result of primarily no-load costs and start-up costs, actual costs are not convex, and the lumpiness creates additional requirements to ensure efficient market design (Elmaghraby et al. 2004; O'Neill et al. 2005). This creates the need for uplift payments so that resources that do not recover their no-load or start-up costs from the LMP, which is typically based solely on incremental energy costs, will get side payments (see Section 4.3.4).

For energy markets, pricing in some market regions is not exactly based on the pure marginal cost. MISO has discussed the extended LMP (ELMP), which is based on the convex-hull pricing concept (Gribik et al. 2011; Gribik et al. 2007). This is similar to the hybrid-dual approach at the NYISO (FERC 2001). A question arose in the past on the correct price given when peaking gas turbines were turned on to meet high energy demands. When 1 MW of additional load must be served, and a peaking plant with a minimum capacity of 20 MW is turned on to provide its needed energy, the next cheaper unit would be backed down by 19 MW (Stoft 2002). In this situation, the marginal cost of energy would be the bid-based cost of the cheaper unit. This means that even though the more expensive unit was needed, the marginal cost of energy (and price) was not increased. The peaking unit would get paid less than its bid-based costs, requiring a make-whole payment, and the rest of the generation fleet would earn lower revenues because the marginal-cost-based price was suppressed. The ELMP and hybrid-dual pricing concepts consider the non-convex aspects of the resource costs and constraints as part of their pricing rules. For certain resources it will include the no-load and/or start-up cost of the resource as part of its total bid cost, meaning these nonconvex costs can influence the price. The benefit to this approach is more transparency in pricing to the more expensive resource by having prices better reflect actual costs. The convex-hull pricing approach is currently an ongoing debate.

Because of the increased variability and uncertainty of VG on the system, additional resources may be committed without being economic according to their energy costs. This could lead to increased times where resources online would not recover their costs because energy prices are based on the marginal cost of energy. These resources would receive uplift payments, reducing the transparency of prices. Future pricing mechanisms, similar to ELMP or hybrid-dual pricing that incorporate these non-convexities into energy prices should be evaluated with increased penetrations of VG.

4.4.5 Flexible Ramping Products

Finally, it is important that the electricity market designs are incentivizing increased flexibility to provide energy when that flexibility is needed. It is debatable whether incentivizing flexibility is being done efficiently in all U.S. markets today. A few areas have been introducing and proposing changes to their electricity market designs to ensure that energy markets are incentivizing the greater flexibility needed from increased penetrations of VG. Other areas are not presently making significant changes to their designs, perhaps believing that the mechanisms described above in Section 4.3 are enough to incentivize flexibility in the energy markets. Some market areas, namely CAISO and MISO, have begun to introduce explicit markets for energy flexibility as a new ancillary service. We discuss these next.

CAISO has performed a number of studies to analyze the impacts of integrating significant levels of VG on its system. Two of the more recent studies analyzed the impacts that VG has on the capacity and ramping needs for its energy markets. The first study determined that the amount of ramping needs would increase by up to 30 MW/min to 40 MW/min with 20% renewables (CAISO 2007). A later study found similar ramp rate increases and determined that the amount of incremental load-following capacity that would be needed as a result of the variability and uncertainty of VG was 845 MW and 930 MW for upward and downward load-following capacity, respectively (CAISO 2010a). Figure 4-3 and Figure 4-4 were taken from that report and show the total capability of load-following up and load-following down, respectively, from the generating fleet. Figure 4-5 shows the same information as Figure 4-4, except that the total load-following down capability is not limited by resources that are self-scheduled and not offering their flexibility to the market. As shown, some of the early morning hours would not be able to meet the total load-following down capability of 930 MW (Figure 4-4). However, if more resources changed from being self-scheduled resources to flexible resources, the requirement could be met easily (Figure 4-5). This can support the idea that further incentives are needed for the self-scheduled resources to offer their flexibility to meet the increased flexibility needs resulting from increased VG.

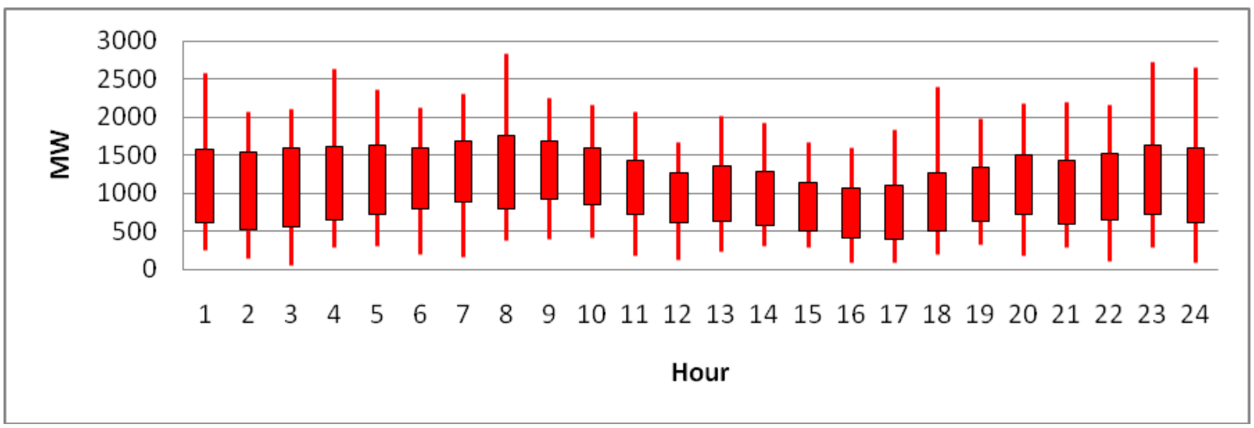

Figure 4-3. Load-following up capability in generating fleet, summer period 2009–2010 (CAISO 2010)

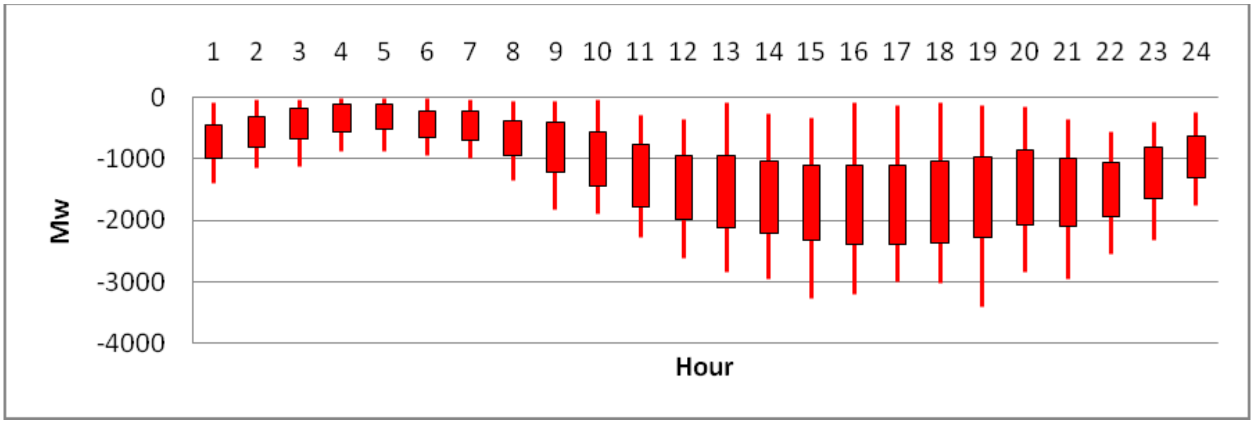

Figure 4-4. Load-following down capability in generating fleet, summer period 2009–2010 (CAISO 2010)

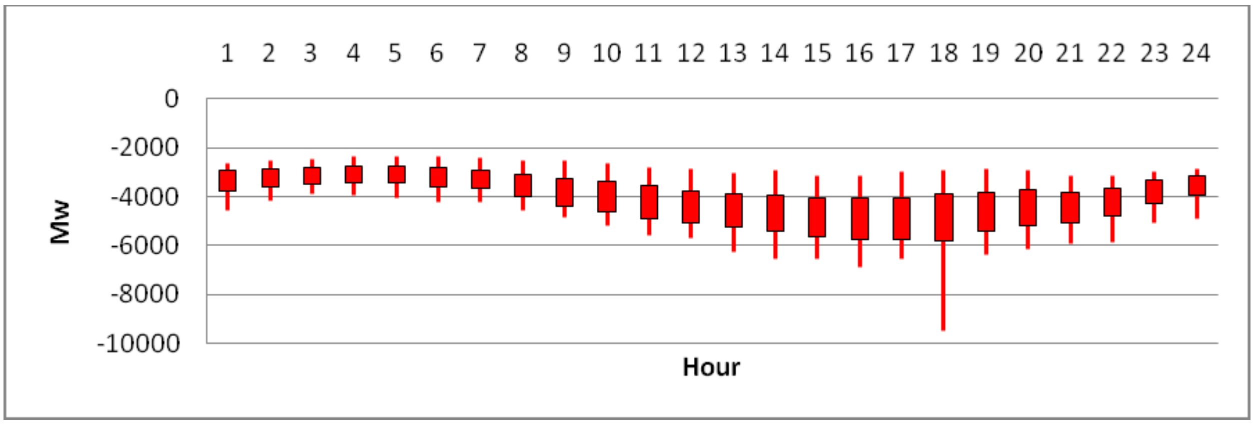

Figure 4-5. Load-following down capability in generating fleet if self-scheduled resources are included, summer 2009–2010 (CAISO 2010)

In August 2011, the CAISO board of governors approved a flexible ramping constraint mechanism in the ISO energy market design (CAISO 2011, Abdul-Rahman et al. 2012). This is an additional constraint added to the market-clearing engine that ensures that sufficient ramping capacity is committed and available in the real-time commitment and real-time dispatch process. The use of this constraint reduces infeasibilities in the dispatch procedure compared to when ramp capability is not committed, reduces the need for reliance on regulation reserve and relying on neighboring balancing errors, and eliminates the need to biased hour-ahead forecasts to prepare for potential variations in real time. At present, the constraint was only for upward ramp capability needs. The amount of ramp capability that is required in the constraint is determined by the CAISO operators based on the (1) expected level of variability for the interval, (2) potential uncertainty as a result of load and VG forecast error, and (3) differences between the hourly, 15-min average net load levels and the actual 5-min net load levels. These levels are determined from historical data, and the total requirements are published for the various market processes.

Similar to other ancillary service products, there is a potential for a lost opportunity cost for resources that are withholding their capacity to meet this ramping constraint. If a resource foregoes profit in the energy market or other ancillary service market to reserve capacity for the ramping constraint, it has a lost opportunity cost for serving the ramp constraint. In the current market design for this flexible ramping constraint, all resources that meet this ramping constraint with capacity that is not being used for other ancillary service products are paid the marginal resource's lost opportunity cost. The value is based on the incremental cost that would be incurred by the system if increasing the ramping need by one unit. Currently, there is no allowance for other costs associated with ramping to be added—i.e., no separate bid for this product—therefore, the only way for a resource to get paid to provide ramping capability under this ramping constraint would be if a lost opportunity cost were incurred. All of the costs associated with the price paid to suppliers selected for the flexible ramping constraint are currently paid by demand.

During the first few years of the flexible ramping constraint, some observations can be made. In 2012, the total cost to this constraint was approximately $20 million, compared to $35 million for the spinning reserve market (CAISO 2012). It was found that during this time period much of the flexible ramping capacity was in the northern part of the system and often unavailable to provide assistance in relieving congestion in the southern part of the system. Table 4-10 shows additional statistics from the first year the constraint was enacted. The table shows total payments, the percentage of intervals in which the constraint was binding (i.e., nonzero, when the constraint required change in dispatch), the percentage of intervals in which the flexibility constraint requirement could not be met (i.e., had a procurement shortfall), and the average price of the constraint when it was binding. The spring time period had the highest prices, as well as the most binding and short intervals. Payments seemed to be greatly reduced by the end of the year. However, the first quarter of the following year 2013, the costs of flexible ramping constraint were $10M, half that of the entire previous year, whereas spinning reserve costs were approximately $6M (CAISO 2013b). This was mostly because the ISO increased the requirement more consistently than it did in 2012. The costs reduced by the end of the 2013.

Table 4-10. Statistics for the First Year of Flexible Ramping Constraint in CAISO (CAISO 2012)

Month	Total Payments to Generators ($M)	Intervals Constraint Was Binding (%)	Intervals with Procurement Shortfall (%)	Average Shadow Price When Binding ($/MW-h)
January	$2.45	17%	1.0%	$38.44
February	$1.46	8%	1.3%	$77.37
March	$1.90	12%	1.0%	$42.75
April	$3.37	22%	1.5%	$39.86
May	$4.11	23%	6%	$79.48
June	$1.49	13%	2.3%	$77.37
July	$1.01	8%	1.4%	$42.75
August	$0.77	7%	1.2%	$39.86
September	$1.03	13%	0.8%	$79.48
October	$0.9	9%	1.0%	$39.19
November	$0.23	4%	0.5%	$53.34
December	$1.09	9%	1.6%	$61.84

After the flexible ramping constraint was approved, the ISO and its stakeholders proposed a full flexible ramping product similar to other ancillary service products. The ISO, along with stakeholders and its board of governors, agreed that greater market effectiveness could be gained by developing market-based products that can better identify, commoditize, and compensate for this flexibility. The main differences in the flexible ramping product from the constraint described above are the inclusion of downward ramping, the change in using the 5-min real-time dispatch interval rather than the 15-min real-time pre-dispatch model, the inclusion of the product in the day-ahead market, and a flexibility demand curve. Because CAISO uses a multi-period market-clearing engine, ramping requirements are already within the model based on the expected change in net load from one interval to the next. This is the minimum ramping requirement that must be met. The ISO then will require additional ramping capability requirements above the expected ramping requirement to meet the unexpected ramping capability requirement, which can be as high as the 97.5th percentile change in net demand (or 2.5th percentile for the downward-ramp capability). Between the minimum and maximum ramp need, there is a stepped demand curve. This will ensure that the ISO will procure a certain amount of ramping capability based on both the need and the additional cost. The penalty costs that are a part of this flexible ramping product demand curve are based on the probability of power balance violations as a result of not having ramping capability and the penalty cost of those violations. The maximum price of ramping capability of 250 $/MWh is set when there is not enough ramping capability to meet the minimum ramping need, i.e., the expected ramping need.

Other market and settlement rules accompany the new flexible ramping product. Bids for ramping capability are only allowed in the day-ahead market. The prices that occur in real time are based on only the lost opportunity cost incurred by ramping units not able to fully participate in the energy or ancillary service markets, or from the penalty prices that are part of the flexible ramping product demand curve. The settlement between the day-ahead and real-time markets is performed similarly to other products, including energy. The quantity of flexible ramping capability available in the day-ahead market is sold at day-ahead flexible ramping prices, and the difference in real time is paid (or bought back) at the real-time flexible ramping capability price. Note that the difference takes into account the interval resolution of the different markets (i.e., because the day-ahead market is hourly, it is divided by 12 to calculate the difference from the real-time market ramping capability, because the real-time market is in a 5-min resolution).

The importance of the flexible ramping product can be illustrated with some short examples from Zu and Tretheway (2012). This importance can be shown simply with the expected ramp capability need before discussing the need from unexpected ramp capability. As discussed, CAISO, like many other ISOs, solves the real-time market using a multi-period market-clearing engine (e.g., multi-period security-constrained economic dispatch). The first example is a two-period dispatch solution with a load of 420 MW in Interval 1 and 590 MW in Interval 2. The second example is the same scenario with a flexible ramping product requirement set to require slightly more ramping capability than the inherent need from the first scenario (170.01 MW in 5 min). Both scenarios use a two-generator set with characteristics shown in Table 4-11. Both generators have zero-cost bids for flexible ramping and minimum generation levels at 0 MW.

Table 4-11. Generator Properties for Flexible Ramp Example

Generator	Bid Cost ($/MWh)	Ramp Rate (MW/min)	Pmax
Gen 1	25	100	500
Gen 2	30	10	500

Table 4-12. Scenario 1, Multi-Interval Dispatch

Generator	Interval 1 (LMP=25$/MWh)		Interval 2 (LMP=35$/MWh)	
	Energy	Flex	Energy	Flex
G1	380	0	500	0
G2	40	0	90	0

Table 4-13. Scenario 2, Flex Reserve Product

Generator	Interval 1 (LMP=$30/MWh, FRP=5$/MWh)		Interval 2 (LMP=30$MWh)	
	Energy	Flex	Energy	Flex
G1	379.99	120.01	500	0
G2	40.01	50	90	0

Both scenarios have essentially identical operational results; however, the prices that result are quite different in each scenario because of the flexible ramp requirement. The prices are based on the marginal cost of the entire period for providing the service in the associated time interval. Finally, with the different distribution of pricing, it turns out that the revenues for both units will be equal in both scenarios when both intervals are settled. The main issue that CAISO describes in the draft summary is that the current market design will consider only the first interval during multi-period dispatch as the binding settlement interval. Because the price of the second interval will change when it is the binding interval, the overall result is that the revenues of the units will not be the same, even though the operational results will still remain identical. The first scenario would not have the same incentive for providing flexibility as would the second scenario.

The example above shows some benefit of the flexible ramp product when expected variability is present. The product will also increase the ramping need above the expected ramp to be able to meet the unexpected ramp need (ramp needs that are not forecasted). This allows for resources to have capacity and ramp available in case a ramping event occurs. Pricing based on the lost opportunity cost allows all resources to be indifferent whether providing flexible ramping or energy either for certain or uncertain ramping events.

MISO has proposed a similar product to CAISO, the up-ramp capability and down-ramp capability (Navid and Rosenwald 2012; Navid et al. 2011). MISO had claimed that the most common reason for scarcity pricing conditions in its area was not caused by limited capacity but by insufficient ramp capability. These scarcity conditions were causing large price spikes in the energy and ancillary service markets. The product would be introduced in both day-ahead and real-time markets to ensure enough ramping capability would be available in the future. The product has similar concepts to CAISO, including a demand curve for insufficient ramp capability, pricing based solely on the lost opportunity costs from other products, and a requirement based on historical information to meet both expected and unexpected ramp requirements. MISO has also included deliverability requirements within the ramp capability conceptual design. The schedules for up-ramp capability and down-ramp capability would be made such that, when combined with energy, the full deployment of that capacity will not violate any transmission constraints assuming pre-defined locations of the variability and uncertainty that caused the ramp capability need. This ensures that the locations of reserving the ramp capability will be able to be deployed without transmission constraints when needed. The product is being filed with FERC, and it is not likely to be introduced into the market until 2015.

One key difference between the CAISO and MISO approaches is how the payments for the ramping capability will be allocated. In the current CAISO proposal, the allocation will be based on a cost causation principle while the MISO proposal will be based on primary beneficiaries of the product. In CAISO, the allocation is proposed to be distributed among loads, suppliers, and fixed ramp resources (e.g., external transactions and internal self-scheduled resources). Loads causing need for ramping will be allocated based on their 10-min movement. Suppliers will be allocated based on their 10-min uninstructed deviations from real-time schedules. Fixed ramp resources will be allocated similarly to load based on 10-min net movement. This allocation proposal from CAISO will be the first which an ancillary service costs are reimbursed through cost causation principles as opposed to it being reimbursed fully by load-serving entities.

Other market operators are not seeing the need for this new ramping product to incentivize flexibility. For example, NYISO has suggested that the DAMAP, make-whole payments, and optimal operating resources at their most efficient production level are sufficient and fully incorporate load-following (flexibility) services into pricing and scheduling outcomes (Pike 2013). An additional payment for this flexibility would be unnecessary. Although ERCOT has a supplemental reserve service in its new ancillary service market redesign, similar to those products in CAISO and MISO, it also notes that such a service will not ultimately be required in the long term (ERCOT 2013). This is contrary to the direction of CAISO and MISO. The differences between these market designs should be further studied. For example, the need in CAISO may be because of the higher amount of self-scheduled bilateral agreements that have been in place since the energy crisis in 2000 and 2001, limiting the amount of flexibility available to the system. MISO has a very large coal fleet, which has limited flexibility. NYISO has a significant amount of flexible natural gas on their system, which could be why they have not seen a need yet for incentivizing further flexibility. One size does not fit all when it comes to electricity market designs, and each area has solved its historical issues through specific designs involving the stakeholders and market participants in each area. The topic of flexibility incentives should take a more holistic view to see what the reasons are for further market design changes, how these changes should occur, and how increasing amounts of variable energy resources may affect these market design changes.

4.5 Summary

This section discussed the importance of short-term flexibility in system operations and how that flexibility is needed to better accommodate the increased variability and uncertainty of VG. Flexibility, though often hard to define, is an important characteristic that should be incentivized so that it can be called upon when it is needed. There are many different forms of flexibility and different ways that market operators can extract that flexibility. Market designers have established numerous traditional mechanisms to incentivize resources to offer their flexibility to the market and new, recent mechanisms that attempt to further incentivize increasing levels of flexibility when that flexibility is needed. Things like make-whole payments and evolving ancillary service markets can incentivize flexibility while self-scheduling and hourly settlement intervals can inhibit needed flexibility.

The various regional markets have not converged on approaches to incentivize flexibility in short-term markets. For example, a number of market regions have proposed new flexible ramping ancillary service products to incentivize further flexibility, whereas others have decided this is unnecessary and current mechanisms may already get the needed flexibility. Whether these differences are because of regional differences in the system characteristics, generating portfolios, or existing market rules or procedures is still undetermined. Further research should study these different designs with varying system characteristics to make conclusions on whether certain market mechanisms do in fact fit different characteristics more than others.

A number of new scheduling software programs have been developed that show ways of providing flexibility at low cost and improved reliability with increased levels of VG. The ways in which the pricing is determined in these new scheduling models is not always as straightforward. When new methods appear to reduce costs or improve reliability, it is important that the resulting prices are analyzed to determine whether the resources providing the flexibility to improve system operations are actually being incentivized to do so and that unintended consequences are avoided. Otherwise, although it looks like costs are reduced or reliability is improved in the short term, the resources may not be incentivized to do as directed; therefore, in the long term, these improvements and projected cost savings may not be realizable.

Due to the complicated nature of the electric power system and the relationship of all market products (e.g., energy, ancillary service, FTR, capacity) with each other, it is important that any new modifications to one design do not adversely impact the other. When new designs are made to improve the way that flexibility is incentivized, careful analysis should determine whether this will affect how other markets incentivize other required attributes. Research that goes into new designs should always account for how it may affect other designs. In addition, metrics that are used to show the benefits of any design should be all encompassing. If one design reduces the system production costs, further evaluation should ensure that it either improves or keeps constant the reliability and incentive structure of the system. Therefore, further research into the new designs that may improve incentivizing flexibility should consider all system metrics to the extent that it can, before promoting a new design to be put into practice.

The electricity designs that have been developed in the United States are very sophisticated due to the intricacies with including the physics of the power system within the market mechanisms. Many of the trends of electricity market design evolution, especially with the further improvements with software computational capabilities, have moved toward greater complexity. Another debated topic is whether this complexity is necessary. Should the energy markets be simple with one-part bids and offers? This question will likely continue to be at the center of all market design changes as the thinking continues to evolve.

Many of the mechanisms described in this section could have more significant impacts when even greater penetrations of VG are integrated onto the system. Designs such as primary frequency response markets, pay-for-performance ancillary services, and convex-hull pricing are all in their infancy, and their impacts on a changing system should be analyzed further. The research performed in these areas should help all of the market areas find some consistency going forward when determining appropriate market designs with these continually changing systems.

5 Conclusions and Next Steps

Expanded adoption of VG will increase the net variability and uncertainty that must be managed by power system operators and therefore will require additional system flexibility. Flexibility that is available must have proper economic signals to incentivize its participation in the relevant operational markets. In addition, sufficient flexibility must be developed long term, in the investment timescale, so that it can be available to the operators when needed. The increase in VG will also result in a shifting mix of resources with low variable operating costs and high fixed capital costs, with less load-carrying capability relative to installed capacity. This changing mix can impact how resources that are needed in the long term recover their costs and are incentivized to stay in the market.

Many of the tools necessary to manage these challenges already exist, but the market structures may not be properly designed to incentivize the resources to offer their full capabilities, the efficient use of these resources, or to elicit sufficient investment. Improper utilization of existing flexibility, or unwillingness of resources to provide flexibility, can lead to efficiency and/or reliability degradation. It can lead to the market operator not having access to enough system flexibility to meet the changing net load, resulting in degraded system reliability. More often, it can lead to higher costs when more expensive flexibility is used instead of the most economic flexibility that is not offered into the market. Proper market designs to incentivize flexibility, both in investment and in operation, are critical to meeting these challenges in an efficient manner. In addition, when there is insufficient revenue for resources—those that are needed for long-term reliability—to recover their variable and fixed costs, it can lead to an unreliable system when those resources choose to leave the market. Reliability issues can also arise from miscalculations of the long-term reliability need and how each resource can contribute to meeting it, both for capacity and other attributes such as flexibility. Proper market designs should allow for a sufficient level of resources needed for long-term reliability to recover their overall costs to remain as a market participant. An additional property of well-designed markets is being technologically neutral—specifying what is needed in the system and allowing different technologies to compete.

With increasing penetrations of VG, several potential market impacts were identified. These impacts include the effects on cleared quantities and prices of energy, ancillary services, FTR, and capacity markets. Schedules and prices may be reduced or increased for different products. This causes revenue streams to change, and the important products that market participants offer may be reprioritized. In many ways, these changes are simply reflective of a changing supply mix and a changing service need. In other ways, the changes could lead to unintended consequences that could then lead to either an inefficient or unreliable electricity system or both.

This report reviewed several historical approaches to meeting the challenges of revenue sufficiency for long-term reliability and incenting flexibility in short-term operation and how the approaches have been evolving in recent years due, in some part, to increased VG penetrations. These were determined to be two of the most critical and complicated market design challenges foreseen. It is generally agreed within the industry that these challenges must be met; however, the way that each market has met the challenges and whether or not additional modifications to the market designs are necessary to ensure they will continue to be met is a very contentious and debated issue. Within the United States, each market is evolving differently. Some areas are making significant changes to their market for these reasons. Others have remained relatively unchanged, potentially because the current design has been determined to already meet these challenges sufficiently. These alternative approaches likely arise because each system is so different, with various generating portfolios, economies, transmission networks, history, and regulatory procedures.

Specific market design elements can have an impact on incentives. To illustrate this principle, we examined the impact of market settlement and self-scheduling and found that a more flexible dispatch can be achieved when a generator responds directly to the 5-min prices with 5-min settlements than would be the case with self-scheduling or hourly settlements.

In our discussion of capacity market mechanisms, we showed the difficulty in using peak periods as a proxy for times of system risk and the difficulty—if not the impossibility—of the ability of a capacity market to capture the salient aspects of resource adequacy and reliability. The regions in the United States that have capacity markets have alternative ways of mapping the reliability target to the capacity acquisition target, though there are significant differences in market timing and questions about the ability to capture the long-term investment process to ensure adequacy.

This report focuses on the ways that the industry has changed wholesale electricity market designs to meet the two challenges of revenue sufficiency and flexibility incentives. From this overview, a number of new research opportunities arise to meet these challenges and improve the ways that the markets are designed. Because the way in which power systems are planned and operated are so closely tied to the market designs, these improvements may also strengthen the reliability and efficiency of the power system. A few examples of research areas are:

- Improved methods to calculate resource adequacy requirements
- Improved methods to calculate how resources contribute to resource adequacy
- Incorporating true probability reliability metrics in markets that are used to meet revenue sufficiency and long-term reliability targets and ensure revenue sufficiency
- Improving resource adequacy metrics to incorporate attributes needed for long-term reliability other than capacity
- Developing appropriate mechanisms that incentivize and ensure that attributes needed for long-term reliability other than capacity will be built or installed
- Improved definitions and metrics for defining flexibility attributes in system operations
- Developing better ways of determining how much flexibility is available at given times
- Improved scheduling strategies or models that can better meet the increased variability and uncertainty of VG and extract the correct amount of flexibility required
- Improved pricing methods that incentivize resources to offer and provide flexibility when flexibility is needed

- Improved settlement rules that better incentivize resources to offer and provide flexibility when flexibility is needed

- Design of new or evolving market products that can incentivize flexibility when flexibility is needed.

We have presented the two challenges of long-term revenue sufficiency and incentivizing flexibility in short-term operations independently from each other; however, these two challenges are linked. Flexibility may become the more significant factor for ensuring power system reliability in future systems, and linking reliability requirements between the short-term and the long-term is crucial. In addition, the way in which revenues and incentives are designed in the short-term markets can dictate how they should be designed in the long-term markets. Maintaining a link between the short-term and long-term incentives and reliability needs must be considered in activities related to market design efforts. Inconsistencies between long-term and short-term markets may lead to an ineffective system design that results in stranded flexibility in real time or insufficient flexibility installed on the system.

In the United States, the electricity markets are designed so that they explicitly meet specified reliability targets, and the physics of the electricity system are a basis for how energy and all supporting services are bought and sold. This is mostly true in the market designs for meeting both long-term and short-term reliability needs. This became very clear when presenting the numerous complex designs that have been proposed in these markets to achieve certain goals. It is clear that the reliability of the system and the design of its electricity markets are inseparable. System operators, planners, and reliability regulators cannot make changes to its mechanisms to meet reliability without some recognition of how the electricity market is designed. Likewise, market designers cannot create new or modified designs without recognizing the reliability needs of the system. The communication between different entities that may have responsibility on either side should continue to expand to ensure that mechanisms are put in place that can lead to efficient and reliable systems simultaneously.

The list above provides some examples of research topics that should be pursued that can help shape the wholesale electricity market design of the future, with a focus on how the characteristics of VG can impact the markets. The two main challenges highlighted in this report are not the only challenges that exist with increased penetrations of VG. VG can have many other impacts on wholesale electricity market designs that should be addressed. Finally, VG such as wind and solar power are not the only technologies causing a paradigm shift for the electric power system and wholesale (and retail) electricity markets. A shift to more distributed resources connected to distribution systems can have a major impact on the electricity market design as well as the traditional utility business model. Also, increased demand response and the introduction of more energy storage technologies will continually challenge the traditional electricity market design. Last, improved communication and information technology as well as computational capabilities can allow for changes to the electricity market that may have never been dreamed of before. Researchers, new and existing market participants, and market designers should continue to evaluate how, why, and if the electricity market design should be improved to continue to meet the reliability and efficiency needs of consumers.

References

Abdul-Rahman, K.; Alarian, H.; Rothleder, M.; Ristanovic, P.; Vesovic, B.; Lu, B. (2012). "Enhanced System Reliability Using Flexible Ramp Constraint in CAISO Market." *IEEE Power and Energy Society General Meeting Proceedings*; July 22–26, 2012, San Diego, California.

Allen, A.; Brown, C.; Hickey, J.; Le, V.; Safuto, R. (2009). "Energy Storage in the New York Electricity Market." Draft white paper. Rensselaer, NY: New York Independent System Operator, December. Accessed May 2014: http://www.nyiso.com/public/webdocs/markets_operations/committees/bic_miwg/meeting_materials/2010-01-21/Energy_Storage_Resources.pdf.

Anderson, K.W. (2013). Memorandum to Public Utility Commission of Texas Regarding Proceeding to Ensure Resource Adequacy in Texas, Aug. 28th 2013.

Astrape Consulting. (2013). "The Economic Ramifications of Resource Adequacy." White paper. Washington, D.C.: Eastern Interconnection States Planning Council and the National Association of Regulatory Utility Commissioners. Accessed May 2014: http://www.naruc.org/grants/Documents/Economics%20of%20Resource%20Adequacy%20WhitePaper_Astrape_Final.pdf.

Bertsimas, D.; Litvinov, E.; Sun, X.; Zhao, J.; Zheng, T. (2013). "Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem." *IEEE Transactions on Power Systems* (28:1) Feb.; pp. 52–63.

Billinton, R.; Allan, R.N. (1996). *Reliability Evaluations of Power Systems.* 2nd Ed. New York: Plenum Press.

Bouffard, F.; Galiana, F. (2008). "Stochastic Security for Operations Planning with Significant Wind Power Generation." *IEEE Transactions on Power Systems* (23:2) May; pp. 306–316.

Bouffard, F.; Ortega-Vazquez, M. (2011). "The value of operational flexibility in power systems with significant wind power generation," *In proceedings 2011 IEEE Power and Energy Society General Meeting,* Detroit, MI.

Brattle Group, The; Astrape Consulting. (2013). *Resource Adequacy Requirements: Reliability and Economic Implications.* Washington, D.C: FERC. Accessed July 2014: https://www.ferc.gov/legal/staff-reports/2014/02-07-14-consultant-report.pdf.

Brisebois, J.; Aubut, N. (2011). "Wind Farm Inertia Emulation to Fulfill Hydro-Quebec's Specific Need." *IEEE Power and Energy Society General Meeting Proceedings*; July 24–28, Detroit, MI.

Brown, R. (1993). *The Regulation of Public Utilities Theory and Practice.* Public Utilities Reports, Inc.; pp. 49–125.

CAISO. (2007). *Integration of Renewable Resources: Transmission and Operating Issues and Recommendations for Integrating Renewable Resources on the California ISO-Controlled Grid.* Folsom, CA. Accessed May 2014: http://www.caiso.com/1ca5/1ca5a7a026270.pdf.

CAISO. (2009). *Settlements and Billing: Business Practice Manual for Configuration Guide—Supplemental Reactive Energy Settlement, Version 5.0.* Folsom, CA.

CAISO. (2010a). *Integration of Renewable Resources: Operational Requirements and generation Fleet Capability at 20% RPS.* Folsom, CA. Accessed May 2014: http://www.caiso.com/2804/2804d036401f0.pdf.

CAISO. (2010b). *Market Issues and Performance Annual Report.* Folsom, CA. Accessed May 2014: http://www.caiso.com/2b66/2b66baa562860.pdf.

CAISO. (2011). "Opportunity Cost of Flexible Ramping Constraint: Draft Final Proposal." Folsom, CA. Accessed May 2014: http://www.caiso.com/Documents/DraftFinalProposal-FlexibleRampingConstraint.pdf.

CAISO. (2012). *Annual Report on Market Issues and Performance.* Folsom, CA. Accessed May 2014: http://www.caiso.com/Documents/2012AnnualReport-MarketIssue-Performance.pdf.

CAISO. (2013a). *Business Practice Manual for Reliability Requirements.* Folsom, CA.

CAISO. (2013b). Q1 2013 Report on Market Issues and Performance. Folsom, CA. http://www.caiso.com/Documents/2013FirstQuarterReport-MarketIssues_Performance-May2013.pdf.

CAISO. (2014). "Flexible Resource Adequacy Criteria and Must-Offer Obligation: Market and Infrastructure Policy Draft Final Proposal." Folsom, CA.

Cramton, P.; Ockenfeld, A. (2012). "Economics and Design of Capacity Markets for the Power Sector." *Zeitschrift für Energiewirtschaft* (36); pp. 113–134.

Cramton, P.; Stoft, S. (2008). "Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency." *Utilities Policy* (16); pp. 194–201.

Ela, E.; Tuohy, A.; Milligan, M.; Kirby, B.; Brooks, D. (2012a). "Alternative Approaches for Incentivizing the PFR Ancillary Service." *The Electricity Journal* (25:4) May; pp. 88–102.

Ela, E.; Edelson, D. (2012b). "Participation of Wind Power in LMP-Based Energy Markets." *IEEE Transactions on Sustainable Energy* (3:4) Oct.

Ela, E.; Kirby, B.; Navid, N.; and J.C. Smith. (2012c). "Effective Ancillary Service Market Designs on High Wind Power Penetration Systems." *Power and Energy Society General Meeting Proceedings*; July 22–26, 2012, San Diego, California.

Ela, E.; O'Malley, M. (2012d). "Studying the Variability and Uncertainty Impacts of Variable Generation at Multiple Timescales." *IEEE Transactions on Power Systems* (27: 3). http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6166377.

Ela, E.; Gevorgian, V.; Fleming, P.; Zhang, Y.C.; Singh, M.; Muljadi, E.; Scholbrook, A.; Aho, A.; Buckspan, A.; Pao, L.; Singhvi, V.; Tuohy, A.; Pourbeik, P.; Brooks, D.; Bhatt, N. (2014). *Active Power Control from Wind Power: Bridging the Gaps.* NREL/TP-5D00-60574. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy14osti/60574.pdf.

Ela, E.; Gevorgian, V.; Tuohy, A.; Kirby, B.; Milligan, M.; O'Malley, M. (2014). "Market Designs for the Primary Frequency Response Ancillary Service—Part I: Motivation and Design." *IEEE Transactions on Power Systems* (29:1) Jan.

Ela, E.; Kirby, B.; Botterud, A.; Milostan, C.; Krad, I.; and Koritarov, V. (2013a). "The Role of Pumped Storage Hydro Resources in Electricity Markets and System Operation." Preprint. Prepared for the HydroVision International Conference, July 23–26. NREL/CP-5500-58655. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy13osti/58655.pdf.

Ela, E.; Milligan, M.; Kirby, B. (2011). *Operating Reserves and Variable Generation: A Comprehensive Review of Current Strategies, Studies, and Fundamental Research on the Impact That Increased Penetration of Variable Renewable Generation Has on Power System Operating Reserves.* NREL/TP-550-51978. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy11osti/51978.pdf.

Ela, E; Diakov, V.; Ibanez, E.; Heaney, M. (2013b). *Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales.* NREL/TP-5500-58274. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy13osti/58274.pdf.

Elmaghraby, W.; O'Neill, R.; Rothkopf, M.; Stewart, W. (2004). "Pricing and Efficiency in 'Lumpy' Energy Markets." *The Electricity Journal* (17:5) June.

EnerNex Corportion. (2011). Eastern Wind Integration and Transmission Study. NREL/SR-5500-47078. Golden, CO: NREL. Accessed July 2014: http://www.nrel.gov/docs/fy11osti/47078.pdf.

ERCOT. (2013). "Future Ancillary Services in ERCOT." ERCOT concept paper, Draft version 1.0. Austin, TX: September. Accessed May 2014: http://www.ercot.com/content/news/presentations/2014/ERCOT%20AS%20Concept%20Paper%20Version%201_0%20as%20of%209-27-13%201745.pdf

FERC. (1997). Promoting Wholesale Competition Through Open Access Non-discriminatory Transmission Services by Public Utilities: Recovery of Stranded Costs by Public Utilities and Transmitting Utilities, Order No. 888, FERC Stats. & Regs., March 4. Accessed May 2014: http://www.ferc.gov/legal/maj-ord-reg/land-docs/order888.asp.

FERC. (2001). Order on Motion to Implement Hybrid Fixed Block Pricing Rule and Requiring Tariff Filing, Acting on Related Requests for Rehearing, and Accepting Preliminary Report, 95 FERC ¶ 61,121, FERC Stats & Regs., April 26.

FERC. (2011a). Demand Response Compensation in Organized Wholesale Energy Markets, 134 FERC ¶ 61,187, Order No. 745, FERC Stats. & Regs., March 15. Accessed May 2014: http://www.ferc.gov/EventCalendar/Files/20110315105757-RM10-17-000.pdf.

FERC. (2011b). Frequency Regulation Compensation in the Organized Wholesale Power Markets, 138 FERC ¶ 61,123, Order No. 755, FERC Stats. & Regs., October 20. Accessed May 2014: http://www.ferc.gov/whats-new/comm-meet/2012/021612/E-3.pdf.

FERC. (2011c). ISO New England Inc., 135 FERC ¶ 61,029, at p. 342, April 13. Accessed May 2014: http://www.iso-ne.com/regulatory/ferc/orders/2011/apr/fcm_%20redesign_order_april_13_2011.pdf.

FERC. (2012). "Energy Primer: A Handbook of Energy Market Basics." Staff white paper; pp. 69, col. 2. Accessed May 2014: http://www.ferc.gov/market-oversight/guide/energy-primer.pdf.

FERC. (2013). "Centralized Capacity Market Design Elements." Staff report, AD13-7-000.

GE Energy, (2005) "The Effects of Integrating Wind Power on Transmission System Planning, Reliability, and Operations: Report on Phase 2," Prepared for The New York State Energy Research and Development Authority, Accessed May 2014: http://distwind.uwig.org/nyserdaphase2.pdf

GE Energy. (2013). *Western Wind and Solar Integration Study Phase 1.* NREL/TP-550.47434. Golden, CO: NREL. Accessed July 21014: http://www.nrel.gov/docs/fy10osti/47434.pdf.

Garver, L.L. (1966). "Effective Load Carrying Capability of Generating Units." *IEEE Transactions on Apparatus and Systems* (8); p. 910–919.

Gribik, P.R.; Chaterjee, D.; Navid, N.; Zhang, L. (2011). "Dealing with Uncertainty in Dispatching and Pricing in Power Markets." *IEEE Power and Energy Society General Meeting Proceedings*; July.

Gribik, P.R.; Hogan, W. W.; Pope, S.L. (2007). "Market-Clearing Electricity Prices and Energy Uplift." Accessed May 2014: https://www.misoenergy.org/Library/Repository/Meeting%20Material/Stakeholder/Workshops%20and%20Special%20Meetings/2010/Convex%20Hull%20Pricing%20Workshops/20091030%20Market-Clearing%20Electricity%20Prices%20and%20Energy%20Uplift.pdf

Hirst, E. (2001). *Real-Time Balancing Operations and Markets: Key to Competitive Wholesale Electricity Markets.* Washington, D.C.: Edison Electric Institute. Accessed May 2014: http://wp.ornl.gov/sci/ees/etsd/pes/pubs/RTMReport.pdf.

Hirst, E.; Kirby, B. (1997). *Creating Competitive Markets for Ancillary Services.* ORNL/CON-448. Oak Ridge, TN: Oak Ridge National Laboratory. Accessed May 2014: http://web.ornl.gov/sci/ees/etsd/pes/pubs/con448.pdf.

Hodge, B.-M.; Lew, D.; Milligan, M.; Holttinen, H.; Sillanpää, S.; Gómez-Lázaro; Scharff, R.; Söder, L.; Larsén, X.G.; Giebel, G.; Flynn, D.; Dobschinski, J. (2012). "Wind Power Forecasting Error Distributions: An International Comparison." Preprint. Prepared for the 11th Annual International Workshop on Large-Scale Integration of Wind Power Into Power Systems as Well as on Transmission Networks for Offshore Wind Power Plants Conference, Nov. 13–15. NREL/CP-5500-56130. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy12osti/56130.pdf.

Hogan, W.W. (1996). "Markets in Real Electric Networks Require Reactive Prices." *The Energy Journal* (14:3).

Hogan, W.W. (1998). *Competitive Electricity Market Design: A Wholesale Primer.* Cambridge, Massachusetts: Harvard University, John F. Kennedy School of Government, Center for Business and Government. Accessed May 2014: http://www.hks.harvard.edu/fs/whogan/empr1298.pdf.

Hogan, W.W. (2005). "On an 'Energy Only' Electricity Market Design for Resource Adequacy." White paper. Cambridge, MA: Harvard University.

Hogan, W.W. (2012). "Electricity Scarcity Pricing Through Operating Reserves: An ERCOT Window of Opportunity." Working paper. Cambridge, MA: Harvard University. Accessed May 2014: http://www.hks.harvard.edu/fs/whogan/Hogan_ORDC_110112r.pdf.

Hogan, W.W. (2013). "Back Cast of Interim Solution B+ to Improve Real-Time Scarcity Pricing." White paper. Austin, TX: Electricity Council of Texas. Accessed May 2014: http://www.hks.harvard.edu/fs/whogan/Back%20Cast%20of%20Interim%20Solution%20B+%20to%20Improve%20Real-Time%20Scarcity%20Pricing%20Whitepaper%20-%20Final_r.pdf.

Huang, S.; Dumas, J.; Gonzalez-Perez, C.; Wei-Jen, L. (2009). "Grid Security Through Load Reduction in the ERCOT Market." *IEEE Transactions on Industry Applications* (45:2) March–April.

Ibanez, E.; Milligan, M. (2012). "Impact of Transmission on Resource Adequacy in Systems with Wind and Solar Power." Preprint. Prepared for the 2012 IEEE Power and Energy Society General Meeting. NREL/CP-5500-53482. Golden, CO: NREL. Accessed July 2014: http://www.nrel.gov/docs/fy12osti/53482.pdf.

IEEE Task Force. (2007). "IEEE Task Force on Large Interconnected Power Systems Response to Generation Governing: Present Practice and Outstanding Concerns." *IEEE Special Publication* 07TP180.

Ingleson, J.; Allen, E. (2010). "Tracking the Eastern Interconnection Frequency Governing Characteristic." *IEEE Power and Energy Society General Meeting Proceedings*; July, Minneapolis, Minnesota.

International Energy Agency. (2014). *The Power of Transformation: Wind, Sun and the Economics of Flexible Power Systems,* OECD/IEA 2014.

Isemonger, A.G. (2009). “The Evolving Design of RTO Ancillary Service Markets.” *Energy Policy* (37:1) Jan.; pp. 150–157.

ISO-NE. (2004). *Impact of Virtual Transactions on New England’s Energy Market.* Holyoke, MA. Accessed May 2014: http://www.iso-ne.com/pubs/spcl_rpts/2004/virtual_transactions_report.pdf.

Keane, A.; Milligan, M.; D’Annunzio, C.; Dent, C.; Dragoon, K.; Bernhard, H.; Holttinen, H.; Samaan, N.; Söder, L.; O’Malley, M. (2011). “Capacity Value of Wind Power.” *IEEE Transactions on Power Systems* (26:2) May; Accessed May 2014: http://www.nerc.com/docs/pc/ivgtf/ieee-capacity-value-task-force-confidential%20(2).pdf.

King, J.; Kirby, B.; Milligan, M.; Beuning, S. (2011). *Flexibility Reserve Reductions from an Energy Imbalance Market with High Levels of Wind Energy in the Western Interconnection.* NREL/TP-5500-52330. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy12osti/52330.pdf.

Kirby, B.; Milligan, M.; Ela, E. (2010). “Providing Minute-to-Minute Regulation from Wind Plants.” *9th Annual International Workshop on Large-Scale Integration of Wind Power into Power Systems and Transmission Networks for Offshore Wind Power Plants Conference Proceedings*; Oct., Quebec, Canada.

Kirby, B.; Milligan, M.; Makarov, Y.; Hawkins, D.; Lovekin, J.; Jackson, K.; Shiu, H. (2004). *California Renewables Portfolio Standard: Renewable Generation Integration Cost Analysis—Phase III: Recommendations for Implementation.* Sacramento, CA: California Energy Commission. Accessed May 2014: http://www.energy.ca.gov/reports/500-04-054.PDF.

Lannoye, E.; Milligan, M.; Adams, J.; Tuohy, A.; Chandler, H.; Flynn, D.; O’Malley, M. (2010). “Integration of Variable Generation: Capacity Value and Evaluation of Flexibility.” *IEEE Power and Energy Society General Meeting Proceedings*; July 25–29, 2010, Minneapolis, Minnesota.

Lannoye, E.; Flynn, D.; O’Malley, M. (2012). “Evaluation of Power System Flexibility,” *IEEE Transactions on Power Systems*, (27:2), p. 922-931.

Levin, T.; Botterud, A. (forthcoming). “Capacity Adequacy and Revenue Sufficiency in Electricity Markets with Wind Power.” *IEEE Transactions on Power Systems.*

Lew, D. Brinkman, G.; Ibanez, E.; Florita, A.; Heaney, M.; Hodge, B.-M.; Hummon, M.; Stark, G.; King, J.; Lefton, S.A.; Kumar, N.; Agan, D.; Jordan, G.; Venkataraman, S. (2013). *Western Wind and Solar Integration Study Phase 2.* NREL/TP-5500-55588. Golden, CO: NREL.

Liang, J.; Grijalva, S.; Harley, R. (2011). “Increased Wind Revenue and System Security by Trading Wind Power in Energy and Regulation Reserve Markets.” *IEEE Transactions on Sustainable Energy* (2:3), July; pp. 340–347.

Likover, R. (2013). "Energy Market Accounting Day-Ahead and Real-Time: Introduction to Wholesale Electricity Markets (WEM 101)." Holyoke, MA: ISO New England. Accessed April 2014: http://www.iso-ne.com/support/training/courses/wem101/12_energy_market_accounting_likover.pdf.

London Economics. (2013). "Estimating the Value of Lost Load: Briefing Paper Prepared for the Electric Reliability Council of Texas, Inc., by London Economics International, LLC." Austin, TX: ERCOT. Accessed July 2014: http://www.ercot.com/content/gridinfo/resource/2014/mktanalysis/ERCOT_ValueofLostLoad_LiteratureReviewandMacroeconomic.pdf

Lu, B.; Shahidehpour, M. (2004). "Short-Term Scheduling of Combined Cycle Units." *IEEE Transactions on Power Systems* (19:3) Aug.; pp. 1,616–1,625.

Lyons, K. et al. (2000). "An Introduction to Financial Transmission Rights." *Electricity Journal* (13:10) Dec.

Ma, J.; Silva, V.; Belhomme, R.; Kirschen, D.S. (2013). "Evaluating and planning flexibility in sustainable power systems," *IEEE Power and Energy Society General Meeting Proceedings:* July 21-25, 2013, Vancouver, British Columbia.

Maggio, D. (2012). "Impacts of Wind-Powered Generation Resource Integration on Prices in the ERCOT Nodal Market." *IEEE Power and Energy Society General Meeting Proceedings*; July 22–26, 2012, San Diego, California.

Mai, T.; Wiser, R.; Sandor, D.; Brinkman, G.; Heath, G.; Denholm, P.; Hostick, D.J.; Darghouth, N.; Schlosser, A.; Strzepek, K. (2012). *Renewable Electricity Futures Study: Volume 1—Exploration of High-Penetration Renewable Electricity Futures.* NREL/TP-6A20-52409-1. Golden, CO: NREL.

Makarov, Y.; Lu, S.; Ma, J.; Nguyen, T.B. (2008). *Assessing the Value of Regulation Resources Based on Their Time Response Characteristics.* PNNL-17632. Sacramento, CA: California Energy Commission, Public Interest Energy Research Program.

Meibom, P.; Barth, R; Larsen, H.; Brand, H; Tuohy, A.; Ela, E. (2011). *Advanced Unit Commitment Strategies in the United States Eastern Interconnection.* NREL/SR-5500-49988. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy11osti/49988.pdf.

Milligan, M.; Clark, K.; King, J.; Kirby, B.; Guo, T.; Liu G. (2013). *Examination of Benefits of an Energy Imbalance Market in the Western Interconnection.* NREL/TP-5500/57115. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy13osti/57115.pdf.

Milligan, M.; Ela, E.; Lew, D.; Corbus, D.; Yih-huei Wan; Hodge, B. (2012). "Assessment of Simulated Wind Data Requirements for Wind Integration Studies." *IEEE Transactions on Sustainable Energy* (3:4) Oct.; pp. 620, 626.

Milligan, M.; Holttinnen, H.; Söder, L.; Clark, C.; Pineda, I. (2012). "Markets to Facilitate Wind and Solar Energy Integration in the Bulk Power Supply: An IEA Task 25 Collaboration." Preprint. Prepared for the 11th Annual International Workshop on the Large-Scale Integration of Wind Power Into Power Systems as Well as on Transmission Networks for Offshore Wind Power Plants Conference. NREL/CP-5500-562. Golden, CO: NREL. http://www.nrel.gov/docs/fy12osti/56212.pdf.

Milligan, M.; Kirby, B. (2010). *Market Characteristics for Efficient Integration of Variable Generation in the Western Interconnection.* NREL/TP-550-48192. Golden, CO: NREL. Accessed July 2014: http://www.nrel.gov/docs/fy10osti/48192.pdf.

Milligan, M.; Kirby, B.; Gramlich, R.; Goggin, M. (2009). *Impact of Electric Industry Structure on High Wind Penetration Potential.* NREL/TP-550-46273. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy09osti/46273.pdf.

Milligan, M.; Kirby, B.; Beuning, S. (2010) "Combining balancing areas' variability: impacts on wind integration in the Western Interconnection," NREL\CP-550-48249.

Milligan, M.; Porter, K. (2008). "Determining the Capacity Value of Wind: An Updated Survey of Methods and Implementation." Preprint. Prepared for the American Wind Energy Association WINDPOWER Conference, June 1–4, 2008. NREL/CP-500-43433. Accessed May 2014: http://www.nrel.gov/docs/fy08osti/43433.pdf.

Mills, A.; Ahlstrom, M.; Brower, M.; Ellis, A.; George, R.; Hoff, T.; Kroposki, B.; Lenox, C.; Miller, N.; Stien, J.; Wan, Y. (2009). *Understanding Variability and Uncertainty of Photovoltaics for Integration with the Electric Power System.* LBNL-2855E. Berkeley, CA: Lawrence Berkeley National Laboratory. Accessed May 2014: http://emp.lbl.gov/sites/all/files/lbnl-2855e.pdf.

Monitoring Analytics, LLC. (2010). *State of the Market Report for PJM: Section II—Energy Market, Part I.* Eagleville, PA. Accessed May 2014: http://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2010/2010-som-pjm-volume2-sec2.pdf.

Navid, N.; Rosenwald, G. (2012). "Market Solutions for Managing Ramp Flexibility with High Penetration of Renewable Resource." *IEEE Transactions on Sustainable Energy* (3:4); pp.784–790.

Navid, N.; Rosenwald, G.; Chatterjee, D. (2011). *Ramp Capability for Load Following in the MISO Markets.* Carmel, IN: MISO Market Development and Analysis. https://www.midwestiso.org/Library/Repository/Communication%20Material/Key%20Presentations%20and%20Whitepapers/Ramp%20Capability%20for%20Load%20Following%20in%20MISO%20Markets%20White%20Paper.pdf.

NERC. (2011). *Methods to Model and Calculate Capacity Contributions of Variable Generation for Resource Adequacy Planning.* Washington, D.C. Accessed May 2014: http://www.nerc.com/docs/pc/ivgtf/IVGTF1-2.pdf.

NERC. (2013). *2013 Long-Term Reliability Assessment.* Washington, D.C. Accessed May 2014: http://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/2013_LTRA_FINAL.pdf.

Milligan, M.; Ela, E.; Hein, J.; Schneider, T.; Brinkman, G.; Denholm, P. (2012). *Exploration of High-Penetration Renewable Electricity Futures. Vol. 4 of Renewable Electricity Futures Study.* NREL/TP-6A20-52409-4. Golden, CO: National Renewable Energy Laboratory.

NYISO. (2009). "NYISO Technical Bulletin 049: Multi-Pass Methodology of Security Constrained Unit Commitment: Revised 1/9/2009." Rensselaer: NY.

O'Neill, R.; Dautel, T.; Krall, E. (2011). *Recent ISO Software Enhancements and Future Software and Modeling Plans: Staff Report.* Washington, D.C.: FERC. Accessed May 2014: https://www.ferc.gov/industries/electric/indus-act/rto/rto-iso-soft-2011.pdf.

O'Neill, R.; Sotkiewicz, P.; Hobbs, B.; Rothkopf, M.; Stewart, W. (2005). "Efficient Market-Clearing Prices in Markets with Nonconvexities." *European Journal of Operational Research* (164:1) July.

Oren, S.S.; Spiller, P.; Varaiya, P; Wu, F. (1995). "Nodal prices and transmission rights: A critical appraisal." *The Electricity Journal*, (8:3) April.

Oren, S.S. (2001). "Design of Ancillary Service Markets." *34th Hawaii International Conference on Systems Sciences Proceedings*; Jan. 5–8, 2001, Maui, Hawaii.

Ortega-Vazquez, M.; Kirschen, D. (2007). "Optimizing the Spinning Reserve Requirements Using a Cost/Benefit Analysis." *IEEE Transactions on Power Systems* (22:1) Feb.; pp. 24–33.

Pfeifenberger, J.P.; Spees, K.; Carden, K.; Wintermantel, N. (2013) "Resource Adequacy Requirement: Reliability and Economic Implications," Report prepared for FERC, Brattle Group/Astrape Consulting, 2013.

Pike, R. (2013). "Incentivizing Flexibility in NYISO Markets." *EUCI Conference on Fast/Flexiramp Products and Policies*; April.

PJM Interconnection (2013), *RTO-Wide Five-Year Selection Process Request for Proposal for Black Start Service,* July 1, 2013, Available: http://www.pjm.com/~/media/markets-ops/ancillary/rto-wide-5-year-selection-process-request-for-proposal-for-black-start-servic.ashx

Porter, K.; Fink, S.; Rogers, J.; Mudd, C.; Buckley, M.; Clark, C. (2012). *PJM Renewable Integration Study: Task Report—Review of Industry Practice and Experience in the Integration of Wind and Solar Generation.* Schenectady, NY: GE Energy. Accessed July 2014: http://pjm.com/~/media/committees-groups/task-forces/irtf/postings/pris-task3b-best-practices-from-other-markets-final-report.ashx.

Price, J.E.; Rothleder, M. (2011). "Recognition of Extended Dispatch Horizons in CA Energy Markets." *IEEE Power and Energy Society General Meeting Proceedings*; July 24–29, Detroit, Michigan.

Rebours, Y.; Kirschen, D.; Trotignon, M. (2007). "Fundamental Design Issues in Markets for Ancillary Services." *The Electricity Journal* (20:11); pp. 26–34.

Rogers, J.; Porter, K. (2012). *Summary of Time Period-Based and Other Approximation Methods for Determining the Capacity Value of Wind and Solar in the United States.* NREL/SR-5500-54338. Work performed by Exeter Associates Inc., Columbia, Maryland. Golden, CO: NREL. Accessed December 16, 2013: http://www.nrel.gov/docs/fy12osti/54338.pdf.

Rose, K. (2011). *An Examination of RTO Capacity Markets,* IPU Working Paper No. 2011-4, Available: http://ipu.msu.edu/research/pdfs/Working-Paper-Rose-Capacity-Markets.pdf.

Rux, M. (1993). "An Incremental Economic Dispatch Method for Cascaded Hydroelectric Power Plants." *IEEE Transactions on Power Systems* (8:3); pp. 1,266–1,273.

Saraf, N.; McIntyre, K; Dumas, J.; Santoso, S. (2009). "The Annual Black Start Service Selection Analysis of ERCOT Grid," *IEEE Transactions on Power Systems* (24:4); p. 1867–1874.

Sarkar, V. et al (2008), "A Comprehensive Assessment of the Evolution of Financial Transmission Rights." *IEEE Transactions on Power Systems* (23:4) Nov.

Schulz, R. (1999). "Modeling of Governing Response in the Eastern Interconnection." *IEEE Power Engineering Society Winter Meeting Proceedings*; Feb., New York, New York.

Shahidehpour, M.; Yamin, H.; Li, Z. (2002). *Market Operations in Electric Power Systems.* New York, NY: John Wiley and Sons.

Shiu, H. (2006). *California Renewables Portfolio Standard Renewable Generation Integration Cost Analysis: Multi-Year Analysis Results and Recommendations—Final Report. Work performed by California Wind Energy Collaborative.* Sacramento, CA: California Energy Commission. Accessed July 2014: http://www.energy.ca.gov/2006publications/CEC-500-2006-064/CEC-500-2006-064.PDF

Smith, J.C., et al. (July 2010). "Impact of Variable Renewable Energy on U.S. Electricity Markets." *IEEE Power and Energy Society General Meeting Proceedings*; July 25–29, Minneapolis, Minnesota.

Stoft, S. (2002). *Power System Economics: Designing Markets for Electricity.* New York, NY: IEEE Press & Wiley-Interscience, John Wiley & Sons, Inc.

Tuohy, A.; Ela, E.; Kirby, B.; Brooks, D. (2012). "Provision of Regulation Reserve from Wind Power: Economic Benefits and Steady-State System Operation Implications." *11th International Workshop on Large-Scale Integration of Wind Power into Power Systems and Transmission Networks for Offshore Wind Power Plants Conference Proceedings*; Lisbon, Portugal.

Ulbig, A.; Andersson, G. (2012). "On operational flexibility in power systems," *In Proceedings IEEE Power and Energy Society General Meeting,* San Diego, CA.

Virmani, S. (1999). "Security Impacts of Changes in Governor Response." *IEEE Power Engineering Society Winter Meeting Proceedings*; Feb., New York, New York.

Walawalkar, R.; Fernands, S.; Thakur, N.; Chevva, K. (2010), "Evolution and current status of demand response (DR) in electricity markets: Insights from PJM and NYISO," *Energy*, (35:4), p. 1553-1560.

Wan, Y. (2011). *Analysis of Wind Power Ramping Behavior in ERCOT.* NREL/TP-5500-49218. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy11osti/49218.pdf.

Wan. Y. (2005). *A Primer on Wind Power for Utility Applications.* NREL/TP-500-36230. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy06osti/36230.pdf.

Wang, J.; Botterud, A.; Bessa, R.; Keko, H.; Carvalho, L.; Issicaba, D.; Sumaili, J.; Miranda, V. (2011). "Wind Power Forecasting Uncertainty and Unit Commitment." *Applied Energy* (88:11); pp. 4,014–4,023.

Wang, J.; Wang, X.; Wu, Y. (2005). "Operating Reserve Model in the Power Market." *IEEE Transactions on Power Systems* (20:1) Feb.; pp. 223–229.

Zhang, J.; Hodge, B.-M.; Florita, A.; Lu, S.; Hamann, H.F.; Banunarayanan, V. (2013). "Metrics for Evaluating the Accuracy of Solar Power Forecasting." Preprint. Prepared for the 3rd International Workshop on Integration of Solar Power Into Power Systems Conference, Oct. 21–22, 2013. NREL/CP-5500-60142. Golden, CO: NREL. Accessed May 2014: http://www.nrel.gov/docs/fy14osti/60142.pdf.

Zhou, Z.; Botterud, A. (2013). "Price Responsive Demand for Operating Reserves in Co-Optimized Electricity Markets with Wind Power." Presentation at FERC Technical Conference, June 25, 2013.

Zhou, Z.; Botterud, A. (2014). "Dynamic Scheduling of Operating Reserves in Co-optimized Electricity Markets with Wind Power." *IEEE Transactions on Power Systems* (29:1); pp.160–171.

Zu, L.; Tretheway, D. (2012). "Flexible Ramping Performance: Second Revised Draft Final Proposal." Folsom, CA: California ISO (CAISO). http://www.caiso.com/Documents/SecondRevisedDraftFinalProposal-FlexibleRampingProduct.pdf.

Appendix A: Analysis of Revenue Sufficiency from WWSIS-2

As a demonstration of the reduction in operating profits resulting from price suppression caused by widespread deployment of wind and solar generation, archived results from WWSIS-2 were examined (Lew et al. 2013).

The purpose of WWSIS-2 was to study the impacts of cycling on system and generator costs and emissions. WWSIS-2 examined five different scenarios with different combinations of wind and solar deployment, as shown in Table AA-1. The different combinations of wind and solar allow for examination of the different impacts of wind and solar on the revenues of the generators in the generation fleet.

Capturing accurate electricity and reserve clearing prices was not a focus of WWSIS-2. As a consequence, the model was tailored to accurately reflect system operations but not price formation. The energy prices and resulting revenues from WWSIS-2 should be viewed with this limitation in mind.

Table AA-1. Wind and Solar Scenarios in WWSIS-2

Scenario	Penetration (% of Annual Load)		
	Wind	Solar	Wind Plus Solar
No Renewables	0	0	0
TEPPC	9.4	3.6	13
High Solar	8	25	33
High Mix	16.5	16.5	33
High Wind	25	8	33

In this analysis, operating profits for each generator were calculated as the revenues from electricity sales less the short-run costs of producing the electricity. Short-run costs include fuel, startup, and variable operations and maintenance costs. Revenues are the product of the electrical generation at each power plant multiplied by the coincident electricity price. Make-whole or uplift payments that are present in most organized markets were not considered in this analysis. Revenues resulting from ancillary services were also not considered in this analysis.

Penalty prices are present in the WWSIS-2 data. Penalty prices are high prices caused by a violation of one or more soft constraints in the unit commitment and dispatch problem. In production-cost models, operational or system constraints (such as supply-demand balance for energy or ancillary services) are often written with soft constraints by adding a penalized slack variable to the constraint. These soft constraints often allow the model to run faster but also introduce penalty prices in which the model accepts a solution that violates a soft constraint instead of continuing to look for a solution that does not violate the soft constraint. This phenomenon occurs in actual market operations, and in actual market operations the penalty prices are often removed by performing a pricing run with relaxed constraints on the violated soft constraints. However, WWSIS-2 did not perform pricing runs, so the penalty prices remain present in the electricity price data.

Figure AA-1 shows the electricity price for two weeks in the Southern California region of the No Renewables scenario. The left image shows the first week in May, when electricity prices were within the expected range of marginal costs. No penalty prices were present during this week. In contrast, the right image shows the first week in August, when five of the seven days showed reserve shortages that caused the energy prices to spike to the penalty price for reserve shortage (approximately $4,000/MW-h above the maximum scale shown on the graph).

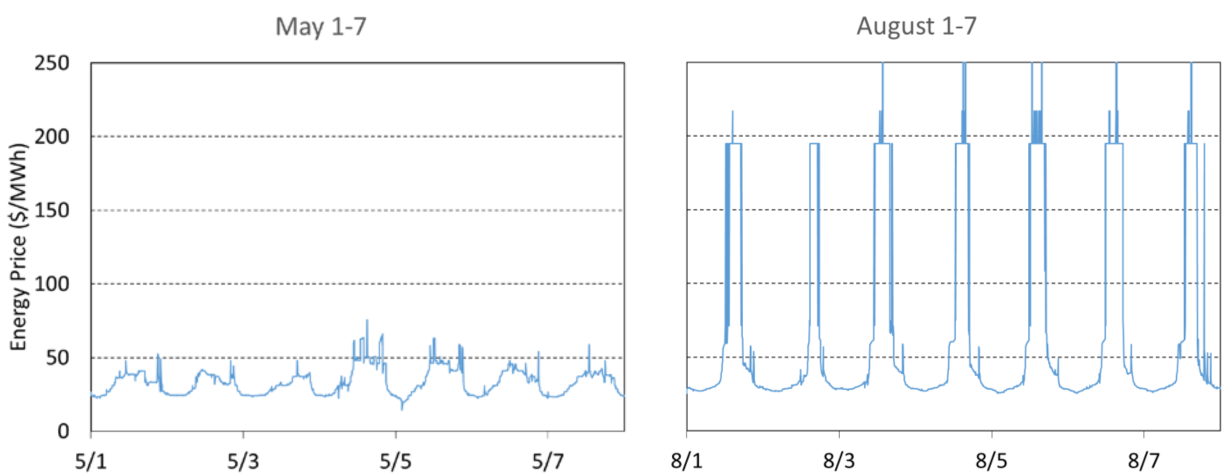

Figure AA-1. Prices for two weeks in one region of the WWSIS-2 results. The prices in the WWSIS-2 indicate penalties for insufficient reserve procurement.

These penalty prices are a normal and expected part of the model results but complicate interpretation of the revenues earned by each generator. Penalty prices can be caused by a true condition in the system being modeled (for example, insufficient capacity to provide a reserve), but they can also be caused by the optimality tolerance given to the solver[17]. Therefore, in many cases a solution may contain small violations of soft constraints that add negligibly to the problem's objective function but cause spikes in the price data as a result of the incurred penalty prices. Differentiating between these two causes is often difficult. In this analysis, the operating profits were calculated first with the penalty prices included and then with all penalty prices replaced with a cap of $250/MWh, which is a proxy for the highest-marginal-cost generation capacity available.

Generator operating profits from the real-time energy market were calculated as the difference between generator revenues and operating costs. Revenues were calculated as the product of each generator's generation and the regional energy price for each 5-min interval. Costs included

[17] To be solved with feasible run times, mixed-integer programs such as the unit commitment and dispatch problem allow the solver to stop searching for a better solution when the solver determines it is within a specified tolerance of a bound on the optimal solution.

fuel, start-up, and variable operations and maintenance costs. WWSIS-2 study included all of the Western Electricity Coordinating Council, so portions of Canada and Mexico were included in the study, but WWSIS-2 expanded only the wind and solar capacities in the portion of WECC in the United States. Therefore, the non-U.S. generators were removed from the data set, and the remaining generators were then grouped according to the following categories: nuclear, coal, natural gas combined cycle (NGCC), natural gas combustion turbine (NGCT), natural gas steam boiler (steam), storage, solar photovoltaic (PV), and wind. Operating profits were then normalized by the installed capacity of all generators in each category. The overall generation expansion plan included estimates of future generation and retirements from the Western Electricity Coordinating Council's Transmission Expansion Planning and Policy Committee (TEPPC).

Figure AA-2 shows the operating profits calculated for each generator category in each of the WWSIS-2 scenarios with the penalty prices included in the price data. The data show that operating profits are as would be expected: low-marginal-cost generators earn more operating profit than higher-marginal-cost generators. The notable exception is the storage category, which has very low operating profits. This phenomenon is because of the system benefits that storage provides—the optimal use of storage for the system is different than for the storage owner (Sioshansi 2010).

When considering the nuclear, coal, and combined-cycle categories, the operating profit appears to drop substantially from the No Renewables scenario to the scenarios that include renewables. However, the price data from the No Renewables scenario includes many more instances of penalty prices that skew the apparent operating profits of the No Renewables case. Figure AA-3 shows the amount of revenue a generator would have earned if it operated at full output during every penalty-price interval (defined in this analysis as prices exceeding $250/MWh) in each of the scenarios, compared to the amount of revenue earned when all penalty prices above $250/MWh were replaced with $250/MWh. The No Renewables scenario contains more than twice as much penalty price revenue as the next highest scenario (HiSolar) and almost eight times as much penalty price revenue as the lowest scenario (HiWind). The large number of penalty prices in the No Renewables scenario is because the thermal fleet was optimized for the TEPPC scenario and is therefore somewhat undersized for the No Renewables scenario.

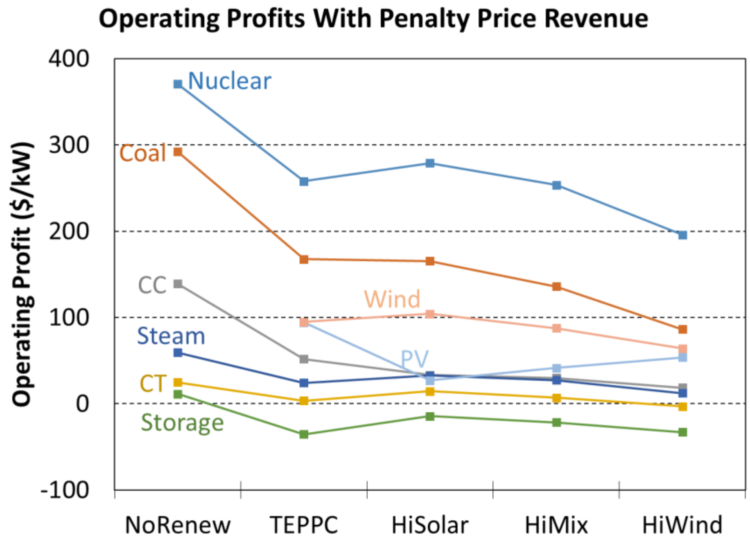

Figure AA-2. Operating profits in each WWSIS-2 scenario, including penalty price revenue

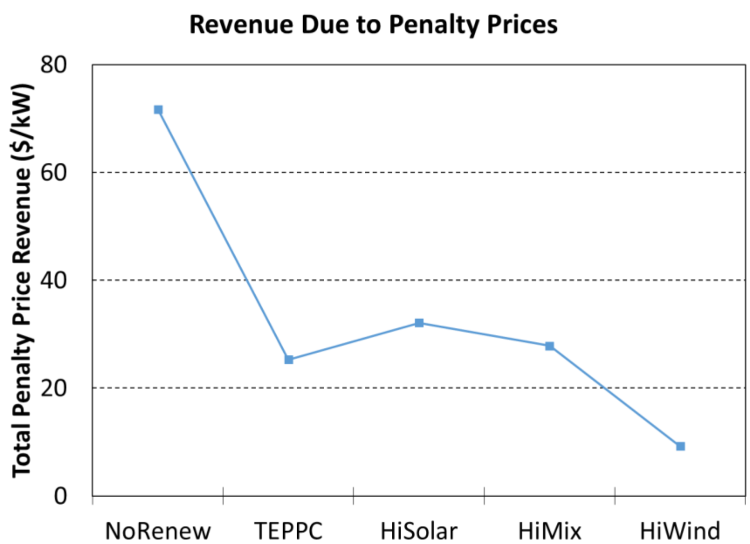

Figure AA-3. Excess revenues to a baseload generator as a result of penalty prices above $250/MWh

Figure AA-4 shows an estimate of the operating profits calculated for each generator category after subtracting the potential penalty price revenue shown in Figure AA-3 from the results shown in Figure AA-2. Note that not all generators would necessarily be generating at full capacity when the penalty prices occur, so reducing their operating profits by the full penalty price revenue amount gives a lower bound on the actual operating profits. In particular, the steam, combustion turbine, and storage categories are all significantly negative. Negative operating profits after subtracting the potential penalty price revenue suggests that these generator categories did not actually earn the full potential penalty price revenue, or had significant startup or non-convex operating costs that would have been addressed through make-whole payments, or both, so these categories require more detailed attention before any conclusions can be drawn about their operating profits.

When comparing the No Renewables scenario to the scenarios with renewables, the nuclear, coal, and combined-cycle categories show decreases in operating profits. Interestingly, the combined-cycle operating profits decrease to approximately zero in all three high-renewables scenarios. Also of note is that the operating profits of the nuclear and coal categories show similar trends to each other, and that the decreases for these two categories depend more on the amount of wind generation in the scenario than on the total wind and solar generation. The HiWind scenario shows the largest reduction for these two categories, with a reduction in operating profits of approximately 38% for the nuclear category and 65% for the coal category.

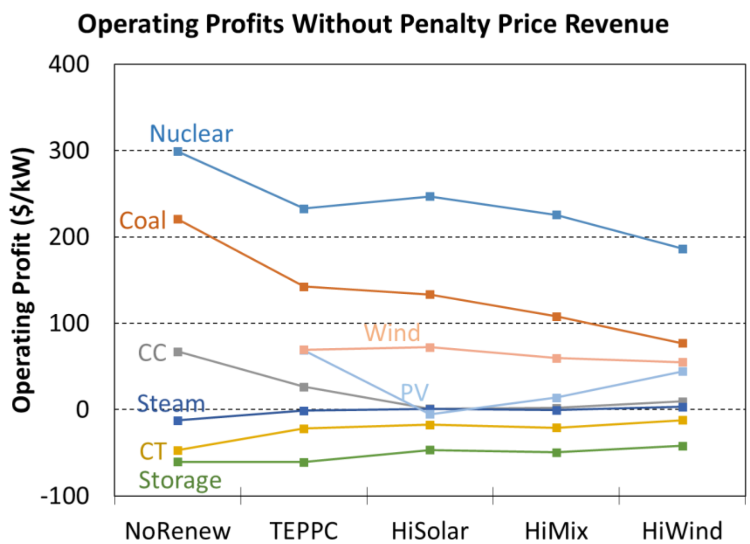

Figure AA-4. Operating profits in each WWSIS-2 scenario, after subtracting penalty price revenue. Note that the estimate of penalty price revenue assumes a 100% capacity factor during penalty intervals, so it is likely that this graph underestimates the operating profits for non-baseload technologies.

The results above do not include any revenue from capacity markets, true scarcity prices, or strategic bidding by market participants. The long-term revenue from these sources will be termed “capacity value revenue” in this section. The long-term capacity value revenue is equal to the annual fixed carrying costs of a peaking generator, and in this analysis it is calculated as the annualized carrying cost of a combustion turbine less the annual operating profit of the combustion turbine category. An estimate for the annual carrying costs for nuclear, coal, combined-cycle, and combustion turbine generators is given in Table AA-2, and the capacity value revenue needed to give net annual revenue equal to estimated annual carrying costs for a new combustion turbine is shown in Figure AA-5.

Table AA-2. Example of Overnight and Annualized Capital Costs

	Nuclear	Coal	Combined Cycle	Combustion Turbine
Overnight Capital Cost ($/kW)[18]	5,429	2,883	1,006	664
Annual Carrying Costs ($/kW)[19]	706	375	131	86

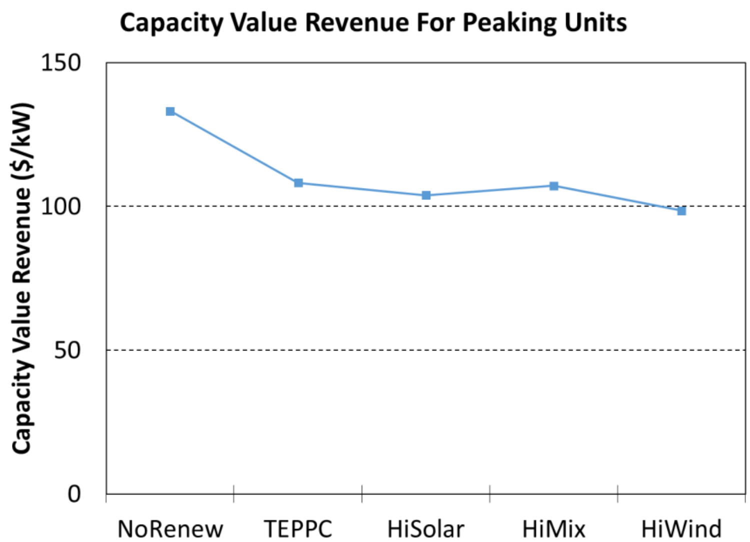

Figure AA-5. Annual capacity value revenue necessary for a new combustion turbine (peaking generator) to cover its annual carrying costs

Figure AA-6 compares the estimates of annual carrying costs for new generating units to the estimated total annual revenues (including operating profits with penalty prices removed and capacity value revenues). The capacity value revenues were constructed so that the CT category annual revenues exactly match the annual carrying costs. The nuclear and coal categories are below the annual carrying costs in all of the scenarios. The combined-cycle category shows annual revenues above the carrying costs in the No Renewables and TEPPC scenarios and below the carrying costs in all three high-renewables scenarios.

---

18 From U.S. Energy Information Agency Assumptions to the Annual Energy Outlook 2013.
19 Assumed annual fixed carrying costs are equal to 13% of overnight capital costs.

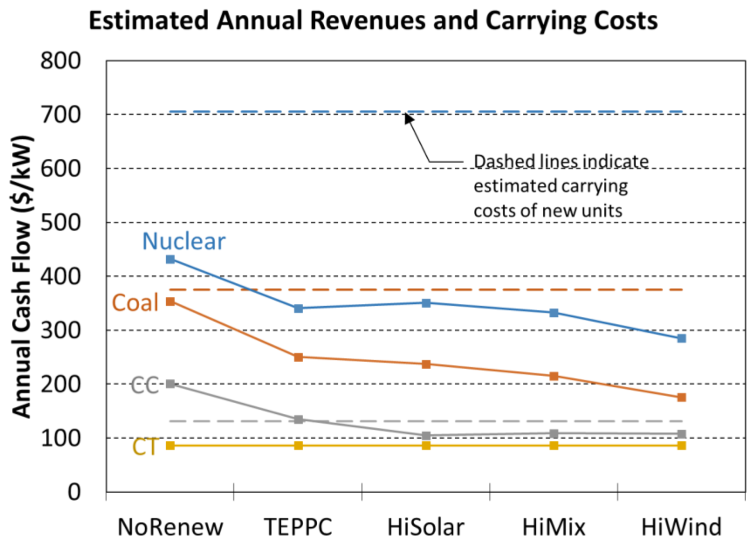

Figure AA-6. Comparison of estimated annual revenues (including both operating profits and capacity value revenue) and estimated annual carrying costs

The results above have important caveats that should be observed. First, WWSIS-2 was not intended to evaluate generator revenues and therefore did not focus on price formation. The results shown in this section are illustrative only, but suggest that many generator categories will see reduced demand for their operation, reduced electricity prices while operating, and therefore reduced operating profits. A future study with the explicit purpose of evaluating generator revenues would be interesting and necessary to determine how revenue sufficiency is affected under high wind and solar penetrations.

Second, in this analysis revenues resulting from penalty prices were estimated rather than calculated explicitly for each generator, so the operating profits after removing the penalty price revenues are also estimates. Further analysis of the existing WWSIS-2 data could identify the actual penalty price revenues and could also determine the make-whole payments and therefore true operational profits for generators with substantial non-convex operational costs (e.g., combustion turbines).

Third, revenue sufficiency requires estimates of the required annual carrying cost of the generators. Category-level estimates were used; however, it is likely that project-specific costs will vary significantly from the estimates used.

Finally, any conclusions regarding revenue sufficiency must be made with the installed capacity by generator type and the long-term equilibrium fleet composition in mind. Even with zero wind or solar generation, perturbations such as low natural gas prices can cause significant revenue sufficiency problems for a fleet optimized for higher natural gas prices. The composition of the generation fleet will respond to a variety of forcing functions, including capital and fuel costs, wind and solar installed capacities, environmental regulations, and others, and will do so over many years. It is likely that analysis of a single year or scenario will be insufficient to draw meaningful conclusions regarding revenue sufficiency.

Appendix A Reference

Sioshansi, R. (2010). "Welfare Impacts of Electricity Storage and the Implications of Ownership Structure." *The Energy Journal* (31:2); pp 173–198.

Appendix B: Market Prices and Revenue Sufficiency under Optimal Thermal Generation Expansion with Wind Power

Introduction

As discussed in this report, it has been recognized for a long time that capacity adequacy is a challenge in electricity markets, mainly because of limited demand response, which prevents efficient pricing, particularly during scarcity situations. Therefore, it is generally agreed upon that some sort of regulatory intervention is needed to ensure sufficient incentives for adequate investments in generation capacity. The energy-only market approach focuses on improved scarcity pricing through administratively set prices in situations when demand for energy and reserves exceeds supply. Another approach is capacity markets or payments, which constitute an additional source of revenue for generators (and possibly demand resources), to make up for insufficient income from energy and ancillary service markets. The objective behind these approaches is to provide enough generation capacity to meet peak load. However, how the rapid expansion of renewables influences the need for and suitability of different market solutions for capacity adequacy is still unknown.

A wide body of literature is devoted to analyzing the impact of variable resource uncertainty on optimal operational commitment and dispatch decisions. Some work has been done on long-term generation expansion planning with renewable energy. However, there is very limited existing work that analyzes the economic impacts to existing thermal generation units from a large-scale expansion of renewables. The analysis in Levin and Botterud (2014) contributes to the existing literature by conducting a quantitative analysis of resource adequacy and revenue sufficiency in systems with increasing wind generation. An improved stochastic generation expansion model is used to analyze how increasing wind penetration influences the optimal investment mix and revenue and profitability of thermal generators. The purpose of the study was not to model the complex dynamics of real-world decision making in electricity markets; rather, a much narrower question was analyzed, i.e., whether the prices for energy and reserve that emerge under an optimal system expansion plan would be sufficient to cover operating and capital costs for thermal generators. The case study, which is summarized below, found that administratively-determined scarcity prices or capacity payments are needed for all thermal generators to make a profit regardless of the wind power penetration. However, baseload plants are more exposed to reductions in average energy prices as more wind is added to the system than are intermediate and peak load plants; therefore, they become more dependent on incentive mechanisms.

In the sections that follow, we give a brief overview of the modeling framework and case study assumptions and results. For more details on relevant literature, the generation expansion model, and the case study, refer to Levin and Botterud (2014).

Model

The generation expansion model is a mixed-integer programming model that determines the lowest-cost expansion of thermal generators and the corresponding commitment and dispatch that meets demand, reserve requirements, and ramping constraints in a given electricity system. The model first utilizes scenario reduction to identify a set of representative wind power scenarios and to limit the computational space of the problem. A two-stage stochastic mixed-integer programing problem is then formulated in which expansion decisions are first-stage decisions that apply universally to all scenarios; whereas unit commitment and dispatch decisions are made at the second stage for each scenario and in each time step. Wind power forecast uncertainty is represented in terms of increased need for operating reserve (i.e., spinning reserve up).

The expansion model in Jin et al. (2014) provides the starting point for the analytical framework; however, several key adjustments have been made. In Jin et al. (2014), binary decision variables were used to represent expansion and commitment decisions for each individual generation unit. The improved model utilizes an integer formulation to reduce computational complexity and enable analysis of larger systems. Candidate generation units are grouped by type, and integer variables represent the number of units of each type that are expanded or committed. This integer formulation of generation units, which is similar to the one in Palmintier et al. (2011), offers substantial time savings without significantly sacrificing accuracy.

The model minimizes the total annual cost for the system, including costs from capital investment, unit commitment, operations and maintenance, fuel, and penalties from unserved load and violations of the reserve requirement. Constraints include load balance, reserve requirements, max/min limits on thermal generators, ramping constraints, wind curtailment, and additional commitment constraints as a function of the expansion state. The full mathematical model formulation is documented in Levin and Botterud (2014). The prices for energy and reserve are calculated by fixing all the integer variables in the model (i.e., for expansion and commitment) and resolving as a linear programming problem. The prices for energy and reserve are then equal to the dual variables of the load and reserve constraints, under the assumption that the two products are co-optimized in the electricity market.

The model builds on a number of simplifying assumptions. It conducts a static least-cost system expansion for one year only, does not include transmission constraints or demand response, and considers only spinning reserve up, ignoring other types of reserve. Minimum up- and down-time constraints for thermal generators are also omitted. Moreover, only one market-clearing stage is modeled (at an hourly time resolution), without considering the impact of variability and forecasting errors on real-time balancing of supply and demand, other than through higher operating reserve requirements.

Case Study

Assumptions

The model was applied to analyze expansion and hourly operational decisions for generation units over the course of a year. Three representative demand profiles were selected to approximate the three different load seasons: summer, winter and spring/fall. Each profile spans one week, or 168 hourly periods. Additionally, three different wind generation profiles, selected with scenario reduction as the most representative ones from the Eastern Wind Integration and Transmission Study wind data in Illinois between 2004 and 2006, were considered for each load season. The results for each representative scenario were weighted probabilistically to determine average characteristics for the system over the course of a year.

The model was run for 16 different wind penetration levels, with wind generation capable of serving a fraction of total demand that varies from 0% to 30% in 2% increments. The representative system has a peak demand level of 25,000 MW. The operating reserve have two components: (1) a fixed fraction of load, assumed to be 10%, to account for generator contingencies and load forecasting errors, and (2) a fraction of the available wind power, which varies with the wind power level. The wind reserve requirement is estimated based on a statistical analysis of hour-ahead forecasting errors between 2004 and 2006, assuming that sufficient wind reserve are kept to compensate for such forecasting errors 99.7% of the time. Hence, the operating reserve requirement increases substantially with the wind penetration.

The system includes four different types of thermal generation units: nuclear, coal, NGCC, and NGCT (Table AB-1). The system was assumed to have an existing generation mix with 20,100 MW of total capacity. The model determines the optimal new generation mix for serving shortfall between load and existing generation for a given year. The default assumption was to use a penalty cost of $1100/MW-h for reserve shortfalls,[20] but sensitivity analyses around this parameter with values of $250/MW-h and $2,500/MW-h were also performed. A sensitivity case with higher fuel prices for natural gas ($7.50/MMbtu) was also analyzed. Generator revenue was calculated based on the prices for energy and reserve under the optimal expansion plan. Although the penalties for energy and reserve shortfalls are always included in the objective function, prices were calculated with and without the inclusion of these scarcity prices. Energy and reserve prices without scarcity prices were calculated by retroactively subtracting the reserve shortfall penalty in each period when a reserve shortfall occurred.

[20] A penalty for load curtailment of $3,500/MWh was applied, but load curtailment did not occur in any of the cases.

Table AB-1. Generation Technology Parameters

	Nuclear	Coal	NGCC	NGCT
Max. Output (MW)	2,200	1,300	400	210
Min. Output (MW)	2,200	520	160	84
Overnight Cost ($/kW)	5,429	2,883	1,006	664
Annualized Investment Cost ($/MW)	443,782	247,371	94,241	62,203
Var. O&M Cost ($/MWh)	2.10	4.39	3.21	10.19
Fixed O&M Cost (#/kW)	92	31	15	7
Heat Rate (btu/kWh)	10,452	8,800	6,430	9,750
Fuel Cost ($/MMbtu)	0.50	2.30	4.50	4.50
Marginal Generation Cost ($/MWh)	7.33	23.52	27.49	46.58
No Load Cost ($/MW)[a]	-	1.09	7.34	12.50
Max. Spinning Reserve (% Max. Output)	0	20	50	80
Ramp Up Limit (% of Max. Output/h)	0	35	50	100
Ramp Down Limit (% of Max. Output/h)	0	35	50	100
Start-Up Cost ($/MW)[a]	-	131.35	61.80	40.60
Shut-Down Cost ($/MW)[a]	-	1.31	0.62	0.41

[a]These costs are in units of dollars per MW nameplate capacity.

Results

The new generation capacity that is constructed for each wind penetration level under each of the four sensitivity cases is shown in Figure AB-1. In all three cases with low natural gas prices, NGCC and NGCT capacity are preferred to coal and nuclear capacity. In the baseline case, total new thermal capacity decreases from 6,920 MW to 6,120 MW as the wind penetration level increases from 0% to 30%. The share of NGCT capacity increases slightly with increasing wind levels, offering more ramping flexibility and reserve capacity. The reduction in thermal capacity from increasing wind penetration is modest in all three cases. This is because wind output during periods of peak load is only moderate, and the reductions in required generation capacity to meet net load are to some extent offset by the need for additional reserve capacity. When the reserve shortfall penalty is reduced to $250/MWh, less new capacity is developed and reserve shortfalls occur more frequently. Conversely, when the penalty is raised to $2,500/MWh, the share of new capacity allocated to NGCT units increases to limit the occurrence of reserve shortfalls. Increasing the price of natural gas to $7.50/MMbtu makes NGCC relatively more attractive than NGCT and the share of new NGCC capacity increases. Additionally, one new coal unit is developed for low wind penetration levels (0% to 6%), because coal is more cost competitive with higher natural gas prices.

The total generation and reserve dispatch by unit type are shown in Tables III and IV. The total generation levels for the first three sensitivity cases are essentially identical, but there is a significant shift from natural gas to coal generation when natural gas prices reach \$7.50/MMbtu. Total reserve are also quite similar for the first three sensitivity cases, with slightly fewer reserve supplied (i.e., more reserve shortfall) when the shortfall penalty is \$250/MWh and slightly more supplied (i.e., less reserve shortfall) when the penalty is \$2,500/MWh. Note that curtailed wind power is allowed to provide reserve. There is a large increase in total wind curtailment, and therefore reserve provision from wind (from 7,352 GWh to 10,689 GWh), when natural gas prices are high, because natural gas units are committed less frequently and are therefore less available to provide reserve.

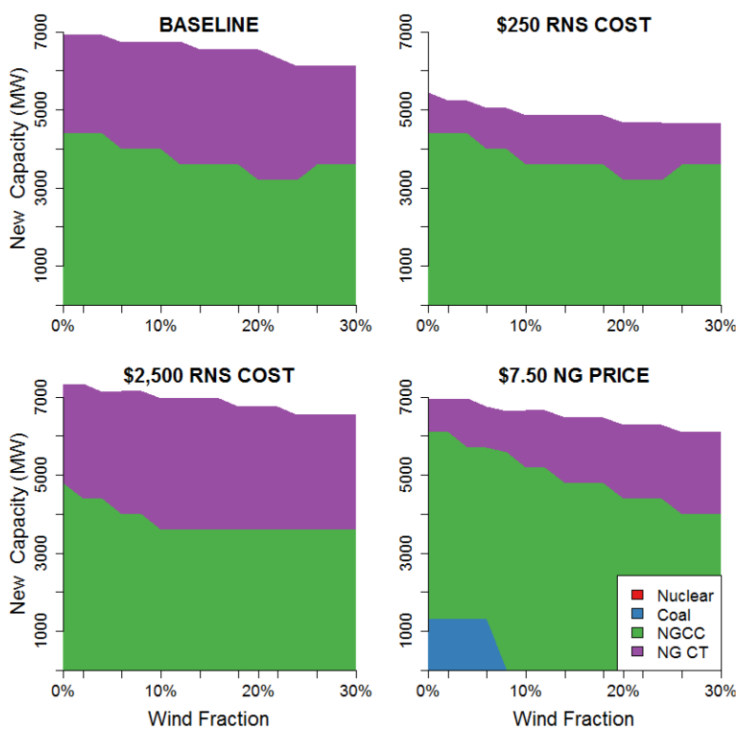

Figure AB-1. New generation capacity as a function of wind penetration for each of the four cases

Table AB-2. Generation by Technology (MWh)

	Wind %	Nuclear	Coal	NGCC	NGCT	Total
Baseline	0%	38,544	65,463	27,604	5,882	137,493
	30%	38,544	42,696	17,664	4,713	103,617
$250 RNS Cost	0%	38,544	65,466	27,608	5,875	137,493
	30%	38,544	42,673	17,681	4,699	103,598
2,500 RNS Cost	0%	38,544	65,425	28,329	5,195	137,493
	30%	38,544	42,648	17,676	4,722	103,589
$7.50 NG Price	0%	38,544	75,869	19,759	3,321	137,493
	30%	38,544	49,118	15,160	4,126	106,948

Table AB-3. Provision of Reserve by Technology (MW-h)

	Wind %	Nuclear	Coal	NGCC	NGCT	Wind Curtail.	Total
Baseline	0%	-	1,071	8,760	3,895	-	13,727
	30%	-	8,170	14,576	5,075	7,357	35,182
$250 RNS Cost	0%	-	1,032	8,803	3,707	-	13,542
	30%	-	8,188	14,565	4,873	7,338	34,965
2,500 RNS Cost	0%	-	838	9,177	3,732	-	13,747
	30%	-	8,183	14,584	5,109	7,330	35,205
$7.50 NG Price	0%	-	2,755	7,837	3,135	-	13,727
	30%	-	8,337	11,255	4,629	10,688	34,910

The impact of increasing wind penetration on prices for energy and reserve is shown in Table AB-4 and Table AB-5. With higher wind penetration levels, the marginal unit of generation shifts away from higher-cost natural gas toward lower-cost coal and also wind power during periods when it is curtailed (which happens as frequently as 23% of the time in the base case with 30% wind). Therefore, the average energy price decreases with higher wind penetration levels. However, more wind capacity also increases demand for reserve, thereby increasing the average reserve price. As a result, the share of revenue obtained from providing reserve (instead of generation) increases substantially with wind capacity. A majority of the reserve is provided by gas-fired units (NGCC and NGCT plants) (Table AB-3), which therefore benefits the most from the increasing prices in the reserve markets. Coal units also provide reserve during some periods, whereas nuclear units are assumed inflexible and can therefore not provide reserve at all. Table AB-4 and Table AB-5 also show that scarcity pricing has a substantial impact on prices. When prices are set to the reserve shortfall penalty during reserves scarcity, prices increase for both energy and reserve substantially, and this is the case for all wind penetration levels.

Table AB-4. Average Energy Prices for Different Wind Penetration Levels (0% to 30%)

		0%	10%	20%	30%
Baseline	No Scarcity	$29.44	$27.47	$24.35	$20.95
	Scarcity	$40.25	$40.95	$34.54	$31.84
$250 RNS Cost	No Scarcity	$29.44	$27.72	$24.36	$20.92
	Scarcity	$40.74	$42.75	$39.86	$33.50
2,500 RNS Cost	No Scarcity	$29.15	$27.75	$24.08	$20.94
	Scarcity	$38.98	$43.35	$37.86	$38.94
$7.50 NG Price	No Scarcity	$36.68	$34.08	$30.53	$27.71
	Scarcity	$57.09	$50.40	$48.65	$49.50

Table AB-5. Average Reserve Price for Different Wind Penetration Levels (0% to 30%)

		0%	10%	20%	30%
Baseline	No Scarcity	$0.00	$0.61	$5.54	$7.51
	Scarcity	$10.80	$14.09	$16.73	$18.41
$250 RNS Cost	No Scarcity	$0.00	$0.93	$5.80	$7.50
	Scarcity	$11.29	$15.95	$21.30	$20.07
2,500 RNS Cost	No Scarcity	$0.00	$0.78	$5.80	$7.58
	Scarcity	$9.82	$16.38	$19.03	$25.57
$7.50 NG Price	No Scarcity	$0.00	$2.69	$9.80	$11.74
	Scarcity	$20.42	$19.02	$27.92	$33.53

The impacts on revenue sufficiency for thermal generators are illustrated in Figure AB-2 through Figure AB-4. The marginal profit per unit generation for each technology tends to drop as wind penetration increases when scarcity pricing is not included in the market price of energy and reserve (Figure AB-2). Hence, without scarcity pricing, the increases in reserve prices for higher wind levels do not make up for the decreases in energy prices. Moreover, none of the generation technologies make a profit. The marginal profits for nuclear, coal, and NGCC units are similar, in the range of $-30 to $-50 per MWh, whereas the marginal profits for NGCT generators are much lower, ranging from $-105/MWh to $-200/MWh in the four cases. This is because the NGCT units operate relatively infrequently and primarily provide reserve during peak periods. Moreover, without scarcity prices NGCTs will never see prices exceed their marginal costs.

The results shown in Figure AB-2 suggest that some form of scarcity pricing or other incentive mechanism is needed to ensure that thermal generators maintain profitability. Scarcity pricing leads to significant increases in profits for all technologies, as illustrated in Figure AB-3. The marginal profits for NGCC and NGCTs fluctuate around zero across the wind penetration levels for all four cases; hence, the two technologies that comprise the optimal expansion mix do recover their total costs in most cases when the reserve shortfall penalties are reflected in energy and reserve prices. This trend of revenue sufficiency for NGCC and NGCT occurs regardless of the level of the reserve shortfall penalty, because a lower penalty gives more frequent reserve shortfalls and vice versa. However, a higher reserve shortfall penalty makes the profits of the NGCT units very volatile.

Figure AB-3 shows that the existing nuclear and coal units still do not break even, even with scarcity pricing reflected in energy and reserve prices. This is largely because of their high capital costs and because they are not part of the optimal expansion portfolio. Marginal profit also tends to drop as a function of the wind penetration level for the baseload technologies, because they are more exposed to the reduced energy prices. Coal plants, and particularly inflexible nuclear units, are also less able to benefit from the increase in reserve prices with wind penetration compared to NGCCs and NGCTs. In the high natural price case, however, coal unit profits are close to zero. In this case, coal is also part of the optimal expansion mix for low wind penetration levels (Figure AB-1).

An alternative mechanism to ensure revenue sufficiency is a capacity payment, which could be provided through a capacity market, as discussed in this report. Figure AB-2 shows the required capacity payment for different technologies to break even in the four cases, under the assumption that scarcity pricing is not reflected in energy and reserve prices. The required capacity payments for NGCC and NGCT plants remain more or less the same for different wind penetration levels and also across the four cases. In contrast, nuclear and coal plants need much higher capacity payments to break even. Moreover, the required capacity payment incentive increases with more wind power. Hence, the analysis of break-even capacity payments also indicates that baseload plants are more exposed to revenue insufficiency than intermediate and peak-load plants as wind penetration increases.

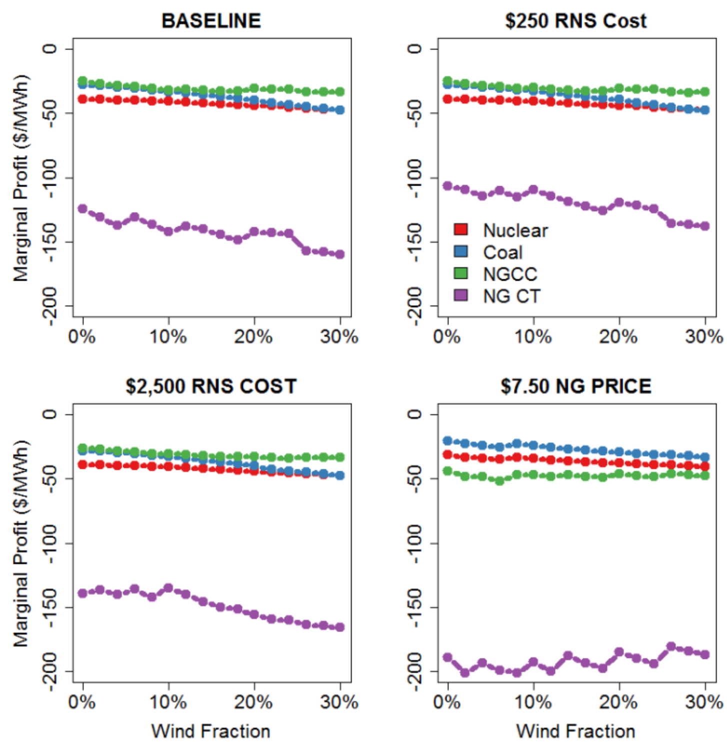

Figure AB-2. Marginal profit of thermal generators without scarcity pricing

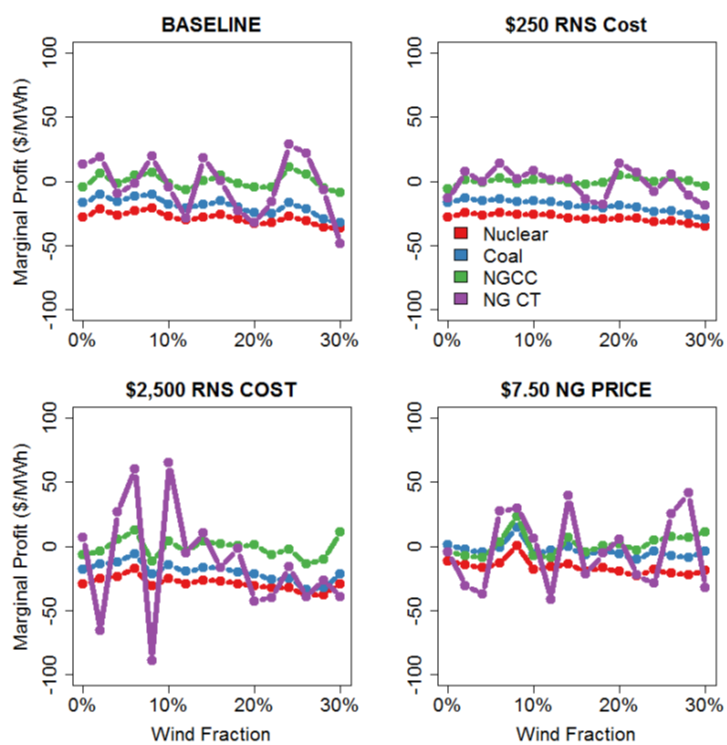

Figure AB-3. Marginal profit of thermal generators with scarcity pricing

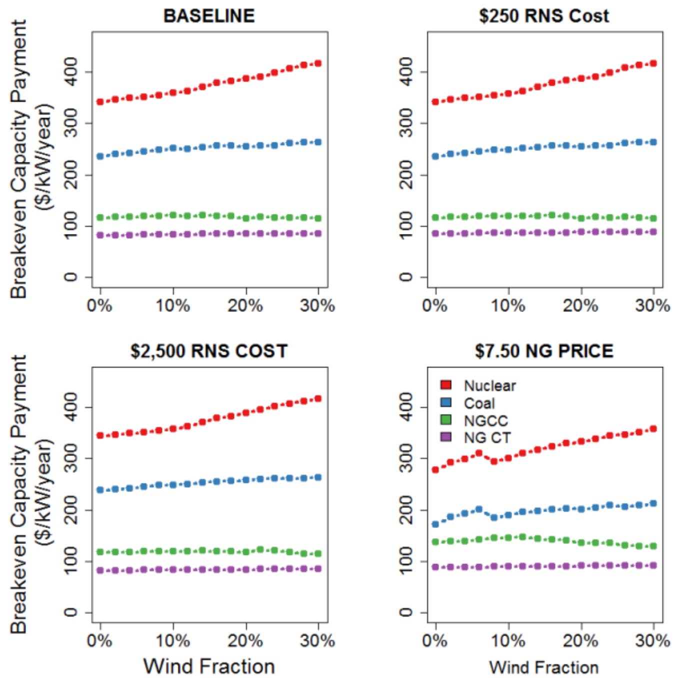

Figure AB-4. Break-even annual capacity payment by technology without scarcity pricing in energy and reserve markets

Conclusions

From the case study summarized in this section, we identify several general trends. First, higher levels of wind generation tend to reduce energy prices, but at the same time increase reserve prices because of higher reserve requirements as a result of the increasing wind power forecast uncertainty. Second, new wind generation primarily displaces generation from baseload power plants, whereas additional peaking generation is required to account for the variable and uncertain wind generation and to provide additional needs for spinning reserve. Third, without scarcity pricing, the profits for all thermal technologies are always negative and decrease with increasing wind penetration. Scarcity pricing through reserve shortfall penalties tends to provide revenue sufficiency for NGCC and NGCT plants at all wind penetration levels, but generator profits are highly dependent on the frequency of reserve shortfalls. In contrast, nuclear and coal plants see reductions in profitability with higher wind penetration, even with scarcity pricing. Fourth, in the absence of scarcity pricing, the capacity payment required to break even does not change as a function of wind penetration level for NGCC and NGCT plants, whereas nuclear and coal plants require higher capacity payments with increasing wind. Overall, these results suggest that it is likely that baseload plants will suffer more revenue sufficiency problems than peaking plants as the penetration of wind power and other renewables increase. Finally, for a more detailed presentation of this analysis and the underlying modeling framework, we refer to Levin and Botterud (2014).

Appendix B References

Jin, S.; Botterud, A.; Ryan, S.M. (forthcoming). "Temporal Versus Stochastic Granularity in Thermal Generation Capacity Planning With Wind Power." *IEEE Transactions on Power Systems*.

Levin, T.; Botterud, A. (forthcoming). "Capacity Adequacy and Revenue Sufficiency in Electricity Markets with Wind Power." *IEEE Transactions on Power Systems*.

Palmintier, B.; Webster, W. (2011). "Impact of Unit Commitment Constraints on Generation Expansion Planning with Renewables." *IEEE Power and Energy Society General Meeting*; July 24–28, Detroit, Michigan; pp. 1–7.