U.S. DEPARTMENT OF ENERGY

Demand Response and Energy Storage Integration Study

March 2016

Notice: *This report is being disseminated by the U.S. Department of Energy (DOE). As such, this document was prepared in compliance with Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554) and information quality guidelines issued by DOE. Though this report does not constitute "influential" information, as that term is defined in DOE's information quality guidelines or the Office of Management and Budget's Information Quality Bulletin for Peer Review, the report was reviewed both internally and externally prior to publication. This report has benefitted from review by the Study Team that includes researchers from Lawrence Berkeley National Laboratory (LBNL), National Renewable Energy Laboratory (NREL), Oak Ridge National Laboratory (ORNL), and Sandia National Laboratory (SNL). In addition, this report also incorporates input from 16 external peer reviewers representing electric utility industry, academia, and the broader stakeholder community.*

Acknowledgments

The Demand Response and Energy Storage Integration Study was sponsored by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy and Office of Electricity Delivery and Energy Reliability. The study represents a joint multi-National Laboratory effort to examine the role of demand response and energy storage in electricity systems with different penetration levels of variable renewable resources and to improve the understanding of associated markets and institutions.

We would like to thank Henry Kelly, Carla Frisch, Bill Parks, and Imre Gyuk for their support. Any errors or omissions are solely the responsibilities of the authors.

Authors

Office of Energy Efficiency and Renewable Energy	Ookie Ma
Office of Electricity Delivery and Energy Reliability	Kerry Cheung

Study Team

Lawrence Berkeley National Laboratory	Daniel J. Olsen, Nance Matson, Michael D. Sohn, Cody M. Rose, Junqiao Han Dudley, Sasank Goli, and Sila Kiliccote
	Peter Cappers and Jason MacDonald
National Renewable Energy Laboratory	Paul Denholm, Marissa Hummon, Jennie Jorgenson, and David Palchak
Oak Ridge National Laboratory	Michael Starke and Nasr Alkadi
Sandia National Laboratories	Dhruv Bhatnagar, Aileen Currier, and Jaci Hernandez
Consultant	Brendan Kirby
University College Dublin	Mark O’Malley

(This page intentionally left blank)

Executive Summary

Motivation and Background

Demand response and energy storage resources present potentially important sources of bulk power system services that can aid in integrating variable renewable generation. While renewable integration studies have evaluated many of the challenges associated with deploying large amounts of variable wind and solar generation technologies, integration analyses have not yet fully incorporated demand response and energy storage resources.

This report represents an initial effort in analyzing the potential integration value of demand response and energy storage, focusing on the western United States. It evaluates two major aspects of increased deployment of demand response and energy storage:

(1) Their operational value in providing bulk power system services
(2) Market and regulatory issues, including potential barriers to deployment.

Key questions include:

- What is the regional and temporal availability of demand response resources in the western United States?
- What is the relative operational value of demand response and energy storage in providing bulk power system services?
- How does the operational value of demand response and energy storage vary as a function of deployment and renewable penetration?
- What are potential barriers to deployment in current market and regulatory environments?

The work contained in this report stems from a number of more detailed studies published by members of the study team. Links to these publications as well as related materials can be found at energy.gov/eere/analysis/demand-response-and-energy-storage-integration-study.

This study does not attempt to:

- Assess all potential mechanisms and sources of demand response
- Assess all possible value streams of demand response and energy storage
- Consider the costs associated with the deployment and integration of demand response or energy storage
- Consider energy storage technologies deployed in distribution systems or behind the meter
- Simulate contingency events or consider the impacts of changing system dynamics
- Determine the optimal sizing or location of demand response or energy storage.

Overview of Demand Response and Energy Storage

Demand response and energy storage resources can be obtained from a number of different technologies. While these technologies can provide a range of value streams to different stakeholders, for the purpose of supporting bulk power system operations, they have the common characteristic of being able to shift energy use in time to help maintain the generation-load balance. As such, demand response and energy storage technologies are evaluated with a common framework in this study.

Demand response encompasses many different strategies by which commercial, residential, municipal, and industrial electricity customers are incentivized to adjust, in the short-term, when they use electricity (in contrast to energy efficiency and energy conservation that seek to reduce total electric load). Similarly, energy storage technologies like pumped storage hydropower, batteries, and flywheels save electricity produced at one point in time for later use, effectively shifting demand. Both demand response and energy storage technologies can be used to provide energy services and/or ancillary services such as frequency regulation and contingency reserves. A key difference between demand response and energy storage is that the use of demand response is inherently tied to specific end-uses with associated temporal and spatial patterns of electricity consumption. This difference also has implications for the availability of demand response resources over time.

Sources of Value

This study focuses on assessing two sources of value that demand response and energy storage can provide to bulk power system operations: energy services and operating reserves. The value of energy services reflects the variable costs of operating the power system, which are primarily fuel costs and variable operations and maintenance costs associated with committing and dispatching a generator. Operating reserves, part of a larger class of services known as ancillary services, are used to ensure electric system balancing for short-term variations between generation and load, including during a contingency (large, sudden, and unexpected loss of generation or transmission capacity). The value of operating reserves reflects a combination of different cost components:

(1) Incremental fixed costs associated with generation equipment required to provide operating reserves
(2) Operational costs associated with operating generators at part load (e.g., efficiency losses) and responding to short-term variations (e.g., increased wear and tear, maintenance, time spent offline for repairs)
(3) Opportunity costs associated with occurrences when a generator withholds energy production in order to supply operating reserves (e.g., generator lost profits).

Three operating reserve products are evaluated in this study: frequency regulation, contingency reserve, and ramping reserve.

Study Approach and Limitations

This study performs a set of power system simulations of the western United States where demand response and energy storage resources are deployed under differing levels of wind and solar penetration. For simplicity, the deployment of these resources was studied without changing the generator composition or transmission capacity in the base scenarios. This approach may lead to different operational values compared to ones with higher or lower reserve margins. Additionally, this study does not consider the upfront capital costs needed to deploy these resources or the benefits of deferred asset investments. The power system simulations hold capacity hour-by-hour for operating reserves but do not explicitly model grid resources providing frequency regulation and responding to contingency events. As a result, this study does not consider the potentially higher value of resources that can more accurately follow frequency regulation dispatch or respond to real-time, unforecasted conditions. The modeled deployments of demand response and energy storage resources are evaluated separately and compared against common base cases.

The operational value of demand response and energy storage is quantified in two ways representing two perspectives. In the first approach, the difference in total cost for operating the system for a study year (i.e., total production costs) between two modeled scenarios is used to estimate the operational value of adding various amounts of demand response or energy storage. Total production cost savings represent a societal value derived from avoided fuel and operations and maintenance costs across the entire system. In the second approach, the operational value of demand response and energy storage is estimated by using short-run marginal costs of production for different services calculated from the simulations. The marginal cost calculation helps disaggregate operational value for the different services, hour-by-hour, and for different geographic locations. We further assume that short-run marginal costs are a proxy for hourly prices for bulk power system services and represent the value a market participant might receive, hypothetically, if selling services into an organized wholesale market.

This study evaluates demand response and energy storage deployed in the Western Interconnection of

Figure ES-1. Study area including 36 balancing authorities (small print) and 12 reserve sharing group (large print) assumptions for the 2020 study year

the United States, illustrated in Figure ES-1. The modeling approach is based on the Western Wind and Solar Integration Study Phase 2 (WWSIS-2) (Lew et al. 2013). This includes two base cases: a low renewable case with 14% of electricity from wind and solar power and a high renewable case with 33% of electricity from wind and solar power. Consistent with the WWSIS-2, the transmission capacity is increased from the low renewable case to the high renewable case to accommodate the different generation profiles. Simulations over the full 2020 study year are conducted in the production cost model PLEXOS. Production cost models calculate various costs of system operation, including fuel, operations and maintenance, and generator starts; however, they do not consider or include capital costs. The results are also based solely on the day-ahead simulation. This approach takes into account wind and solar forecast errors between the day-ahead and real-time dispatches by holding additional operating reserves (frequency regulation and ramping reserve) to meet anticipated increases in variability and uncertainty. However, the real-time dispatch is not simulated.

Demand Response and Energy Storage Deployment Scenarios

In this study, we model one demand response deployment scenario and a set of deployment scenarios for two general classes of energy storage technologies. The two energy storage technology classes include an operating reserves-only device and one that can be co-optimized for both energy and operating reserves. Energy storage technologies are assumed to be connected at the transmission level. Customer-sited electric energy storage (e.g., batteries) is not considered in this analysis, while customer-sited thermal energy storage (e.g., electric water heaters, building thermal capacity) is categorized as demand response resources. These deployment scenarios are modeled independently but use the same analysis framework. For simplification, optimal placement, optimal sizing (capacity and duration), and capital costs of these resources are not considered. While these limitations impact the accuracy of simulated operational values, the observed trends are indicative of expectation for scenarios with higher penetration of variable renewable resources.

We develop a scenario for the available demand response resource based on analysis of different end-uses across commercial buildings, residential buildings, municipal functions, and the industrial sector. The estimated resource is calculated by assessing the fraction of each end-use electric load that can be utilized for demand response based on assumptions regarding the physical constraints of the underlying end-use appliance and equipment systems. Service constraints, the deployment of suitable control systems, and historic rates of retail customer participation in demand response programs were also factored into the scenario development. End-uses included in the analysis constitute 30% of total electricity use in the western United States. However, after accounting for anticipated participation rates, only 4% of total electricity use is assumed to be enrolled in demand response programs. This leads to a study scenario with an annual cumulative availability of 11.3 TW-h, or 1.4% of total electricity use. Imposing only the physical and service constraints, assuming full participation and adoption of necessary control systems, increases the estimated resource size by a factor of 8. In our approach, this larger quantity represents the technical potential for demand response in the study scenario.

The first class of energy storage technology modeled is an operating reserves-only device that resembles a short duration battery with 1 hour of energy storage at rated capacity and 80% roundtrip efficiency (similar to a lithium-ion battery). This first class of modeled devices is sized to meet 50% of the average hourly operating reserve requirement within each region. One set of scenarios implements frequency regulation-only devices at a total deployment of 487 MW in the low renewable case and 597 MW in the high renewable case. The other set of scenarios implements contingency reserve-only devices at a total deployment of 1,314 MW in both renewable cases.

The second class of energy storage technology modeled is a device capable of providing both energy and operating reserves, resembling a battery with 8 hours of storage capacity at rated power output and 75% roundtrip efficiency (similar to a sodium-sulfur battery). The simulations include scenarios where the devices provide only energy and also scenarios where the devices are co-optimized to provide both energy and operating reserves. Energy storage in these scenarios were sized to equal about 3.3% of average load in each region, resulting in a total modeled deployment of 3,045 MW.

Modeling Results

Value of Energy and Operating Reserves

The value of demand response and energy storage is assessed relative to the system characteristics in the baseline scenarios. In the base system without additional demand response or energy storage, the annual operational costs determined from production cost modeling are about \$11.5 billion and \$6.5 billion in the low and high renewable cases, respectively. Production costs are lower in the high renewable case because the renewable generation has no direct fuel costs.

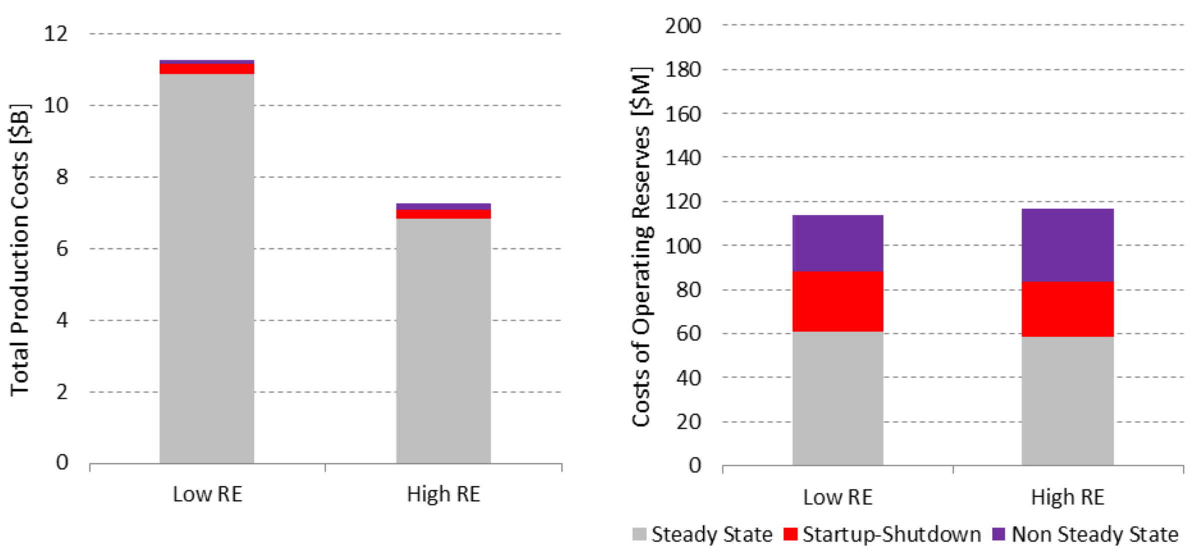

Figure ES-2. For the Western Interconnection model, total production costs (left) and production costs associated with just the provision of operating reserves (right)

Figure ES-2 illustrates that most of the simulated production costs are associated with the provision of energy. Only about 1.0% (\$115 million) of these costs in the low renewable case and 1.8% (\$118 million) in the high renewable case are associated with the provision of operating reserves. Despite the lower energy production costs in the high renewable case, total costs for operating reserves are higher due to the greater operating reserve requirements needed to accommodate the increased variability and uncertainty.

Production costs associated with operating reserves can be distinguished by generator operating mode. Steady state costs are associated with fuel and variable operations and maintenance costs for generators producing electricity at a fixed operating point; startup and shutdown costs are costs for starting and stopping a generator; and non-steady state costs are additional operations and maintenance costs to provide frequency regulation. While cost components like generator startup and non-steady state operation are relatively small compared to total production costs, they are significant components for the production costs associated with operating reserves. Avoiding these high cost components is an important source of value for demand response and energy storage.

Operational Value of Demand Response

Figure ES-3 shows the operational values associated with the demand response resource study scenario deployed for the low and high renewable cases in the study year. The total production cost savings (i.e., the difference in production costs between scenarios with and without added demand response resources) is shown on the left and the implied market value (i.e., marginal cost of every service multiplied by the amount provided by demand response resources, aggregated across every hour and every region) is shown on the right. In both renewable cases and for both perspectives, the operational values are normalized by the estimated cumulative availability of demand response resources used for bulk power system services.

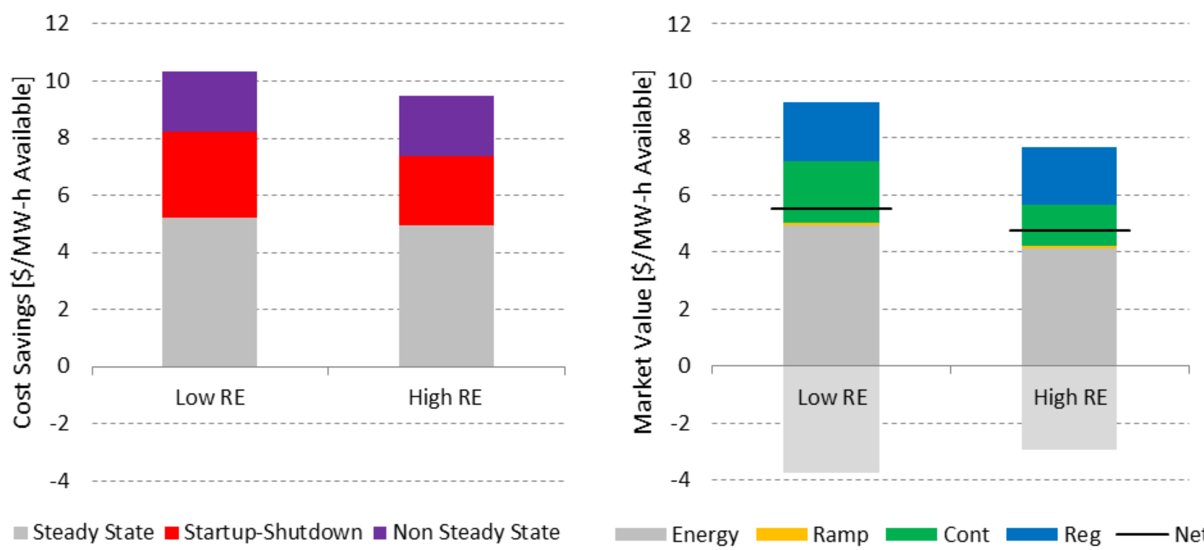

Figure ES-3. Operational value of demand response resources in terms of total production cost savings (left) and the implied market value (right) in the low and high renewable cases.

Figure ES-3 (left) shows that the majority of operational savings achieved through the use of demand response comes from avoiding steady state costs. However, a high fraction of savings also comes from avoided generator startups and shutdowns and avoided non-steady state operational costs. The gray bars in Figure ES-3 (right) show the implied market value for energy shifting in terms of load sheds (positive energy value) and load recovery (negative energy value) at the assumed level of demand response deployment. While energy shifting constitutes the majority of gross market value, the provision of operating reserves constitutes the majority of net value after considering the cost of load recovery (approximately one-third of gross market value). From both the perspective of total production cost savings and the implied market value, the per-unit operational value of demand response is lower for the higher renewable case. This trend stems partly from the reduction in energy production cost at increased levels of renewable generation, which decreases steady state savings and implied energy market value, and partly from the increased variability associated with wind and solar, which leads to fewer savings achievable from avoiding generator starts and stops needed to meet the larger operating reserve requirements.

The demand response resource in the study scenario shifts up to 0.8% of average daily energy use and reduces up to 1.9% of load during the top 100 load hours in the study year. In the model, demand response resources also provide a large fraction of operating reserves, meeting 39% of frequency regulation requirements in the low renewable case and 33% in the high renewable case, and 22% of contingency reserve requirements in both the low and high renewable cases.

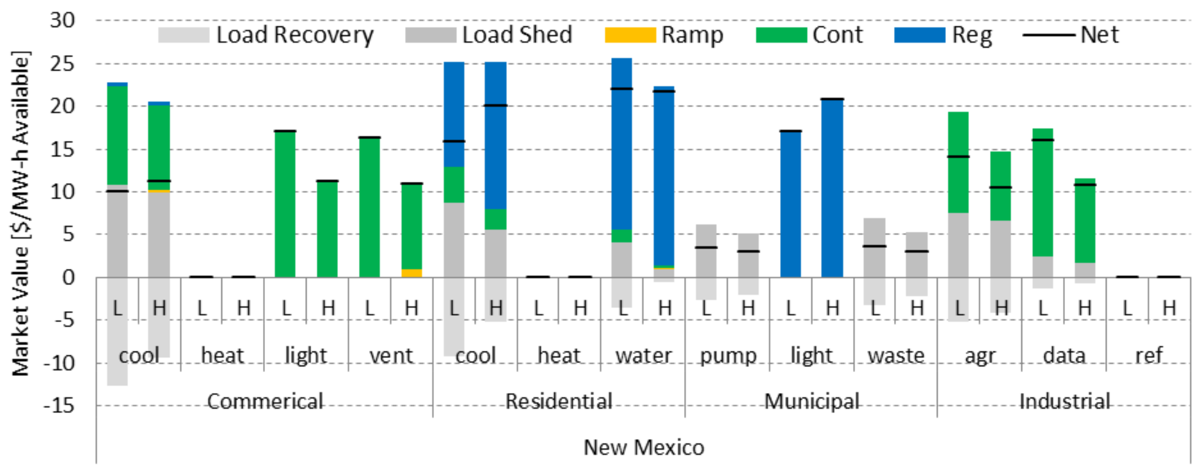

Figure ES-4. Operational value of different types of demand response resources in New Mexico Zone at low (L) and high (H) levels of wind and solar generation, based on marginal costs

System-specific issues such as local generation mix, load profiles, resource capabilities, and transmission capacity cause significant variations in operational values by demand response resource type and by region. Figure ES-4 provides the implied market value of various demand response resource types for different services in the modeled New Mexico zone. The operational value of each demand response resource type depends on the correlation of resource availability with times of high production costs as well as the flexibility in scheduling energy use. With increased penetration of wind and solar generation, certain drivers impact net value in both directions. For example, Figure ES-4 shows that some resources, such as residential space cooling, see increased net value. Other resources, such as data centers, see decreased net value. Greater penetration of wind and solar generation tend to reduce the marginal costs of contingency reserve (green bars) and increase the marginal costs of frequency regulation (blue bars). The addition also alters the marginal costs of energy during times of peak loads, which can have important implications for the value of energy shifting (gray bars). Load sheds tend to have reduced value in the high renewable case; however, the costs of load recovery also tend to be lower. The net energy arbitrage value can be higher or lower depending on the individual resource type and the region.

Operational Value of Energy Storage

Figure ES-5 shows the operational value of the first class of energy storage devices modeled for the low and high renewable cases. In one set of scenarios, energy storage devices can provide only frequency regulation. In the other set, energy storage devices can provide only contingency reserve. The left of Figure ES-5 shows the operational value of the energy storage deployed as total production cost savings, and the right represents the implied market value based on simulated marginal costs. In both renewable cases and for both perspectives, the operational values are normalized by the energy storage capacity deployed for the respective scenarios.

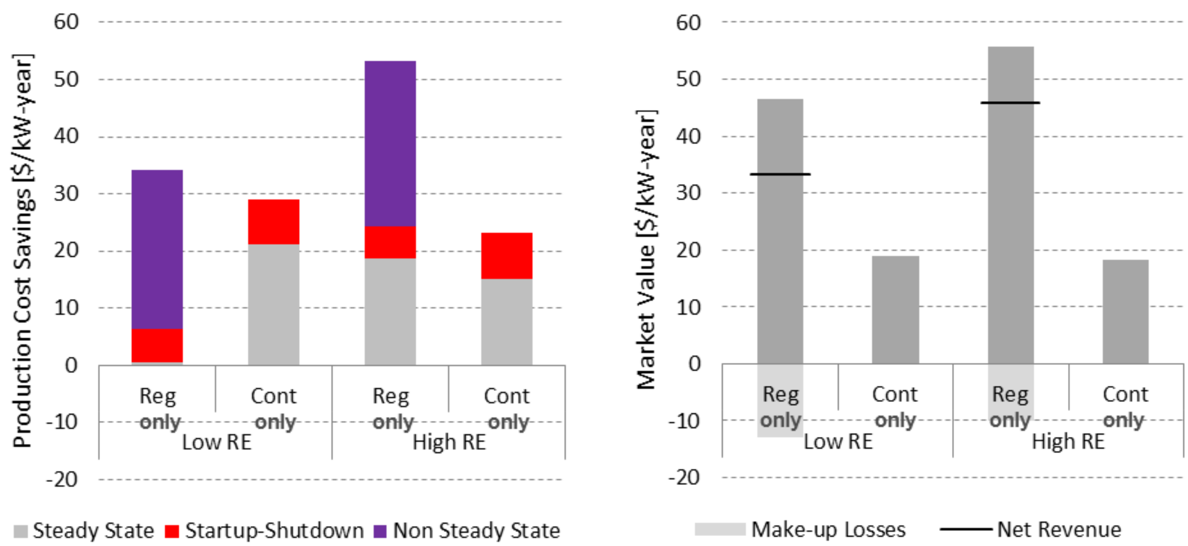

Figure ES-5. The operational value of energy storage in two sets of models runs, one in which energy storage can provide only frequency regulation and another in which energy storage can provide only spinning contingency reserves. Value is determined by total production cost savings to the system (left), and the implied market value calculation based on simulated marginal costs (right). Make-up losses are energy costs associated with the storage efficiency losses.

Compared to the low renewable case, the results for the high renewable case show an increase in the per-unit value for the frequency regulation-only storage devices but a decrease for contingency-only storage devices for both perspectives. The trends observed are partly from lower energy production costs at increased levels of renewable generation, which decrease steady state savings in the contingency-only scenario, and partly from the increased need for frequency regulation, which leads to greater savings in the frequency regulation-only scenario.

Figure ES-6 provides the simulation results for the second class of energy storage devices modeled in energy-only and co-optimized scenarios, for both the low renewable and high renewable cases. As in Figure ES-5, the left panel provides the total production cost savings associated with the addition of this class of energy storage devices normalized by the amount deployed while the right panel provides the implied market value normalized by the amount deployed. As shown in the results, adding the ability to co-optimize energy and operating reserves increases the operational value of energy storage devices.

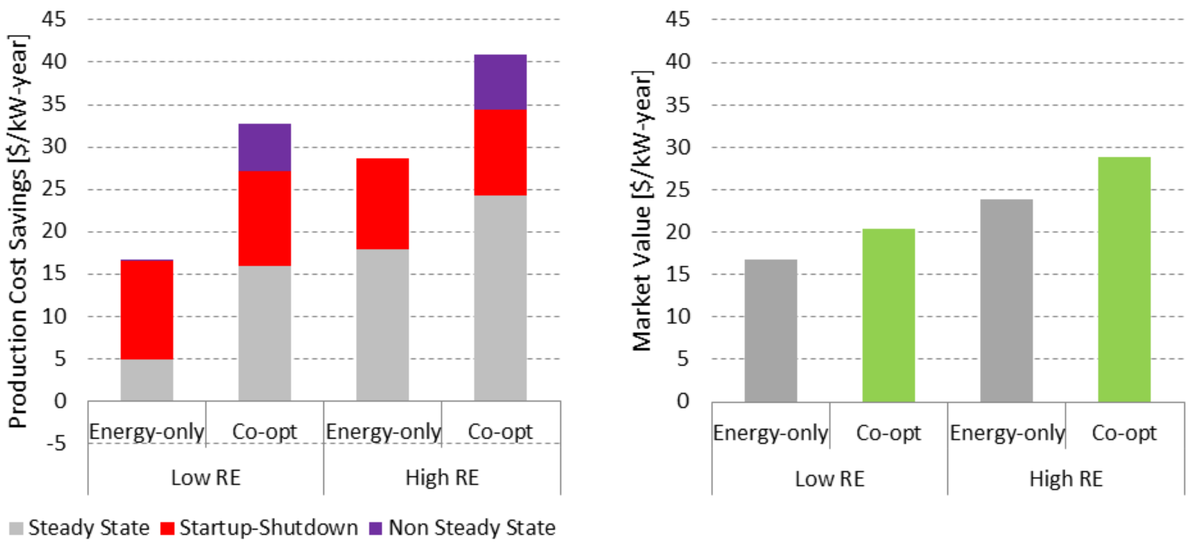

Figure ES-6. The operational value of energy storage (co-optimized for energy and operating reserves) in terms of total production costs savings (left) and the implied market value (right) in the low and high renewable cases

Comparing Figures ES-5 and ES-6 shows that the per-unit value of the co-optimized devices is less than the frequency regulation-only devices. This counterintuitive result is partly due to the different amounts of energy storage deployed in the different scenarios. The total deployment of co-optimized devices is much greater in size than the frequency regulation-only devices. Because the incremental value of energy storage tends to decline with additional deployment (as with any resource, including demand response), the per-unit value tends to be smaller for scenarios with larger deployments of energy storage. Figure ES-7 shows the modeling results of the co-optimized devices for three levels of energy storage deployment, with an approximate curve fit to illustrate this relationship.

Furthermore, the large energy capacity (8-hour duration at the rated power) of the co-optimized devices introduces additional drivers that impact the modeled operational value. While the devices in the frequency regulation-only scenario (sized to meet 50% of the average hourly operating reserves) are nearly fully utilized providing 49% of the system requirement, the devices in the co-optimized scenario, although significantly larger, provide only 27% of the frequency regulation requirement. When providing frequency regulation, the co-optimized devices can provide other services like contingency reserves and energy shifting with the remaining capacity. When discharging energy, the energy storage device suppresses the marginal cost of energy. Displaced generation can offer capacity as operating reserves, which also suppresses the marginal cost of operating reserves. As a result, the larger co-optimized devices modeled provide a combination of services whose per-unit value is actually less than the much smaller frequency regulation devices.

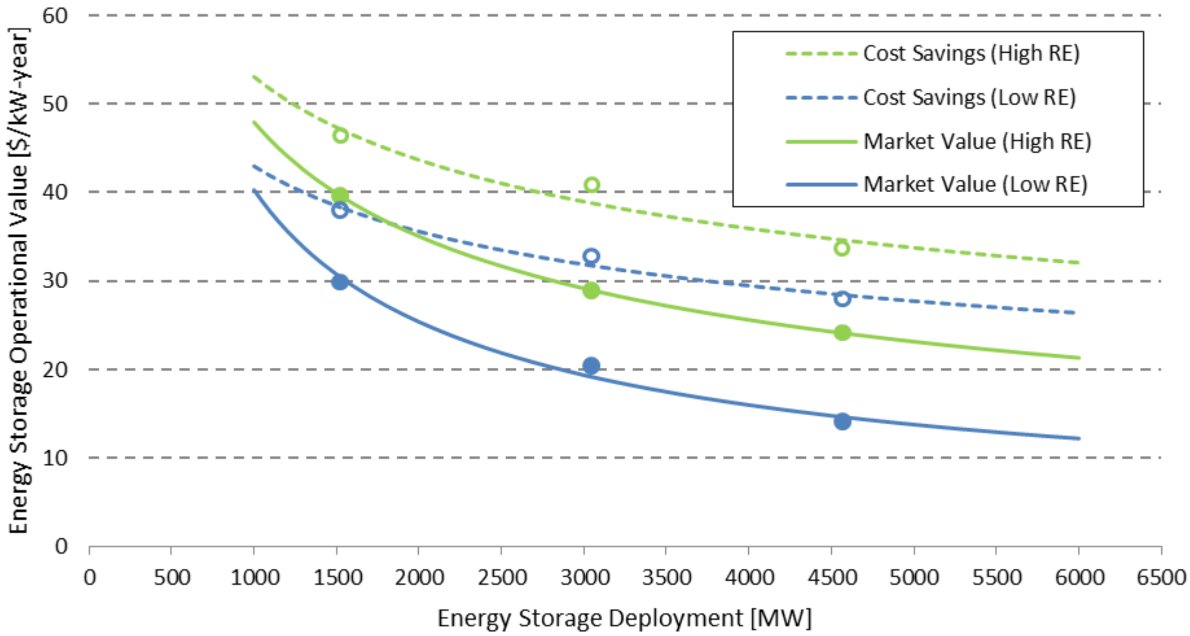

Figure ES-7. The per-unit operational value of the co-optimized energy storage devices as a function of total deployment in the low and high renewable cases. Cost savings represent total production cost savings; market value represents the implied market value based on marginal costs of production.

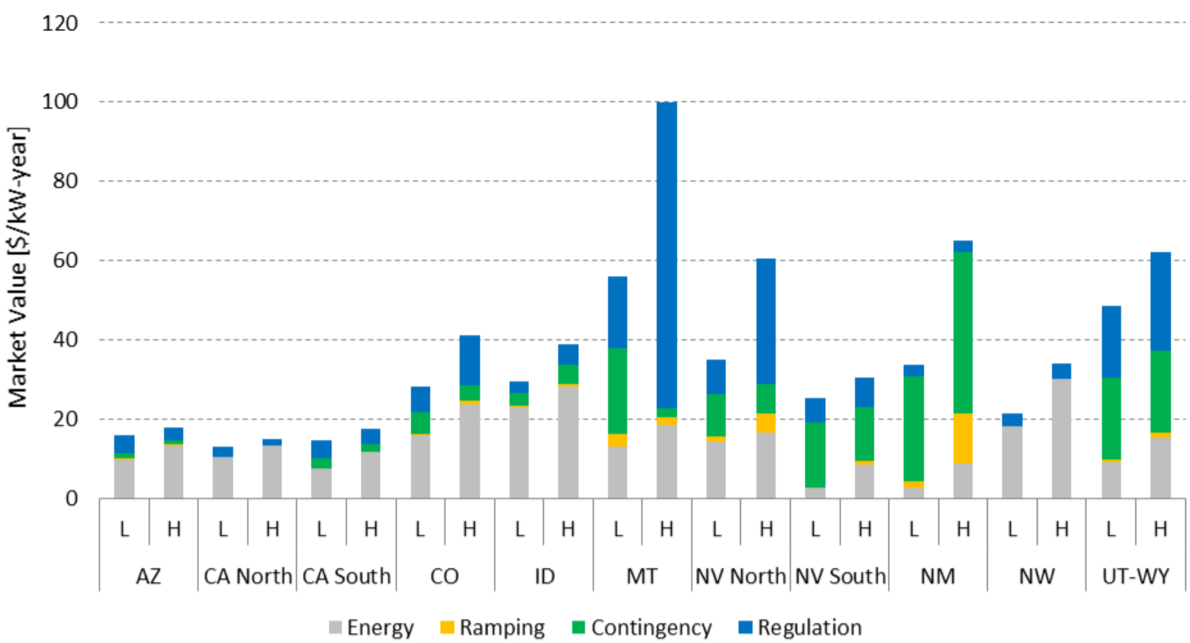

Figure ES-8. The operational value of the co-optimized energy storage devices for the various services in the different regions based on marginal costs of production

As demonstrated by the demand response results, there can be significant regional variation in operational value for energy storage in both the low and high renewable cases. Figure ES-8 provides an indication of this variation by showing the implied market value for various services calculated by using simulated marginal costs of production. The analysis results indicate an increase in value for energy storage in the presence of greater deployment of wind and solar power. The increase in energy storage value observed may be attributed to the greater need for operating reserves and the tendency of renewable generation to suppress off-peak energy marginal costs. Lower off-peak energy production costs increase the value of energy shifting, as can be seen by the greater steady state cost savings in Figure ES-6. This second factor is less evident for many demand response resources that have constraints on availability. For instance, there is limited thermal inertia in buildings, so cooling loads may not be able to shift to hours with the lowest energy costs.

The Importance of Capacity Value

Production cost models can provide estimates of the operational value of demand response and energy storage but do not consider the value of providing system capacity. The ability to replace conventional generation is a potential source of value and, in some cases, may be substantially greater than the operational value. This study does not attempt to directly quantify the value of capacity for demand response or energy storage resources due to model limitations and assumptions used in the deployment scenarios. Additionally, the issue of persistence and availability of demand response over time complicates the discussion of capacity value. Production cost modeling does not dispatch resources for contingency events and do not consider long-term planning needs for system adequacy.

In the Western Interconnection model, the average production cost savings resulting from energy storage devices (co-optimized for energy and operating reserves) ranges from \$33–\$41/kW-year in a scenario where 3.0 GW of energy storage is deployed. For the sake of a simple first-order comparison, capacity values for non-generation resources, such as demand response or energy storage, can be approximated based on a proxy generation resource like a natural gas-fired combustion turbine; though the capacity value may not be equivalent. There is a large range of estimates for the annualized cost of a new combustion turbine, with examples ranging from a lower value of \$77/kW-year (taken from a Colorado filing) to a higher value of \$212/kW-year (taken from a California filing). This implies the capacity value of an energy storage device could be considerably higher than its operational value, assuming the device can provide capacity benefits and is deployed in a region deficient in capacity.

Similarly, the capacity value of demand response could be significantly greater than its operational value. The total production cost savings of the modeled demand response resource is about \$110 million per year. The total estimated reduction in peak load from utilizing the demand response resource is about 3.1 GW (based on the ability for demand response to shift energy use during the top 100 hours of each balancing authority area), which would translate to \$240–\$660 million per year in avoided capacity, assuming the proxy resource is a natural gas-fired combustion turbine.

While capacity value for non-generation resources can be quite significant, realizing this value may require capacity providers to be capable of several hours of response duration—6–10 hours in some regions. Additionally, the availability and persistence of the resource may need to be assured across multiple years to address concerns regarding resource adequacy. However, many types of demand response resources and energy storage technologies have duration limits that may make it difficult to serve as a capacity resource. Meanwhile, other bulk power system services, such as operating reserves, require significantly less response duration. Thus, a non-generation resource capable of providing 1 MW of capacity over several hours may alternatively be used to provide many more MWs of a service like contingency response over those same hours.

Market and Regulatory Issues

Market and regulatory barriers for demand response and energy storage participation in bulk power system services fall into several categories, including issues associated with eligibility, cost, and revenue capture. Eligibility barriers relate to how regional reliability councils and balancing authorities define bulk power system services. These definitions can explicitly include or exclude certain classes of resources or implicitly exclude them by defining services that only certain classes can provide. Cost barriers relate to how regional reliability councils and balancing authorities define the attributes of performance and the required enabling infrastructure necessary for participation in wholesale markets. While these issues do not directly prevent demand response and energy storage from providing bulk power system services, they can strongly impact the costs associated with their provision, and thereby indirectly limit participation.

Revenue capture barriers represent the potentially limited ability of demand response and energy storage to be compensated appropriately for the value that they could provide to the grid. While these barriers exist for all resources, they are more challenging for non-generation resources such as demand response and energy storage. Revenue capture barriers also include challenges associated with smaller providers that would require an intermediary to bring their resources to the wholesale market. The intermediary, either the retail electricity provider or a third party aggregator, may not see a business case to do so, thus preventing these smaller providers from participating in the provision of bulk power system services.

Recent North American Electric Reliability Corporation (NERC) standards and Federal Energy Regulatory Commission (FERC) Orders (e.g., Order 755, Order 784) have sought to address a number of these deployment barriers. As an example, pay-for-performance market revisions can better compensate demand response and energy storage for the capability of faster and more accurate response when providing frequency regulation, compared with conventional generation. Furthermore, some balancing authorities have addressed cost barriers for smaller demand response providers that face high per-unit enablement costs. Costs can be prohibitive if each resource is required to invest in the same monitoring and communications equipment and connection to a dedicated communication network typically used with large generators.

Conclusions

This study conducts a preliminary assessment of the potential for demand response and energy storage resources to provide bulk power system services for the Western Interconnection in scenarios with low and high penetration levels of wind and solar. It also examines the market rules and regulations that impact the use and deployment of these non-generation resources nationally. Overall, these efforts yield a number of key findings:

- A significant fraction of operational value attributable to demand response and energy storage resources are the avoided costs associated with generator startups/shutdowns and reduced costs associated with generators modulating output while providing frequency regulation. These costs are nominal when looking at total production costs but represent a significant fraction of costs for operating reserves.
- Due to the limited temporal flexibility of demand response resources, the provision of operating reserves has more market value than energy shifting services, assuming prices are based on marginal costs of production. However, the availability of these resources to provide energy shifting services helps to optimize the operation of the broader system.
- Energy storage provides greater value in scenarios with higher renewable penetration due to the increased need for operating reserves and the greater opportunities for energy arbitrage through the storage of low-cost, off-peak electricity. Additionally, co-optimization of energy and operating reserves results in greater value than just energy shifting services alone.
- Market structures can limit the ability of any new entrant, including demand response and energy storage, to be compensated commensurate with savings they create. Existing markets do not include generator startup costs in price formulation, so they may not necessarily compensate demand response or energy storage for reducing these costs along with other value streams.
- Marginal costs for operating reserves include lost opportunity costs from generators (forgone profit of selling energy). However, the lost opportunity cost component is defined as zero for demand response and energy storage resources providing only operating reserves (forgone profit of selling energy is zero). Large penetration of these resources can saturate the market for operating reserves and drive down market clearing prices.
- While capacity value of demand response and energy storage was not studied in detail, a simple calculation based on the assumption of a proxy generation resource suggests that capacity value could be several times larger than the operational value. However, realizing this value may require resources to provide many hours of response duration (e.g., 6–10 hours), generally increasing the cost of providing these services.
- While there are multiple challenges to deploying large customer demand response and energy-limited storage resources in wholesale markets, smaller customer resources that seek to provide bulk power system services face additional barriers. First, they may require an aggregator that might not see a business case to provide those services. Second, communications and control requirements imposed on individual large providers can be cost-prohibitive if applied equally to smaller providers. However, these requirements could be modified and technical hurdles overcome to reduce implementation costs without compromising system reliability.

(This page intentionally left blank)

Table of Contents

1 Introduction

Grid modernization and advances in technologies are enabling resources like demand response and energy storage to support a wider array of electric power system operations. Historically, thermal generators and hydropower combined with the transmission and distribution infrastructure have been mostly adequate to serve customer load reliably and with sufficient power quality. While demand response and energy storage are potential alternatives and complements to these assets in some applications, work is needed to better understand the extent to which these non-generation resources can provide cost-effective technical benefits and whether market rules and regulations are in place to facilitate their use.

Answering these questions is challenging. The U.S. electric power sector is complex from an engineering perspective and from an institutional perspective. It is highly fragmented across regions with differing resource compositions as well as differing market and regulatory environments. Furthermore, regional characteristics of the system are changing and various technology options are evolving in availability, cost, and performance. Because there is generally no single solution that can meet all the challenges of the grid, both technical and non-technical solutions will compete based on their cost-effectiveness and the institutional difficulty in their implementation.

Demand response and energy storage can have both operational value, through reducing costs associated with using grid assets, and capacity value, through offsetting the costs of procuring new assets. Cost savings can stem from efficiency gains by facilitating improved generator scheduling, more efficient generator operating points, and higher utilization of grid assets. Other value streams include increased system flexibility and the ability to support greater penetration of variable renewable resources. However, deployment can be limited by capital costs, integration costs (e.g., monitoring, communications, and controls), as well as challenges with multi-year resource availability in the case of demand response and energy losses from parasitic loads and round-trip efficiency in the case of energy storage.

Utilizations of demand response and energy storage have long histories, but experiences are predominantly limited to a few specific applications. Electric utilities have historically used demand response to manage peak prices (typically during generation shortfalls) and for curtailment in emergency situations during times of high loss of load probability (Cappers, Goldman, and Kathan 2010). Pumped storage hydropower units have extensively provided daily and weekly load balancing, following set schedules of pumping water uphill when electricity demand is low and releasing it through turbines when demand is high. These plants were often installed in conjunction with nuclear power plants to help manage the system because nuclear power plants were not intended to modulate output to meet time-varying demand (Elzinga et al. 2012). However, the dramatic changes occurring in the power system, from generators to loads, present significant opportunities to optimize the use of demand response and energy storage, advance research and development in these technologies, and support grid modernization.

Demand response and some energy storage resources have an inherent shortcoming with regard to response duration. However, the importance of speed and accuracy over duration can make demand response and energy storage well-suited to the technical requirements for many power system applications, especially with greater deployment of renewable energy and other clean technologies that can be limited or curtailed due to their inherent variability. Newer technology implementations for these resources have enhanced capabilities in response speed, ramp rates, accuracy for following system operator instructions and responding to frequency deviations, fast switching between charging and discharging, and low to zero minimum loading points.

This study evaluates two major aspects of increased deployment of demand response and energy storage: their operational value in providing bulk power system services, and related market and regulatory issues, including potential barriers to deployment. The report is organized as follows:

Section 2 provides an overview of the three bulk power systems services needed by the grid to provide reliable electricity: capacity, energy, and operating reserves. It also introduces how demand response and energy storage can provide many of those services.

Section 3 analyzes the potential operational value of demand response and energy storage in providing grid services. It describes a set of grid simulations of the western United States, where demand response and energy storage resources are deployed under differing levels of wind and solar penetration. This approach adapts and extends state-of-the-art techniques used in renewable integration analyses to better understand the value of demand response and energy storage.

Section 4 describes market and regulatory issues associated with the deployment of demand response and energy storage broadly. Specifically, it describes the different market and regulatory environments at both the retail and wholesale levels that impact the deployment of these technologies.

Section 5 summarizes major findings in the report.

Section 6 discusses areas for future work that could bring greater insight into the potential opportunities for demand response and energy storage to provide grid services in evolving market structures and act as enabling technologies for greater efficiency of grid operations under changing compositions of the generator fleet.

2 Bulk Power System Services

This study focuses on bulk power system services supplied by electric generators and their alternative provision by demand response and energy storage resources. Bulk power system services include capacity, energy, and ancillary services. There are a number of types of ancillary services (FERC 2006) (FERC 2007), but the present work looks only at a subset, called operating reserves (Ela, Milligan, and Kirby 2011). The relationships between capacity, energy, and operating reserves are shown conceptually in Figure 2-1.

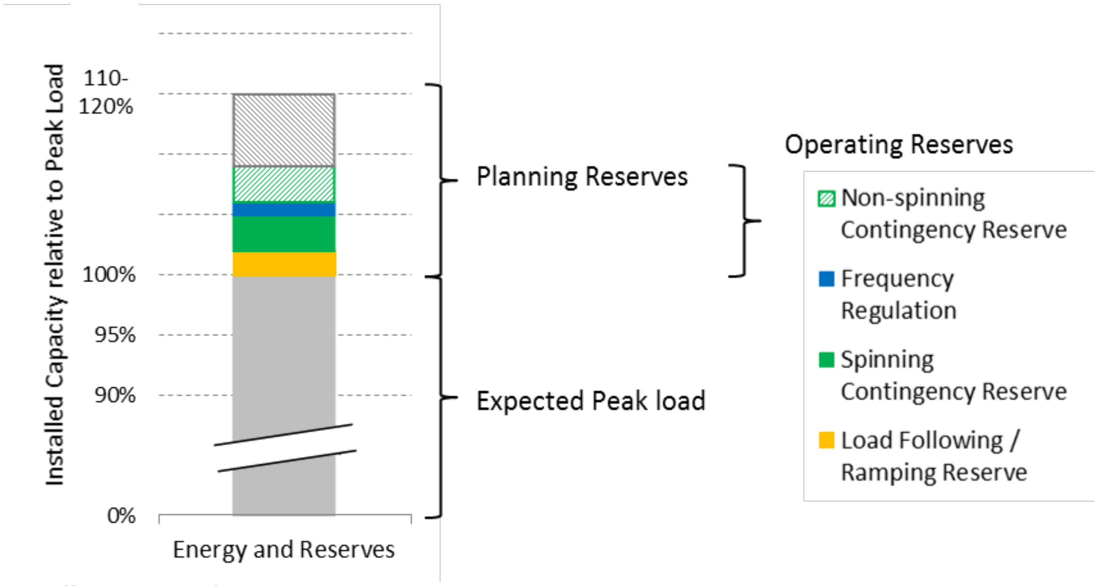

Figure 2-1. Different types of bulk power system services included in this study and the relationship between the resources needed to provide energy and the amount needed to provide operating reserves during times of peak load. Planning reserves are needed to meet system reliability standards overall (i.e., from year to year), while operating reserves are needed to ensure reliability within operational time frames that are much shorter (i.e., hourly).

Capacity, energy, and operating reserves have different units of measure. Electricity service, when referring to energy, is sold in units like kilowatt-hours (kWh) or megawatt-hours (MWh). However, capacity and operating reserves involve the commitment of resources to offer energy during set times. This can, thereby, be measured in units of power (i.e., kW or MW) times the service duration, signified (in this report) by a 'dash' (i.e., kW-h or MW-h). While energy involves the physical buying and selling of electricity, capacity and operating reserves provide the insurance that energy will be available when and where it is needed.

Requirements for bulk power system services are enforced by balancing authorities (i.e., those responsible for maintaining electric generation-load balance within reliability standards) that oversee balancing authority areas, which are composed of one or more electricity service territories covered by load serving entities (i.e., electric utilities that provide electricity service to retail customers). These load serving entities rely on a combination of electricity resources, such as utility-owned generation and energy storage, utility-run demand response programs, bilateral contracts, and purchases in wholesale markets to serve their customers. The following sections describe each of the three major categories of services needed by the bulk power system (capacity, energy, and operating reserves) and discuss important differences in their provision by generation as compared with non-generation resources like demand response and energy storage.

2.1 Capacity

Power systems need adequate resources to meet electricity demands with high levels of reliability. Capacity resources are those that can be called upon to deliver energy (or reduce the use of energy), during times of high risk for unserved load (i.e., the inability to balance electricity supply and demand). A common metric for resource adequacy is the planning reserve margin, which is the fraction of available capacity in excess of the expected peak electricity demand. Calculations on the level of planning reserve margin necessary to meet reliability requirements vary by balancing authority area, based on local conditions and regulations (NERC 2013).

In some areas served by an independent system operator (ISO) or regional transmission organization (RTO), the balancing authorities establish capacity markets to facilitate procurement of sufficient capacity, at the right locations, to meet forecasted electricity demand plus the planning reserve margin. Table 2-1 provides some example historical capacity market prices. In regions without capacity markets, capacity transactions occur only bilaterally through individual contracts between load serving entities and capacity providers (Griffith 2008). Further discussion of capacity value is provided in Section 3.6.

Table 2-1. Forward Capacity Market Clearing Prices for ISO New England and PJM Interconnection in Terms of $/kW-year. PJM has locational capacity prices and congested zones tend to have higher prices than the overall PJM system. Both operate one-year forward capacity markets three years ahead of contract delivery (Kirby 2013).

Delivery Year	'07-'08	'08-'09	'09-'10	'10-'11	'11-'12	'12-'13	'13-'14	'14-'15	'15-'16	'16-'17
ISO New England				$54.00	$43.20	$35.40	$35.40	$38.52	$41.16	$37.80
PJM System	$14.88	$43.56	$37.20	$63.60	$42.84	$6.00	$10.08	$48.96	$49.68	$21.72
PJM Most Congested Zone	$72.12	$81.72	$86.64	$67.92	$42.84	$81.12	$90.24	$87.48	$130.32	$79.92

2.2 Energy

The cost of capacity is based primarily on the fixed costs of generation. These cover the carrying costs of the capital investments plus fixed operations and maintenance costs under anticipated conditions. By contrast, the cost of energy is based primarily on the variable costs of operating the power system, which are primarily fuel costs but also include variable operations and maintenance costs. The scheduling of energy supply from power plants is based on these variable costs, with lower cost units given preference over higher cost units to minimize the total operational costs (also called production costs). Example historical ISO/RTO energy market prices are given in Table 2-2 and hourly marginal energy costs in the Public Service of Colorado area are shown in Figure 2-2.

Table 2-2. Selected Real-Time Hourly Energy Prices in Several ISO/RTO Markets from 2002 to 2012 (Potomac Economics 2014)

	Average Real-Time Hourly Energy Prices ($/MWh)									
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
ERCOT	44.3	44.6	72.8	55.2	56.4	77.2	34.0	39.4	53.2	28.3
NYISO (New York City)				71.3	77.2	96.3	47.1	60.5	55.4	41.7
NYISO (West New York)								43.4	40.7	33.2
PJM (load weighted avg.)	41.2	44.3	63.5	53.3	61.7	71.1	39.1	48.4	45.9	35.2
ISO-NE (New England Hub)	48.6	52.1	79.7	59.7	66.7	80.6	42.0	49.6	46.7	36.1

Energy scheduling is complicated by a number of operational constraints, like generator start up times, generator minimum loading points, limited transfer capacity of the transmission system (i.e., congestion), and the need to hold operating reserves (discussed later in Section 2.3). These complications require power system operators to make forecasts of electricity demand (from minutes to months ahead) and then schedule resources in advance of when they are needed. Energy scheduling occurs in discrete time intervals that are as short as 5 minutes in some balancing authority areas and as long as 1 hour in others. When these constraints are binding, like during times of transmission system congestion, higher operational cost generators may need to be scheduled compared to a lower cost alternative without the presence of the constraint (Wood and Wollenberg 2013). Further discussion of energy is provided in Section 3.2.

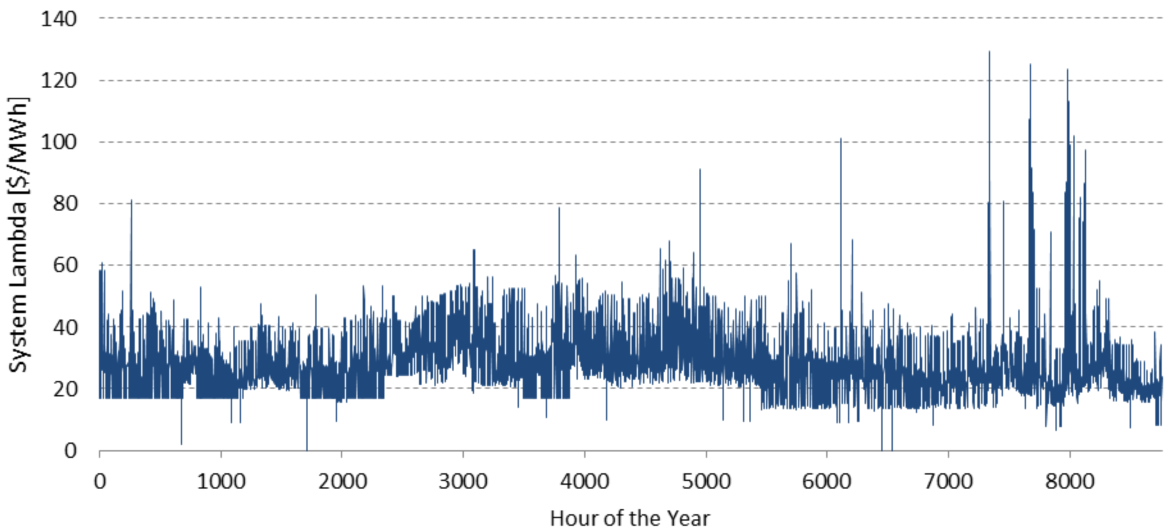

Figure 2-2. Historical 2011 hourly system lambda for the Public Service of Colorado. System lambda is a proxy for hourly energy prices. The annual average is about $28/MWh, ranging between $0/MWh and $130/MWh throughout the year.

2.3 Operating Reserves

Power system operators balance electric load and generation resources primarily through energy scheduling. But below the shortest energy scheduling time interval (varying from 5 minutes to 1 hour), the holding of operating reserves ensures there are enough resources to meet moment-by-moment balancing, meet changes between intervals, and respond to contingencies, like the sudden loss of a large generator or major transmission line. Operating reserves are distinct from energy services; while the costs may involve small amounts of energy, their real value is in the capacity held in reserve and the ability to respond quickly and reliably to maintain balance.

Operating reserves include frequency regulation, load following reserve, and contingency reserve (Hirst and Kirby 1996); recently in the Mid-Continent ISO and California ISO, there is a proposed ramping (also called flexibility) reserve (Navid and Rosenwald 2013; Xu and Tretheway 2012). Frequency regulation responds to random, minute-by-minute variations in aggregate system load that are too fast to be followed by the economic dispatch of energy. Contingency reserves respond to sudden but infrequent supply disruptions and must also be procured separately from energy scheduling. While both frequency regulation and contingency reserves are held for every hour of the year, system operators continually adjust frequency regulation (through the automatic generation control signal) under normal conditions but deploy contingency reserves only when associated events occur. Load following and the proposed ramping reserve service are closely linked. Both ensure there is sufficient operating range and ramping capability available to meet the daily net load curve (load minus the contribution from variable generation like wind and solar power).

Table 2-3. Selected Ancillary Service Tariffs and Requirements in Non-ISO/RTO Balancing Authority Areas (OATI, Inc. 2013). Note that ancillary service tariffs are similar to capacity payments in these regions.

	Tariff ($/kW-year) and Requirement (% of system peak)†			
Balancing authority	Regulation		Spinning Contingency	
AEP West Zone	31.68	1.20%	42.72	2.10%
Arizona Public Service	88.92	1.17%	75.12	3.19%
Duke Energy Carolina Power & Light	47.52	1.20%	47.52	1.77%
El Paso Electric	37.20	0.87%	37.20	1.75%
Florida Power & Light	57.84	1.35%	61.92	0.43%
Idaho Power	78.36	1.50%	78.36	2.86%
PacifiCorp West	93.60	4.24%	105.60	1.75%
Portland Gas & Electric	80.40	1.30%	77.40	3.50%/2.50%*
Public Service of Colorado	80.88	1.50%	82.56	3.50%/2.50%*
Public Service of New Mexico	103.20	1.50%	112.32	3.50%
Southern Company	50.40	1.15%	50.40	2.00%
Tucson Electric	145.20	1.29%	145.08	3.50%

† The requirement is not necessarily that of the aggregate balancing authority, just the requirement imposed on users of the balancing authority area transmission system.

* Some balancing authorities have separate requirements for the load served by thermal and hydropower generators.

Operating reserves are further distinguished as either spinning or non-spinning. For conventional generators, spinning refers to resources that are connected and synchronized with the power system, and non-spinning refers to those that are available and ready to be connected and synchronized within a specified amount of time (usually from 10 to 30 minutes). For non-generation resources providing operating reserves, the fundamental difference lies in response speed rather than the existence of a physical rotating mass, as the term suggests (Kirby 2006). Frequency regulation always comes from spinning resources, but load following and ramping reserve can come from a combination of spinning and non-spinning resources. Contingency reserves include spinning and non-spinning components, but many regions require a minimum percentage to be spinning (e.g., 50% for balancing authorities under the WECC [NERC 2011] and 40% in the Mid-Continent ISO [MISO 2013]).

In non-ISO/RTO balancing authority areas, transmission providers charge their transmission customers for the amount of operating reserves they do not self-supply or procure through third-party suppliers (FERC 2013). Operating reserves are typically quoted on a monthly basis; costs are generally settled according to the transmission customers' contribution to the transmission system peak load (see Table 2-3). Where there are organized wholesale markets, the ISO/RTO balancing authority runs a competitive hourly market for the supply of operating reserves. Historic operating reserve prices are available from market clearing prices posted on the individual system operator websites (see Table 2-4 for an example). In contrast to non-ISO/RTO markets, rates paid for operating reserves vary by hour (or shorter) and display strong daily and seasonal variations (MacDonald et al. 2012). Further discussion of operating reserves is provided in Section 3.2.

Table 2-4. Selected Ancillary Service Prices in Several ISO/RTO Markets from 2002 to 2012. Contingency reserve, as described in the text, is sometimes called spinning and non-spinning reserve, responsive reserve, or 10-minute and 30-minute reserve. (Milligan and Kirby 2010), updated. (Multiply numbers in table by 8.76 to convert to $/kW-year.)

	Average Market Clearing Price $/MW-hour										
Operating Reserve	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
	California ISO										
Regulation (Up + Down)	26.9	35.5	28.7	35.2	38.5	26.1	33.4	12.6	10.6	16.1	10.0
Spinning	4.3	6.4	7.9	9.9	8.4	4.5	6.0	3.9	4.1	7.2	3.3
Non-spinning	1.8	3.6	4.7	3.2	2.5	2.8	1.3	1.4	0.6	1.0	0.9
Replacement	0.90	2.9	2.5	1.9	1.5	2.0	1.4				
	Electric Reliability Council of Texas (ERCOT)										
Regulation (Up + Down)		16.9	22.6	38.6	25.2	21.4	43.1	17.0	18.1	31.3	9.2
Responsive		7.3	8.3	16.6	14.6	12.6	27.2	10.0	9.1	22.9	9.1
Non-Spin		3.2	1.9	6.1	4.2	3.0	4.4	2.3	4.3	11.8	6.7
	New York ISO (east)										
Regulation	18.6	28.3	22.6	39.6	55.7	56.3	59.5	37.2	28.8	11.8	10.4
Spinning	3.0	4.3	2.4	7.6	8.4	6.8	10.1	5.1	6.2	7.4	6.0
Non-spinning	1.5	1.0	0.3	1.5	2.3	2.7	3.1	2.5	2.3	3.9	3.8
30 Minute	1.2	1.0	0.3	0.4	0.6	0.9	1.1	0.5	0.1	0.1	0.3
	Midwest ISO (day-ahead)										
Regulation								12.3	12.2	10.8	7.8
Spinning								4.0	4.0	2.8	2.3
Non-spinning								0.3	1.5	1.2	1.4
	ISO New England										
Regulation + mileage			54.6	30.2	22.7	12.7	13.8	9.3	7.1	7.2	6.7
Spinning					0.3	0.4	1.7	0.7	1.8	1.0	1.7
10-Minute					0.1	0.3	1.2	0.5	1.6	0.4	1.0
30-Minute					0.0	0.1	0.1	0.1	0.4	0.3	1.0

2.4 Comparison of Generation and Non-Generation Resources

Bulk power system services have historically been provided almost exclusively by large centrally operated generators, and as such, bulk power system service definitions and the associated attributes of performance have evolved around their characteristics and capabilities. However, these services can also be provided by non-generation resources like demand response and energy storage.

Demand response encompasses many different strategies by which commercial, residential, municipal, and industrial electricity customers are paid or incentivized to adjust when they use electricity. Similarly, energy storage technologies like pumped storage hydropower, batteries, and flywheels save electricity produced at one point for use at a later time—maybe just moments later or hours to days later. Both demand response and energy storage have historically been utilized by utilities to support system operations. However, evaluation of demand response and energy storage in the provision of a more comprehensive set of services or comparison between these resources and conventional generators in wholesale electricity markets is more complex due in part to several fundamental differences.

The most significant difference between these non-generating resources and conventional generators is the limited response duration of both demand response and energy storage. Generators do not typically have response duration limits (due to the availability of fuel supply) and can continuously provide energy and operating reserves. In contrast, energy storage resources have physical energy limits based on the size of the storage device implemented (rated power and energy capacity) as well as its state of charge. Demand response resources also have duration limits that are driven by a combination of the effective storage within the end-use appliance or equipment systems and the customer willingness to curtail load (Kirby 2006; Rubinstein, Xiaolei, and Watson 2010). These limitations have implications for all three bulk power system services described previously.

For example, the limited response duration presents a scheduling problem that is largely unique to demand response and energy storage. Because of this limitation, both demand response load reductions and energy storage output need to be timed to provide energy services during periods of highest value; similarly, any load recovery or charging need to occur during periods of lowest cost, in order to maximize value. This adds complexity to analyzing demand response and energy storage value. These energy limits also provide challenges to evaluating the ability of demand response and energy storage to provide operating reserves, especially when considering the expected duration of contingency events or actual energy flows that occur when providing frequency regulation.

The challenges associated with response duration are further complicated when evaluating demand response. A key difference between demand response and energy storage is that demand response is inherently tied to end-use loads with associated daily and seasonal electricity consumption patterns (Piette, Kiliccote, and Dudley 2012). This affects when demand response is available to provide capacity, energy, and operating reserves. For example, residential air conditioning is typically in use during hot summer afternoons and can be available to provide response during those times, but air conditioning will have little to no availability for response during colder seasons. Further, demand response availability may be difficult to forecast across different time scales, from operational time frames to multi-year planning horizons.

Another important difference between non-generating sources and conventional generators is the scale of the individual resources and the associated challenge of measuring and verifying response; integration of monitoring and communication technologies for measurement and verification of these small resources can be cost-prohibitive. Effective utilization of individual small demand response resources relies on aggregation to ensure a reliable response. The aggregate response could be predictable and reliable even if that of any individual end-use load is not (Kirby 2006). Alternatively, aggregation can enable a greater range of services that could be provided, such as frequency regulation from a portfolio of end-use loads that are individually incapable of providing such services (Enbala 2011; Callaway 2009).

As discussed later in Section 3.3, the flexibility, duration, and location of demand response resources vary greatly. In many cases, flexible end-use loads capable of providing demand response have effective storage components (such as thermal and cooling inertia) that augment their ability to be dispatched when needed (Rongxin et al. 2010; Goli, McKane, and Olsen 2011). Furthermore, concentrations of industries, facility types, and customer usage patterns vary regionally (Watson et al. 2012). On the other hand, energy storage devices have limitations depending on the specific technology used. Some energy storage technologies, such as pumped storage hydropower and compressed air energy storage, may be constrained geographically, while many others technologies, such as batteries and flywheels, have fewer siting obstacles.

The provision of operating reserves from demand response and energy storage poses additional challenges in wholesale electricity markets. The cost of providing operating reserves in these market regions is calculated by the system operator and based on the opportunity cost that results from withholding energy provision (discussed further in Section 3.1). However, for non-generation resources, these costs are difficult to calculate and, in the case of demand response, challenging to verify for individual entities. Each end-use load is distinct, and the value of interrupting its operation imposes different costs. These costs can vary over tie and include some factors that are ultimately subjective to the customer being served by that load. Energy storage devices capable of both energy and operating reserves can have opportunity costs similar to generators. However, unlike generators whose opportunity costs can be assessed coincident with their provision of operating reserves, opportunity costs for energy storage resources are inherently inter-temporal; their opportunity costs will depend on the energy prices when charging and discharging occurs.

Finally, the operational costs of demand response and energy storage can be quite different from conventional generators. The main operational cost for energy storage is charging energy, plus parasitic loads and round-trip efficiency losses. For demand response, the main operational cost is lost opportunity of use or value of lost service (Woolf et al. 2013). In some instances, it may not matter to the customer when an appliance or piece of equipment is used, and the demand response strategy would fall within customer tolerances, resulting in a low opportunity cost. At other times, the demand response request may be disruptive to the customer, with opportunity costs increasing with response duration (Callaway and Hiskens 2011).

3 Modeling and Results

Large Western Interconnection-wide renewable integration studies have evaluated many of the challenges associated with deploying large amounts of variable wind and solar generation technologies. Integration studies use high-resolution time series analysis covering a year or more of system data with unit commitment and economic dispatch modeling (EnerNex 2011; GE Energy 2010) to evaluate operational impacts associated with increasing the supply of variable generation. These studies also examine benefits of improved wind and solar resource forecasting and trade-offs between institutional changes (like increasing balancing area cooperation) and technical changes (like installing new flexible generation).

There has been continuous advancement in the tools, analytic approaches, and data sets: improved meso-scale modeling for time series weather data synchronized with load data (Potter, et al. 2008), realistic short-term resources forecasts, and improved treatment of conventional generator ramping and cycling costs (Lew et al. 2012; IEA 2013). Demand response and energy storage resources present important sources of bulk power system services and can aid in the integration of variable generation into the grid; however, integration analyses have not yet fully incorporated these resources explicitly into grid simulation models.

The modeling approach taken in this study is based on the Western Wind and Solar Integration Study Phase 2 (WWSIS-2), incorporating associated recommendations from its technical review committee (Lew et al. 2013). The approach taken did not attempt to:

- Assess all the potential mechanisms and sources of demand response
- Assess all the possible value streams of demand response and energy storage
- Consider the costs associated with the deployment and integration of demand response or energy storage
- Consider energy storage technologies deployed in distribution systems or behind the meter
- Simulate contingency events and consider the impacts of changing system dynamics
- Determine the optimal sizing or location of demand response or energy storage.

Simulations over the full 2020 study year are conducted in the production cost model PLEXOS (Energy Exemplar 2013) in order to assess the value of demand response and energy storage in system operations. Capital costs associated with the deployment of demand response and energy storage were not considered in the determination of value. Production cost models are utility planning tools that model unit commitment and economic dispatch processes. They are commonly used in renewable integration analyses because they can mimic many of the decisions faced by power system operators. Further, production cost models output a number of useful metrics such as estimates of different types of operational costs and power plant emissions. As typical with production cost models, the model can calculate the cost of holding operating reserves, but frequency regulation dispatch and contingency events are not simulated. Simulation of shorter term grid dynamics is an important area of future research, requiring models capable of simulating detailed generator and non-generator resource response (Ela and O'Malley 2012; Varadan et al. 2012).

As in the WWSIS-2 study, we assume that grid assets are committed and dispatched according to a least-cost optimization. In reality, there are numerous (typically proprietary) contractual agreements (for both power and fuel), as well as system operator behaviors and market behaviors, that result in differences between modeled and actual outcomes. We also assume larger reserve sharing groups, reflecting potential improvements to inter-balancing authority area cooperation in the 2020 study year; within a reserve sharing group, operating reserves may come from any member balancing authority area (see the map in Figure 3-1). Lastly, we model the transmission system zonally (using linearized DC power flow) with transfer limits between balancing authority areas, greatly simplifying network constraints. As a result, the modeling results do not capture localized congestion that could be relieved from the deployment of small-scale energy storage or enablement of demand response, which could provide an additional potential source of value (i.e., deferral of transmission and distribution asset investments). Additional details on transmission system assumptions can be found in the WWSIS-2.

The results are also based solely on the day-ahead unit commitment and economic dispatch. This approach takes into account wind and solar forecast errors between the day-ahead and real-time dispatches by holding additional operating reserves (frequency regulation and ramping reserve) to meet anticipated increases in variability and uncertainty. However, the real-time dispatch is not simulated. The production cost model simulations begin with the use of two scheduling models to determine outage scheduling and allocate certain limited energy resources (primarily hydropower). The production cost model then performs chronological unit commitment and economic dispatch modeling for the day-ahead market run. The optimization time window for the day-ahead unit commitment is 48 hours, rolling forward in 24-hour increments. The extra 24 hours in the unit commitment horizon (for a full 48-hour window) are necessary to properly commit generators with high startup costs and to properly dispatch resources with energy constraints.

3.1 Western Interconnection and Colorado System Models

To make quantitative estimates of the value of demand response and energy storage, we simulate power system operations on two systems, a model of the Western Interconnection (see Figure 3-1) and a smaller Colorado system. Our base assumptions are derived from the Transmission Expansion Planning Policy Committee (TEPPC) 2020 Common Case (WECC 2011) with modifications from the WWSIS-2. The generation mix and electricity consumption for the modeled systems are provided Figure 3-2 and in Table 3-1, respectively. In both systems, the high renewable base case has greater quantities of wind and solar deployed compared to the low renewable base case, but each case has the same amount of conventional resources. This assumption implies greater system reliability in the high renewable case due to excess capacity and higher total system capital costs (which are not considered in this study).

Figure 3-1. Study area including 36 balancing authorities (small print) and 12 reserve sharing group (large print) assumptions for the 2020 study year

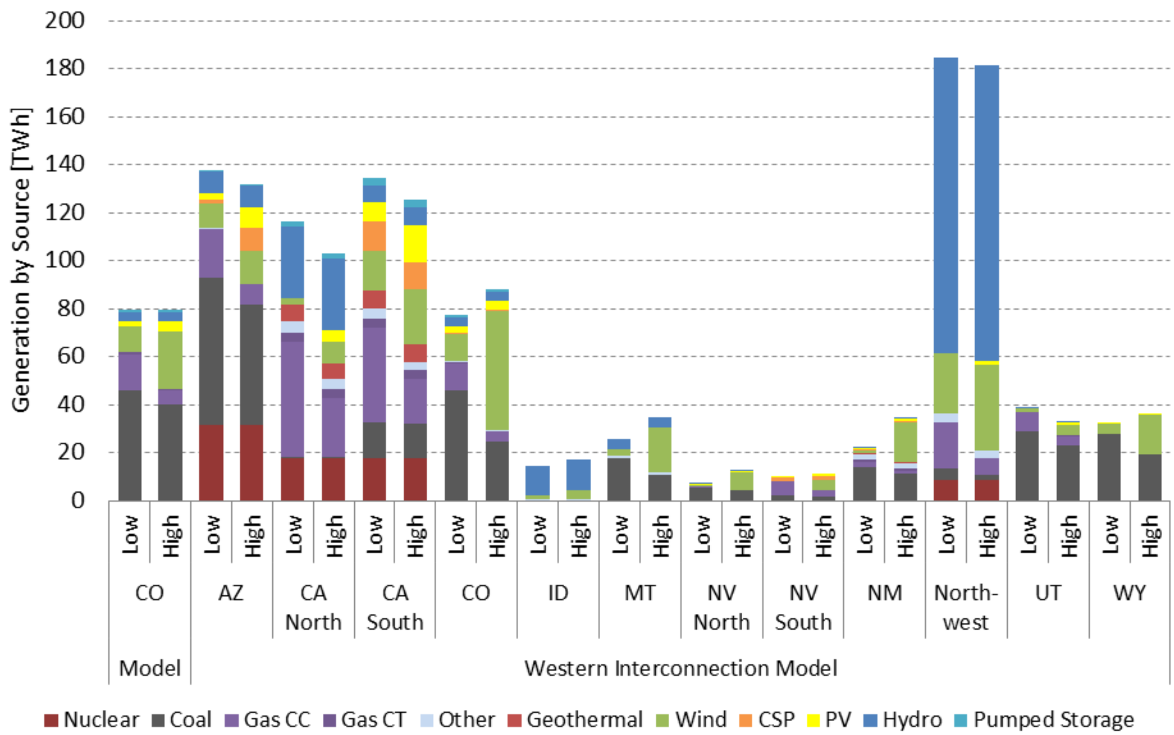

Figure 3-2. Electricity generation by source in the low and high renewable base cases of the Colorado and Western Interconnection models

Table 3-1. Generation Resource Mix in the Low and High Renewable Base Cases for the Two Model Systems, a Colorado System Model and a Western Interconnection Model

	Colorado Test System Model		Western Interconnection Model	
Generator Type	Low Renewable	High Renewable	Low Renewable	High Renewable
Coal Steam	6,178 MW	6,178 MW	30.26 GW	30.26 GW
Combined Cycle	3,724 MW	3,724 MW	56.65 GW	56.65 GW
Gas Turbine/Gas Steam	4,045 MW	4,045 MW	23.00 GW	23.00 GW
Hydropower	777 MW	777 MW	52.30 GW	52.03 GW
Pumped Storage	560 MW	560 MW	6.30 GW	6.30 GW
Onshore Wind	3,347 MW	7,216 MW	27.61 GW	65.86 GW
Solar Photovoltaic	986 MW	2,301 MW	7.07 GW	20.27 GW
Concentrating Solar Power	0 MW	0 MW	4.79 GW	7.19 GW
Other†	513 MW	513 MW	7.41 GW	7.41 GW
Total	20,130 MW	25,314 MW	215.39 GW	268.97 GW

† Other generation includes oil- and gas-fired internal combustion engines.

The operating reserve requirements utilized for each reserve sharing group in both test systems are based on statistical analysis of load, wind, and solar variability within each of the member balancing authority areas (Ibanez et al. 2012). The resulting annual average requirements for frequency regulation, contingency reserves, and ramping reserves for each reserve sharing group are provided in Table 3-2. The amount of frequency regulation and ramping reserves required in the high renewable case is generally larger than the low renewable case due to increased variability and uncertainty. However, contingency reserve requirements are based on the size of the single largest contingency (i.e., N-1 reliability criteria) which should not change between the two cases. In the Western Interconnection system model, the requirement for each of the reserve sharing groups is assumed to be 6% of load, with at least 3% from spinning resources. The non-spinning portion of this requirement was not modeled due to historically low opportunity costs and the abundance of non-spinning resources in the model. Additionally, we augment the generator characteristics in the TEPPC dataset to reflect part-load heat rates and startup costs prepared by Intertek/APTECH for the WWSIS-2 (Kumar et al. 2012) to more accurately reflect the additional costs associated with providing operating reserves.

Table 3-2. Average Reserve Requirements by Reserve Sharing Group as a Percentage of Average Electric Load in Each Region

	Regulation Reserve		Contingency Reserve		Ramping Reserve	
Reserve sharing group	Low RE	High RE	Low RE	High RE	Low RE	High RE
Colorado System	1.3 %	1.8 %	4.5 %	4.5 %	0.6 %	1.2 %
Western Interconnection						
Arizona	1.1 %	1.1 %	3.0 %	3.0 %	0.6 %	0.7 %
California North	1.0 %	1.0 %	3.0 %	3.0 %	0.2 %	0.3 %
California South	1.0 %	1.0 %	3.0 %	3.0 %	0.5 %	0.6 %
Colorado	1.1 %	1.8 %	3.0 %	3.0 %	0.7 %	2.2 %
Idaho	1.1 %	1.3 %	3.0 %	3.0 %	0.5 %	1.0 %
Montana	1.6 %	5.6 %	3.0 %	3.0 %	1.5 %	7.6 %
Nevada North	1.2 %	2.6 %	3.0 %	3.0 %	0.5 %	3.1 %
Nevada South	1.0 %	1.4 %	3.0 %	3.0 %	0.1 %	1.0 %
New Mexico	1.2 %	2.7 %	3.0 %	3.0 %	0.6 %	3.4 %
Northwest	1.1 %	1.1 %	3.0 %	3.0 %	0.7 %	0.8 %
Utah	1.0 %	1.2 %	3.0 %	3.0 %	0.3 %	0.8 %
Wyoming	1.6 %	3.3 %	3.0 %	3.0 %	1.5 %	4.3 %

The Colorado system model is derived from the characteristics of two western balancing authority areas, Public Service of Colorado and Western Administration Colorado Missouri. While most assumptions are inherited from the larger Western Interconnection system model, the Colorado system model is developed by turning off generation and load outside of these two balancing authority areas. In the parent model, the Colorado system meets some of its electric load using generation resources outside its geographic footprint. To make up this difference in the isolated system, natural gas combined cycle units are added into the modeled region to meet a 15% planning reserve margin. This smaller system enables numerous controlled experiments to be modeled that would be prohibitive in the larger system due to computational time (Denholm et al. 2013).

A potentially important source of value for both demand response and energy storage is the provision of frequency regulation and avoiding costs associated with non-steady state operation of thermal generators. Our study builds on the TEPPC and WWSIS-2 datasets to better understand this value through refined assumptions. First, there are added costs associated with non-steady state operation. Plants providing frequency regulation incur additional wear and tear and heat rate degradation (PJM 2012). However, actual performance data related to an individual generator's ability to provide operating reserves are not widely available. We assume a regulation cost adder based on observations of ISO/RTO market generator bid data (Denholm et al. 2013) given in Table 3-3.

Table 3-3. Additional Wear and Tear and Heat Rate Degradation Costs for Providing Frequency Regulation

Generator type	Cost ($/MW-h)
Supercritical coal	15
Subcritical coal	10
Combined cycle	6
Gas/Oil steam	4
Hydropower	2
Pumped storage hydropower	2

While parameters like ramp rates and operating limits determine a generator's ability to provide energy and operating reserves, only a subset of generators have the necessary equipment to follow the automatic generation control signal for the provision of frequency regulation. We assume that combustion turbines are not instrumented to provide frequency regulation (WWSIS-2 Technical Review Committee 2012) and that only 60% of the ramp capability of the remaining generators can provide frequency regulation (CAISO 2011). We also make the assumption that nuclear, geothermal, and fixed schedule hydropower plants are not able to provide any operating reserves. Which specific generators, in actuality, are capable of providing frequency regulation is proprietary information. In practice, there may be fewer generators that provide this service than in our assumptions, which would result in higher value for frequency regulation. Sensitivity analyses on the cost of operating reserves based on generator availability is provided in Hummon et al. (2013).

3.2 Cost Savings and Market Value

Calculating the operational value of demand response and energy storage can be approached two ways. First, we can compare the total costs of operating the system (i.e., production cost) between two scenarios. By making incremental changes to the system, changes to production costs can be attributed to the addition or removal of grid assets or to changes in system operation. Differences in production costs represent the total operational value to the modeled system, irrespective of how that value is distributed to different entities. Second, we can utilize the calculated short-run marginal costs of production generated in PLEXOS to determine an implied market value in each balancing authority area for each service during each hour. In ISO/RTO regions, marginal costs equate to market clearing prices for individual services and indicate the expected revenue that resources would earn as market participants (Hogan 1998). In non-ISO/RTO regions, marginal costs are analogous to a *system lambda*, and relate to a vertically integrated utility's avoided cost for providing the associated service (Booth and Rose 1995).

Production costs are all costs aside from capital and fixed operations and maintenance (fixed costs), which ultimately reduce to the sum of fuel costs and variable operations and maintenance costs (variable costs). It is illustrative to assign these basic cost components to different generator operational modes such as startup-shutdown, steady state operation, and non-steady state operation. Startup costs are those associated with bringing an offline resource online. For thermal generators, these costs may depend on when the unit was last operated. Steady state costs come from holding a single operating point; non-steady state costs come from modulating around that operating point, as in the case of providing frequency regulation. Figure 3-3 shows the production cost results from the base case (i.e., no additional demand response or energy storage) for the low and high renewable cases of the Western Interconnection model. As shown in Figure 3-3, steady state costs constitute the majority of total production costs (left), but startup-shutdown and non-steady state operation are significant components of the production costs associated with operating reserves (right).

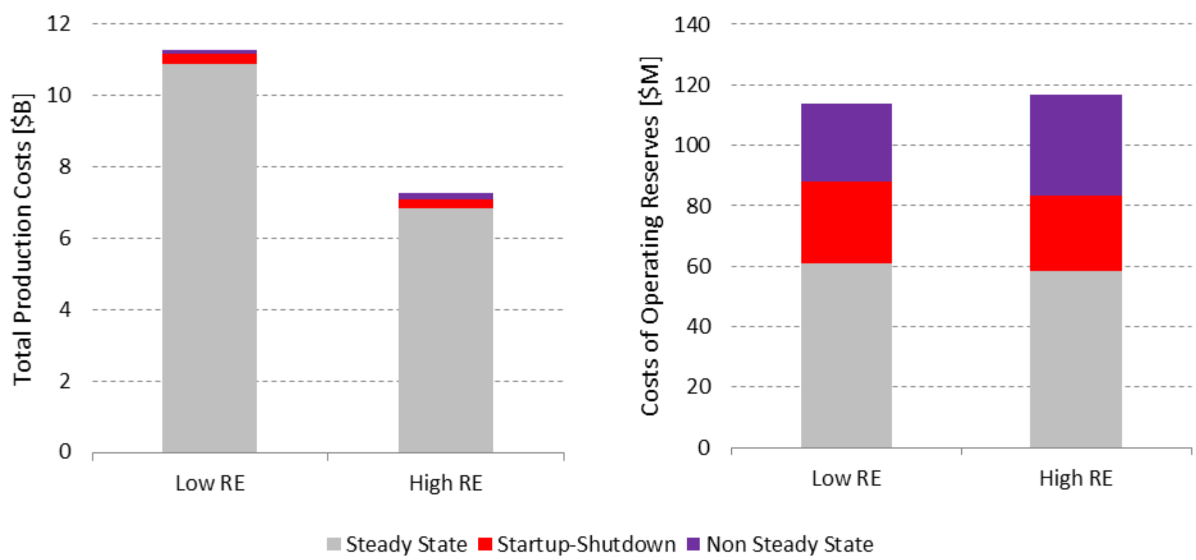

Figure 3-3. Production costs in the Western Interconnection model disaggregated by different operational modes. On the left are total costs for combined energy and operating reserves, and on the right are costs associated with just operating reserves. The costs of operating reserves is calculated by taking the difference in production costs between the base cases and cases where the operating reserves requirements are set to zero.

Short-run marginal costs of production are the change in total production costs resulting from a small change in electric load or operating reserves requirements. For instance, the marginal cost of energy is the incremental change in total production costs if the demand for electricity decreases (or increases) by one unit (i.e., the cost to serve the last unit of energy). These costs would be equivalent to the market clearing prices in an ISO/RTO market if generators bid only their true variable costs (i.e., do not engage in strategic bidding). Of note, startup-shutdown costs that are reflected in the total costs of production are not represented in short-run marginal costs.

For operating reserves, the short-run marginal costs of production are composed of two components: a lost opportunity cost and any additional operational costs. Power system markets and operations calculate lost opportunity costs for generators that provide operating reserves. The lost opportunity cost is the difference between the system marginal cost for energy and the generator's marginal cost for energy; in other words, forgone profit. Generators may also incur additional operational costs (e.g., additional fuel and operations and maintenance) for providing operating reserves, as discussed in Section 3.1 with our assumed values in Table 3-3. Thus, the sum of the lost opportunity cost and any additional operational costs for the generator serving the last unit of operating reserves sets the system marginal cost for operating reserves (Isemonger 2009). The distribution of marginal costs within each region for the various bulk power system services over the study year are provided in Figure 3-5 for the base cases in the Western Interconnection system model.

Figure 3-4 illustrates what the implied market size would be in the Western Interconnection if resources are paid at the marginal costs of production. In other words, the market size is equal to the total demand for energy and operating reserves for each region during each hour multiplied by the marginal cost for each service in that hour for that region and summed over the entire study year. It should be emphasized that this implicitly assumes that the marginal costs generated by a production cost model are equal to the prices in a market setting and does not consider the potentially significant impact of generator bidding strategies and other factors that could cause market prices to deviate from a generator's actual variable costs.

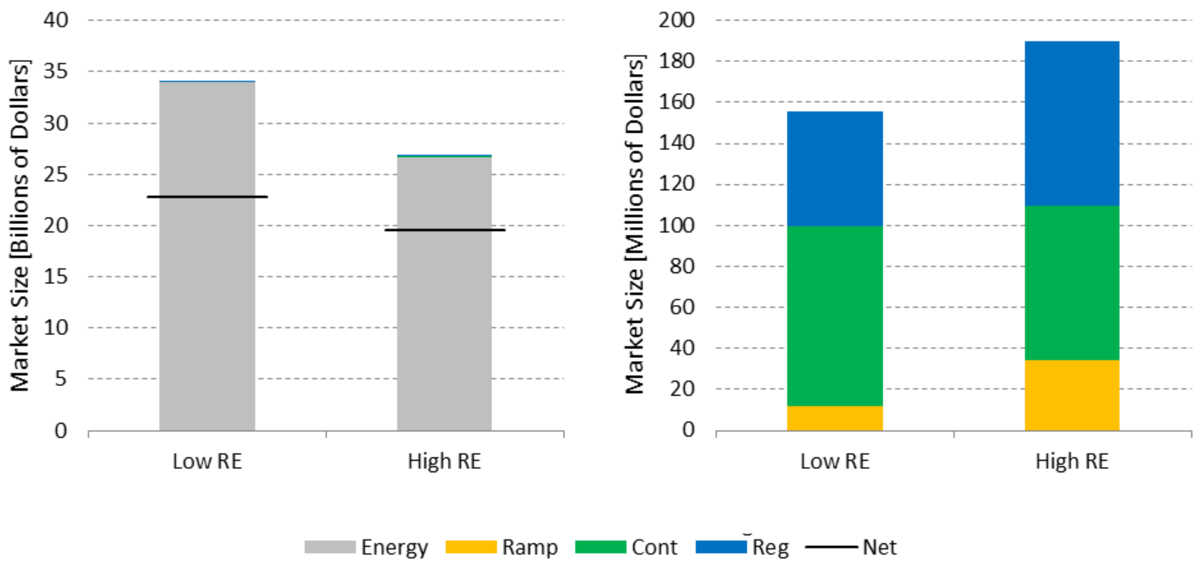

Figure 3-4. Implied market size for the Western Interconnection model if resources are paid at the marginal cost of production for energy and operating reserves. On the left is the total market size for combined energy and operating reserves and on the right is the market size associated with just operating reserves. The black lines (Net) show the net market size after subtracting out the production costs (shown in Figure 3-3).

Finally, power system market and operations (which production cost models attempt to mimic) are built around a number of hard and soft constraints. Hard constraints are those that systems operations must obey, whereas soft constraints are those that system operations can violate if meeting the constraint is more expensive than a penalty price. Often, penalty prices are set sufficiently high such that soft constraints are rare. In ISO/RTO markets, these penalty prices are set higher than what market participants are allowed to bid. However, the specific implementations of markets and operations differ among balancing authorities. Even in the case of hard constraints, the modeling software may not always find a solution within the allowable tolerance. In these cases, the various constraints may be relaxed sequentially, in order of presumed system criticality, until a solution is reached. These events, often referred to as scarcity (Hogan 2012), have a number of effects on market prices, which are difficult to quantify without introducing assumptions that can bias the final results.

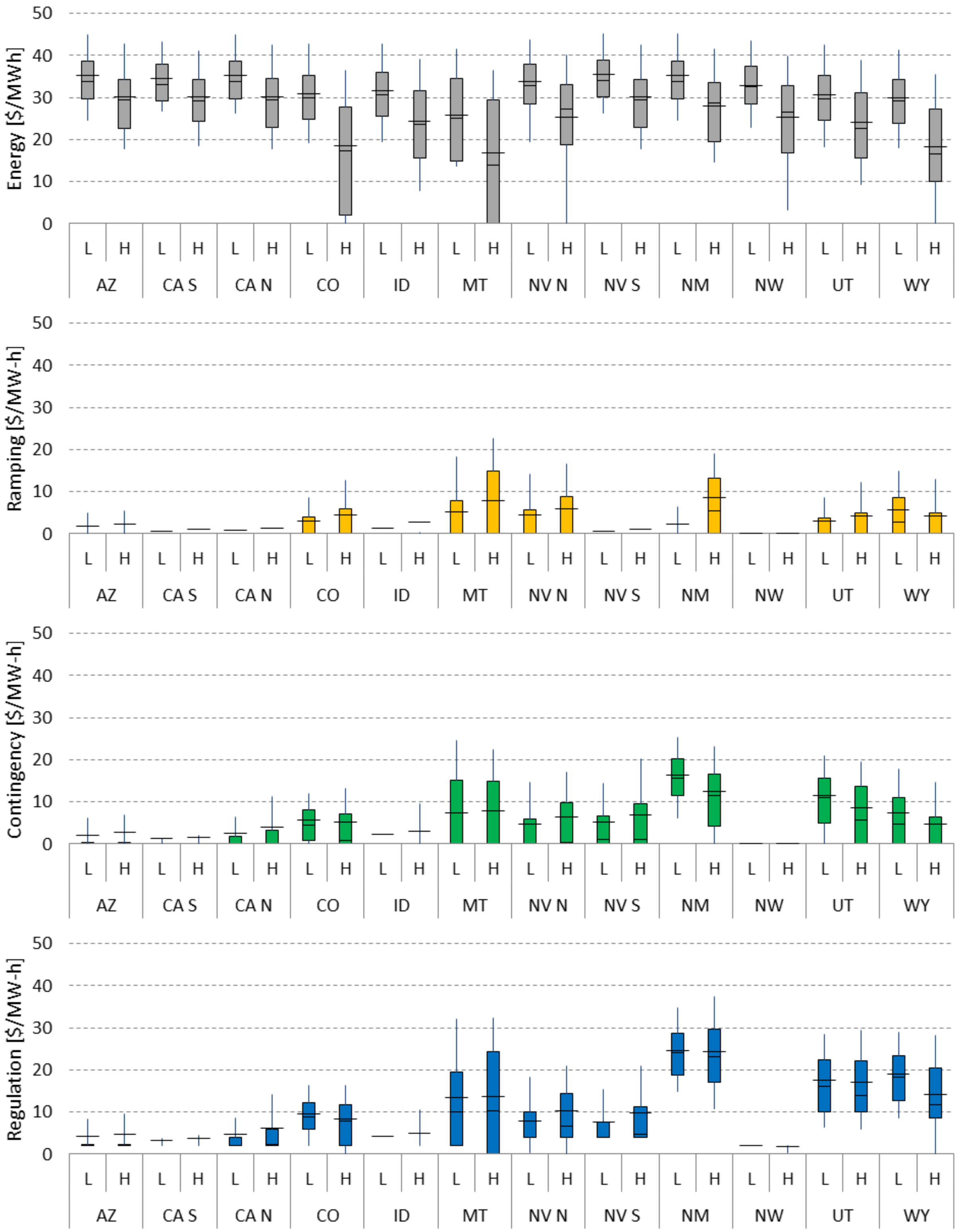

Figure 3-5. Marginal cost of energy, ramping reserve, contingency reserves, and frequency regulation in the Western Interconnection model in the base cases for both the low (L) and high (H) renewable cases. Vertical lines show the 10% and 90% values in the distribution, the boxes provide the quartile values (i.e., 25%, median, and 75%), and the horizontal lines show the average values.

3.3 Demand Response

There are two broad categories of demand response implementation, distinguished by whether or not the demand response resources are explicitly dispatched by the power system operator (FERC 2010). Under some control strategies, retail electricity customers respond to retail pricing tariffs. For instance, customers could change their consumption based on time-varying electricity prices (Zhou and Botterud 2013). In other strategies, demand response resources are treated similarly to generators or energy storage (Hummon et al. 2013) and the load response is characterized by operational constraints including maximum response duration, ramp rates, and timing of load recovery (when the energy utilization lost during a load shed increases load at an earlier or later time). While there are advantages and disadvantages to different approaches, we utilize the second approach to explore the use of demand response for the provision of a broader range of grid services. However, the underlying mechanisms by which retail electricity customers are enrolled in demand response programs to provide these services is outside the scope of this study. A number of alternative programs and implementation strategies are discussed in Chuang (2009).

3.3.1 End-Use Loads

To estimate the potential operational value of demand response, we need to estimate the availability of demand response resources and implement that estimate in production cost models. We first begin by generating regional profiles for different loads that may be able to shift energy use and provide operating reserves. We start with projected hourly profiles generated by TEPPC for each balancing authority area in the year 2020 and approximate the relative contributions from appliances and equipment systems in different economic sectors. In the resource assessment, we include end-uses across commercial buildings, residential buildings, municipal functions, industrial non-manufacturing, and industrial manufacturing. While we discuss all five of these sectors in this report for completeness, the data from the industrial manufacturing resource assessment was not completed in time for the production cost modeling. These end-use loads, while important demand response resources, are not analyzed for their potential operational value in our simulations. End-uses included in the resource assessment are given in Table 3-4, and the associated electricity demands are given in Figure 3-7.

Table 3-4. End-Uses Included in the Demand Response Resource Assessment across Different Economic Sectors

Commercial Buildings	Residential Buildings	Municipal Functions	Industrial Facilities
Indoor lighting	Space cooling	Freshwater distribution	Agricultural water pumping
Space cooling	Space heating	Road and garage lighting	Cold storage in refrigerated warehouses
Space heating	Water heating	Wastewater pumping	Data center servers and equipment cooling
Ventilation			Manufacturing†

†Representing 28 manufacturing processes. Further breakdown given in Table 3-6.

For end-use loads that are sensitive to weather, we normalize the resource data sets to the year 2006 using a linear regression on historical load, temperature, and rainfall data. As the most recent wind and solar resource data is from the year 2006, this normalization captures some of the correlations between electricity demand and wind and solar generation (Lew et al. 2013).

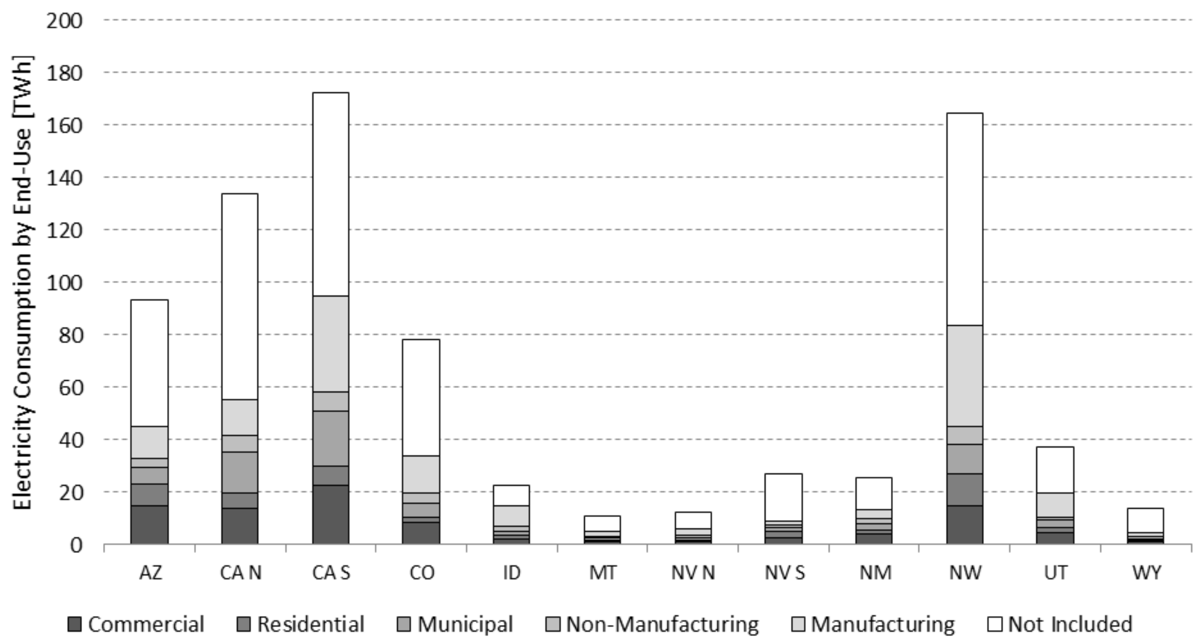

Figure 3-7. Electricity consumption by various end-use types included in the demand response resource assessment. The white bar signifies the remaining electricity consumption by end-uses not included in the assessment.

There are a number of additional resources that are not included in this assessment. First, projecting the changing composition of end-use loads is outside the scope of this project (e.g., electric vehicles could be a significant draw on future electricity production). Second, there are many smaller loads that could be aggregated (e.g., miscellaneous building plug loads) for demand response, but their individual characteristics vary significantly, complicating their analyses. Generally, existing data sources for these loads lack the necessary fidelity to determine their potential contribution, preventing their inclusion in this study. Lastly, demand response strategies can be coupled with on-site generation to provide services jointly. These may include back-up generators, combined heat and power systems, and microgrids. While distributed generation may play a role in providing bulk power system services, this area of investigation is outside the scope of this study.

The availability of different types of end-use appliance and equipment systems (potentially used for demand response) varies significantly across the Western Interconnection. Each region has different drivers for electricity demand depending on factors such as population growth, local climate, consumer behavior, and commercial and industrial economic activity. For instance, Figure 3-8 shows the locations of manufacturing facilities for Textile Mill Products and Rubber and Miscellaneous Plastics Products and associated clusters of economic development. Quantification of the demand response resource starts with understanding the regional distribution of different end-use appliances and equipment systems and their patterns of use. A flexibility analysis (discussed in the next section) is then applied to electricity consumption data of each end-use type to estimate response capabilities for different bulk power system services (Olsen et al. 2013; Starke, Alkadi, and Ma 2013).

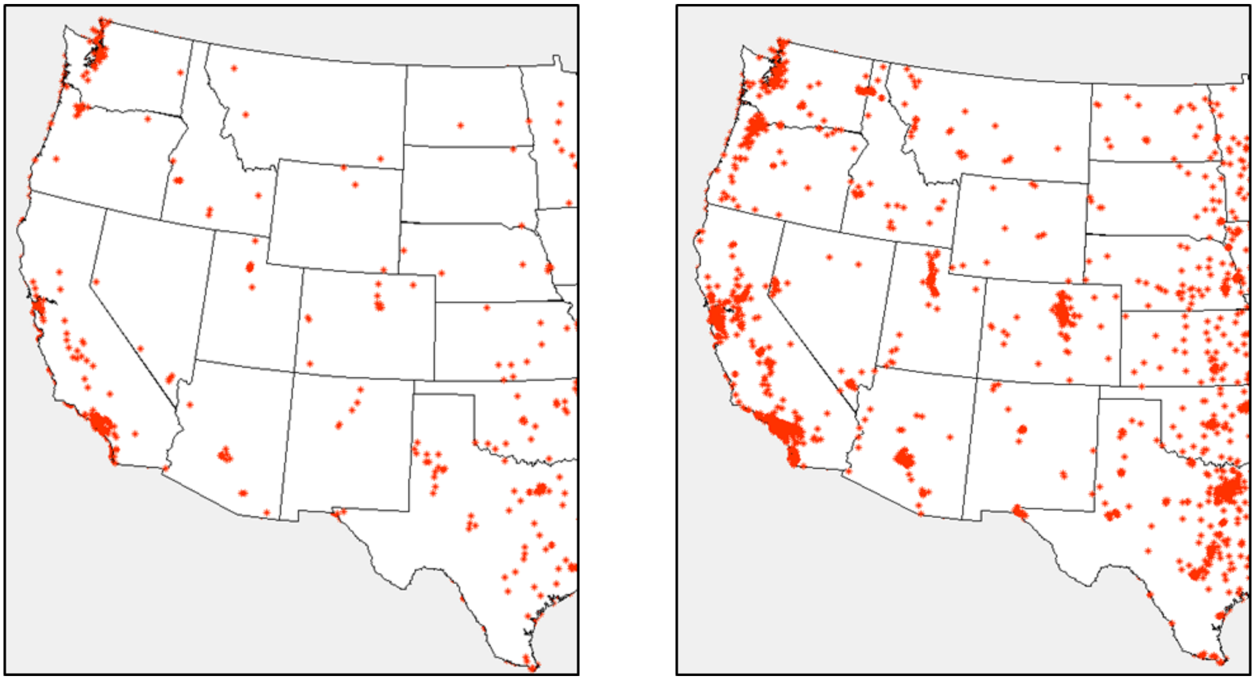

Figure 3-8. Red dots shows plants locations in the western United States affiliated with the Textile Mill Products manufacturing subsector (left) and the Rubber and Miscellaneous Plastics Products manufacturing subsector (right).

3.3.2 Demand Response Services

The analysis described in the previous section establishes the load profiles of end-use appliances and equipment systems assumed to be capable of participating in demand response. However, only a fraction of the load may be available for shedding or shifting to provide bulk power system services. Common definitions for these services do not exist across all regions; but for this study, we make some generalizations based on physical requirements (see Table 3-5). Combined with flexibility analyses of the different end-uses, these generalized product definitions determine the assumed eligibility of potential demand response resources for different services (see Tables 3-6 and 3-7). This methodology also helps determine associated strategies for the demand response resources to participate in power system operations (Ma et al. 2013).

The production cost model treats energy and operating reserves differently. Demand response energy provision involves the shifting of energy use from one energy scheduling interval to another. This is accomplished explicitly in the model, including constraints like the load shed duration (i.e., the maximum number of hours of a shed) and timing of load recovery. Depending on the demand response strategy, load recovery may occur prior to the load shed or may occur following the load shed. These two strategies, using the analogy with energy storage, correspond with pre-charging and recharging the effective storage capability utilized for demand response. During load recovery, some fraction of the avoided energy use will be consumed. The timing and profile of load recovery has important implications for value to the power system as well as to the retail electricity customers providing the demand response service. For instance, thermal loads must maintain temperature within tolerances, limiting the time difference between a load shed and a load recovery and limiting the potential benefits of arbitraging between times of different energy costs (Hummon et al. 2013).

Table 3-5. Generalized Bulk Power System Product Definitions Used to Assess the Capabilities of Different Types of Demand Response Resources. Actual requirements implemented in different regions may differ from these assumptions.

Bulk Power System Product	General Definition	Physical Requirements: How fast to respond	Length of Response	Time to respond fully	How often called
Frequency Regulation	Response to random unscheduled deviations in scheduled net load	30 seconds	Energy neutral in 15 minutes	5 minutes	Continuous within bid period
Spinning Contingency Reserve	Rapid and immediate response to a loss in supply	5 minutes	≤ 30 minutes	≤ 10 minutes	Less than once per day
Ramping (Flexibility) Reserve	Load following reserve for large un-forecasted wind and solar ramps beyond that for daily load	5 minutes	1Hour	20 minutes	Continuous within bid period
Energy	Shift energy utilization from one time period to another	5 minutes	≥ 1Hour	10 minutes	1-2 times per day, 4-8 hour notification

Table 3-6. Assumed Eligibilities for Residential and Commercial Buildings, Municipal Functions, and Industrial Non-Manufacturing End-Uses to Provide Different Types of Bulk Power System Services

Resources	Bulk Power System Services: Frequency Regulation	Ramping Reserve	Contingency Reserve	Energy
Commercial Buildings				
Cooling	X	X	X	X
Heating				X
Lighting		X	X	
Ventilation	X	X	X	
Residential Buildings				
Cooling	X	X	X	X
Heating	X	X	X	X
Water heating	X	X	X	X
Municipal Functions				
Outdoor lighting	X	X	X	
Freshwater pumping				X
Wastewater pumping				X
Industrial Non-Manufacturing				
Data centers			X	X
Agricultural pumping			X	X
Refrigerated warehouses				X

For operating reserves, the constraints on the demand response resource are implicit in the eligibility for providing services. Production cost models do not simulate the actual utilization of operating reserves. They merely reserve sufficient ramp capability and operating range from grid resources to ensure that operating reserve requirements could be met in the resulting unit commitment and economic dispatch. Operating reserves are mostly compensated by the amount of capacity offered in each time period, rather than the amount of energy delivered (or curtailed in the case of demand response) when that capacity is dispatched. However, the frequency of use in operating reserves can have significant impacts on the utilization of demand response. For instance, spinning contingency reserve may require resources that can respond almost instantly, ramp to full response in less than 10 minutes, be called as frequently as daily, and must be capable of sustaining a curtailment for up to 30 minutes. Only demand response resources able to do so can qualify as a spinning contingency reserve resource in our assessment. The actual frequency of use (also referred to as dispatch-to-contract ratio) varies significantly by operating reserve product.

Table 3-7. Assumed Eligibilities for Industrial Manufacturing Processes to Provide Bulk Power System Services Sorted by Standard Industrial Classification (SIC)

		Bulk Power System Services			
Industry SIC	Dominant Process	Frequency Regulation	Ramping Reserve	Contingency Reserve	Energy
Food and Kindred Products					
20	Packaging				X
20	Chiller	X	X	X	X
Textile Mill Products					
22	Wrapping				X
22	Weaving				X
Apparel, Finished Products from Fabrics and Similar					
23	Wrapping				X
23	Weaving				X
Lumber and Wood Products, Except Furniture					
24	Sawing				X
24	Planning				X
Furniture and Fixtures					
25	Sawing				X
25	Planning				X
Paper and Allied Products					
26	Chipper				X
26	Dewatering Press	X	X	X	X
Printing, Publishing and Allied Industries					
27	Chipper				X
27	Dewatering Press	X	X	X	X
Chemicals and Allied Products					
28	Electrolysis	X	X	X	X
28	Compressor	X	X	X	X
28	Grinding				X
Petroleum Refining and Related Industries					
29	Catalytic Cracking	X	X	X	X
Rubber and Miscellaneous Plastic Products					
30	Mixing	X	X	X	X
30	Mill				X
Leather and Leather Products					
31	Mixing	X	X	X	X
31	Mill				X
Stone, Clay, Glass, and Concrete Products					
32	Electric Furnace	X	X	X	X
32	Crushing				X
Primary Metal Industries					
33	Electrolysis	X	X	X	X
33	Crushing and Classifying				X
Transportation Equipment					
37	Metal Cutting				X
37	Final Assembly	X	X	X	X

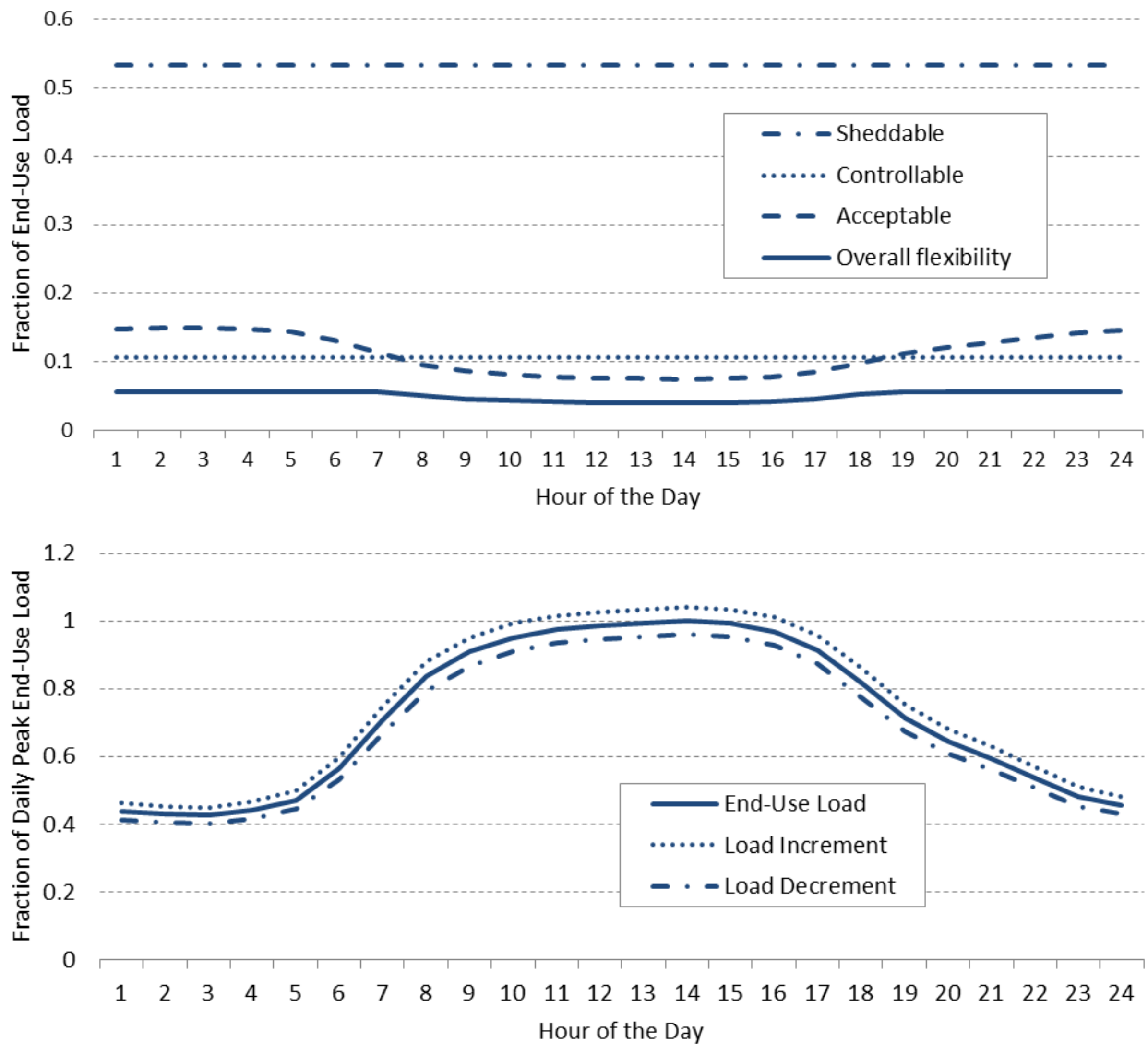

Figure 3-9. Example flexibility calculation for ramping reserve provided by commercial building ventilation. The combination of sheddable, controllable, and acceptable filters determine the overall flexibility (top) and when multiplied by the electric load, yield a maximum available response (bottom). The overall flexibility is only a small fraction (i.e., 4–6%) of the overall electric end-use load.

Based on assumed product definitions and dispatch-to-contract ratios, the ability for end-use loads to provide demand response can be represented by three filters that, in combination, estimate the subset and fraction of load that is available to be responsive. These filters include sheddability, controllability, and acceptability. Sheddability relates to physical constraints of the end-use equipment and is equal to the percentage of the load which could be shifted or shed by a demand response strategy. Controllability refers to the percentage of load that has the controls in place necessary to achieve this shift or shed. Acceptability relates to end-user attributes like building occupant comfort and employee work schedules. This last filter is particularly difficult to assess because it refers to the percentage of end-use load that reflects the end-users willingness to accept an actual or perceived reduction in the level of service to participate as a power system resource. In Figure 3-9, an example application of these three filters for commercial lighting load is given. The top of Figure 3-9 shows the three filters and their values over a 24 hour period. The bottom of Figure 3-9 shows the application of these filters to an end-use load profile. The resulting fraction of end-use load that is able to be utilized for demand response is the difference between the solid and dotted lines. Additional discussion of these filters is given in the text box on pages 34-35 (Olsen et al. 2013).

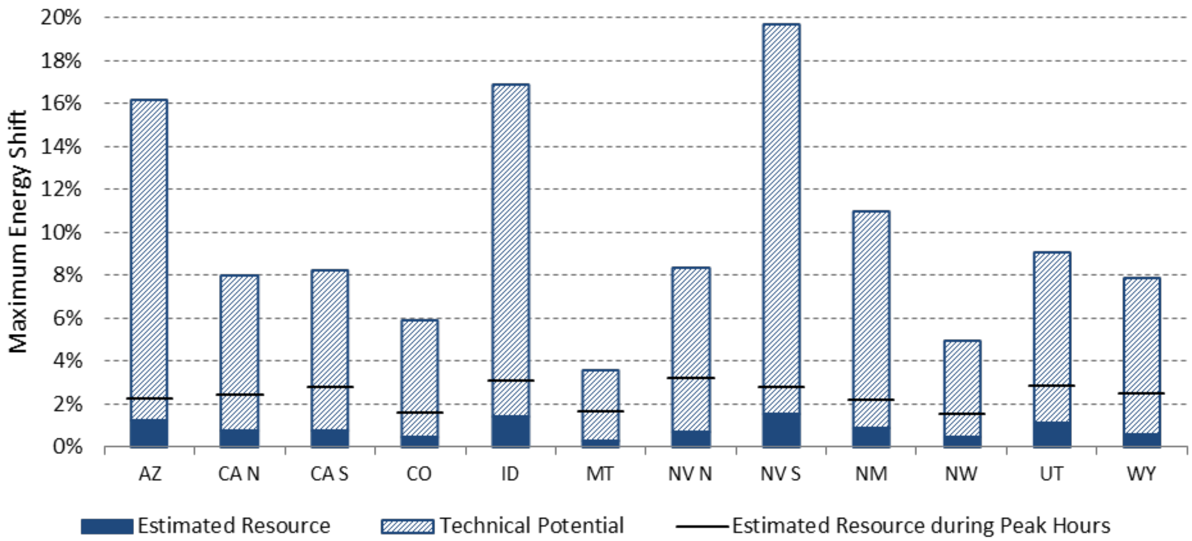

Figure 3-10. Average max daily energy shifting potential in regions of the Western Interconnection for the estimated resource (from partial participation) and the technical potential (with full participation). Horizontal lines indicate the average energy shifting potential during the top 100 load hours.

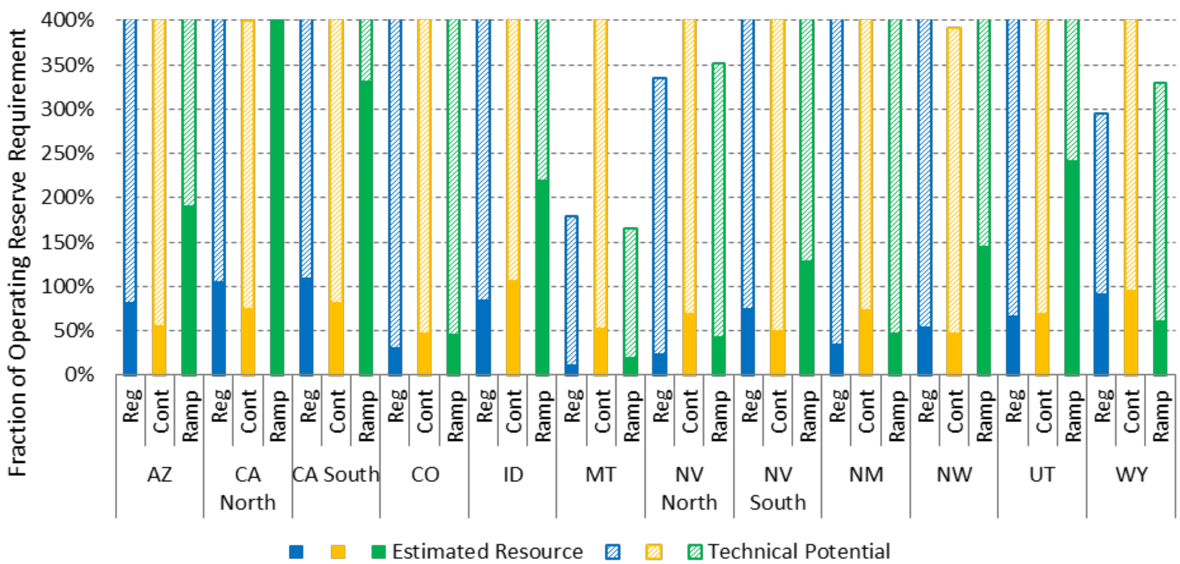

Figure 3-11. Estimated resource (from partial participation) and technical potential (with full participation) for demand response to provide operating reserves in regions of the Western Interconnection. The vertical axis is limited to 400% for clarity, though in many cases the technical potentials are higher. Percentages are based on model assumptions and not actual operating reserve requirements.

The flexibility determination of different types of loads is derived from a number of existing studies, engineering-based analyses of the appliance and equipment systems, historical participation data in demand response programs, and survey data. Our 2020 scenario for the available demand response resource is based on expectations of communications and control systems in place due to factors outside of demand response programs (e.g., the “Internet-of-things”), as well as extrapolations of existing demand response program participation rates to new programs. The technical potential of demand response in this study assumes full participation (i.e., application of only the sheddability filter to the end-use load profiles). Figures 3-10 and 3-11 summarize the estimated resource (from partial participation) and technical potential (with full participation) for demand response providing energy and operating reserves.

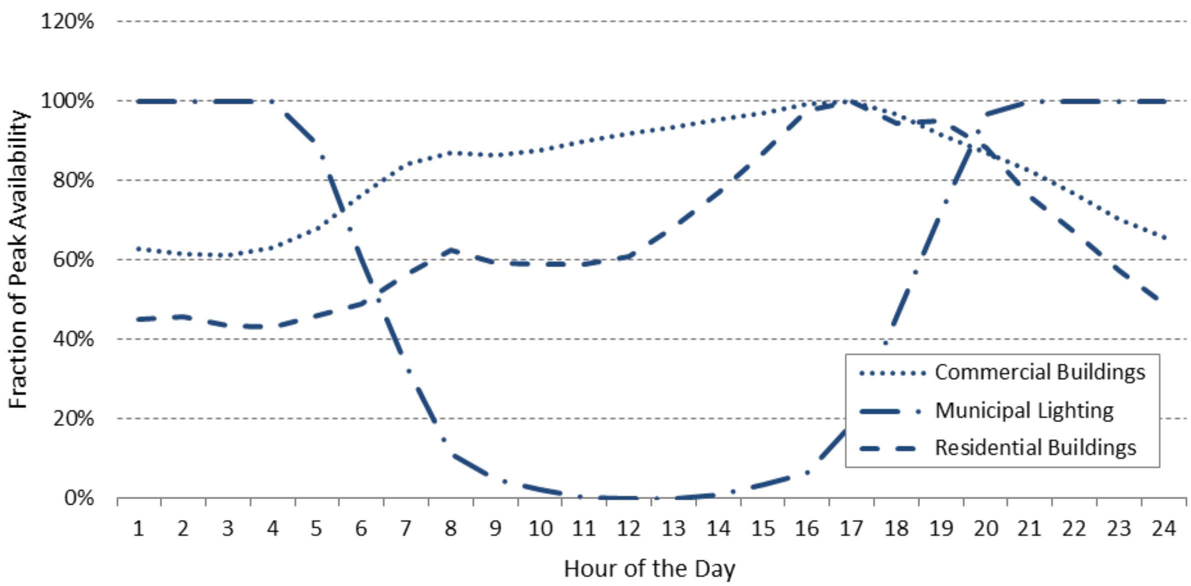

Figure 3-12. Average hourly availabilities for ramping reserves from different types of demand response resources in the Colorado system model compared with the average system load

The estimates for demand response resources, from a bottom-up calculation methodology, are based on characterization of appliance and equipment systems that comprise of just under half of the total electricity use in U.S. regions of the Western Interconnection (49% of TEPPC 2020 forecast). The remaining 51% of electricity use represents other end-uses (such as miscellaneous building plug loads) that are difficult to characterize and are excluded from the present work. The estimated demand response resource in 2020, based on assumptions using an extrapolation from current rates of retail customer participation (i.e., 6% of total electricity use), indicates that the resources could provide about one-third of the operating reserve requirement in a scenario with 33% wind and solar generation in the western United States, and shift 1.0% of the average daily energy use in the 2020 study year. The technical potential, which represents full participation (i.e., 49% of total electricity use), is about 10 times larger and could shift about 9% of the average daily energy use. Note that in the power system simulations, the demand response resources from industrial manufacturing are not included. The demand response resource in the simulated scenarios represents 4% of total electricity use and can shift up to 0.8% of average daily energy use. Each of the different demand response resources has different hourly, daily, weekly, and seasonal availabilities. For instance, the availability of commercial and residential heating tends to peak between 7 a.m. and 9 a.m., but availability of municipal lighting tends to peak between 8 p.m. and 4 a.m. The availabilities for agricultural pumping, data centers, municipal pumping, and wastewater pumping remain fairly flat over the course of a day. For a seasonal perspective, agricultural pumping, commercial cooling, residential cooling, and refrigerated warehouses have peak availabilities in the summer; municipal lighting, commercial heating, and residential heating have peak availabilities in the winter; and data centers, residential water heating, commercial lighting, commercial ventilation, municipal pumping, and wastewater pumping loads have fairly constant availabilities over the year, with regular weekday-weekend usage patterns. The average hourly availabilities for ramping reserves from a few estimated demand response resources in the Colorado system model are shown in Figure 3-12, and corresponding monthly averages for operating reserves are shown in Figure 3-13.

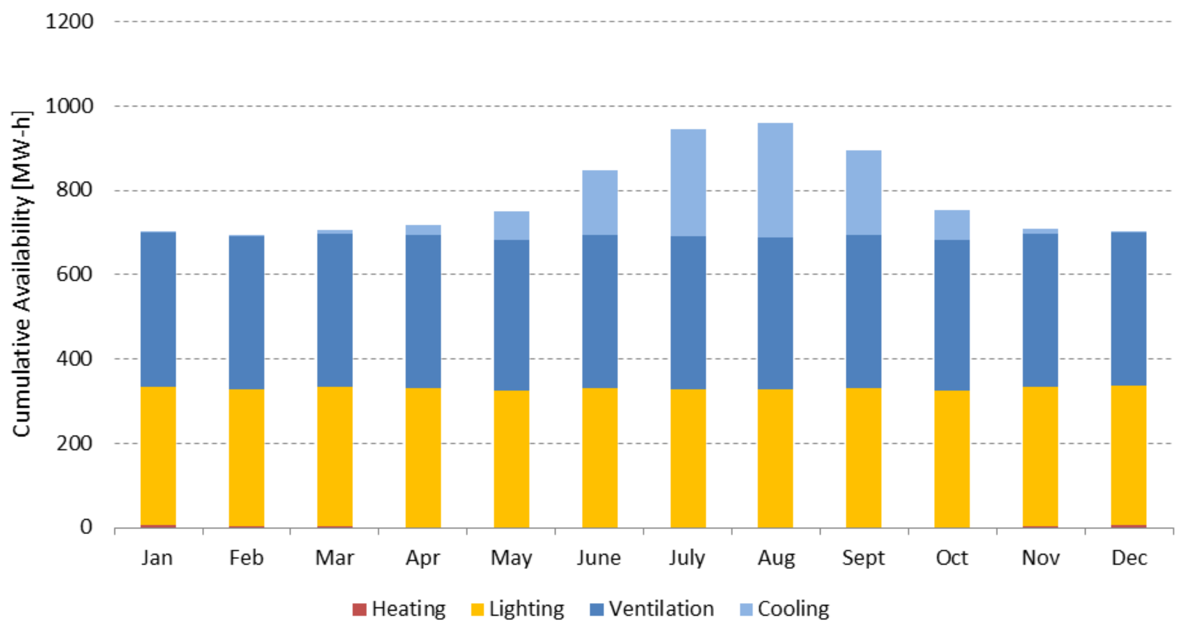

Figure 3-13. Cumulative monthly availabilities for ramping reserves from different types of commercial building demand response resources in the Colorado system model

3.3.3 Demand Response Operational Value

This subsection examines the operational value of demand response in the low renewable and high renewable cases of the Western Interconnection model. In the low renewable case, 14% of electricity production comes from wind (≈10%) and solar (≈4%) power; in the high renewable case, 33% of electricity production comes from wind (≈25%) and solar (≈8%) power. The results presented reflect the implementation of demand response resources estimated from commercial, residential, municipal, and industrial non-manufacturing end-uses in the production cost model (industrial manufacturing end-uses are not modeled). End-uses included in the modeling constitute 30% of total electricity use, whereas 49% of total electricity use was assessed. After accounting for participation assumptions in demand response programs, only 4% of total electricity use is available to support power system operations.

Ultimately, an annual cumulative availability of 11.3 TW-h for demand response (1.4% of total electricity use) was used in the modeled scenarios. This subsection looks at general findings from our simulations and Section 3.5 looks in more detail at the drivers of value under increasing levels of wind and solar power.

The value of demand response can be examined from the perspective of savings in production costs or from the perspective of implied market value, as discussed at the beginning of Section 3.1. Production cost savings refer to differences in fuel and operations and maintenance costs between scenarios with and without added demand response resources. Implied market value is associated with the sum of payments (and charges) to providers (and consumers) for energy and operating reserves in every hour and across all regions. Payments are based on calculated marginal costs in the modeling framework and we equate these marginal costs to market clearing prices.

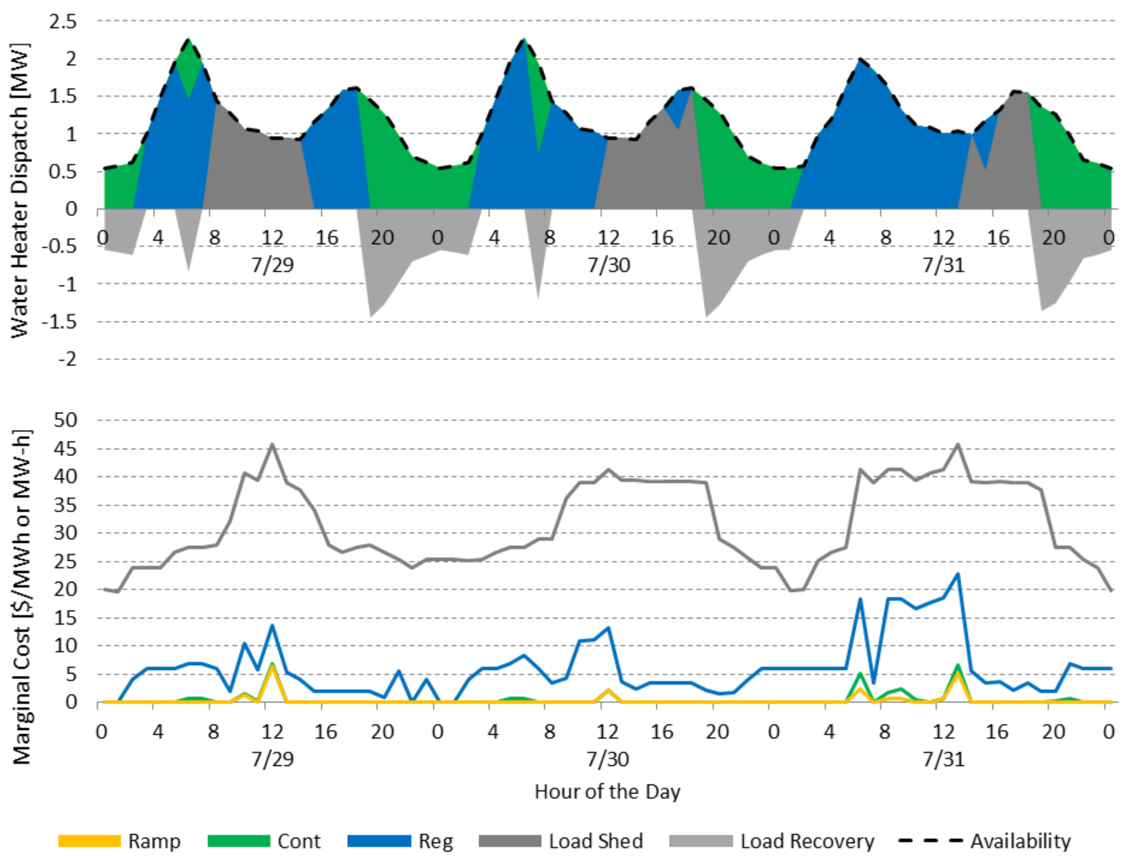

Figure 3-14. Example dispatch of water heating demand response resources in the Colorado model system (top) and associated hourly marginal costs for various bulk power system services (bottom)

In the modeling, we treat demand response resources similarly to energy storage deployed on the bulk power system. The model co-optimizes demand response to provide both energy and operating reserves alongside conventional generation. An example dispatch for water heating demand response resources and the associated hourly marginal costs for different bulk power system services is given in Figure 3-14. As indicated by the dashed line in the top of Figure 3-14, demand response resources that can provide more than one service are constrained such that the total capacity provided in each hour cannot exceed the single largest individual availability for energy, frequency regulation, contingency reserve, or ramping reserve. The bottom of Figure 3-14 gives the hourly marginal costs for the different grid services generated by the production cost model that are used in the market value calculations. As with energy storage, energy transactions with demand response include both a load shed (similar to discharging) and a load recovery (similar to charging). When calculating market value, demand response resources are assumed to be paid the marginal cost of energy when shedding load and are assumed to pay the marginal cost of energy when recovering that load. In this study, we assume that energy shifting with demand response is net-energy neutral. In other words, there are no losses (or gains) from shifting energy from one time period to another. In actuality, end-use loads may incur an efficiency penalty such that the energy used for load recovery is greater than the amount used for load shedding, similar to the effect of round trip efficiency for energy storage. Other end-use loads may experience the opposite, i.e., a conservation effect. There is insufficient information at this time to make either assessment comprehensively across our resource assessment.

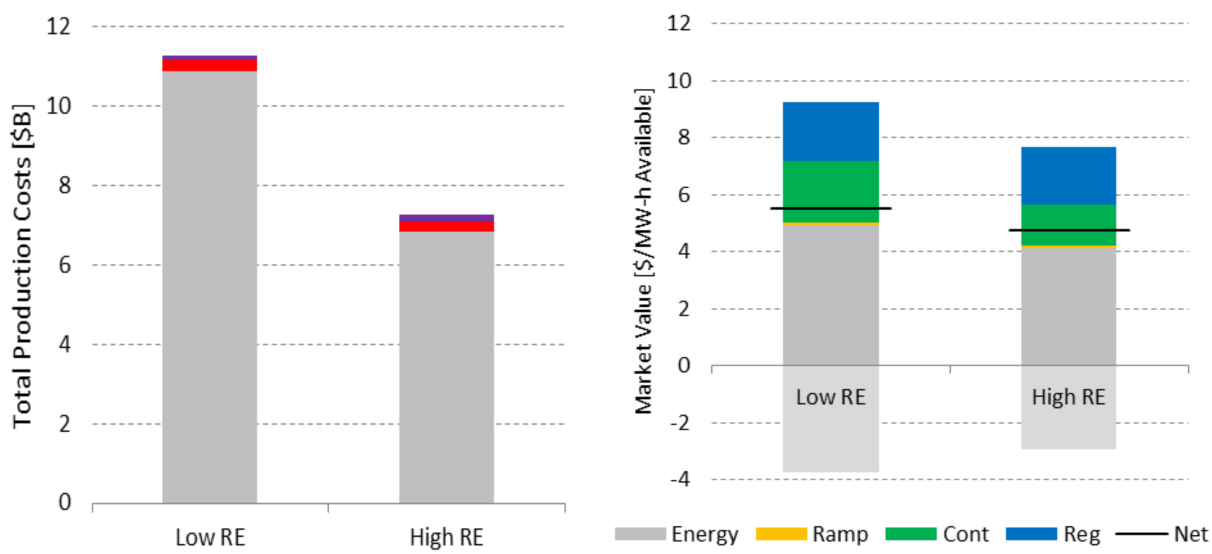

Figure 3-15. Operational value of demand response resources in terms of production cost savings (left) and implied market value if resources are paid at the marginal costs of production (right) in the low and high renewable cases of the Western Interconnection model.

Figure 3-15 compares the production cost savings associated with demand response resources (left) for the low renewable case (14%) and the high renewable case (33%), as well as the implied market value if demand response resources participated in a market environment (right). The operational values shown are normalized by the maximum cumulative availability from demand response resources that can be used for bulk power system (see text box on pages 34-35 for more detail on defining a metric to measure demand response value). The total reduction in production cost is $117 million in the low renewable case and $107 million in the high renewable case, with a maximum cumulative availability from demand response of 11.3 TW-h, leading to savings of $10.4 per MW-h of availability in the low renewable case and $9.5 per MW-h in the high renewable case. Furthermore, this figure disaggregates cost savings into several components. Savings in steady state costs represent the increased use of generators with lower variable costs (i.e., fuel) and improved heat rates from generators operating closer to their peak efficiencies. Savings in startup-shutdown costs represent the reduced need to start and stop generators to meet system requirements, thus avoiding the associated startup fuel costs, increases in operations and maintenance, and any additional wear and tear. Lastly, savings in non-steady state costs represent decreased need for generators to modulate output when providing frequency regulation. The assumed unit costs for non-steady state operations are provided in Table 3-3.

On the right of Figure 3-15, the implied market value is shown with and without the need to purchase energy (at the system marginal cost) for load recovery. The net value, if the purchase cost is subtracted out, is represented by a black horizontal line. The cost of load recovery is a substantial fraction, about one third, of the gross market value for the modeled demand response resources. While energy shifting constitutes the majority of gross market value, the provision of operating reserves constitutes the majority of net market value. Total production costs are dominated by steady-state operational costs (see left side of Figure 3-3); however, the operational savings achieved through the use of demand response (see left side of Figure 3-15) comes from a high fraction of avoided generator startups and shutdowns as well as avoided non-steady state operational costs. One consequence (which will be shown more explicitly in the discussion of energy storage value in Section 3.4) is that the value of demand response (or energy storage) declines rapidly with the level of deployment. This results from the small overall opportunity to avoid these specific high operational cost items and the relatively infrequent spikes in the costs for energy.

In addition to examining the total operational value of demand response resources, the study also examines the operational value of resources by region and by type (see Figures 3-16 through 3-19). Implied market values based on marginal costs are used to disaggregate the value of the discrete services; production costs are not distinct on a regional basis due to significant inter-balancing authority area transactions that occur in the modeling. In general, the regional differences in values for demand response resources mostly follow the regional patterns of operating reserve marginal costs. For instance, examining Figure 3-5, the average marginal cost of frequency regulation is six times higher in New Mexico compared with Arizona leading to similarly large differences in implied market value. The modeled operational values by region are a good indicator of how the availabilities of each type of demand response resource correlate with times of high system costs. Those with higher correlations tend to have higher values and those with lower correlations tend to have lower values.

A number of other observations can be drawn from the data presented in Figures 3-16 through 3-19. Across regions, cooling loads consistently have low net values for energy transactions (gray bars in figures). Both commercial and residential cooling end-uses have restrictions regarding the timing of load shedding and load recovery (i.e., limited flexibility) due to the assumed limits on thermal inertia of buildings and residences. While cooling loads and their associated demand response availabilities are correlated with times of high energy costs, they are limited in their ability to take advantage of low cost off-peak energy for load recovery. For these end-use loads, the value of load shedding is nearly canceled out by the costs of load recovery. Even though the energy arbitrage value of cooling load-based demand response is low, the production cost model still utilizes these resources in order to reduce generator startups. From a market perspective, these demand response resources may earn limited revenues, but they can still help lower overall production costs.

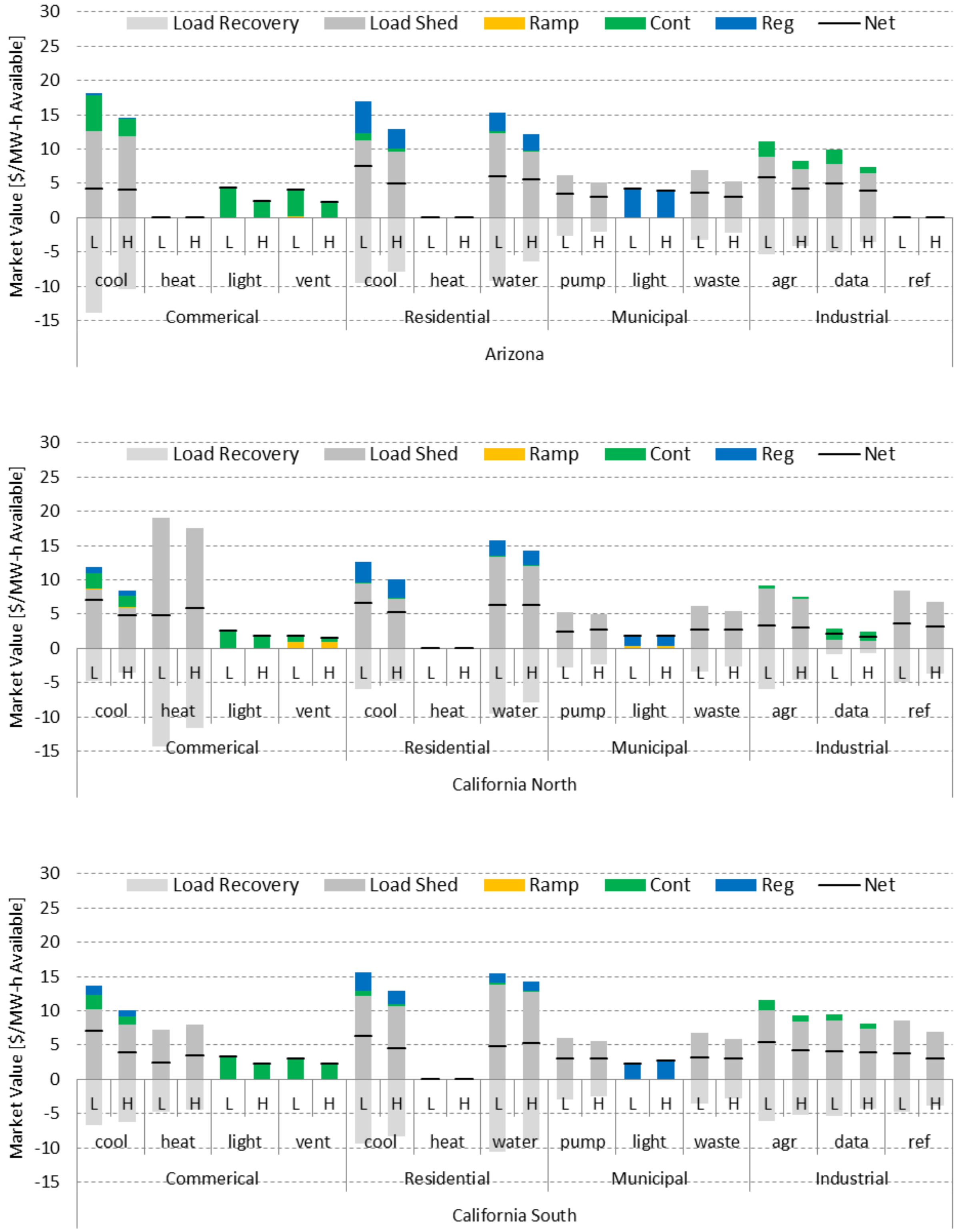

Figure 3-16. Operational value (implied market value) based on marginal costs of production for different types of demand response resources in the Western Interconnection model at low and high levels of wind and solar generation

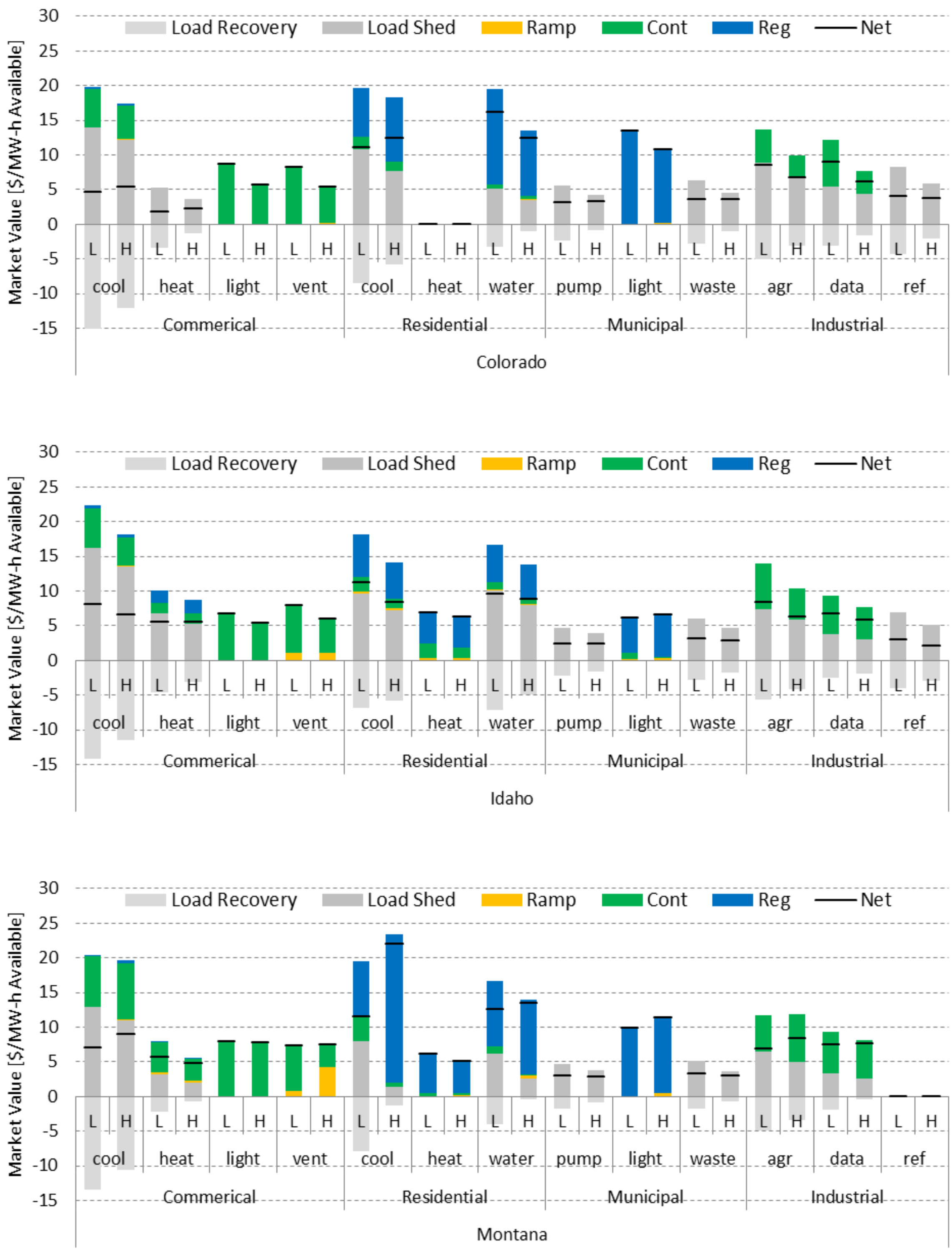

Figure 3-17. Operational value (implied market value) based on marginal costs of production for different types of demand response resources in the Western Interconnection model, at low and high levels of wind and solar generation

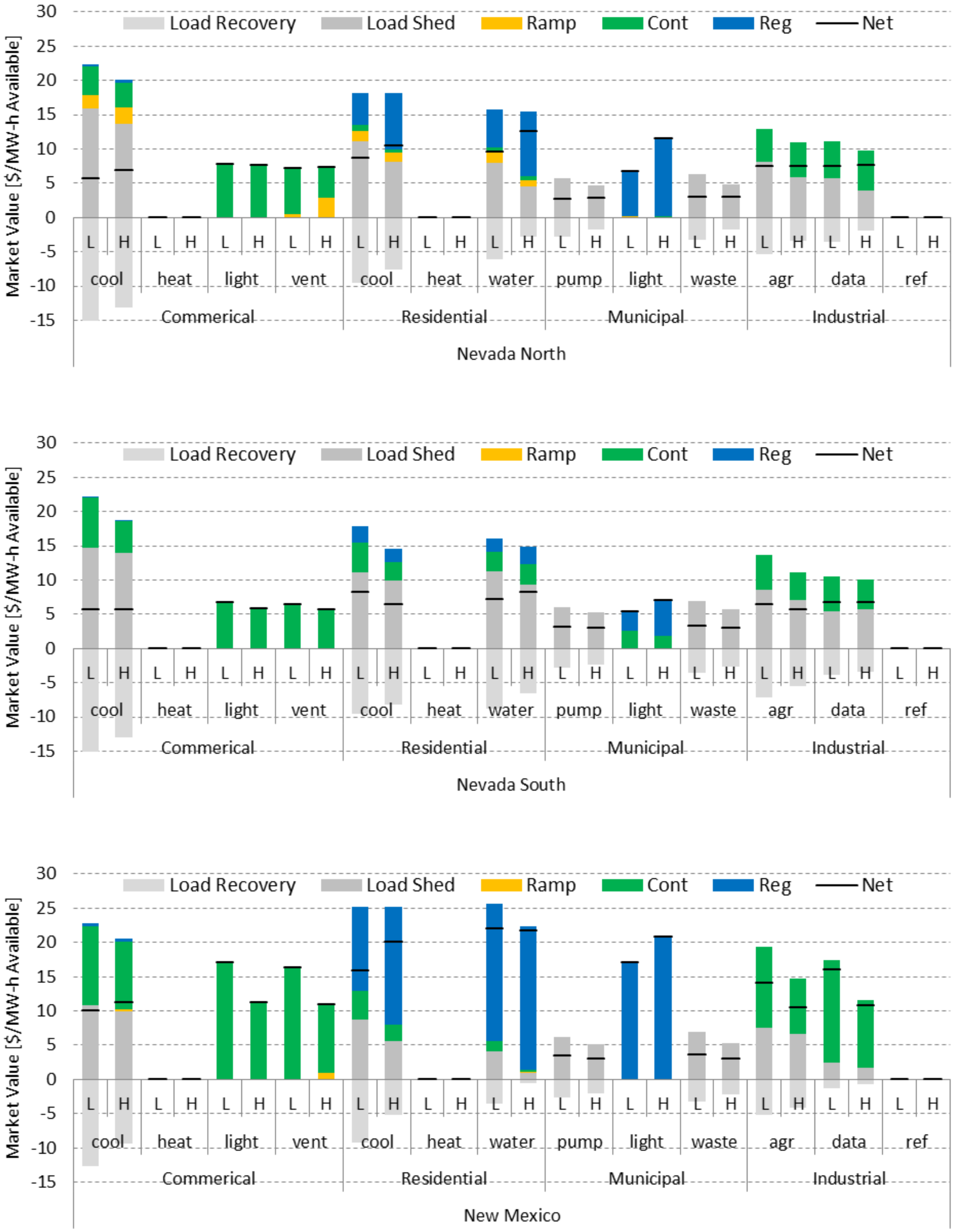

Figure 3-18. Operational value (implied market value) based on marginal costs of production for different types of demand response resources in the Western Interconnection model, at low and high levels of wind and solar generation

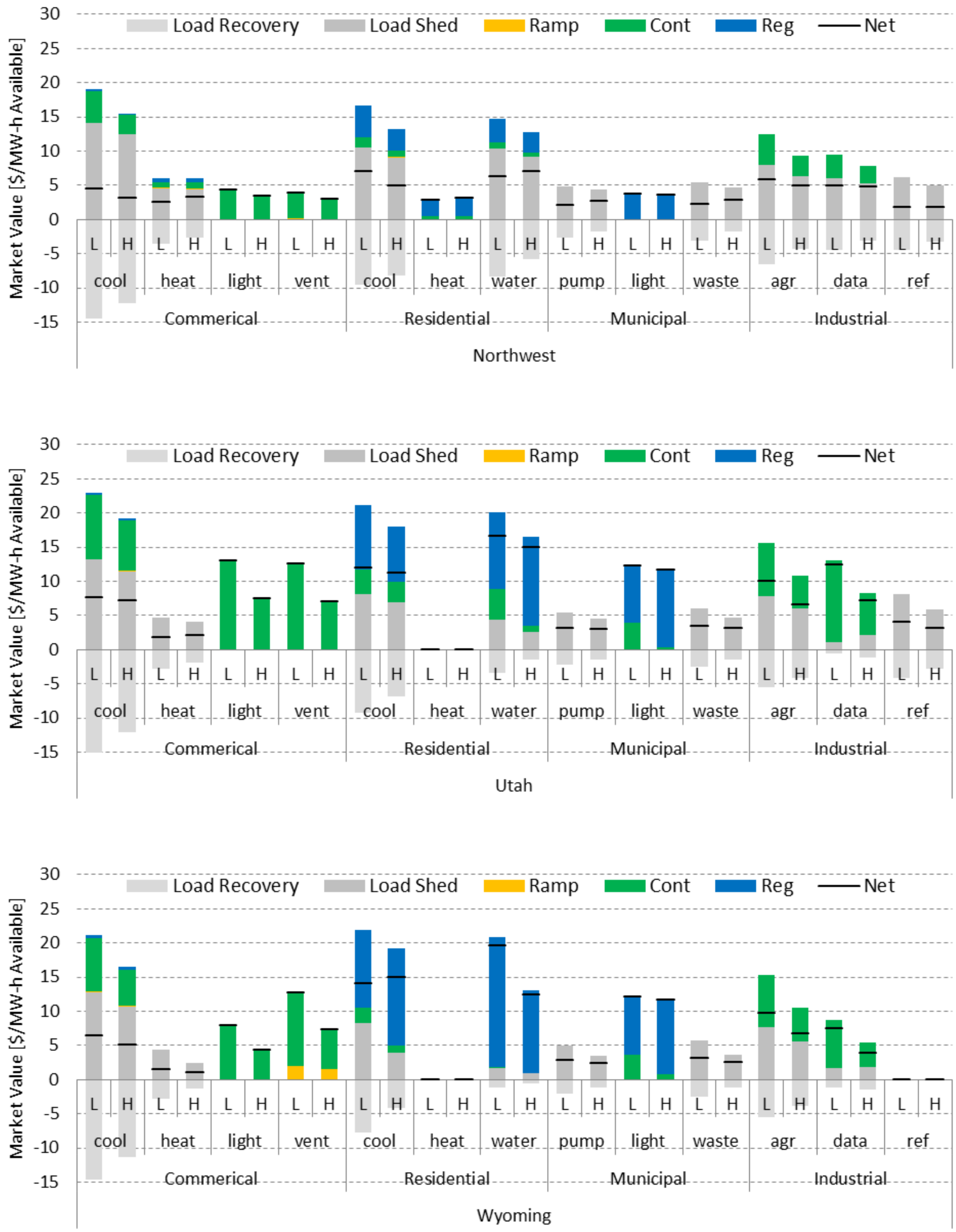

Figure 3-19. Operational value (implied market value) based on marginal costs of production for different types of demand response resources in the Western Interconnection model, at low and high levels of wind and solar generation

Text Box: Measuring Demand Response Value

Unlike power plants that have standard measures, there appears to be no consistent and clear measure for demand response. The measure by which demand response is valued (i.e., dollar per *what*) has important implications for understanding the relative value across different resource types and scenarios. For conventional generators (and energy storage), the denominator is often kW-year. This gives the value per kW of a generator summed over the year. The size of conventional generators and energy storage, in terms of kW, is well-defined and provides a good measure for the device cost (in the case of energy storage it requires knowledge of the number of hours of storage). A similar metric could be used for demand response; however, there are several important considerations. For instance, the capacity rating for demand response could be based on the peak enabled capacity or the average load of the enabled end-use resource.

As discussed, the demand response resource (R) is defined as the flexibility (F) times the end-use load profile (L) and the flexibility is the sheddability (S) times the participation rate (P):

$$R_{l,p}(t) = F_{l,p}(t) \cdot L_l(t) = \left(S_{l,p}(t) \cdot P_{l,p}(t)\right) \cdot L_l(t)$$

which is time-dependent (t) and specific to each type of load (l) and each type of bulk power system product (p). The participation rate is defined as:

$$P_{l,p}(t) = \min\left(C_{l,p}(t), A_{l,p}(t)\right)$$

where the participation rate equals the minimum of controllability (C) and acceptability (A). We can then define the peak enabled capacity ($\hat{C}$) as:

$$\hat{C}_{l,p} = \max_t \left\{F_{l,p}(t) \cdot L_l(t)\right\} = \max_t \left\{R_{l,p}(t)\right\}$$

We can also define the average enabled end-use load ($\bar{E}$) as:

$$\overline{E_{l,p}} = \frac{1}{T} \cdot \sum_{t=0}^{T-1} \left(P_{l,p}(t) \cdot L_{l,p}(t)\right) = \frac{1}{T} \cdot \sum_{t=0}^{T-1} \frac{R_{l,p}(t)}{S_{l,p}(t)}$$

where (T) is the number of time intervals (i.e., 8,784 hours in the year 2020). Figure 3-20 shows these values for commercial building cooling (top) and commercial building lighting (bottom) end-uses. Cooling loads have prominent peaks in both load and demand response availability, leading to the enabled peak capacity being much larger than the average enabled end-use load. The opposite is true for lighting loads.

Each of these metrics leads to difficulties in comparing value across different end-use types. Thus, this study uses maximum cumulative availability (A) for demand response:

$$A_{l,p} = \sum_{t=0}^{T-1} \max_{l,p} \left\{F_{l,p}(t) \cdot L_l(t)\right\}$$

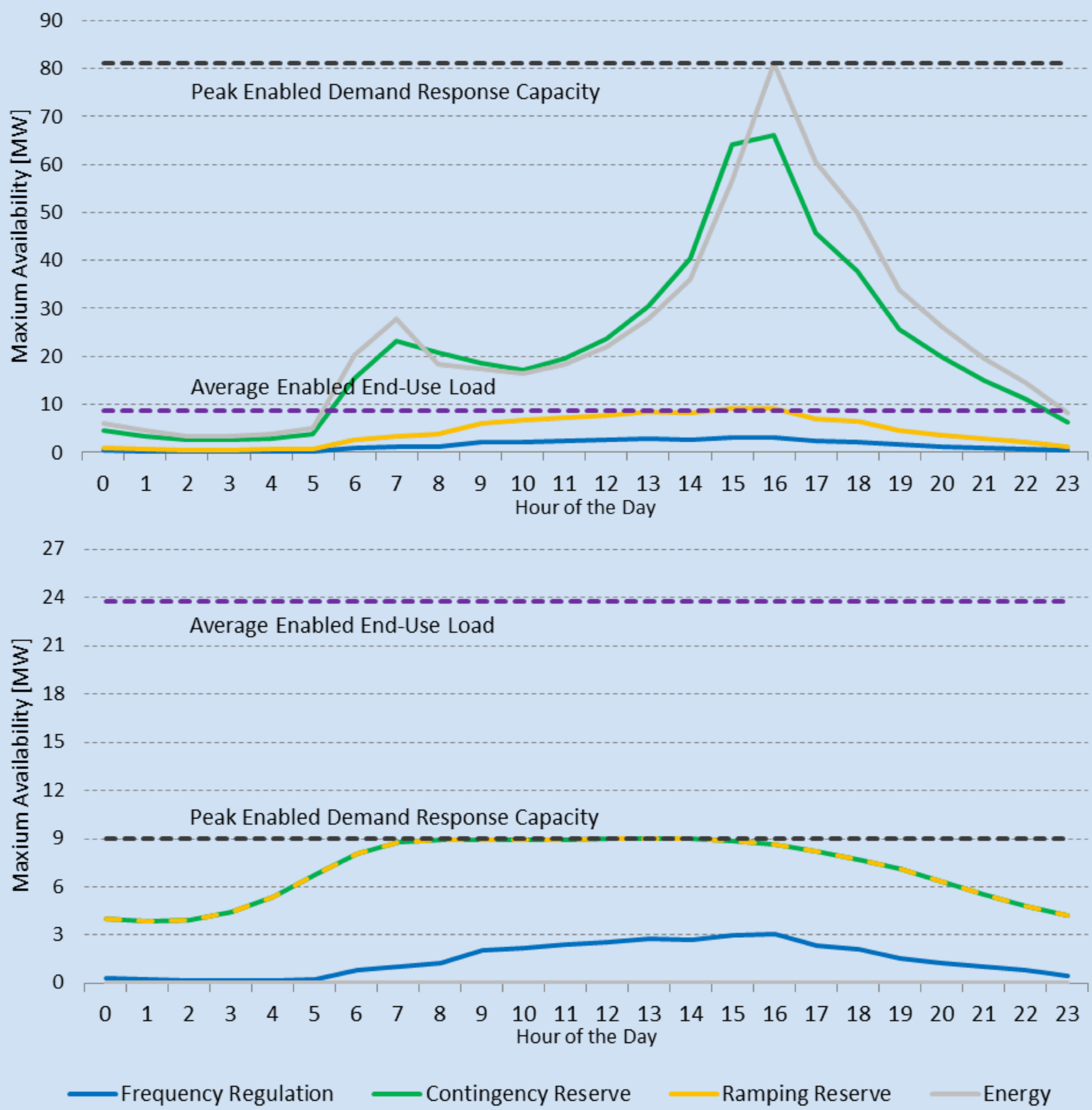

Figure 3-20. Maximum hourly availabilities for energy and operating reserves using commercial building cooling loads (top) and commercial building lighting loads (bottom) in the Northwest region, showing a comparison of peak enabled demand response capacity and average electric load for the enabled demand response resource

The maximum cumulative availability is the maximum availability for all bulk power system products at each hour and summed over the study year. This provides an assessment of how the availability of different demand response resources matches system needs. Within an individual region, those resources more correlated with high production cost times will have greater values in terms of $/MW-h of availability; those less correlated will have lower values. However, none of the three discussed measures provides a good indication of the costs associated with enabling the demand response resource. Enablement costs may have large fixed costs that scale with the number of enabled sites, rather than the total size or timing of the resource availability. Additional work is needed to better assess demand response enablement costs.

In contrast, other end-use loads, such as municipal freshwater and wastewater pumping, have more energy scheduling flexibility and are assumed to be capable of shifting energy throughout a 24-hour period. These loads provide, on average, twice as much value for load shedding than they pay in load recovery because they are capable of arbitraging between the highest and lowest energy cost hours of the day. Figure 3-21 provides an illustration of how different resources vary in their ability to take advantage of energy arbitrage opportunities. The figure compares the value of energy shifting achieved in the simulation to a theoretical value derived from the difference between the two highest hourly energy costs and the two lowest hourly energy costs daily. Two-hour arbitrage value is chosen for comparison because no resource is found to be capable of exceeding this benchmark. Municipal loads provide nearly the maximum value, whereas commercial and residential cooling loads provide only a small fraction of it. Of note, our results are based on static estimates of energy scheduling flexibility, so price signals could alter the rates and patterns of customer participation, changing flexibility.

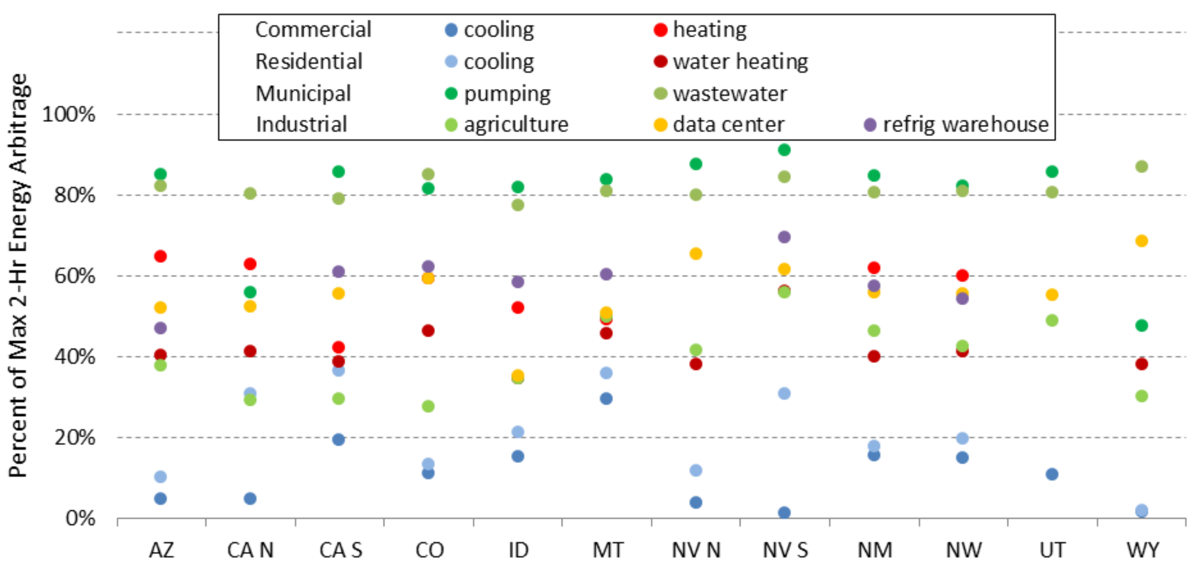

Figure 3-21. Comparison of the average net value that resources can provide for energy shifting to a theoretical value based on the difference between the two highest and two lowest hourly marginal costs for energy for each day. A 2-hour energy arbitrage reference was selected as no resource is found to be capable of exceeding this benchmark (i.e., the y-axis has no resource greater than 100%).

The modeled demand response resources provide greater value from the provision of operating reserves over the shifting of energy use. The calculated net market value for operating reserves is about 3 times greater than that for energy shifting. The greater value stems partly from the fact that operating reserves are held every hour of the year, but energy shifting is a daily activity with significant operational constraints. However, this value will be impacted if actual dispatch-to-contract ratios differ significantly with study assumptions (i.e., frequency regulation mileage). As shown in Figure 3-22, resources capable of providing operating reserves are typically utilized close to their maximum availabilities. However, energy shifting-only resources have utilization factors of less than 20%. Overall, the demand response resources modeled provide about 25% of the total operating reserves required in the Western Interconnection for the study year 2020; in some regions, the share for specific requirements can be as high as 50% or less than 1% (see Table 3-8). Table 3-8 also shows that ramping reserve provisions by demand response are low in the modeled scenarios, reflecting the low value of ramping reserves compared with frequency regulation and contingency reserves. Ramping reserve is assumed to have less stringent requirements than other operating reserves (see Table 3-5) and would generally be less expensive to procure (Hummon et al. 2013).

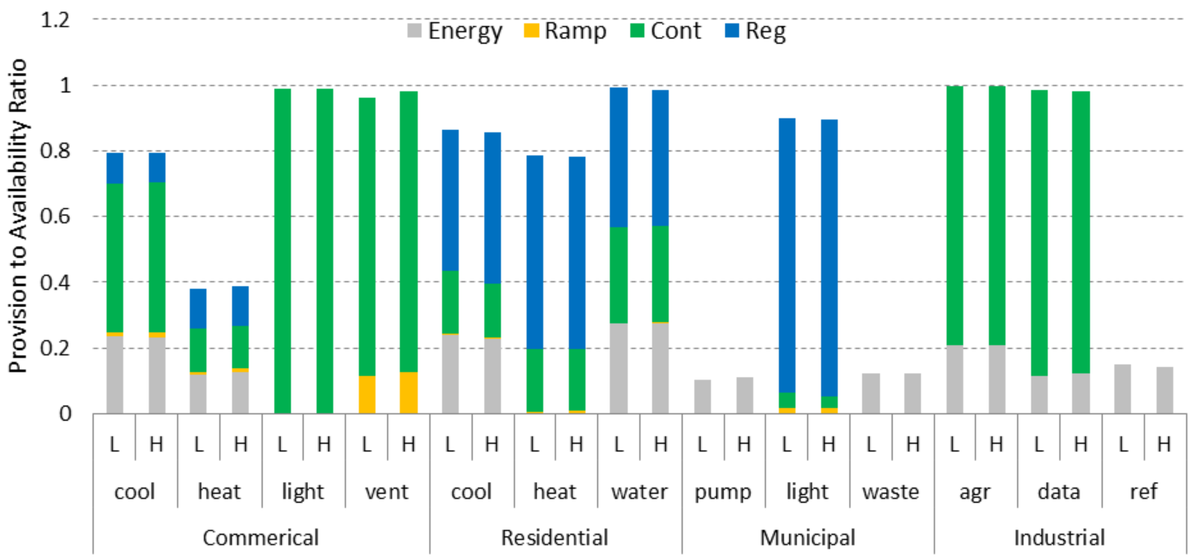

Figure 3-22. Utilization of demand response resources in the Western Interconnection model based on the ratio of demand response provisioned to the maximum cumulative availability.

Table 3-8. Fraction of Demand Response Provision in the Low and High Renewable Cases of the Western Interconnection Model. For operating reserves (frequency regulation, contingency reserve, and ramping reserve), the fraction provision is calculated by dividing the amount of service provided by the total amount needed. For energy, the fraction provision is calculated by dividing the amount of energy shifted through demand response by the total electricity usage.

	Freq. Regulation		Contingency Reserve		Ramping Reserve		Energy	
	Low RE	High RE	Low RE	High RE	Low RE	High RE	Low RE	High RE
Arizona	48%	46%	15%	14%	1.2%	0.8%	0.3%	0.3%
California North	49%	48%	24%	24%	16%	13%	0.1%	0.1%
California South	41%	42%	36%	36%	7.6%	4.9%	0.2%	0.2%
Colorado	30%	16%	14%	14%	0.1%	0.4%	0.1%	0.1%
Idaho	37%	32%	20%	20%	1.8%	1.2%	0.3%	0.3%
Montana	22%	4.6%	7.6%	6.6%	0.2%	0.4%	0.1%	0.1%
Nevada North	26%	12%	24%	22%	8.6%	1.6%	0.2%	0.2%
Nevada South	37%	34%	19%	17%	1.9%	0.6%	0.3%	0.3%
New Mexico	49%	24%	23%	22%	0.1%	0.4%	0.2%	0.2%
Northwest	36%	35%	15%	15%	11%	9.0%	0.1%	0.1%
Utah	38%	43%	23%	19%	0.1%	0.1%	0.1%	0.1%
Wyoming	9.3%	5.5%	6.0%	4.8%	0.4%	0.2%	0.1%	0.1%

Lastly, there are only a few demand response resources types (reflected in Figures 3-16 through 3-19) for which the net value of demand response changes appreciably between the low and high renewable cases. On average, from the perspective of both production cost savings or implied market value, the difference between the low and high renewable cases is less than 10%. For most services, the average operational value is similar between the low and high renewable cases. However, the average operational value for spinning contingency reserves generally declines. This results from the changes in the utilization of conventional dispatchable generation under increased levels of wind and solar power in the modeled scenarios. Due to increased variability, more conventional dispatchable generation may be utilized at part-load, increasing the supply of operating reserves available inherently through energy scheduling. The role of various drivers of operational value under changing levels of wind and solar penetration is discussed further in Section 3.5.

3.4 Energy Storage

Each energy storage technology, including batteries, flywheels, compressed air, and pumped storage hydropower, has different capabilities and cost structures. While there are numerous technical differences among and within these technology types, we consider the general characteristics of energy storage rather than attempting to determine the values of individual energy storage technologies.

3.4.1 Storage Characterization

In this study, we examine two general classes of energy storage technologies; the first is an operating reserves-only device, which resembles a high-power, short duration battery. We assume the device is not ramp-constrained and can provide its full output range for operating reserves instantly. For a device providing spinning contingency reserves, we assume that the device simply provides up to its full discharge capacity without incurring any operational costs, and we do not consider the energy component of a contingency event. For frequency regulation, we also assume the device can provide up to its full capacity and that the service is net-energy neutral in each one-hour simulation interval.

Even if the frequency regulation service is net-energy neutral over time, there will be energy losses. This will produce a net consumption of energy by the storage device, which is the product of two factors: the fraction of capacity used to provide energy (sometimes referred to as the dispatch-to-contract ratio) (Kempton 2005) and the device efficiency. The first factor depends on the amount of energy that flows through the device when called to provide frequency regulation. This energy is multiplied by the loss rate to produce the amount of energy consumed by the storage device. A dispatch-to-contract ratio of 14.2% is chosen (Ferreira 2013) and an efficiency loss rate of 20% was selected, based on a net round-trip trip efficiency of 80% (similar to a lithium-ion battery) (Akhil et al. 2015). In other words, for each hour, a storage device providing 100 MW of regulation consumes 2.8 MWh of energy. The cost of this make-up energy is assumed to be at the marginal cost of energy and loss calculations are handled outside the production cost model. This approach was taken because deployment of provisioned operating reserves is not handled in the modeling software, as discussed in Denholm et al. (2013).

The second general class of energy storage technologies is one that can provide both energy and operating reserves, a device that resembles a high-energy, long duration battery. The simulations include scenarios in which the devices provide only energy and scenarios in which the devices are co-optimized to provide both energy and operating reserves (which allows for disaggregation of the value of the two classes of services). The modeled devices can provide both energy and operating reserves simultaneously, but only up to the rated capacity of each device. We assume a 75% round-trip efficiency (similar to a sodium-sulfur battery) (Akhil et al. 2015) and that the devices can ramp over its entire range within a single energy scheduling interval (with no minimum generation level and the ability to switch between charging and discharging within a single scheduling interval). We also assume constant efficiency as a function of load and state of charge, no minimum up or down times, and 8 hours of energy storage at the rated capacity. Lastly, no fixed or variable operations and maintenance costs were assigned to either class of energy storage device for simplicity.

Two sets of scenarios are investigated to analyze the value of energy storage, and how the value changes as a function of the amount deployed and the level of renewable penetration. As with the demand response simulations (see Section 3.3), we evaluate a Western Interconnection model where various amounts of energy storage are deployed in the different regions. Energy storage devices were added without replacement of existing generators in the regions. However, due to the long modeling run times, only a limited number of scenarios could be evaluated with the entire Western Interconnection model. As a result, we also analyze a more comprehensive set of modeling scenarios in the Colorado system model (see Section 3.5) to examine, by extension, the general trends in energy storage value as a function of the amount deployed and the level of renewable penetration observed in the Western Interconnection model.

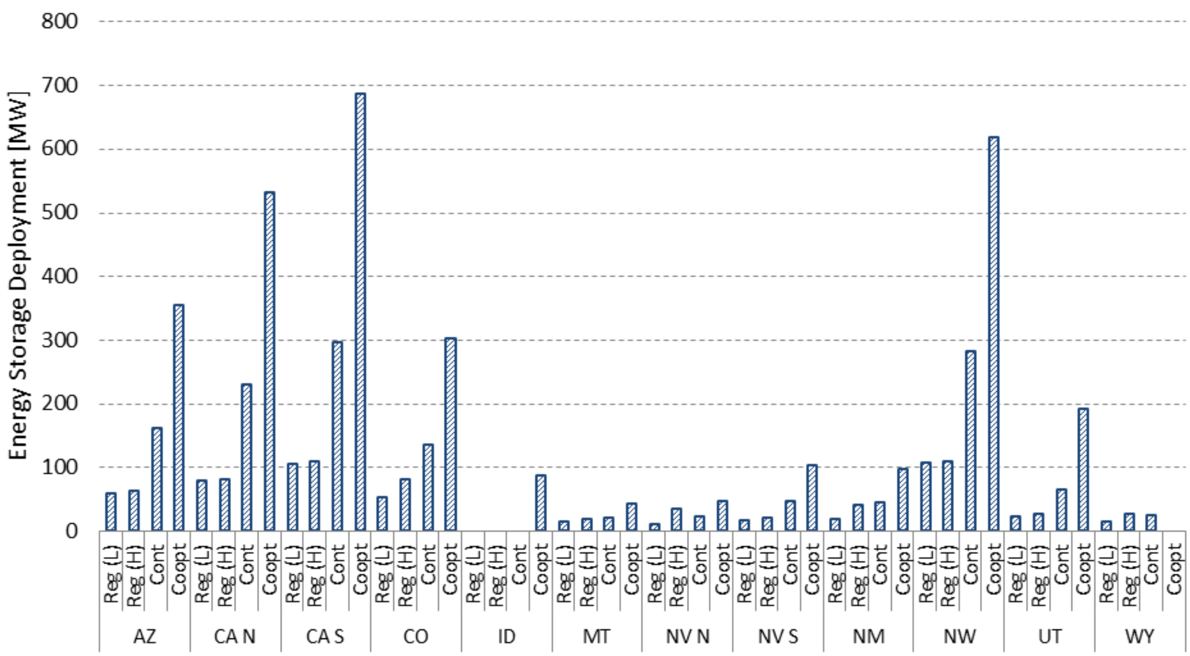

Figure 3-23. Deployment of energy storage in the different scenarios. The total deployment of regulation-only devices is 487 MW in the low renewable case (L) and 597 MW in the high renewable case (H). Contingency-only devices total 1,314 MW and co-optimized energy and operating reserve devices total 3,045 MW in both the low and high renewable cases.

Figure 3-23 shows the modeled deployment of energy storage in each region. For the operating reserves-only scenarios, modeled devices are sized to meet 50% of the average hourly operating reserve requirements within each region (see Table 3-2) and are distributed to the various regions in the Western Interconnection based on these requirements. While somewhat arbitrary, this size is large enough to represent a significant contribution within the system, but not so large as to saturate the operating reserves requirement. Requirements differ between frequency regulation and contingency reserves and lead to a total modeled deployment in the Western Interconnection model of 487 MW (597 MW in the high renewable case) and 1,314 MW, respectively.

For the energy-only and co-optimized scenarios, larger energy storage deployments are chosen equal to about 3.3% of average load in each region, resulting in a total modeled deployment of 3,045 MW. This is substantially larger than the operating reserve-only scenarios to account for the potential additional revenue streams associated with energy transactions. In addition, we examine two other levels of deployment to understand the impact of energy storage penetration on its value: one in which each modeled energy storage device is 50% smaller and one in which each is 50% larger.

3.4.2 Energy Storage Operational Value

This subsection examines the operational value of energy storage in the low renewable and high renewable case of the Western Interconnection model, as described in Section 3.1. We apply a similar approach as used to calculate the operational value of demand response in Section 3.3.3. We first calculate the total production cost in the base system without the deployment of additional energy storage devices. We then add the amounts and types of energy storage (summarized in Figure 3-23) to the various regions and calculate the difference in production costs. Any reduction in operational cost is attributed to the deployment of energy storage, representing its system operational value. In addition, we use the marginal costs of energy and operating reserves to estimate the revenue that the resources could expect to receive in a market environment. As discussed previously in Section 3.2, this assumes that the market clearing prices correspond to the marginal costs calculated by the production cost model.

The discussion begins with an assessment of the scenarios with operating reserves-only energy storage devices and is followed by an assessment of the scenarios with larger devices that are capable of being co-optimized for energy and operating reserves.

Figure 3-24 provides the operational value for energy storage devices providing only operating reserves for the low renewable case and the high renewable case. The bar charts on the left represent the production cost savings normalized by the amount of energy storage devices deployed for the specific scenarios (regulation-only and contingency reserve-only) in each case. In Figure 3-24 (low renewable case, left), introduction of the frequency regulation-only devices reduce total operational costs by $16.7 million. Dividing this value by the total installed storage capacity (487 MW) produces an average value of about $34/kW-year. The figure further breaks this savings into the categories tracked by the production cost model. Steady state savings represents the fuel and operations and maintenance costs avoided by more efficient dispatch that results when thermal plants are not required to provide operating reserves. For frequency regulation, the steady state benefits are somewhat offset by the required make-up energy assumed in this analysis. Startup and shutdown savings represents the ability of energy storage to reduce costs associated with plant starts. For regulation-only devices, the largest net cost savings stem from avoided non-steady state operation costs associated with thermal plants following a regulation signal. This includes both additional maintenance costs as well as heat rate degradation captured in the regulation cost described in Table 3-3. For the scenarios with devices that provide contingency reserves only, the per-unit value is less than that of frequency regulation-only devices. The calculated savings from use of contingency reserves-only devices is mostly associated with reduced part-load operation and more efficient overall commitment and dispatch of the generator fleet.

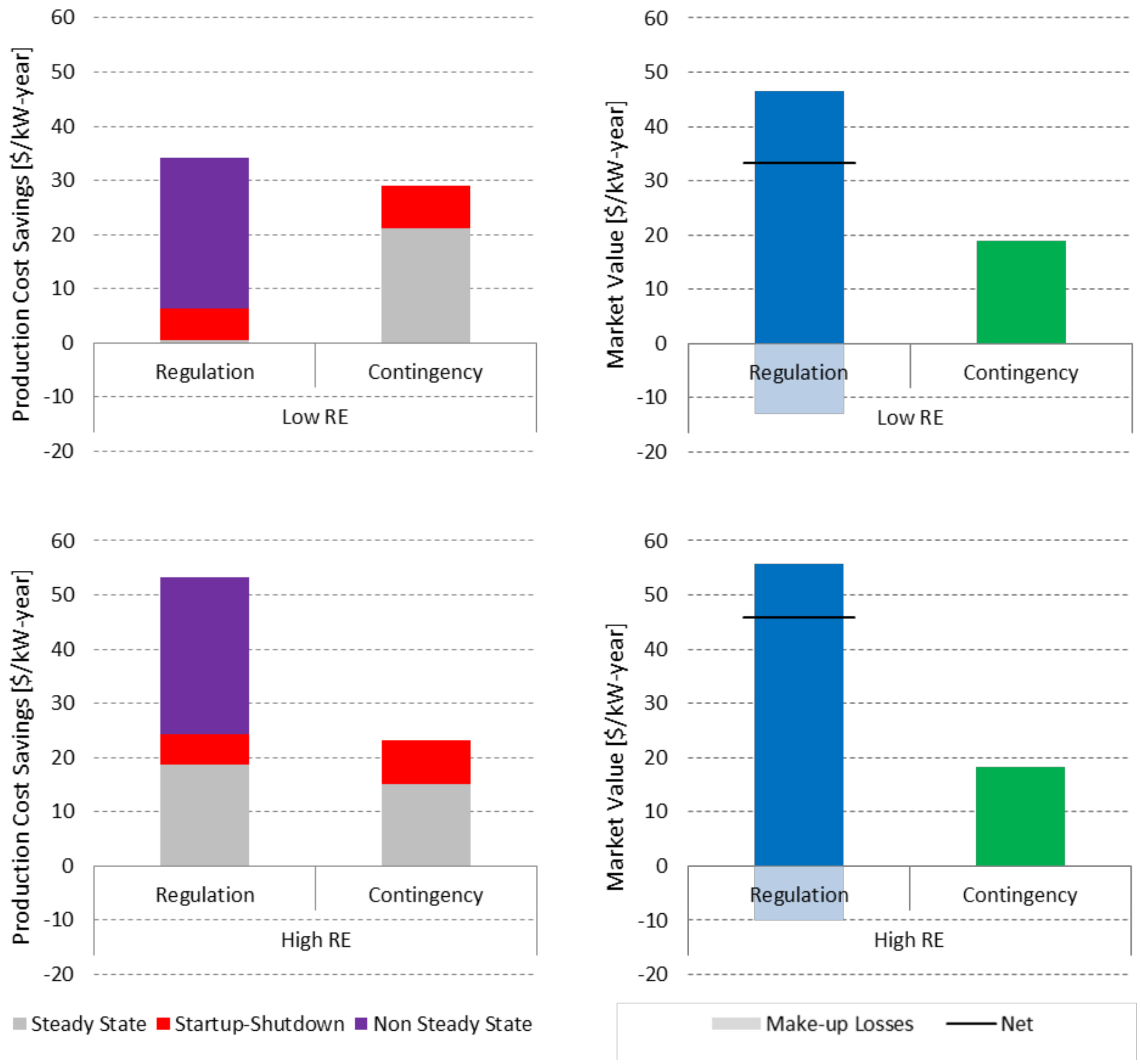

Figure 3-24. The operational value of energy storage providing either frequency regulation only or spinning contingency reserves only. Value is based on production cost savings to the system (left) and implied market value based on marginal costs of production (right).

Figure 3-24 (right) represents the implied market value based on operating reserves being paid at the hourly marginal costs of production (as calculated by the production cost model). For the regulation devices, we include the savings achieved through the provision of reserves as well as the assumed cost of make-up energy, with the net market value represented by the black line. The implied market values are less than the total production cost savings due to two factors. First, the costs of generator startups are not captured in marginal costs of production. Second, the addition of energy storage into the system will generally reduce marginal costs (i.e., market clearing prices), which can have a significant impact on the calculated value (EPRI 2013). The high renewable scenarios show a modest increase in the value of energy storage compared to the low renewable scenarios. This increase can be attributed to the larger operating reserves requirements in the high renewable case and the lower cost of energy used for make-up. These factors are discussed in more detail in Section 3.5.

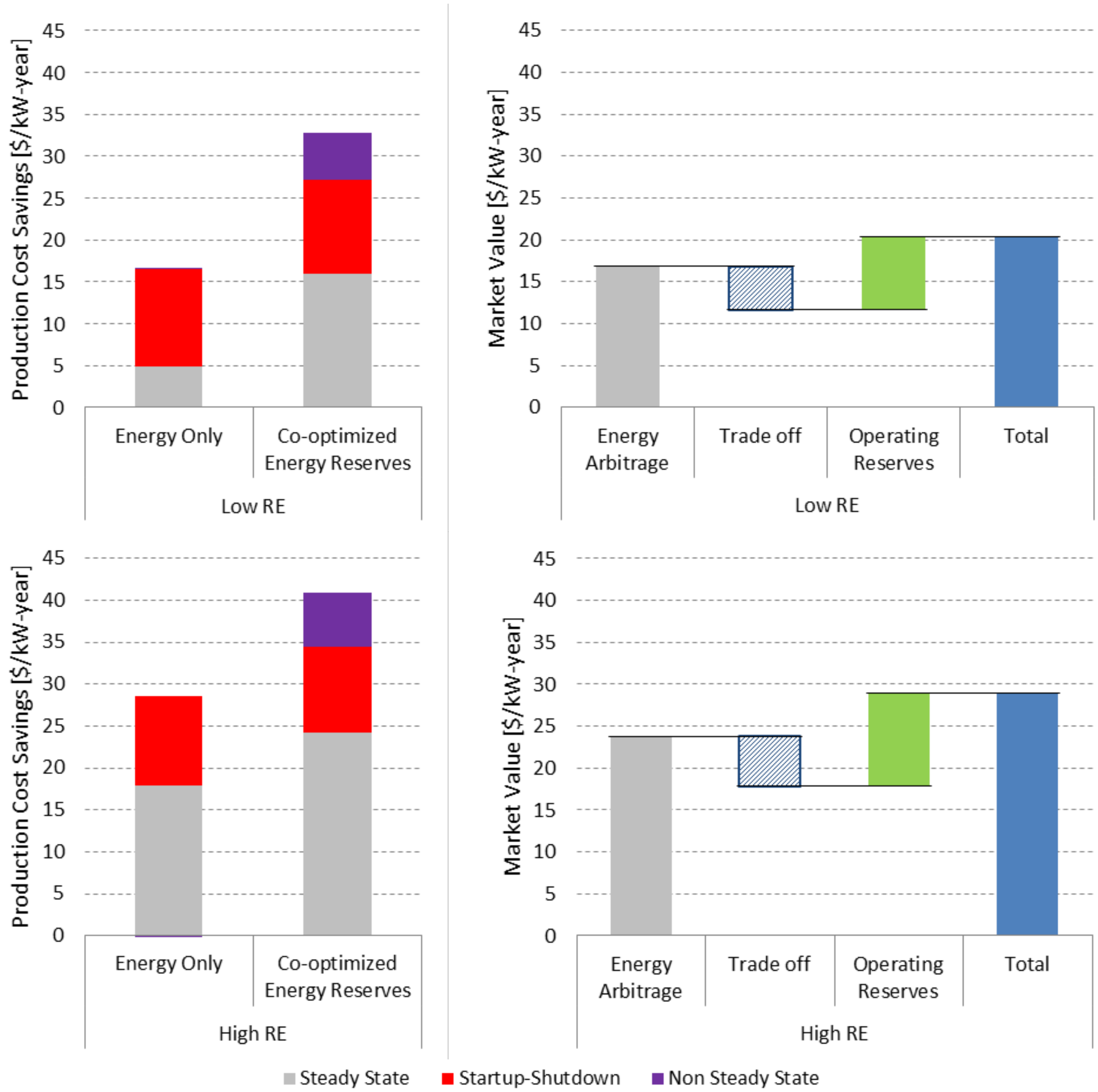

Figure 3-25. The operational value of energy storage (co-optimized for energy and operating reserves) in terms of production costs savings (left) and implied market value (right), in the low renewable (top) and high renewable (bottom) cases of the Western Interconnection model. The 'trade off' represents the reduction in implied market value achieved from energy arbitrage alone due to energy storage capacity used for the provision of operating reserves, leading to greater net value through co-optimization.

Figure 3-25 provides the results for the energy-only and co-optimized energy storage scenarios. As in Figure 3-24, the left panel provides the production cost savings associated with the addition of energy storage for the low and high renewable cases, normalized by the amount deployed, and the right panel provides the implied market value.

The production cost model records how much energy and operating reserves are provided by the co-optimized devices in the energy-only and co-optimized scenarios. Using the hourly energy storage plant dispatch data and the calculated marginal costs, we can acquire a sense of the tradeoffs among possible services energy storage can provide. This is demonstrated in the charts on the right side of Figure 3-25. By participating in energy transactions, a co-optimized energy storage device foregoes some of its ability to acquire value from providing operating reserves, and vice versa. Comparing the energy-only and co-optimized scenarios, the co-optimized devices gain about double what they lose in the trade-off so, in aggregate, they provide more operational value.

A comparison of Figures 3-24 and 3-25 shows that the per-unit value of the co-optimized devices is less than the frequency regulation-only devices. This counter-intuitive result is partly due to the different amounts of energy storage deployed in the different scenarios modeled. The total deployment of co-optimized devices is more than four times greater than the frequency regulation-only devices. Because the incremental value of energy storage tends to decline with additional deployment (as with demand response and other resources added to an unconstrained system), the per-unit value tends to be smaller for larger deployments of energy storage devices. Furthermore, the large energy capacity (8-hour duration at the rate power) of the co-optimized devices introduces an additional driver that impacts value. While the devices in the frequency regulation-only scenario are nearly fully utilized providing 49% of the requirement (the scenario deployed enough to meet 50% of the requirement), the devices in the co-optimized scenario, while significantly larger, provide only 27% of the frequency regulation requirement. Despite the lower provision of frequency regulation by the co-optimized devices, an indicator that the frequency regulation requirement is not saturated by the devices, the average marginal costs for frequency regulation are lower than that in the frequency regulation-only scenario. When providing frequency regulation, the co-optimized devices can provide other services like contingency reserves and energy shifting with the remaining capacity. When discharging energy, the energy storage device suppresses the marginal cost of energy, which in turn also suppresses the marginal cost of operating reserves (note that the model does not capture this effect for the makeup losses when energy storage provides frequency regulation). As a result, the larger co-optimized devices modeled provide a combination of services whose per-unit value is actually less than the much smaller frequency regulation-devices.

Consequently, quantifying the value of energy storage cannot be disentangled from the assumed level of deployment. Figure 3-26 shows the results of the co-optimized scenario for the three levels of energy storage deployment modeled, with an approximate curve fit to illustrate this relationship. The curve shows the operational value based on production costs savings as well as the implied market value assuming resources are paid at the marginal costs (i.e., market clearing prices). From this result, it is evident that the operational values of energy storage in different regions would change substantially if the assumed deployment levels differed from our modeled scenarios.

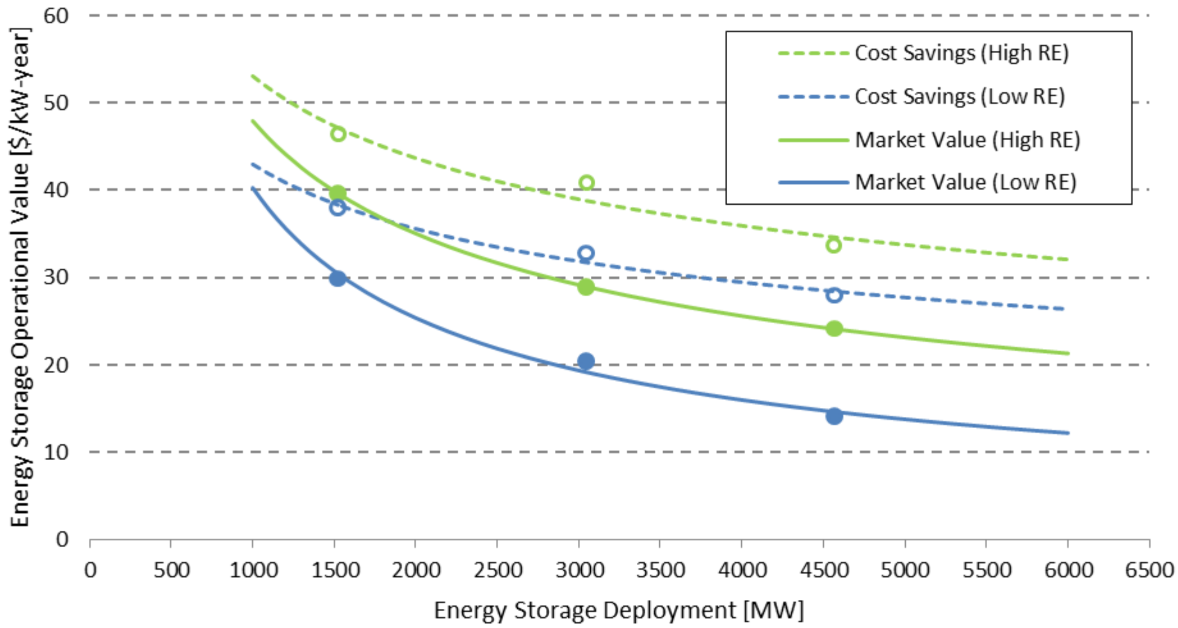

Figure 3-26. The operational value of energy storage (co-optimized for energy and operating reserves) as a function of total deployment in the low and high renewable cases of the Western Interconnection model. The value calculated from both perspectives decline with greater deployment levels.

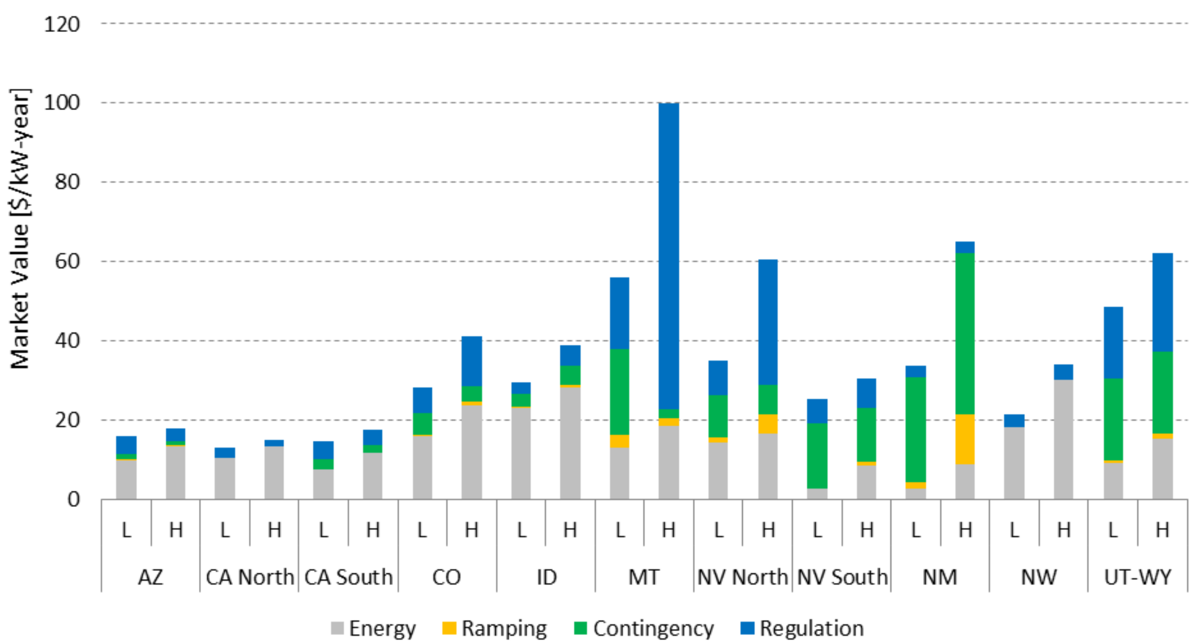

Figure 3-27. Implied market value of energy storage (co-optimized for energy and operating reserves) in the low and high renewable cases, assuming prices are based on the marginal costs of production

3.5 Operational Value with Variable Generation

Sections 3.3 and 3.4 provide an indication of the operational value of demand response and energy storage in a large system (i.e., the Western Interconnection model). However, they do not provide a detailed understanding of why the value of these resources changes with increased penetration of variable generation.

The operational values of both demand response and energy storage devices (capable of providing both energy and operating reserves) are largely driven by two primary factors: the required amount of operating reserves in a system and the associated cost, and the time-varying cost of energy in that system, including the difference between on-peak and off-peak marginal costs of energy. Due to system-wide effects, there are competing drivers whose net effect can increase or decrease the value of demand response or energy storage within a power system. As a result, while generalizations drawn from a limited set of scenarios cannot be applied to specific market environments, the analysis can identify important drivers of value and their expected directionality in impact.

To explore these issues, the study uses a set of scenarios with increasing levels of variable generation and calculates the changes to total production costs and implied market value based on marginal costs of production associated with energy and operating reserves generated from the production cost model, as discussed in Section 3.2. The scenarios are based on the Colorado system model described in Section 3.1. Its relatively small size allows for simulation of many combinations of generator compositions and characteristics with significantly shorter computational times. We first examine the impact of renewable penetration levels on the cost of operating reserves, a key source of value for both demand response and energy storage. Scenarios vary from 15% to 35% renewables (by energy) in approximately 5% increments. Reserve requirements for each level of renewable generation are based on the methods in Ibanez et al. (2012). We then calculate the cost of system operation with and without holding reserves (the difference being the total cost to the system for procuring reserves) as well as the marginal cost (i.e., price) of each operating reserve product.

Table 3-9. Fraction of Renewable Generation from Wind and Solar Resources in the Colorado System Model and the Associated Model Requirements for Operating Reserves

Generation from Renewables	Wind Generation (GWh)		Solar PV Generation (GWh)		Cumulative Operating Reserve Requirement		
					Regulation (GW-h)	Contingency (GW-h)	Ramping (GW-h)
15%	10,705	(13.6%)	1,834	(2.3%)	1,050	3,548	502
20%	13,838	(17.4%)	2,556	(3.2%)	1,134	3,548	600
25%	18,097	(22.8%)	3,168	(4.0%)	1,281	3,548	769
30%	21,433	(27.0%)	3,750	(4.7%)	1,364	3,548	855
35%	23,752	(29.9%)	4,260	(5.4%)	1,422	3,548	918

Table 3-9 summarizes the quantity of operating reserves required in the various scenarios and Table 3-10 gives cost values associated with those reserves. For these scenarios, the frequency regulation requirement increases from 1.1 TW-h to 1.4 TW-h and the ramping reserve requirement increases from 0.5 TW-h to 0.9 TW-h. Despite the large increases in operating reserve requirements, there is a relatively modest increase in the costs of providing these operating reserves. The relatively small increase in costs is partly due to decreases in energy production cost in a system with increased levels of wind and solar power as well as energy schedule changes that reduce conventional generation output, releasing capacity to provide operating reserves.

Table 3-10. Total Costs and Marginal Costs Associated with Operating Reserves at Different Levels of Wind and Solar Generation in the Base Colorado System Model without Additional Demand Response or Energy Storage

Generation from Renewables	Total Production Cost ($M)	Operating Reserve Total Costs		Operating Reserve Marginal Costs ($/MW-h)		
		Total Costs ($M)	Unit Costs ($/MW-h)	Regulation Mean / Median	Contingency Mean / Median	Ramping Mean / Median
15%	1,427	27.4 (2.0%)	5.4	15.48 / 13.81	6.15 / 3.32	1.62 / 0
20%	1,309	29.1 (2.3%)	5.5	16.31 / 14.61	6.14 / 3.09	2.05 / 0
25%	1,170	32.3 (2.8%)	5.8	16.95 / 14.58	5.88 / 2.81	2.21 / 0
30%	1,072	32.1 (3.1%)	5.6	16.81 / 14.52	5.31 / 2.61	2.3 / 0
35%	1,003	31.2 (3.2%)	5.3	16.53 / 14.52	5.04 / 2.51	2.18 / 0

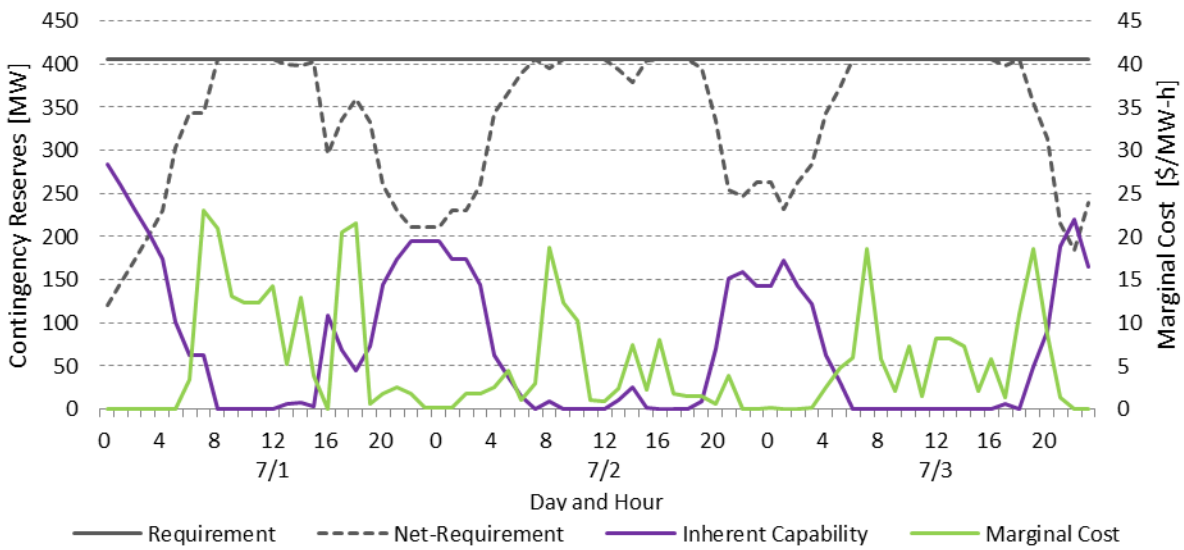

Figure 3-28. Example days in the Colorado system model showing the contingency reserves requirement (gray solid line) compared to the marginal cost of contingency reserves (green solid line). The marginal cost of contingency reserve falls when there is sufficient excess operating range and ramping capability (purple solid line) to meet the operating reserve requirement solely through energy scheduling.

Energy scheduling alone leads to some residual operating range and ramping capability that provides an inherent level of operating reserves. Figure 3-28 illustrates how the inherent capability of the system to provide (spinning) contingency reserve (purple line) varies over three model days. Generators have various constraints such as startup times, minimum operating points, and minimum up times. Consequently, even if there were no uncertainty in net load, the optimal scheduling of generators to follow net load would still result in some hours with unused spinning capacity (Hummon et al. 2013). Any remaining need for operating reserves must be met through a re-dispatch of the system (i.e., a change from the optimal energy schedule), which subsequently incurs additional costs (i.e., lost opportunity costs). The level of inherent capability is an important factor in the value of demand response and energy storage. As shown in Figure 3-28, the marginal cost of contingency reserves (green line) tends to be low when the net requirement (dashed black line) is also low, signifying that the system can supply a large fraction of contingency reserves without incurring lost opportunity costs.

Understanding inherent capability is particularly important to understand how changes in the composition of the generator mix, especially in cases with increasing levels of variable renewable generation as discussed in this section, impact the costs of operating the system. A system that is able to accommodate variable generation will necessarily have more inherent operating range and ramping capability due to energy scheduling alone. Simulations of the Colorado system model with the addition of an energy storage device demonstrate this effect in more detail. We add a 100-MW frequency regulation-only device to the system and calculate its value as a function of renewable penetration (with scenarios up to 55% renewable generation).

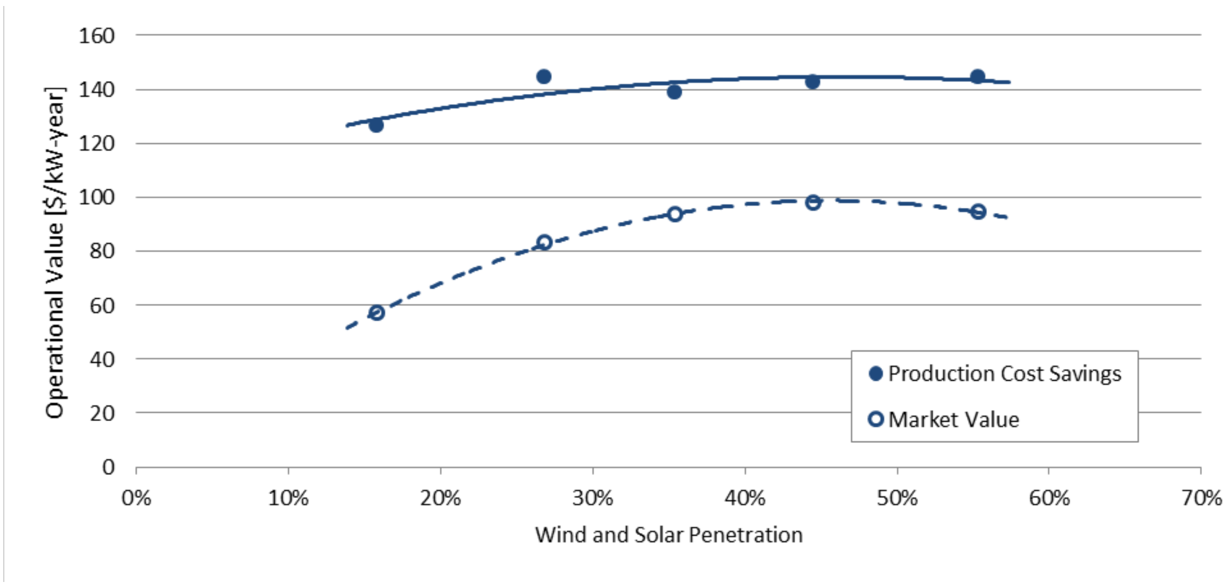

Figure 3-29. Value of a 100-MW frequency regulation-only energy storage device as a function of renewable generation in the Colorado system model. Production cost savings and implied market value based on marginal costs of production for the Colorado system are shown.

Figure 3-29 provides the results for the frequency regulation-only device, with operational value in terms of production cost savings and implied market value. It shows a modest increase and then decrease in operational values with increasing penetration of wind and solar power. This behavior stems from a competition between the increasing frequency regulation requirement at higher levels of wind and solar penetration (which leads to increased operational value of the energy storage device) and the increasing level of inherent capability (which leads to decreased operational value of the energy storage device). The impact of these factors can be seen more clearly in Table 3-10, which provides the total costs and marginal costs of operating the base system at different levels of wind and solar generation without additional demand response or energy storage. Note that the inherent capability leads to zero marginal costs for ramping reserve for more than half of all hours.

As with previous comparisons, Figure 3-29 also shows a lower quantity for the implied market value compared with the operational value based on production cost savings. This result is driven largely by suppression of the marginal costs of frequency regulation (as shown in Figure 3-30) and other operating reserve products due to saturation from excess capacity. In the scenario with the lowest amount of renewable penetration, the energy storage device provides about 82% of the system's regulation requirement and reduces the average marginal cost of regulation by about 60%, from $16/MW-h to $6.5/MW-h.

In Figure 3-29, the production cost savings are relatively flat with changing penetration of wind and solar power, but the market value increases significantly between scenarios in which wind and solar power increase from 15% to 44% of generation. The contrast between the two measures of operational value stem from several drivers. At higher levels of wind and solar power, the modeled frequency regulation requirement is also larger. Because the size of the energy storage device is constant between scenarios, at 100 MW, the energy storage device provides a progressively smaller fraction of the requirement as renewable penetration increases. By providing a smaller fraction, there is an associated smaller suppression of marginal costs for frequency regulation. This suppression occurs because energy storage (as well as demand response) provides the regulation service at zero lost opportunity cost. This is demonstrated by Figure 3-30, which shows the average marginal cost for frequency regulation before and after the addition of the frequency regulation-only device.

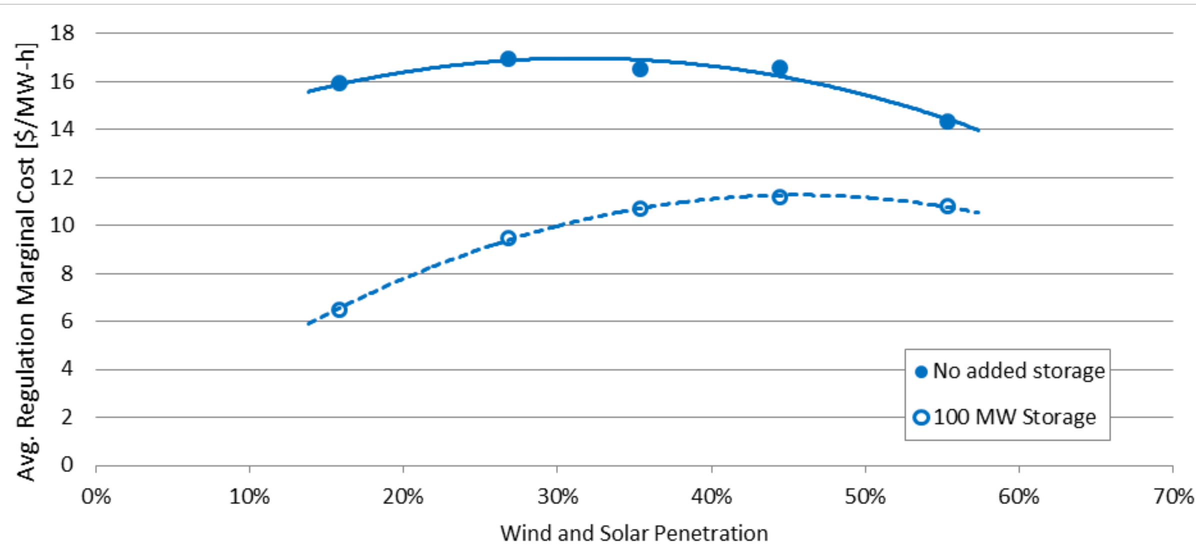

Figure 3-30. Average marginal cost of frequency regulation as a function of renewable penetration in the Colorado system model at different levels of renewable generation and the impact of deploying energy storage

The second issue evaluated is the relationship between renewable generation and marginal costs of energy. The introduction of any zero marginal cost source of generation into an otherwise static generation mix will tend to reduce operational costs during the period of generation. The impact on energy costs will largely depend on the timing of the renewable generation. A renewable generation source that generates predominantly during peak periods could act to reduce peak marginal costs and reduce potential operational value for demand response or energy storage providing energy shifting.

Alternatively, renewable generation during off-peak periods will decrease off-peak marginal costs, reducing the costs of energy storage charging (or load shifting) and increase potential operational value. Ultimately, it is the overall impact of renewable generation on on-peak and off-peak marginal costs that will affect the potential opportunities for demand response and energy storage to provide operational value to the system.

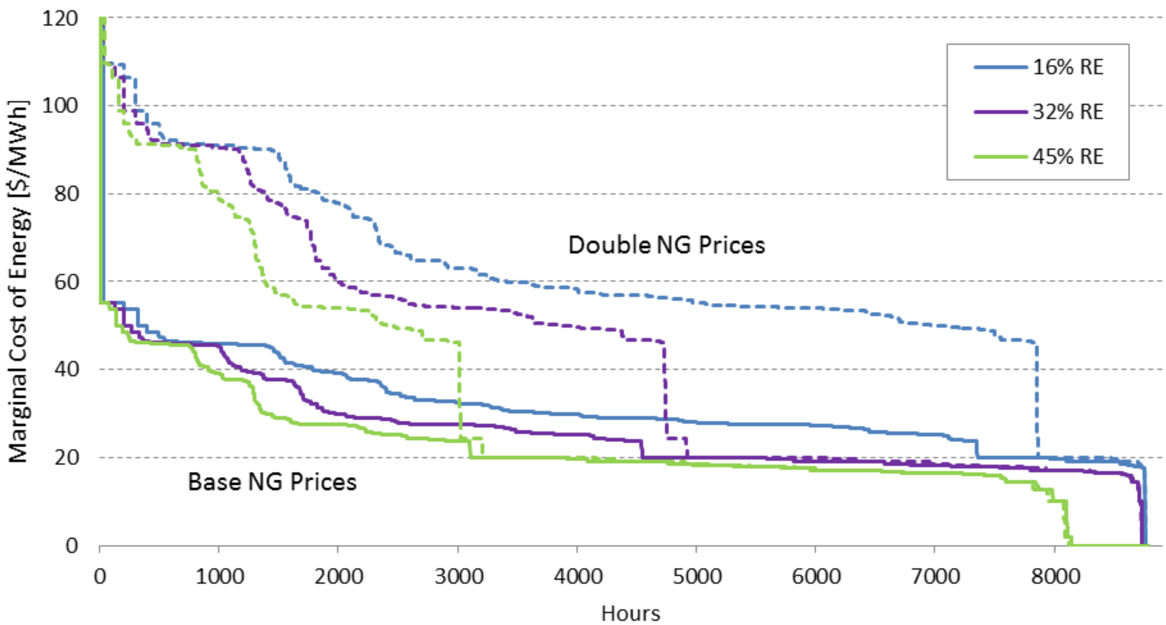

Figure 3-31. Marginal cost duration curve for energy in the Colorado system model for renewable penetrations of 16%, 32%, and 45% and two natural gas prices (about $4.1/MMBTu and $8.2/MMBTu).

Some indication of energy shifting value in the Colorado system model can be derived from the marginal cost duration curves for energy illustrated in Figure 3-31. Curves are shown for three levels of renewable penetration and two natural gas prices (at about $4.1/MBtu and $8.2/MBtu). The higher gas price leads to larger differentials in the marginal costs for energy. In the lowest renewable case (blue lines), this differential for many hours of the year is based on the difference between the operating cost of coal units (generating at about $20/MWh) and natural gas combined cycle units (generating at about $25–$35/MWh in the low gas price scenario or about $45–$70/MWh in the high gas price scenario). If energy storage is used for shifting energy produced by coal units to displace energy produced by natural gas units in the low renewable, low natural gas price scenario (solid blue lines), the net operational cost savings will be small, especially considering losses associated with the round trip efficiency of energy storage devices. The net operational savings will be greater in the low renewable, high natural gas price scenario (dashed blue line), but there are relatively few hours during which energy from off-peak coal units is available for charging energy storage devices. When energy from coal units is not available, energy storage devices must charge with energy from natural gas combined cycle generation to displace higher cost, lower efficiency peaking units (e.g., natural gas- or oil-fired combustion turbine (CT), internal combustion, or steam units). These units have a wide range of marginal costs, as illustrated by the lumpiness on the left-hand side of the price duration curves. However, there are still relatively few hours of very high marginal costs for energy during which there are significant opportunities for operational cost savings from energy shifting. The increased penetration of renewables results in lower overall marginal costs for energy, but with an apparent greater impact on off-peak marginal costs for energy (right side of curves in Figure 3-31). Moving from the 16% case to the 32% case greatly increases the number of hours off-peak coal is available (hours during which marginal costs are less than $20/MWh). At the higher renewable penetration case (45%), there are more hours of zero-cost wind and solar available to charge energy storage devices, presenting greater opportunities for operational cost savings from energy shifting.

The impact of renewable penetration on the value of energy storage is further examined with the results shown in Figure 3-32, which illustrates the operational value of a 300-MW co-optimized energy storage device in the Colorado system. The left panel shows the potential operational value as measured by the reduction in production costs. It shows the general increase in value associated with both steady-state operation (largely associated with energy shifting) and non-steady-state operation (largely associated with providing frequency regulation) as renewable penetration levels increase.

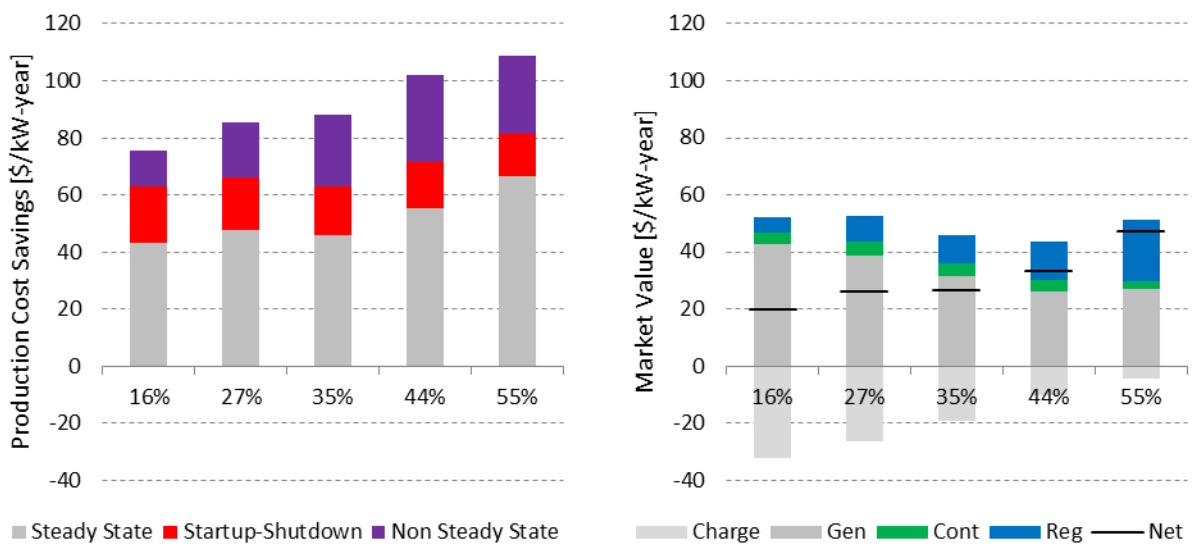

Figure 3-32. Comparison of operational cost savings and implied market value for a 300-MW co-optimized energy storage device in the Colorado system model. Operational values are provided for 5 levels of wind and solar generation, from 16% to 55% of energy.

The right panel in Figure 3-32 provides the system implied market value, assuming the energy storage device is paid at the marginal cost of energy and operating reserves (and pays the marginal cost of energy when charging). These results demonstrate how the energy shifting value of energy storage increases due to reductions in the cost of charging energy (better energy arbitrage). In each hour, the revenue or cost associated with energy shifting is calculated by multiplying the amount of charging/discharging by the marginal price. In our model system, as variable renewable penetration increases, there is generally a decrease in value from sourcing (discharging) energy as there are fewer hours of high price energy (shown by the "Gen" bar in the figure). However, there is a greater reduction in the cost of charging, illustrated by the negative values that decrease as a function of renewable penetration. This produces an overall increase in net market value for the co-optimized devices (black lines). Additionally, the provision and value of regulation increases as a function of renewable penetration as shown earlier (Figure 3-29), despite a suppression of marginal costs compared to the base case without energy storage.

3.6 Capacity Value

The production cost modeling reveals a number of important drivers of operational value for demand response and energy storage resources. However, it does not consider their potential capacity value, which is the ability of demand response and energy storage to offset the need for other new capital investments. Electric system planners utilize a variety of approaches to assess how new investments in grid assets can contribute to capacity needs and the monetary value of that capacity. In the short-run, the value of capacity can vary substantially depending on whether the system has excess or is deficient in meeting its target planning reserve margin (i.e., the level of capacity necessary for meeting reliability standards). In the long-run, the value of capacity may tend toward what is often called the cost of new entry (or cost of best new entrant). A comprehensive analysis of capacity value is outside the scope of this report. The study approach assumes a capacity value of zero because no generators were replaced or retired from the base cases with the deployment of demand response and energy storage resources. However, rough estimates can be made to assess the relative magnitudes of operational and capacity values.

As a simplification, the potential capacity value of demand response or energy storage can be based on a proxy resource, like a natural gas-fired combustion turbine used commonly for meeting peak electricity demands. There is a large range of estimates for the annualized cost of a new combustion turbine, based on equipment costs, location, and financing terms. Examples of this range include a lower value of $77/kW-year (PSCO 2011) and a higher value of $212/kW-year (CAISO 2012). Strategic placement of energy storage or enablement of demand response, such as in areas of transmission and distribution congestion or where it would be difficult to site traditional sources of peaking generation, offer additional value for these non-generation resources. The value of this flexibility could be difficult to quantify, though it might be reflected by a higher value of capacity.

Alternatively, ISO/RTO capacity markets prices can be used as points of comparison; however, these prices may not be directly applicable to the outcomes in this report because the present analysis neglects the role of scarcity pricing and strategic bidding on the part of ISO/RTO market participants. These mechanisms provide an additional measure of capital cost recovery through transactions in wholesale energy and ancillary service markets, which are influenced by many factors. As an example, market values for capacity in recent years have averaged $34/kW-year in the PJM RTO to $82/kW-year in the PJM's most congested areas (see Table 2-1). These market values are less than the full annualized cost of a new combustion turbine, partly because PJM has utilized demand response to provide capacity already.

Energy storage devices such as pumped storage hydropower and underground compressed air energy storage typically have many hours of energy storage at the rated capacity (e.g., 8 or more hours). The amount of capacity they provide to the system tends to be close to their peak output, after considering their expected forced and unforced outage rates. Energy storage devices with fewer hours of storage capability may result in reduced availability for the provision of energy during hours of peak demand (Sioshansi, Madaeni, and Denholm 2013). Alternatively, energy-limited storage devices (i.e., those capable of providing only operating reserves) may also contribute to system capacity planning by providing operating reserves during peak times (Kirby 2006). However, detailed calculations of loss of load probabilities and other reliability metrics are necessary for verification and are outside the scope of this study.

Comparing the capacity value of demand response resources to conventional generation is also challenging. Even if demand response resources are capable of responding during peak times, the underlying demand response programs may have limits with regards to response duration and the maximum number of calls. System planners would also need to be sure that the demand response resources would persist and be available over many years. Additionally, other issues such as local voltage and system frequency stability requirements would need to be met. These considerations would reduce their capacity value relative to generators or energy storage with no such restrictions. On the other hand, the ability of demand response to provide operating reserves during peak times can contribute to addressing some system capacity issues.

In the Western Interconnection model, the value of energy storage devices (co-optimized for energy and operating reserves) ranges from $33 to $41/kW-year in production cost savings, assuming an additional energy storage deployment level of 3.0 GW. Utilizing the simplistic assumption of a proxy resource discussed above, the estimated capacity value for this class of energy storage devices is comparable in magnitude to the operational value but can be several times greater. Using a similarly simplistic assumption, the capacity value for demand response resources could also be several times larger than its operation value. The total production cost savings of demand response deployed in the model (assuming 1.4% of average load is available for demand response) is about $110 million per year. The total estimated resource reduction in peak load is about 3.1 GW (based on the ability for demand response to shift energy use during the top 100 hours of each balancing authority area), which translates to roughly $240 million to $660 million per year in avoided capacity (based on the proxy combustion turbine costs).

4 Market and Regulatory Issues

In addition to technical issues, non-technical issues impact the potential deployment of demand response and energy storage resources through the adoption and application of market rules and regulations that may (1) limit or outright forbid these resources from providing certain bulk power system services, (2) increase the costs for providing these services, (3) decrease the revenue potential for providers, or (4) decrease the ability to capture profits. While engineering-based analyses and power system modeling can quantify economic and environmental benefits from utilizing these resources, market design and the regulatory process impact the cost-benefit assessment and introduce other barriers with differing perspectives from various stakeholders. Technical cost-benefit metrics may be only one of several important decision factors. Identifying and understanding the other non-technical issues that impact the development of demand response and energy storage resources can be important to satisfy cost-effective public policy and societal objectives.

4.1 Market and Regulatory Environments

The U.S. electric power sector is highly heterogeneous; the characteristics of wholesale and retail electricity markets, and the regulatory structure in which electricity providers operate, strongly impact the financial opportunities for providers of resources like generation, demand response, and energy storage. Within these environments, there are a number of influential entities, institutions, and resource options, as illustrated in Figure 4-1. At the wholesale level, some regions are served by an ISO/RTO balancing authority, and in others, wholesale trading occurs only through power exchanges and private bilateral transactions. At the retail level, electric utilities have experienced varying degrees of restructuring, leading to different ownership models for grid assets, governance structures for electricity providers, and relationships between electricity providers and retail customers (FERC 2012).

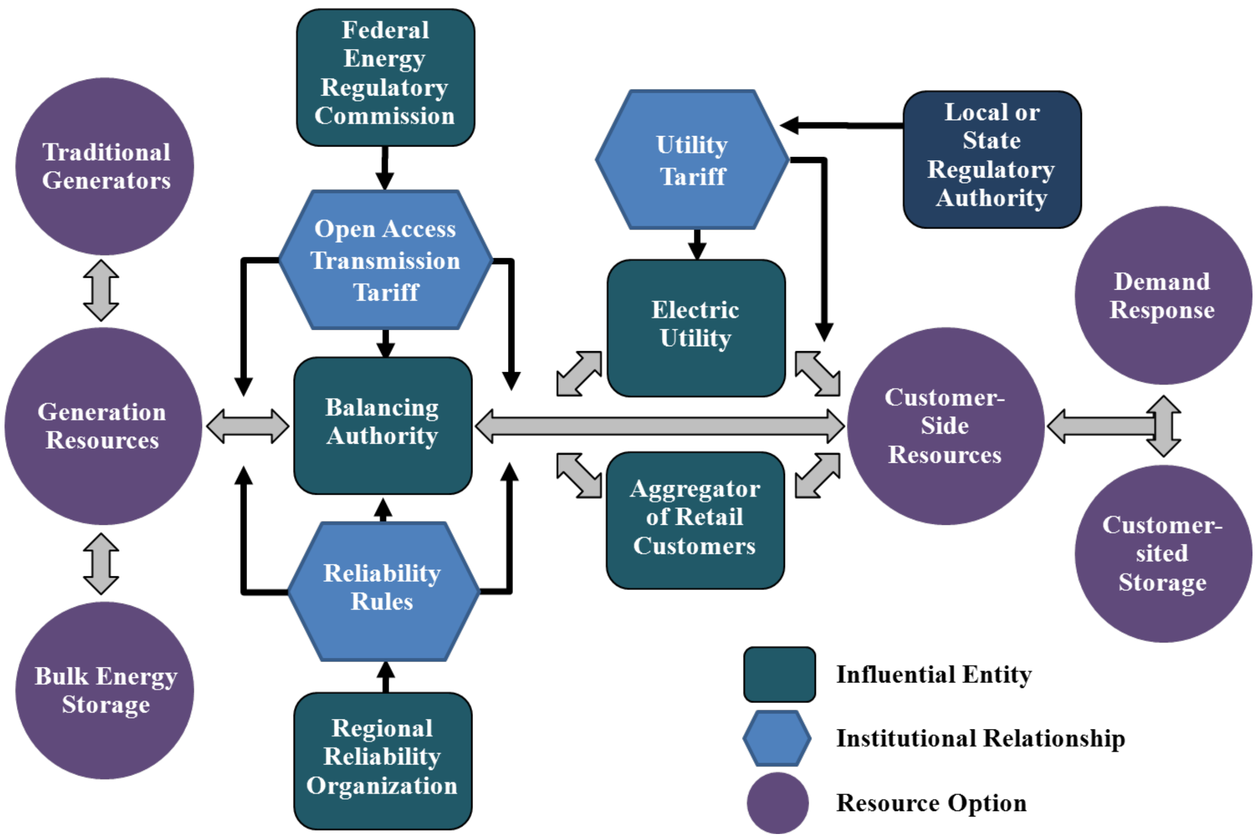

Figure 4-1. Relationships between influential entities, institutions, and resource options (adapted from Cappers, MacDonald, and Goldman [2013])

In each of the various environments, different types of grid investments (e.g., generation, demand response, and energy storage) have different regulatory treatments for cost recovery and return-on-investment (i.e., profit). These compensation mechanisms are either cost-based or market-based. Under cost-based mechanisms, resource providers earn a regulated rate of return on their capital investments, either as part of an electric utility's regulated asset base or as part of tariffs approved by the FERC. Under market-based mechanisms, the prices for services are determined through an organized wholesale electricity market or through a bilateral contract resulting from a competitive solicitation. Unlike cost-based mechanisms, market-based mechanisms (while subject to regulatory oversight) do not necessarily reflect the outcomes of an explicit cost of service analysis. Generally, grid resource investments have received cost-based or market-based treatments but not both simultaneously.

4.1.1 Retail Environments

At the retail level, electric utilities are differentiated primarily on how they are owned and governed. Publically-owned utilities are owned and governed by the municipalities that they serve. Cooperatively-owned utilities are owned by their member ratepayers and governed by boards of directors. Lastly, investor-owned utilities are owned by their shareholders and governed by corporate entities. For this report, we focus solely on this third type, investor-owned-utilities. Here, an elected or state-appointed group of utility commissioners regulate the investor-owned utility in a number of ways, including how it sets retail electricity rates and earns profit through capital investments (The Regulatory Assistance Project 2011).

Some states in the United States have undergone electricity restructuring, but the implementations are different. For instance, in New Jersey and Texas, investor-owned utilities have been required to divest their generation assets. In these states, retail electricity customers are able to choose among competitive providers. In New Jersey, incumbent electric utilities continue to serve customer load as default service providers and thereby have supply obligations. However, in Texas, the incumbent utilities are only transmission and distribution (i.e., wires-only) companies and do not serve customers directly. By contrast, in other states like Colorado and Wisconsin, electricity service providers are predominantly vertically integrated utilities that own generation, transmission, and distribution as well as serve retail customer load. In still other states, retail electricity restructuring is only partial. For instance, in California, some of the investor-owned utilities have divested some of their generation assets, with the remaining balance of their supply obligations met through bilateral contracts (i.e., power purchase agreements). Further, only a limited number of California electricity customers have the option of selecting a competitive provider.

Different types of electric utilities have different opportunities for investing in and profiting from supply resources. For instance, a vertically integrated utility can propose generation, demand response, or energy storage solutions to its state regulatory commission for cost-recovery and return-on-investment through retail electricity rates. On the other hand, a restructured utility acting as default service provider typically signs a full service supply contract (i.e., capacity, energy, and ancillary services) with third parties. As such, these utilities do not make detailed decisions regarding the composition of the underlying supply portfolios. These decisions are generally made by the power marketers and independent power producers that service those supply contracts. Furthermore, utilities that provide only distribution service have no responsibilities with regards to electricity supply procurement.

4.1.2 Wholesale Environments

Wholesale markets are differentiated primarily by whether or not they are centrally administered by an ISO/RTO balancing authority. In areas served by an ISO/RTO, there are transparent prices for many bulk power system services, and resource providers can participate in wholesale markets if they satisfy standardized requirements. This allows resources like generation, demand response, and energy storage to readily monetize the resource providers' provision of bulk power system services but does not provide future income certainty due to the lack of long-term contracts (which are separate from the ISO/RTO market). Outside ISO/RTO balancing authority areas, wholesale transactions occur primarily between individual buyers and sellers. This transaction limits price transparency but provides contracted income certainty, typically for multiple years. While energy pricing indices are available at power exchanges, these are limited to a few types of energy products and do not include ancillary services. In contrast to ISO/RTO regions, the value of bulk power system services is largely derived through cost savings internal to vertically integrated utilities (i.e., the utilities' avoided cost for the alternative supply option), rather than a market price.

In both ISO/RTO and non-ISO/RTO regions, most electricity supply is planned in advance to ensure adequacy—days, months, and years ahead of when needed. Closer to real time, balancing authorities oversee unit commitment and economic dispatch processes to ensure that electricity flows within the physical limits of the power system (as discussed in Section 3) and that any energy imbalances (from issues like load forecasting errors, generators not following schedules, and contingency events) can be handled within acceptable reliability standards. While the basic control algorithms are similar between ISO/RTO and non-ISO/RTO balancing authority areas, established processes have different implications in terms of wholesale transactions and financial settlements for service providers.

In ISO/RTO balancing authority areas, system operators run organized wholesale electricity markets, bringing together multiple buyers and sellers, and establishing hourly (or shorter) locational based market clearing prices for energy and ancillary services. Whereas an ISO/RTO balancing authority is an independent entity that solely operates electricity markets, a non-ISO/RTO balancing authority is typically a vertically integrated utility. In these non-ISO/RTO regions, electric utilities schedule resources to balance supply and demand, including their provision of ancillary services. Any residual energy imbalances or other deficiencies in ancillary service commitments are made up by assets operated by the electric utility serving as the balancing authority. Some ISO/RTO markets co-optimize the provision of energy and ancillary services from the same resources but vertically integrated utilities outside these regions may rely on dedicated generation built to provide specific services (WWSIS-2 Technical Review Committee 2012).

In addition to energy and ancillary services, balancing authorities also administer capacity requirements to ensure there will be sufficient resources available, at the right locations, to meet forecasted electricity demand plus a reserve margin (see Figure 2-1 in Section 2.1). Some ISO/RTOs have established capacity markets to allow market participants to sell any available capacity in excess of their needs. These markets can also help address concerns in regions where energy and ancillary service markets may not provide sufficient incentives for participants to develop and maintain resources to meet system adequacy within their planning horizons. In non-ISO/RTO balancing authority areas and those ISO/RTO areas without capacity markets, capacity transactions occur bilaterally.

4.2 Barriers to Deployment

A number of studies have examined various deployment barriers to the utilization of specific types of grid resources (FERC 2009; Cappers, MacDonald, and Goldman 2013;Sioshansi, Denholm, and Jenkin 2012;Bhatnagar et al. 2013), though these studies have taken only a limited look across these multiple types of resources to assess barriers in an integrated fashion. Many of the barriers to demand response and energy storage at the wholesale level are similar and could be addressed jointly. However, some of the barriers at the retail level are distinct and require different approaches. A holistic and comprehensive framework may be useful for understanding how generation and non-generation resources, like demand response and energy storage, are able to compete side-by-side as providers of bulk power system services.

There are a broader range of issues that can adversely affect investment decisions in the deployment of demand response and energy storage resources; however, we attempt to confine the discussion in the following subsections to only those issues that (1) relate specifically to bulk power system services and (2) are distinct to demand response and energy storage compared with generation resources. Some issues that fall outside the scope of these two categories are highlighted below.

One issue area is that demand response and energy storage can have a number of applications at the customer-level and at the distribution-level, separate from the bulk power system. For instance, energy storage can provide back-up power to increase the reliability and resiliency of electricity supply to the customer. Additionally, demand response and energy storage could be used to help alleviate congestion and losses in the distribution system. These applications, in addition to the provision of bulk power system services, can improve the overall value proposition for investment (Akhil et al. 2015). However, assessing these values streams are outside the scope of this study.

Another issue area is that some barriers are common to both generation and non-generation resources. For instance, current markets do not cover all services needed by the bulk power system. As an example, there is no market-based compensation mechanism for providers of primary frequency response (Ela et al. 2013). In fact, there may be a disincentive for generators to provide the service (Eto 2010). Generation, demand response, and energy storage could provide this service (i.e., virtual inertia). However, the inability to capture revenue from providing this service by itself is not necessarily a barrier that is discriminatory to non-generation resources.

The following discussion focuses only on the barriers that may be discriminatory to demand response and energy storage. These barriers are primarily a result of market rules and regulations designed around the specific characteristics of resources that have historically provided bulk power system services (i.e., generation). A level playing field would take a systems perspective instead, focusing on the essential physical characteristics and qualities required for the provision of bulk power system services.

4.2.1 Market and Regulatory Barriers

Market and regulatory barriers to demand response and energy storage participation in bulk power system services fall into several categories. To start, regional reliability councils and balancing authorities establish bulk power system product definitions that may explicitly include or exclude certain classes of resources, or exclude them implicitly by defining services that only certain classes can provide. Even if demand response and energy storage resources are eligible to provide bulk power system services, other issues can impact the value proposition of using these resources by affecting the costs and revenue for providing services. Furthermore, smaller providers face an additional barrier because they require an intermediary to bring their resources to the wholesale market. The intermediary, either the retail electricity provider or a third party aggregator, may not have a profit motive to do so, preventing these smaller providers from participating in the provision of bulk power system services.

A number of actions by the NERC and FERC have sought to eliminate many of the barriers facing demand response and energy storage in market and regulatory environments. For instance, a number of decisions have targeted issues associated with eligibility (FERC 2008; FERC 2011). Yet, there are some specific implementations of bulk power system service definitions that remain as barriers. For instance, the WECC, the regional reliability council for the Western Interconnection, has not yet adopted NERC standards and does not allow demand response and some types of energy storage to provide services classified as spinning (see Section 2.2). Also, ISO New England does not allow demand response resources to provide frequency regulation, though they have administered a pilot program to investigate their performance.

Other FERC actions have addressed issues associated with revenue capture. Many energy storage technologies can ramp substantially faster and deliver more accurate response than thermal power plants (Makarov et al. 2008), and demand response resources can ramp nearly instantly (Kirby et al. 2008), depending on communications and control equipment (Kiliccote et al. 2012). By contrast, generators have mechanical and thermal inertia that limits their ability to ramp and quickly respond. In response to FERC Order 755, several ISO/RTO balancing authorities have implemented performance-based payments (FERC 2011). Following this order, FERC Order 784 extends the treatment of faster resources to non-ISO/RTO balancing authority areas and also alleviates barriers to third-party provision of operating reserves in non-ISO/RTO areas (FERC 2013).

While many barriers are being addressed by policy makers, the following subsections discuss several remaining issues that impact the ability of demand response and energy storage to provide bulk power system services alongside conventional generation.

4.2.2 Issues that Impact Costs

Regional reliability councils and balancing authorities define the attributes of performance and the required enabling infrastructure necessary for participation in wholesale markets. While these issues do not directly prevent demand response and energy storage from providing bulk power system services, they can strongly impact the costs associated with their provision and thereby indirectly limit participation.

Demand response and some energy storage technologies have inherent shortcomings with regard to response duration (i.e., energy-limited) that impact costs. For instance, a short demand response event for many types of end-uses may not be noticeable, but a longer multi-hour event may lead to significant disruption in service quality. Achieving the longer response, while meeting service demands, may require acquisition of more customers and enablement of substantially more end-use loads to spread the impact of interruptions. Similarly, some energy storage technologies such as batteries, flywheels, and above ground compressed air energy storage have capital costs that are roughly proportional to response duration; devices with longer response durations will cost more.

Requirements around response duration typically have little impact on generators but can substantially increase costs for demand response and energy storage. As an example, NERC reliability rules only dictate that contingency reserves are restored within 105 minutes, though various markets require resources to be capable of sustaining the response from 30 minutes to up to 2 hours. Depending on region, contingency reserves are deployed as often as every couple of days or as seldom as every couple of weeks; the average duration is 10 minutes, but reserves rarely last longer than 30 minutes (Kirby et al. 2008). While the essential qualities of the service may not change, the 2-hour requirement can cost four times more than a 30-minute requirement for demand response and energy storage resource providers.

Similarly, energy-limited resources are capable of providing frequency regulation since the service is, in principle, energy-neutral over a set time interval. However, demand response and energy storage resources need sufficient ability to sustain the response for the maximum energy utilization risk in a single scheduling interval (e.g., 5 minutes to 1 hour). All things being equal, larger scheduling intervals lead to larger energy utilizations and larger associated costs for demand response and energy storage to provide this service, which is not the case for conventional generators.

For the case of demand response, individually smaller resource providers face high per-unit costs if balancing authorities require the same monitoring and communications equipment and connection to a dedicated communication network typically used with large generators. These enabling infrastructure requirements ensure that balancing authorities can monitor and verify, in real time, a resource's compliance with dispatch instructions. However, utilization of large aggregations of small resources may have different reliability concerns compared with those of a single large generator. The aggregated response of the various loads and not that of any one individual end-use may be more important to the power system (Kirby 2006). Some ISO/RTOs have relaxed these requirements for demand response resources. For instance, PJM allows demand response providing contingency reserve (i.e., 10-minute synchronized reserve) to submit, within two business days of the event day, historical metering data (at a 1-minute scan rate) for measurement and verification (PJM 2013). Additionally, ERCOT has implemented rules allowing aggregation of small providers without expensive real-time monitoring (ERCOT 2014).

4.2.3 Issues that Impact Revenue Capture

In addition to cost issues associated with participation in bulk power system services, potential demand response and energy storage resource providers may also have varying abilities to capture revenue. In this context, revenue may stem from market-based sales of bulk power systems services, or they may stem from avoided cost savings internal to an electric utility. Thus, within this broad definition, revenue reflects a value stream that may be monetized in order to recoup costs of implementing demand response or energy storage resources as well as to provide profits to resource providers.

In some cases, ISO/RTO market rules can impact revenue for demand response and energy storage providers. For instance, some ISO/RTO balancing authorities require resources to bid jointly into energy and operating reserve markets. Demand response and energy storage resources that are capable of meeting performance attributes for operating reserves may not be capable of sustaining the response for one or more hours as an energy resource. In order to participate, they may need to hedge themselves through some sort of financial instrument (with a concurrent economic loss). If they are unable to take this risk, they will need to withdraw from the operating reserves market (Todd 2008). This issue has been addressed in the PJM RTO, New York ISO and Mid-Continent ISO, specifically for limited energy storage resources (PJM 2013; NYISO 2013; MISO 2011), and is under consideration in the California ISO for both limited energy storage and dispatchable demand response (CAISO 2012).

Furthermore, demand response and energy storage resources may have differing abilities to provide symmetric response for frequency regulation (i.e., the same amount of up and down movement), limiting revenue to the amount based on the more limited ability. This issue has been alleviated in some balancing authority areas. For instance, the New York ISO has modified its automatic generation control signal to dispatch limited energy storage resources with consideration of the state of charge (i.e., decreased capability in one direction when reaching a fully charged or fully discharged state) (NYISO 2013). Similarly, the Mid-Continent ISO has adjusted its 5-minute real-time energy dispatch to maximize regulation capacity for energy storage resources (Chen et al. 2011). Other balancing authorities, including the California ISO and ERCOT, have separate up and down regulation services, thereby allowing resources to bid asymmetrically.

Even if market rules enable revenue capture for demand response and energy storage, the prices for bulk power system services can be volatile, leading to increased risk for potential providers. This is particularly challenging for providers that can only offer operating reserves. First, there are no organized long term contracting mechanisms for operating reserves to allow providers a means to manage this risk. Second, entry of demand response and energy storage resources in operating reserve markets can collapse market clearing prices for those services. Current ISO/RTO electricity markets are designed such that generators are mostly profit-neutral to the provision of energy or operating reserves. In other words, the market guarantees these resources recover lost opportunity costs when they reserve capacity for operating reserves rather than generate electricity (i.e., provide energy). Even if generators bid zero costs for operating reserves, they would still receive their lost opportunity cost if they are economic as energy resources. This is not necessarily the case for many demand response and energy storage resources seeking to provide only operating reserves. For these resources, the lost opportunity cost component could be taken as zero by unit commitment and economic dispatch processes (note that this is different from the lost opportunity cost associated with deferred electricity usage by customers).

Demand response and energy storage resources can also impact wholesale energy market prices when shifting energy. All supply resources that successfully bid into energy markets will tend to suppress the market clearing price and reduce revenue for all providers, in accordance with general economics. However, energy-shifting resources face the dual problem of elevating prices during charging and depressing prices during discharging, thereby reducing the value of energy arbitrage. This issue is particularly acute for pumped storage hydropower plants. For this technology, project economics rely on economies of scale. These systems cannot be sized optimally when considering the elasticity of energy prices in a specific location or at specific times. While historical energy price differentials between peak and off-peak times may provide sufficient revenue for investment in these plants, their entrance into the market can substantially erode that revenue opportunity.

4.2.3.1 Issues Specific to Demand Response Aggregators

Large electricity customers seeking to act as demand response providers can register directly with the ISO/RTO or negotiate individual contracts with their electricity providers. However, smaller electricity customers cannot meet minimum response size requirements for ISO/RTO market participation (e.g., between 100 kW and 1 MW) nor obtain individualized contracts. These smaller customers must participate in a demand response program offered by their retail electricity provider or go through a third-party aggregator to facilitate entry into ISO/RTO markets. Even if there are sufficient revenue streams for development of small demand response resources, these intermediaries must have a profit motive to provide the aggregation services. Otherwise, it falls solely on state regulatory commissions to intervene and require electric utilities under their authority to offer such programs.

Development of small demand response resources can conflict with current investor-owned utility business models. For vertically integrated utilities, demand response can provide operational value through savings in fuel and operations and maintenance costs and reducing the need to purchase services from the wholesale market, particularly at times of high market prices. However, current regulations prevent an investor-owned utility from retaining those savings as profit. Similarly, if demand response resources enable the electric utility to increase off-system sales of bulk power system services, they may be able to keep only a fraction of the revenues as profits, at best. For instance, the state of Wisconsin does not permit the utility to retain any of these revenues (Wisconsin Administrative Code 2012), while the Public Service of Colorado allow utilities to retain up to 20% of net proceeds (Colorado PUC 2009). Lastly, there may be a number of soft costs associated with developing demand response resources (i.e., customer acquisition and backend services), for which an investor-owned utility may receive cost recovery but not a rate of return. Many of these business model issues are similar to those for energy efficiency investments (Goldman et al. 2010).

For other potential demand response providers that are not vertically integrated utilities, issues are different. Wires-only utilities (i.e., those that do not serve retail customers) have no financial incentives to offer demand response programs and are driven-only by regulatory intervention that compels them to do so. Competitive retail providers may not see value in investing in the enablement of their customers to provide demand response. The short-term contracts for retail electricity service may not be long enough to achieve sufficient return-on-investment thresholds for enabling customers to provide demand response. Alternatively, a third-party who is not the retail electricity provider can serve as the aggregator of small demand response resources. However, in some cases, state regulatory commissions, like in Wisconsin, have prohibited third-party aggregators. In others, regulators have required utilities to outsource demand response programs to a third-party, like in Colorado. In still others, like Texas and New Jersey, third party aggregators are allowed to operate freely and sell services in their respective ISO/RTO wholesale markets (Cappers, MacDonald, and Goldman 2013).

5 Conclusion

This study assesses the potential for demand response and energy storage resources to provide bulk power system services and examines the market rules and regulations that impact the use of these resources. In this report, we provide findings from simulations of a Western Interconnection model and a smaller Colorado test model. The simulations explore scenarios with deployment of new demand response and energy storage resources under both low and high levels of wind and solar power. The market and regulatory assessment takes a national view. It looks across different environments for demand response and energy storage deployment with specific examples from select wholesale and retail markets, representative of the diversity of the electric utility sector in the United States.

Overall, these efforts yield a number of key findings.

- A significant fraction of operational value attributable to demand response and energy storage resources are the avoided costs associated with generator startups/shutdowns and reduced costs associated with generators modulating output while providing frequency regulation. These costs are nominal when looking at total production costs but represent a significant fraction of costs for operating reserves.
- Due to the limited temporal flexibility of demand response resources, the provision of operating reserves has more market value than energy shifting services, assuming prices are based on marginal costs of production. However, the availability of these resources to provide energy shifting services helps to optimize the operation of the broader system.
- Energy storage provides greater value in scenarios with higher renewable penetration due to the increased need for operating reserves and the greater opportunities for energy arbitrage through the storage of low-cost, off-peak electricity. Additionally, co-optimization of energy and operating reserves results in greater value than energy shifting services alone.
- Market structures can limit the ability of any new entrant, including demand response and energy storage, to be compensated commensurate with the savings they create. Existing markets do not include generator startup costs in price formulation, so they may not necessarily compensate demand response or energy storage for reducing these costs along with other value streams.
- Marginal costs for operating reserves include lost opportunity costs from generators (forgone profit of selling energy). However, the lost opportunity cost component is defined as zero for demand response and energy storage resources providing only operating reserves (forgone profit of selling energy is zero). Large penetration of these resources can saturate the market for operating reserves and drive down market clearing prices.
- While capacity value of demand response and energy storage was not studied in detail, a simple calculation based on the assumption of a proxy generation resource suggests that its capacity value could be several times larger than its operational value. However, realizing this value may require resources to provide many hours of response duration (e.g., 6–10 hours), generally increasing the cost of providing these services.
- While there are multiple challenges to deploying large customer demand response and energy-limited storage resources in wholesale markets, smaller customer resources that seek to provide bulk power system services face additional barriers. First, they may require an aggregator that might not see a business case to provide those services. Second, communications and control requirements imposed on individual large providers can be cost prohibitive if applied equally to smaller providers. However, these requirements could be modified and technical hurdles overcome to reduce implementation costs without compromising system reliability.

6 Future Work

This study represents an initial effort toward a broad set of research goals to quantify the potential resource availability of demand response providing bulk power systems services and the operational value of demand response and energy storage under varying levels of wind and solar generation. This effort further seeks to understand potential market and regulatory barriers that inhibit generation and non-generation resources from competing side-by-side as service providers. As summarized in Section 5, the present work provides a number of insights toward these broader goals.

There are a number of areas for future work.

- While this study examines the operational value of demand response and energy storage, it does not look for optimal deployments of demand response and energy storage or the economic viability of developing those resources. For instance, additional work is necessary to quantify the costs associated with different types of demand response resources providing different types of bulk power system services. This study simply implements, in the production cost model, the estimated amount of select demand response resources. In an optimal solution, more valuable resources will have greater incentives to offer their capabilities to the power system and thereby may have higher availabilities and different technical attributes.
- The unit commitment and economic dispatch of the demand response resources, through the production cost modeling, has not been tested against detailed examinations of the resources' capabilities. Because we are modeling the aggregate response capabilities of many individual providers, it is not clear how many customers need to be enrolled in order to meet aggregated response utilization and whether the initial assumptions on availability make efficient use of those enrolled.
- This report focuses exclusively on bulk power system services including energy transactions and operating reserves (i.e., frequency regulation, contingency reserve, and ramping reserve). There are a number of other applications for demand response and energy storage such as frequency response and voltage support. Further, there are a number of applications on the customer side and services on distribution systems, including relieving congestion on both the distribution network and local transmission system, which are omitted in this assessment. A comprehensive evaluation of grid services and their provision by demand response and energy storage is needed to quantify their full value.
- There are many potential demand response resources omitted in this assessment. First, projecting the changing composition of end-use loads is outside the scope of this report. For instance electric vehicles could be a significant demand for future electricity production. Second, there are many smaller loads that could be aggregated (e.g., appliances, miscellaneous building plug loads, and pool pumps) for flexible response, but their individual characteristics vary significantly, complicating their analyses. Furthermore, demand response strategies may be coupled with on-site generation like combined heat and power systems to provide bulk power system services. Lastly, the demand response resource assessment for industrial manufacturing process end-use loads was not completed in time to conduct the production cost modeling.

- This study provides an initial estimate of available demand response resources for the Western Interconnection for the provision of operating reserves. A complete national inventory of potential demand response resources (residential, commercial, industrial, agricultural) by location for energy and each of the ancillary services would be very useful to aid regulators, utilities, and developers of grid resources. The inventory should include the cost of enabling response.
- In addition to examination of the Western Interconnection, this work could be implemented for other U.S. grids. For instance, expansion of the analysis to the Eastern Interconnection could leverage modeled scenarios developed through Eastern Renewable Generation Integration Study (Bloom et al. 2015).
- Methods to increase cost-effective demand response participation could be investigated to help reduce the dramatic order of magnitude difference between demand response technical potential and the estimated resource.
- There are several potential extensions of the grid simulations. Results presented here are from modeling day-ahead unit commitment and economic dispatch processes. Future work is necessary to investigate operations closer to real-time and sub-hourly scheduling. This will enable greater insights into forecast errors and the utilization of the ramping reserve product. Additionally, resources beyond demand response and energy storage, such as combustion turbines and smart inverters, can be simulated to assess their capabilities with a similar framework.
- There are numerous questions regarding wholesale market design that are not included in the analysis. For instance, we have not investigated alternative market designs, like performance based rates for frequency regulation per FERC Orders 755 and 784. Additionally, we have not examined in depth the role of scarcity pricing and the intersection with capacity markets.
- The results provided in this report represent only a few example scenarios; as such, the results cannot be applied universally. A variety of generator compositions could occur under increasing renewable penetration scenarios. Specifically, increased penetration of variable renewable generation will result in decreased capacity factors and market revenues of existing generators. This could incentivize additional plant retirements, changing the overall system composition. More comprehensive evaluation of different generation mixes, including combinations of wind and solar penetration levels, will provide additional insights into the evolving value of demand response and energy storage.
- The values of demand response and energy storage are driven greatly by the cost of providing operating reserves, which itself depends greatly on many assumptions regarding the operational flexibility of the generation fleet, in particular the assumed ramp rates and the fraction of fleet available to provide operating reserves. In addition, a large fraction of the marginal cost of frequency regulation in these simulations is derived from the assumed cost of generators operating at a non-steady state while providing regulation reserves. Because performance data related to an individual generator's ability to provide reserve are not widely available, reproducing the cost of operating reserves in a production cost model involves significant uncertainty without better data.
- Production cost models, like the PLEXOS model used in this study, can calculate the cost of holding operating reserves, but they do not simulate explicitly frequency regulation dispatch and contingency events. Simulation of shorter term grid dynamics is an important area of future

research, requiring models capable of simulating detailed generator, demand response, and energy storage response. This may become particularly relevant if variable generation like wind and solar power becomes a large contributor to electricity supply.

- The optimal sizing and location of demand response and energy storage were not considered in this study. Additionally, generator retirements were not considered in the scenarios with higher renewable penetration. Expanding the study to factor in these business operating principles or with scenarios that have similar levels of reliability (e.g., without excess capacity) can produce more realistic and meaningful results.

References

AESO. 2013. "Energy Storage Initiative Issue Identification."

Akhil, A., G. Huff, A. B. Currier, B. C. Kaun, D. M. Rastler, S. B. Chen, A. L. Cotter, D. T. Bradshaw, and W. D. Gauntlett. 2013. *DOE/EPRI 2013 Electricity Storage Handbook in Collaboration with NRECA.* SAND2013-5131.

Akhil, A., G. Huff, A. B. Currier, B. C. Kaun, D. M. Rastler, S. B. Chen, A. L. Cotter, D. T. Bradshaw, and W. D. Gauntlett. 2015. *DOE/EPRI Electricity Storage Handbook in Collaboration with NRECA.* SAND2015-1002.

Bhatnagar, D., A. B. Currier, J. Hernandez, O. Ma, and B. Kirby. 2013. *Market and Policy Barriers for Energy Storage Deployment.* SAND2013-4747.

Bloom, A., A. Townsend, D. Palchak, J. King, E. Ibanez, C. Barrows, M. Hummon, and C. Draxl. 2015. *Eastern Renewable Generation Integration Study.* NREL/TP-6A20-64472.

Booth, W. C., and J. L. Rose. 1995. *Using Hourly System Lambda to Gauge Bulk-Power Prices.* Fortnightly Magazine.

CAISO. 2012. "2011 Annual Report on Market Issues and Performance."

CAISO. 2012. "California ISO Non-Generator Resource Energy Management Project Implementation Plan - Version 2.1."

CAISO. 2011. *Summary of Preliminary Results of 33% Renewable Integration Study.* CPUC LTPP Docket No. R.10-05-006.

Callaway, D. S., and I. A. Hiskens. 2011. "Achieving Controllability of Electric Loads." *Proceedings of the IEEE* 99 (1).

Callaway, D. 2009. "Tapping the Energy Storage Potential in Electric Loads to Deliver Load Following and Regulation, with Application to Wind Energy." *Energy Conversion and Management* 50 (5).

Cappers, P. 2012. *insert text.* LBNL.

Cappers, P., C. Goldman, and D. Kathan. 2010. "Demand Response in U.S. Electricity Markets: Empirical Evidence." *Energy* 35 (4): 1526-1535.

Cappers, P., J. S. MacDonald, and C. Goldman. 2013. *Market and Policy Barriers for Demand Response Providing Ancillary Services in U.S. Markets.* LBNL-6155E.

Cappers, P., Mills, A., Goldman, C., Wiser, R. and Eto, J. H. 2011. *Mass market demand response and variable generation integration issues: A scoping study.* LBNL-5063E.

Chen, Y., M. Keyser, M. H. Tackett, and X. Ma. 2011. "Incorporating Short-Term Stored Energy Resource into Midwest ISO Energy and Ancillary Service Market." *Transactions on Power Systems* (IEEE) 26 (2): 829-838.

Chuang, A. 2009. "Demand-side integration for customer choice through variable service subscription." *IEEE Power & Energy Society General Meeting, 2009.*

Colorado PUC. 2009. "Decision No. C09-1446: Order Addressing Phase I and ECA Issues [Docket No. 09AL-299E]."

Denholm, P., J. Jorgenson, M. Hummon, T. Jenkin, B. Kirby, O. Ma, M. O'Malley, and D. Palchak. 2013. *The Value of Energy Storage for Grid Applications.* NREL/TP-6A20-58465.

Ela, E., and M. O'Malley. 2012. "Studying the Variability and Uncertainty Impacts of Variable Generation at Multiple Timescales." *IEEE Transactions on Power Systems* 27: 1324-1333.

Ela, E., M. Milligan, and B. Kirby. 2011. *Operating Reserves and Variable Generation.* NREL/TP-5500-51978.

Ela, E., M. Milligan, and M. O'Malley. 2011. "A Flexible Power System Operations Simulation Model for Assessing Wind Integration." *Power and Energy Society General Meeting.* NREL/CP-5500-50641.

Ela, E., V. Gevorgian, A. Tuohy, M. Milligan, and M. O'Malley. 2013. "Market Designs for the Primary Frequency Response Ancillary Service-Part I: Motivation and Design." *Transactions on Power Systems* (IEEE) PP (99): 1-11.

Elzinga, D., J. Dillon, M. O'Malley, and J. Lampreia. 2012. "The Role of Electricity Storage in Providing Electricity System Flexibility." In *Electricity in a Climate Constrained World*. IEA.

Enbala. 2011. "Presentation to DOE Workshop October 25-26, 2011." Presented at 2011 Department of Energy Load Participation in Ancillary Services Workshop.

Energy Exemplar. 2013. *Plexos Integrated Energy Model.* Accessed November 2013. www.energyexemplar.com.

Energy Information Administration. 2010. *Status of Electricity Restructuring by State.* Accessed January 2014. www.eia.gov/electricity/policies/restructuring/index.html.

EnerNex. 2011. *Eastern Wind Integration and Transmission Study.* NREL/SR-5500-47078.

EPRI. 2013. "Quantifying the Value of Hydropower in the Electric Grid."

ERCOT. 2014. "Requirements for Aggregate Load Resource Participation in the ERCOT Market."

Eto, J. 2010. *Use of Frequency Response Metrics to Assess the Planning and Operating Requirements for Reliable Integration of Variable Renewable Generation.* LBNL-4142E.

FERC. 2009. *A National Assessment of Demand Response Potential: Staff Report.* FERC.

FERC. 2012. *Energy Primer A Handbook of Energy Market Basics.* FERC.

FERC. 2010. "National Action Plan on Demand Response."

FERC. 2008. "Order 719, Wholesale Competition in Regions with Organized Electric Markets."

FERC. 2011. "Order 745, Demand Response Compensation in Organized Wholesale Energy Markets."

FERC. 2011. "Order 755, Frequency Compensation in Organized Wholesale Power Markets."

FERC. 2012. "Order 764, Integration of Variable Renewable Energy Resources."

FERC. 2013. "Order 784, Third-Party Provision of Ancillary Services; Accounting and Financial Reporting for New Electric Storage Technologies."

FERC. 2006. "Order 888, Promoting Wholesale Competition through Open Access Non-Discriminatory Transmission Services by Public Utilities; Recovery of Stranded Costs by Public Utilities and Transmitting Utilities."

FERC. 2007. "Order 890, Preventing Undue Discrimination and Perference in Transmission Service."

—. 2013. *Pro forma Open Access Transmission Tariff.* Accessed November 2013. www.ferc.gov/industries/electric/indus-act/oatt-reform.asp.

Ferreira, S. R., Rose, D. M., Schoenwald, D. A., Bray, K., Conover, D., Kintner-Meyer, M., Viswanathan, V. 2013. *Protocol for Uniformly Measuring and Expressing the Performance of Energy Storage Systems.* SAND2013-7084.

GE Energy. 2010. *Western Wind and Solar Integration Study.* NREL/SR-550-47434.

Goldman, C., M. Reid, R. Levy, and A. Silverstein. 2010. *Coordination of Energy Efficiency and Demand Response.* LBNL-3044E.

Goli, S., A. McKane, and D. Olsen. 2011. "Demand Response Opportunities in Industrial Refrigerated Warehouses in California." *ACEEE Summer Study on Energy Efficiency in Industry.* LBNL-4837E.

Griffith, M. 2008. "Capacity Markets Demystified." *Fortnightly Magazine* (Public Utilities Fortnightly).

Hirst, E., and B. Kirby. 1996. *Electric-Power Ancillary Services.* ORNL/CON-426.

Hogan, W. 1998. *Competitive Electricity Market Design: A Wholesale Primer.* Harvard University.

Hogan, W. 2012. *Electricity Scarcity Pricing Through Operating Reserves.* Whitepaper.

Hummon, M., P. Denholm, J. Jorgenson, D. Palchak, B. Kirby, and O. Ma. 2013. *Fundamental Drivers of the Cost and Price of Operating Reserves.* NREL/TP-6A20-58491.

Hummon, M., P. Denholm, J. Jorgenson, D. Palchak, O. Ma, D. Olsen, M. D. Sohn, et al. 2013. *Grid Integration of Aggregated Demand Response, Part 2: Modeling Demand Response in a Production Cost Model.* NREL/TP-6A20-68492.

Ibanez, E., G. Brinkman, M. Hummon, and D. Lew. 2012. *Solar Reserve Methodology for Renewable Energy Integration Studies Based on Sub-Hourly Variability Analysis: Preprint.* NREL Report No. CP-5500-56169.

IEA. 2013. "State of the Art in Renewable Integration Studies."

Isemonger, A. G. 2009. "The Evolving Design of RTO Ancillary Service Markets." *Enegy Policy* 37: 150-157.

ISO/RTO Council. n.d. Accessed December 30, 2013. http://www.isorto.org/site/c.jhKQIZPBImE/b.2604471/k.B14E/Map.htm.

Kiliccote, S., P. Price, M. A. Piette, G. C. Bell, S. Pierson, E. Koch, J. Carnam, H. Pedro, J. Hernandez, and A. K. Chiu. 2012. *Field Testing of Automated Demand Response for Integration of Renewable Resources in California's Ancillary Services Market for Regulation Products.* LBNL-5556E.

Kirby, B., J. Kueck, T. Laughner, and K. Morris. 2008. "Spinning Reserve from Hotel Load Response." *The Electricity Journal* 21 (10): 59-66.

Kirby, Brendan. 2006. *Demand Response for Power System Reliability: FAQ.* ORNL/TM-2006/565.

Kumar, N., P. Besuner, S. Lefton, and D. Agan. 2012. *Power Plant Cycling Costs.* NREL/SR-5500-55433.

Lew, D., G. Brinkman, E. Ibanez, A. Florita, M. Heaney, B.-M. Hodge, M. Hummon, et al. 2013. *The Western Wind and Solar Integration Study Phase 2.* NREL/TP-5500-55588.

Lew, D., G. Brinkman, N. Kumar, P. Besuner, D. Agan, and S. Lefton. 2012. "Impacts of Wind and Solar on Fossil-Fueled Generators." *Power and Energy Society General Meeting.* IEEE.

Ma, O., N. Alkadi, P. Capper, P. Denholm, J. Dudley, S. Goli, M. Hummon, et al. 2013. "Demand Response for Ancillary Services." *IEEE Transactions on Smart Grid* PP (99).

MacDonald, J., P. Cappers, D. Callaway, and S. Kiliccote. 2012. "Demand Response Providing Ancillary Services: A Comparison of Opportunities and Challenges in the US Wholesale Markets." *Grid-InterOp 2012.* LBNL.

Makarov, Y. V., J. Ma, S. Lu, and T. B. Nguyen. 2008. *Assessing the Value of Regulation Resources based on their Time Response Characteristics.* PNNL-17632.

Marks, G., E. Wilcox, D. Olsen, and S. Goli. 2013. *Opportunities for Demand Response in California Agricultural Irrigation: A Scoping Study.* LBNL-6108E.

Milligan, M., and B. Kirby. 2010. "Utilizing Load Response for Wind and Solar Integration and Power System Reliability." *WindPower 2010.* NREL/CP-550-48247.

Milligan, M., E. Ela, B.-M. Hodge, B. Kirby, and D. Lew. 2011. "Cost-Causation and Integration Cost Analysis for Variable Generation." *The Electricity Journal* 24 (9): 51-63.

MISO. 2013. "Business Practices Manual Energy and Operating Reserve Markets."

MISO. 2011. "Business Practices Manual: Energy and Operating Reserve Markets."

Navid, N., and G. Rosenwald. 2013. *Ramp Capability Product Design for MISO Markets.* Whitepaper.

NERC. 2004. "Frequency Response Standard Whitepaper."

—. 2013. *Planning Reserve Margin.* http://www.nerc.com/pa/RAPA/ri/Pages/PlanningReserveMargin.aspx.

—. 2011. *WECC Standard BAL-002-1 - Contingency Reserves.* Accessed November 2013. www.nerc.com/files/BAL-002-WECC-1.pdf.

NYISO. 2013. "Ancillary Services Manual."

NYISO. 2013. "Ancillary Services Manual."

OATI, Inc. 2013. *Open Access Same-time Information System.* Accessed November 2013. www.oati.oasis.com.

Olsen, D., M. D. Sohn, N. Matson, J. Dudley, S. Goli, S. Kiliccote, M. Hummon, et al. 2013. *Grid Integration of Aggregated Demand Response Part 1: Load Availability Profiles and Constraints for the Western Interconnection.* LBNL-6417E.

Olsen, D., S. Goli, D. Faulkner, and A. McKane. 2012. *Opportunities for Automated Demand Response in Wastewater Treatment Facilities in California - Southeast Water Pollution Control Plant Case Study.* LBNL-6056E.

Piette, M. A., S. Kiliccote, and J. H. Dudley. 2012. *Field Demonstration of Automated Demand Response for Both Winter and Summer Events in Large Buildings in the Pacific Northwest.* LBNL-6216E.

PJM. 2013. "Manual 12: Balancing Operations."

PJM. 2012. "Manual 15: Cost Development Guidelines Rev. 20."

PJM. 2013. "PJM Manual 11: Energy and Ancillary Serices Market Operations Rev. 62."

Potomac Economics. 2014. *State of the Markets.* Accessed January 2014. http://www.potomaceconomics.com.

Potter, C. W., D. Lew, J. McCaa, S. Cheng, S. Elchelberger, and E. Grimit. 2008. "Creating the Dataset for the Western Wind and Solar Integration Study." *Wind Engineering* 32 (4): 325-338.

PSCO. 2011. "2011 Electric Resource Plan. Volume II Technical Appendix."

Rongxin, Y., P. Xu, M. A. Piette, and S. Kiliccote. 2010. "Study on Auto-DR and Pre-Cooling of Commercial Buildings with Thermal Mass in California." *Energy and Buildings* 42 (7): 967-975.

Rubinstein, R., L. Xiaolei, and D. Watson. 2010. *Using Dimmable Lighting for Regulation Capacity and Non-Spinning Reserves in the Ancillary Services Market: A Feasibility Study.* LBNL-4190E.

Sioshansi, R., P. Denholm, and T. Jenkin. 2012. "Market and Policy Barriers to Deployment of Energy Storage." *Economics of Energy and Environmental Policy* 1 (2): 47-63.

Sioshansi, R., S. H. Madaeni, and P. Denholm. 2013. "A Dynamic Programing Approach to Estimate the Capacity Value of Energy Storage." *Transactions on Power Systems* (IEEE) PP (99): 395-403.

Starke, M., N. Alkadi, and O. Ma. 2013. *Assessment of Industrial Load for Demand Response across U.S. Regions of the Western Interconnection.* ORNL/TM-2013/407.

The Regulatory Assistance Project. 2011. *Electricity Regulation in the U.S.: A Guide.* The Regulatory Assistance Project.

Todd, D., B. Caulfield, M. Helms, M. Starke, B. Kirby, and J. Kueck. 2009. *Providing Reliability Services through Demand Response: A Preliminary Evaluation of the Demand Response Capabilities of Alcoa, Inc.* ORNL/TM 2008/233.

Todd, D., Caufield, M., Helms, B., Generating, A. P., Starke, I. M., Kirby, B. and Kueck, J. 2008. *Providing reliability services through demand response: A preliminary evaluation of the Demand Response Capabilities of Alcoa Inc.* ORNL/TM-2008/233.

Varadan, S., G. Freddo, H. Todus, J. Thiemsuwan, Ke Chen, Khoi Vu., D. Hawkins, and S. Shen. 2012. "A New Approach to Studying the Impact of Intermittent Renewable Resources." *Power and Energy Society General Meeting.* IEEE.

Watson, D., N. Matson, J. Page, S. Kiliccote, M. A. Piette, K. Corfee, B. Seto, et al. 2012. *Fast Automated Demand Response to Enable the Integration of Renewable Resources.* LBNL-5555E.

WECC. 2011. *Assumption Matrix fro the 2020 TEPPC Dataset.* Accessed 2013 November. www.wecc.biz/library/StudyReport/Documents/Assumptions%20Matrix%20for%20the%202020%20TEPPC520Dataset.pdf.

Wisconsin Administrative Code. 2012. "Wisconsin Administrative Code." PSC §117.05.

—. 2012. "Wisconsin Administrative Code." PSC 117.05.

Wood, A. J., and B. F. Wollenberg. 2013. *Power Generation, Operation, and Control.* John Wiley & Sons, Inc.

Woolf, T., E. Malone, L. Schwartz, and J. Shenot. 2013. *A Framework for Evaluating the Cost-Effectiveness of Demand Response.* LBNL.

WWSIS-2 Technical Review Committee. 2012. (November).

Xu, L., and D. Tretheway. 2012. *Flexible Ramping Products, Second Revised Draft Final Proposal.* CAISO.

Zhou, Z., and A. Botterud. 2013. "Price Responsive Demand for Operating Reserves and Energy in Electricity Markets." *Power and Energy Society General Meeting.* IEEE.

(This page intentionally left blank)

U.S. DEPARTMENT OF ENERGY

For more information, visit: energy.gov

DOE/EE-1282 • March 2016