ISO/RTO-managed Electric Energy Regions

James F. Ellison, Leigh S. Tesfatsion, Verne W. Loose, Raymond H. Byrne

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

James F. Ellison
Energy Storage and Transmission Analysis Department
Sandia National Laboratories
MS 1140, PO Box 5800
Albuquerque, NM 87185-1140
Email: jelliso@sandia.gov

Leigh S. Tesfatsion
Department of Economics, Iowa State University
Ames, IA 50011-1070

Verne W. Loose
Energy Storage and Transmission Analysis Department
Sandia National Laboratories
MS 1140, PO Box 5800
Albuquerque, NM 87185-1140

Raymond H. Byrne
Energy Storage and Transmission Analysis Department
Sandia National Laboratories
MS 1140, PO Box 5800
Albuquerque, NM 87185-1140

Abstract

This survey was undertaken as part of the initial phase of a larger market design project. The goal of the project is to design a wholesale electric power market such that electric energy storage (EES) resources are permitted to participate and receive compensation that is commensurate with the benefits they provide to the grid. This survey compares and contrasts
operating reserve markets in the seven U.S. ISO/RTO-managed electric energy regions. The reserve market categories employed in each energy region are placed into a common framework. The terminology used for reserve markets in each region, as well as the characteristics of these markets, is discussed. Finally, the market procedures currently in place in the seven energy regions for the procurement, settlement, and allocation of costs for reserves are examined.

**keywords:** Power market design, ancillary services, operating reserve, frequency control, and electrical energy storage.
Acknowledgment

This research was sponsored by the Office of Energy Delivery and Energy Reliability’s Energy Storage Program at the U.S. Department of Energy. The authors would like to thank Dr. Imre Gyuk and his colleagues at the Energy Storage Program at the U.S. Department of Energy for their funding and support of this project.

At Iowa State, Deung-Yong Heo, Andrew Nguyen, and Dong-Jin Pyo compiled information as to the reserves market mechanics in the seven regions, and helped with overall fact-checking. At Sandia, Cesar Silva Monroy provided valuable feedback on the draft paper, and Ross Gutromson provided project guidance. Any shortcomings in the report are the responsibility of the authors alone.
## Contents

1 Introduction .......................................................... 9

2 Operating Reserve Market Terminology and Characteristics ................. 12
   2.1 An Introduction to Operating Reserve .................................. 12
   2.2 A Common Terminology for Operating Reserve ......................... 13
   2.3 Additional Characteristics of Operating Reserve Markets ............ 16

3 Procurement, Settlement, and Cost Allocation for Operating Reserve .......... 19
   3.1 Overview .......................................................... 19
   3.2 Acquisition of Operating Reserve: Summary Regional Comparisons .... 19
   3.3 PJM Operating Reserve Acquisition .................................. 22
   3.4 ISO-NE Operating Reserve Acquisition ................................ 25
   3.5 ERCOT Operating Reserve Acquisition ............................... 27
   3.6 MISO Operating Reserve Acquisition ................................ 29
   3.7 NYISO Operating Reserve Acquisition ................................ 30
   3.8 CAISO Operating Reserve Acquisition ............................... 31
   3.9 SPP Operating Reserve Acquisition (Planned) .......................... 32

4 Direction for Future Research .......................................... 33

5 Concluding Remarks .................................................. 35

References .............................................................. 37
Figures

1 ISO/RTO Two-Settlement System Activities on a Typical Day D-1. ................. 20
2 ISO/RTO Two-Settlement System Activities on Successive Days D-1 and D. .... 21
3 ISO-NE Locational Forward Reserve Market, High-Level Process Flow .......... 26
4 Operating reserve requirements in the frequency domain. ....................... 34
Tables

1  The Seven U.S. ISO/RTO-Managed Electric Energy Regions .................................. 10
2  Reserve Market Terminology Currently in Use by U.S. ISOs/RTOs. ......................... 14
3  Selected Characteristics of Reserve Markets in U.S. ISOs/RTOs.............................. 17
1 Introduction

This survey compares and contrasts how the seven U.S. ISOs/RTOs manage their needs for operating reserve through centralized reserve markets, and was undertaken as the initial phase of a larger market design project. The goal of the project is to design a wholesale electric power market in such a way that electric energy storage (EES) resources, as well as other types of resources, are permitted to participate fully in the market and receive compensation that is commensurate with the services they provide to the grid, without bias towards or against a particular technology.

In considering a new market design, it is first necessary to document how existing markets are designed, and to determine to what extent existing markets offer the incentives needed for the efficient, reliable operation of the grid even as the resource mix on the grid changes and new resources emerge.

Operating reserve is defined by NERC\(^1\) as “that capability above firm demand required to provide for regulation, load forecasting error, equipment forced and scheduled outages, and local area protection.” NERC further states that operating reserve “consists of spinning and non-spinning reserve” [32]. While there are differences in how the ISOs/RTOs define operating reserve, this paper will adhere to the NERC definition.

In general, the seven ISOs/RTOs require companies that service loads (i.e., the energy requirements of end-use customers) to provide reserves in proportion to their loads. Typically these companies are able to choose whether to provide their own reserves, to secure bilateral contracts for these reserves, or to purchase reserves in reserve markets centrally organized by the ISO/RTO. This paper will concentrate on these centrally-organized reserve markets. For the purposes of this survey, we will define operating reserve markets to include all reserve markets organized by an ISO/RTO that operate on a daily basis. Before comparing these markets, it is first necessary to discuss the role operating reserve plays in facilitating the operation of the power grid.

A power system must operate within a narrow frequency range to avoid system collapse. In North America, for example, the nominal (targeted) value for frequency is set at 60 Hz. North American power systems are required to operate within narrow bands around this nominal value.

Frequency is controlled by ensuring the balance of a power system at any given time, meaning that the amount of active power injected into the system matches the amount of active power withdrawn or dissipated through loads, interchange net outflows, and transmission line losses.\(^2\) Resources must be available to maintain system balance at all times through appropriate generation and/or load adjustments. Balancing services are provided by generating and/or demand response resources that are capable of and stand ready to provide them. These balancing services are an important form of ancillary service for power systems, generally referred to as operating reserve.

Deviations from balance can arise for a variety of reasons. For example, large conventional generating units can unexpectedly fail (termed a forced outage), loads can deviate from contracted

\(^{1}\)The North American Electric Reliability Corporation

\(^{2}\)Balance is also often referred to as system adequacy.
levels, and variable generation from smaller generating units (such as solar and wind units) can differ from what was forecast. In particular, the increase in variable generation required by the renewable portfolio standards of many states raises questions as to whether power systems will have sufficient operating reserve to compensate for this variability.

Vertically-integrated utilities (VIUs) combine generation, transmission, and distribution operations within a single company. Consequently, a VIU can determine a level of operating reserve deemed sufficient for its system and operate its generating units to maintain this level of reserve. There is no need to separate the cost of providing operating reserve from the cost of providing energy.

In contrast, for the seven U.S. electric energy regions managed by Independent System Operators (ISOs) or Regional Transmission Organizations (RTOs), generation has been unbundled from transmission and distribution operations, and is owned by multiple entities. In these regions, if the variable cost of generation for a unit is below the marginal price of that market, then a generation unit owner who reduces output of active power in order to provide operating reserve is foregoing the opportunity to make additional profit. For a generator owner to be willing to forgo the additional profit that could be earned from selling more energy, the owner must anticipate receiving at least this same level of profit from the provision of operating reserve. Similarly, a demand response resource owner that allows load to be increased or decreased in order to provide balancing services forgoes the opportunity to consume energy at optimal times or in optimal amounts. To willingly provide operating reserve, a demand resource owner must also anticipate receiving compensation that at least compensates for this foregone opportunity.

An ISO/RTO can most efficiently operate a power system by ensuring fair and open access and full compensation to all providers of services. In particular, any resource capable of providing operating reserve services should be permitted to do so, and these resources should be compensated based on the net benefits they provide to the system and not on the basis of their physical form.

Table 1. The Seven U.S. ISO/RTO-Managed Electric Energy Regions

<table>
<thead>
<tr>
<th>System</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>California ISO</td>
<td>CAISO</td>
</tr>
<tr>
<td>Electric Reliability Council of Texas</td>
<td>ERCOT</td>
</tr>
<tr>
<td>New England ISO (RTO)</td>
<td>ISO-NE</td>
</tr>
<tr>
<td>Midwest ISO (RTO)</td>
<td>MISO</td>
</tr>
<tr>
<td>New York ISO</td>
<td>NYISO</td>
</tr>
<tr>
<td>PJM Interconnection (RTO)</td>
<td>PJM</td>
</tr>
<tr>
<td>Southwest Power Pool (RTO)</td>
<td>SPP</td>
</tr>
</tbody>
</table>

3Table 1 lists these seven energy regions, along with their abbreviations. These regions serve about two-thirds of the electricity consumers in the United States [18].
Finally, an ISO/RTO must be concerned about future as well as current needs for operating reserve. Whereas regulated VIUs can make the case to their regulator that additional resources are needed to provide operating reserve as needs for reserve grow over time, an ISO/RTO must insure that payments for operating reserve are sufficient to attract new resources into their energy region.
2 Operating Reserve Market Terminology and Characteristics

2.1 An Introduction to Operating Reserve

Operating reserve is generally partitioned into three distinct categories of frequency control: primary, secondary, and tertiary. Each responds faster than the next. Moreover, the mechanism activating reserve in each of these categories is different [8, 44, 45].

Primary frequency control is a local automatic control that rapidly (within seconds) adjusts generator output or load to offset large changes in frequency. The adjustment of generator output is termed governor response, as it is provided by controllable synchronous generators fitted with a speed governor. An important aspect of primary frequency control is that, even when fitted with a speed governor, a generator can provide additional power (to oppose frequency drops) only if it is operating at less than full capacity. The amount of spare capacity on an operating generating unit is generally termed headroom.

Primary frequency control acts to arrest a sharp drop or spike in frequency. It is designed to keep the frequency within specified limits in response to the forced outage of a generator or the loss of a large load. Without this action, generators would be forced to go offline to avoid suffering the severe damage that would result from operating at frequencies outside of their design limits. In a situation where generation is insufficient to meet load, having additional generators disconnecting from the grid would only exacerbate the problem and could lead to a blackout. Therefore, primary frequency control is an indispensable service for reliable grid operations.

In the past, virtually all generators in North America were required to provide primary frequency control. Currently, however, governor response is not required of nuclear plants, most wind turbines, and some combustion turbines [14]. Loads can also be designed to vary based on frequency changes and hence provide primary frequency control. This is called frequency-responsive demand response, and is a relatively newly recognized resource.

In contrast to primary frequency control, primary frequency response is the combination of primary frequency control and system inertia acting to arrest frequency decline. System inertia is a term describing the ability of a power system to resist changes in frequency, and is measured in MW-seconds. All generators synchronized to the system contribute to system inertia. In the event of a contingency, frequency will fall faster in a system with less inertia than in one with more inertia. Therefore, the amount of system inertia impacts the speed at which the primary frequency control reserve must act to arrest the frequency at the same minimum point. The required amount of primary frequency control reserve depends on the size of a power system and its system inertia, the size of the anticipated contingency, and the lowest acceptable frequency level on that system [14].

Primary frequency control can arrest a frequency drop or spike, but it is not designed to restore

---

4 Frequency-responsive demand response is “agreed-to load shedding by end-use customers that complements governor response” [14].
frequency to its nominal value. Rather, restoration is the purpose of another form of operating reserve called secondary frequency control. Secondary frequency control is a central automatic control that acts to adjust active power production to restore the frequency and power interchanges with other systems to their nominal levels following an imbalance [44]. This automatic process, generally termed automatic generation control (AGC) in North America, acts on a time frame of several seconds to counteract frequency deviations. While secondary frequency control can serve to restore frequency following a contingency or the loss of a large block of load, it cannot serve to limit the magnitude of the initial frequency swing following such an event [27].

Finally, tertiary frequency control consists of manual changes in scheduled unit commitment and dispatch levels in order to bring frequency and/or interchanges back to nominal values when secondary frequency control is unable to perform this task. While primary and tertiary frequency controls are essential for reliable grid operations, secondary frequency control is not. Smaller power systems can be operated using only primary frequency control and manual tertiary control. All large interconnected systems, however, use secondary frequency control because manual control cannot remove transmission line overloads quickly enough [44, 45].

2.2 A Common Terminology for Operating Reserve

Terminology currently in use for operating reserve in U.S. ISO/RTO-managed electric energy regions can lead to confusion in that different terms are sometimes used to refer to the same service. In some instances, operating reserve is divided into a larger or smaller number of categories in such a manner that the services listed under the same category across different regions are not exactly the same. Furthermore, some reserve services are neither obtained through a centralized market, nor compensated for outside of a market, despite their value to the functioning of the power grid.

Table 2 presents a compilation of the terminology used by U.S. ISO/RTO-managed electric energy regions for their reserve markets, categorized in accordance with the primary, secondary, and tertiary frequency control descriptions provided above. It is important to caution that these regions are constantly evolving their operational definitions and procedures.

The shading in Table 2 depicts how each region defines its operating reserve. Three consider secondary frequency control (or regulation) reserve to be part of their operating reserve and four do not. For comparison, both the North American Electric Reliability Corporation (NERC) and the Western Electricity Coordinating Council (WECC) define operating reserve as including regulation [51, 32].

As can be seen in Table 2, none of the energy regions has a market for primary frequency control reserve. Some energy regions require a synchronized generator to be capable of primary frequency control as a condition for interconnecting with the grid. However, none of the regions sets specific hourly targets for the amount of primary frequency control desired, nor accepts offers

---

5Note that the regions use differing terminology in referring to this reserve. For example, NYISO and PJM use the term “frequency response”, ISO-NE uses the terms “primary control” or “governor response,” and ERCOT uses “primary frequency response.”
### Table 2: Reserve Market Terminology Currently in Use by U.S. ISOs/RTOs.

<table>
<thead>
<tr>
<th>ISO/RTO</th>
<th>Regulation Up</th>
<th>Regulation Down</th>
<th>Ten-Minute Spinning Reserve</th>
<th>Ten-Minute Non-Spinning Reserve</th>
<th>Thirty-Minute Spinning Reserve</th>
<th>Thirty-Minute Non-Spinning Reserve</th>
<th>Quick Start Spinning Reserve</th>
<th>Quick Start Non-Spinning Reserve</th>
<th>Ten-Minute Spinning Reserve</th>
<th>Ten-Minute Non-Spinning Reserve</th>
<th>Quick Start Spinning Reserve</th>
<th>Quick Start Non-Spinning Reserve</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAISO</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
<tr>
<td>ERCOT</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
<tr>
<td>ISO-NE</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
<tr>
<td>MISO</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
<tr>
<td>NYISO</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
<tr>
<td>PJM</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
<tr>
<td>SPP</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
<td>no market</td>
</tr>
</tbody>
</table>

Sources: [6, 10, 24, 35, 42, 40, 49]

Note: Italics denote a reserve market category, and regular text denotes a reserve market. Shaded text denotes what each energy region defines as its operating reserve markets. Information on SPP reflects terminology used in their intended new Integrated Market Protocols, not in their current markets.
for generation capacity to be reserved for primary frequency control. While primary frequency control reserve is provided by all resources with autonomous governor response that are synchronized (and have the headroom to increase generation), none of the regions provides compensation for this reserve. As the amount of power provided by variable renewable generation increases, the fraction of on-line generation capacity offering primary frequency control will decrease. In the future, market mechanisms may be needed to ensure sufficient provision of primary frequency control reserve.

Table 2 also indicates that a common name for secondary frequency control in the seven energy regions is regulation, or regulation up and regulation down. Three of the seven regions determine separate prices for regulation up and regulation down, each of which is categorized as a secondary frequency control reserve. In the remaining four regions, no distinction is made between regulation up and regulation down; rather, each megawatt of regulation is priced the same, regardless of direction.

This is an important distinction among regions. For example, wind generation can be high at night, when load is at its lowest level. In this case, if power systems are to avoid curtailing wind power, they need to be able to decrease the output of their conventional generating units to compensate for unexpected increases in wind generation. At night, however, some conventional generating units are operating at their minimum capacity levels and may not have the ability to reduce generation further. In these circumstances, regulation up is likely to be plentiful, and regulation down in short supply. Having separate markets for regulation up and regulation down better reflects system conditions, and therefore might be more efficient than a single market for regulation.

Finally, as reported in Table 2, spinning and non-spinning reserve are the terms most commonly used by the energy regions for the two fastest acting categories of tertiary frequency control reserve. To provide spinning reserve, a resource must be synchronized to the grid and must be able to reach the declared output level within a short time interval: e.g., ten minutes for a ten-minute spinning reserve. In contrast, ten-minute non-spinning reserve can be offered by an off-line resource, but it must be able to be synchronized to the grid and brought up to the declared output level within ten minutes.

Three of the energy regions specifically use the term contingency reserve as a category that includes both spinning reserve and non-spinning reserve. The remaining four energy regions provide for both spinning and non-spinning reserve, but they do not formally define them as being contingency reserve.

The term contingency reserve implies a reserve that exists solely in the event of a contingency, such as the forced outage of a generating unit or a transmission tie line failure. The amount of contingency reserve is often set at the size of the largest single contingency (and sometimes higher for an additional margin of safety), which is often the size of the largest generating unit on the system. This amount is then divided between spinning reserve and non-spinning reserve. Many energy systems require that at least 50% of their contingency reserve consist of spinning reserve.

However, spinning and non-spinning reserve can provide benefits beyond contingency re-

---

6 A tie line is a circuit that connects two Balancing Authority Areas [32].
response. They are also used as load following reserve. Load following is the “action to follow the general trending load patterns within the day,” and is usually performed by the economic dispatch of spinning reserve, but can also involve the dispatch of quick-start units that are not on-line [8]. Regulation reserve (secondary frequency control reserve in the framework presented here) is used in conjunction with load following reserve to take care of the second-to-second mismatches between generation and load. Note that a system operator needs to insure that any load following reserve used does not diminish the amount of contingency reserve available – otherwise, it would not be possible to adequately respond to a contingency event.

The seven energy regions also provide for another category of reserve, less flexible than ten-minute spinning and non-spinning reserve, which is called “supplemental” by PJM, “replacement” by ERCOT, “operating” by ISO-NE, and “30-minute” by NYISO. This reserve is generally defined as resources that are either synchronized or non-synchronized to the grid and that can be brought up to the declared level of output within thirty minutes. The purpose of this reserve is to restore the ten-minute spinning and non-spinning reserve after a contingency has occurred. This frees up the spinning and non-spinning generating units to again provide ten-minute spinning and non-spinning reserve, allowing the system to be ready for a second contingency. One energy region, NYISO, formally divides its 30-minute reserve category into two sub-categories: namely, a 30-minute spinning reserve, and a 30-minute non-synchronized reserve.

2.3 Additional Characteristics of Operating Reserve Markets

Table 3 describes more detailed aspects of operating reserve markets in the seven energy regions that cannot be deduced simply by examining reserve categorizations.

Two of the regions, ERCOT and MISO, require all resources that provide regulation to be autonomously frequency-responsive, meaning that they also supply primary frequency control reserve. ERCOT goes a step further, requiring that resources providing spinning reserve (or Responsive Reserve Service in the ERCOT terminology) must also be autonomously frequency-responsive. These are the only two U.S. energy regions that formally incorporate primary frequency response into their market criteria.

ERCOT and PJM are the only regions to specify a minimum capacity that can be offered into their regulation reserve markets. Generally, each region has minimum rated capacity requirements for participation in its energy or reserve markets. However, once a resource is qualified to participate, the resource owner can offer any quantity desired into the regulation reserve market – except in the case of ERCOT and PJM. ISO-NE is the only region to state a minimum ramp rate for regulation reserve market participation.

PJM is the only market that has two categories of spinning reserve (called Synchronous Reserve by PJM): Tier 1 and Tier 2. Tier 1 reserve is provided by any resource that is on-line, following

---

7There are some exceptions to this generalization. CAISO, for example, provides an exception from its 1 MW minimum capacity rule for a generating unit providing “Ancillary Services and/or Imbalance Energy through an aggregation arrangement approved by CAISO.” [7]
<table>
<thead>
<tr>
<th>Function</th>
<th>Product</th>
<th>Characteristics</th>
<th>CAISO</th>
<th>ERCOT</th>
<th>ISO-NE</th>
<th>MISO</th>
<th>NYISO</th>
<th>PJM</th>
<th>SPP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>Regulation</td>
<td>Governor control necessary for participation? yes no yes no no no no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Frequency Reserve</td>
<td></td>
<td>Separate markets for up/down regulation? yes yes no no no no yes yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td>AGC signal required? yes yes yes yes yes yes yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Response</td>
<td></td>
<td>Max time to deliver nominated capacity? 10-30** 10 5 5 5 5 5 60 60 60</td>
<td>10-30**</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Min duration to maintain nominated output? 60 60 60 60 60 60 60 60 60</td>
<td></td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Min ramp rate? (MW/min) 1 1 1 1 1 1 1 1 1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
|                  |           | Min capacity offered? 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW | 1 MW  | 1 MW  | 1 MW   | 1 MW | 1 MW | 1 MW | 1 MW *
| Tertiary         | Spinning  | Governor control necessary for participation? no yes no no no no no no | no    | yes   | no     | no   | no    | no  | no   |
| Frequency Reserve|           | Max delay to deliver nominated capacity? 10 10 10 10 10 10 10 10 10 | 10    | 10    | 10     | 10   | 10    | 10  | 10   |
| Control          |           | Min duration to maintain nominated output? 30 60 60 60 60 60 60 60 60 | 30    | 60    | 60     | 60   | 60    | 60  | 60   |
| Response         |           | Min capacity offered? 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW | 1 MW  | 1 MW  | 1 MW   | 1 MW | 1 MW | 1 MW | 1 MW |
|                  |           | Two-tiered market structure? no no no no no yes no yes yes yes | no    | no    | no     | no   | yes   | no  | no   |
| Non-Spinning     | Reserve   | Max delay to be synchronized and at nominated capacity? 10 30 10 10 10 10 10 10 10 | 10    | 30    | 10     | 10   | 10    | 10  | 10   |
|                  |           | Min duration to maintain nominated output? 30 60 60 60 60 60 60 60 60 | 30    | 60    | 60     | 60   | 60    | 60  | 60   |
|                  |           | Min capacity offered? 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW 1 MW | 1 MW  | 1 MW  | 1 MW   | 1 MW | 1 MW | 1 MW | 1 MW |
| Supplemental     | Reserve   | Separated into synchronized and non-synchronized reserve markets? N/A no no N/A yes no N/A | N/A   | no    | no     | N/A  | yes   | no  | N/A  |
|                  |           | Max delay to be synchronized and at nominated capacity? N/A agreed upon N/A 30 N/A 30 30 N/A | N/A   | agreed| upon   | N/A  | 30    | 30  | N/A  |
|                  |           | Min duration to maintain nominated output? N/A agreed upon N/A agreed upon N/A | N/A   | agreed| upon   | N/A  | N/A   |     | N/A  |

Table 3. Selected Characteristics of Reserve Markets in U.S. ISOs/RTOs.

Note: All units are in minutes, unless otherwise specified. N/A means no centralized daily supplemental reserve market. *Information on SPP reflects characteristics of their intended new Integrated Market Protocols, not of their current markets. **CAISO specifies a value between 10 and 30 minutes, which is used to calculate the maximum regulation a resource can offer. Sources: [6, 10, 24, 30, 35, 42, 49]
economic dispatch, and capable of increasing output. It is not reserved through a market, but instead is simply the amount of reserve present on the system due to economic dispatch. No capacity payments are made to resources providing Tier 1 reserve capacity. If Tier 1 reserve is insufficient, additional reserve capacity is acquired through a centralized market; this additional reserve capacity is termed Tier 2 reserve. Tier 2 reserve capacity is dispatched only after the Tier 1 reserve is exhausted [43].

Most of the regions require that a resource be able to provide continuous output for some specified duration of time in order to qualify as a reserve provider. For example, CAISO requires that spinning and/or non-spinning reserve resources be able to maintain a constant level of power output for a minimum of 30 minutes, whereas ISO-NE and MISO require such resources to be able to maintain a constant power output for a minimum of 60 minutes. This means that a flywheel, which currently is able to provide constant power output at its rated capacity for about 15 minutes, is effectively excluded from participating in the provision of spinning and/or non-spinning reserve in these regions because it can only participate at one half (with a 30-minute duration test) to one quarter (with a 60-minute duration test) of its rated capacity.

Other regions do not specify minimum durations for reserve resources in their market rules, but have other provisions that effectively cap the duration. ERCOT, for example, specifies that the regulation control signals cannot request resource performance beyond the High Sustained Limit (HSL), which is based on averaged telemetered output during 30 minutes of constant output [50]. In this case, evaluating a 15-minute resource (such as a flywheel) with a 30-minute test means that the resource will only be able to participate in the regulation market with up to half of its power output capacity. If ERCOT’s regulation control signals require up to 30 minutes of continuous output in one direction (either an increase or a decrease in energy production), then this restriction may be justified. Otherwise, the restriction is artificial and unnecessary. In recognition of this and other lost opportunities for service, all seven energy regions are currently reviewing their constant output requirements, with a view towards ensuring their reserve markets are not biased against electricity storage resources. Some of them have developed exceptions for energy storage devices. Dealing with such exceptions in a comprehensive manner is beyond the scope of this paper.

ERCOT, ISO-NE, NYISO, and PJM all have a 30-minute reserve category. NYISO further subdivides this category into two sub-categories: 30-minute spinning reserve and 30-minute non-synchronized reserve. ERCOT has the distinction of being the only region that has a daily market for a Replacement Reserve Service; this allows resources that may not always be able to respond in a 30-minute time frame to participate. Such resources are available for service during the particular hours for which they have been notified in advance. In contrast, SPP (for their planned new design), CAISO, and MISO have no 30-minute reserve category; only 10-minute reserve categories are available in these markets.

---

8 Beacon Power’s Smart Energy 25 flywheel can deliver “25 kWh of extractable energy at a 100 kW power level for 15 minutes” [2].

9 Both MISO and NYISO have created a storage-specific resource category for their regulation markets. MISO terms this resource a Stored Energy Resource (SER), and NYISO calls it a Limited Energy Storage Resource (LESR).
3 Procurement, Settlement, and Cost Allocation for Operating Reserve

3.1 Overview

This section compares and contrasts U.S. ISO/RTO practices with regard to the acquisition of operating reserve as defined by [32], which comprises the secondary and tertiary frequency control categories appearing in Table 2 and Table 3. Here we use the term acquisition to refer broadly to the procurement, settlement, and allocation of costs for operating reserve.

As clarified in earlier sections, operating reserve is needed to ensure power system reliability for the benefit of all power system participants. The acquisition of operating reserve is handled through a complicated mixture of market and administrative rules. For guidance through this thicket of rules, heavy use is made of the many business practice manuals (BPMs) provided by each ISO/RTO.

Although it is extremely helpful to have BPMs available for understanding ISO/RTO business practices, four factors complicate their use for ISO/RTO comparisons. First, the business practices in question are extremely complex, covering the operations of multiple interconnected markets operating at multiple time scales. Second, the BPMs are lengthy documents written in highly legalistic language, which hinders their readability. Third, the BPMs are continually being updated to include changes in rules of operation. Fourth, with one exception [30, Attachment B], the ISOs/RTOs do not include in their BPMs the precise forms of the optimization problems (objective functions plus constraints) that are used to determine their price, commitment, and dispatch solutions for energy and reserve.

The purpose of this section is to glean from the BPMs and other source materials the general similarities and differences among the ISOs/RTOs regarding the economic aspects of their operating reserve practices without getting bogged down in technical details. The interested reader is referred to the source materials listed in our reference section for more detailed explanations.

3.2 Acquisition of Operating Reserve: Summary Regional Comparisons

In April 2003, FERC released a white paper proposing the adoption of a common design for U.S. wholesale electric energy markets [15]. As depicted in Fig. 1, the basic building block of this recommended design was a two-settlement system entailing the parallel operation of day-ahead and real-time energy markets, each using locational marginal pricing (LMP) to price energy by the loca-

---

10In economics, a market is defined to be a collection of buyers and sellers that determines the price of one or more products on the basis of their demand bids and/or supply offers [38]. A buyer’s demand bid reflects the maximum price the buyer is willing to pay for the purchase of each successive product unit and a seller’s supply offer reflects the minimum price the seller is willing to accept for the sale of each successive product unit. The key feature distinguishing a market from a purely administrative process is that a market determines prices as a function of demand and supply interactions rather than through administrative decree.
tion and timing of its injection into or withdrawal from the transmission grid. This two-settlement system now constitutes the backbone of the seven U.S. ISO/RTO-managed electric energy regions.

![Figure 1. ISO/RTO Two-Settlement System Activities on a Typical Day D-1.](image)

However, in this white paper FERC did not recommend common means for the provision of ancillary services needed to support electric energy trading, such as balancing services supplied by operating reserve resources. Instead, FERC called on the states within each energy region to determine their own ancillary service practices [15, p. 11].

In consequence, starting from the common foundation of a two-settlement system, each of the seven U.S. ISO/RTO-managed energy regions has evolved its own reliability assessment processes for assessing the need for operating reserve. As indicated in Fig. 2, these assessment processes take place in advance of each operating hour H in order to ensure adequate operating reserve during hour H. Indeed, many types of forward contracts are entered into many days in advance of an operating hour in order to ensure the availability of operating reserve capacity for this hour.

The following subsections provide detailed descriptions of current ISO/RTO rules governing the acquisition of operating reserve through market-based processes. As will be seen, convergence has taken place for some aspects of these rules while substantial divergence remains for others.

For example, all seven ISOs/RTOs allocate the costs of operating reserve to wholesale buyers of energy. Specifically, entities servicing regional and external loads, typically referred to as Load-Serving Entities (LSEs), are assigned reserve requirements on the basis of their relative shares of these loads.\(^\text{11}\) The LSEs are then responsible for meeting these reserve requirements through

\(^{11}\)An LSE is an organization that has an obligation under Federal, State, or local law, or under long-term contracts, to provide electrical power to end-use (retail) customers, or to other LSEs with end-use customers, but that does not
self-supplied reserve, through bilateral arrangements with other market participants capable of providing reserve, and/or through purchases of reserve in some form of ISO/RTO-managed reserve market.

On the other hand, the ISOs/RTOs exhibit significant differences in their procurement and settlement practices for operating reserve. While all seven “co-optimize” energy and reserve, in some cases this co-optimization takes the form of “integrated co-optimization” whereas in other cases it takes the form of “coupled” or “decoupled” co-optimization.

By **integrated co-optimization** we mean that energy and reserve prices and scheduled dispatch levels are determined simultaneously as the solution to an optimization problem that involves the maximization or minimization of a single objective function subject to a single set of constraints. By **coupled co-optimization** we mean that separate optimizations are carried out for energy and reserve, but the optimizations contain coupled constraints. Typically these coupled constraints take the following form: The energy optimization includes inequality constraints reflecting the ISO/RTO’s estimated system-wide and/or local reserve requirements, and the reserve optimization takes into account the opportunity cost (lost benefit) at the margin when capacity is withheld for reserve rather than used to generate energy. Finally, by **decoupled co-optimization** we mean that separate parallel optimizations are carried out for energy and reserve without the imposition of coupled constraints.

The seven ISOs/RTOs can be roughly grouped in accordance with the extent to which they

---

Figure 2. ISO/RTO Two-Settlement System Activities on Successive Days D-1 and D.
support an integrated co-optimization approach to energy and operating reserve, from least to most support, as follows: (i) PJM; (ii) ISO-NE, (iii) ERCOT; and (iv) MISO, NYISO, CAISO, and SPP (planned for 2014). PJM supports a coupled co-optimization for energy and reserve in forward markets together with continuous real-time adjustments on the basis of real-time conditions. ISO-NE supports a decoupled co-optimization for energy and Operating Reserve in forward markets, a coupled co-optimization for energy and Operating Reserve in real time, and a decoupled co-optimization for energy and Regulation in real time. ERCOT supports an integrated co-optimization of energy and reserve in a day-ahead market and a coupled co-optimization of energy and reserve in real time. Finally, the group (iv) energy regions support an integrated co-optimization of energy and reserve in both day-ahead and real-time markets.

The current ISO/RTO settlement practices for regulation service have recently come under particular scrutiny from FERC. The ISO-NE makes payments for regulation service that reflect the amount of service provided. Resources receive a “mileage” payment reflecting the amount of ACE correction provided [24]. NYISO has a penalty that reflects how accurate the resource is in following the AGC signal [36]. However, the remaining five ISOs/RTOs compensate regulation service based purely on the amount of capacity accepted by the market. This does not take into account how the resource providing regulation service actually performs. For example, flywheels are must faster and can follow an AGC signal much better than other resources; yet 1 MW of flywheel regulation capacity is compensated the same as 1 MW of conventional generator regulation capacity. This lack of performance-based compensation was the impetus for FERC’s Order 755, which requires the energy regions to compensate for regulation based on actual service provided [16]. All seven ISOs/RTOs are therefore in the process of reviewing how they will each modify their regulation markets to incorporate resource performance.

The next seven subsections provide a more detailed characterization of the current operating reserve procurement, settlement, and cost allocation practices in each of the seven ISOs/RTOs.

### 3.3 PJM Operating Reserve Acquisition

PJM secures energy and reserve through coupled market-based processes [37, 39, 41, 42]. Specifically, PJM’s scheduling of energy and reserve for each operating day D is handled by means of a forward Day-Ahead Energy Market (DAEM), a Real-Time Energy Market (RTEM), a forward Regulation Market, a forward Synchronized Reserve Market, a forward Day-Ahead Scheduling Reserve Market (DASRM), and an hourly re-scheduling process. Each of these processes will now be more carefully described.

The DAEM, which produces energy prices and energy commitment and dispatch levels for each hour H of an operating day D, closes at hour 12 on day D-1 [42, Section 2]. Up to this close, market participants can submit energy demand bids and energy supply offers for each hour of day D. After this close, PJM performs analysis to clear the DAEM. The market-clearing prices for the DAEM are LMPs calculated as the shadow prices for nodal energy balance constraints.
The RTEM is a balancing mechanism in which market-clearing LMPs for imbalance energy\textsuperscript{12} are calculated every five minutes based on actual system conditions \cite[Section 2]{42}. Separate daily accounting settlements are performed for the DAEM and RTEM markets. The DAEM settlement is based on scheduled hourly quantities and on day-ahead hourly prices, whereas the RTEM settlement is based on actual hourly quantity deviations from day-ahead scheduled quantities and on hourly real-time LMPs calculated from the five-minute real-time LMPs determined for each hour.

PJM also manages forward offer-based Regulation and Synchronized Reserve Markets for the determination of price, commitment, and dispatch schedules for Regulation and (Tier 2) Synchronized Reserve for each hour \( H \) of each operating day \( D \) \cite[Sections 3-4]{42}. PJM determines the hourly zonal demands for Regulation and Synchronized Reserve in these markets in accordance with NERC reliability requirements. Reserve supply offers and operational parameters for hour \( H \) of operating day \( D \) must be submitted to PJM prior to hour 18 on day \( D-1 \). However, market participants can adjust their operational parameters up until 60 minutes prior to the start of hour \( H \), and PJM can adjust its estimates of reserve availabilities up until 30 minutes prior to the start of hour \( H \). PJM then simultaneously determines the least expensive set of resources needed to provide energy, Regulation, and Synchronized Reserve for hour \( H \), taking into account any resources self-scheduled to provide these services. These self-scheduled resources can be the result of bilateral reserve transactions, e.g., between a reserve supplier and a load with a reserve obligation.

In addition, PJM manages a forward offer-based DASRM for Synchronous, Quick-Start, and Supplemental Reserve for each hour of each operating day \( D \) that runs in parallel with the Regulation and Synchronized Reserve Markets \cite[Section 11]{42}. A resource can submit a reserve supply offer to both the DASRM and the Regulation Market, or to both the DASRM and the Synchronized Reserve Market, and it will be compensated for both participations. However, a resource cannot simultaneously submit reserve supply offers to both the Regulation and Synchronized Reserve Markets. PJM determines Operational Reserve requirements for the DASRM based on PJM load forecasts for day \( D \) together with NERC reliability requirements. Resources wishing to participate in the DASRM for day \( D \) must also submit an energy offer for the DAEM for day \( D \)\textsuperscript{13} and they must submit their DASRM reserve supply offers and operational data to PJM by hour 12 of day \( D-1 \), following the same timeline as for the DAEM. PJM participants can also report self-scheduled bilateral Operating Reserve transactions to PJM for inclusion in the DASRM.

The DAEM, RTEM, Regulation Market, Synchronized Reserve Market, and the DASRM together constitute a coupled co-optimization process for energy and reserve. The market clearing prices ($/MWh) determined for Regulation and Synchronized Reserve for each hour of each operating day \( D \) are locational (zonal) prices based on Regulation and Synchronized Reserve supply offers, self-scheduled reserve, zone reliability requirements, and the opportunity costs incurred by marginal cleared supply offers. These opportunity costs, measured in terms of actual (ex-post) real-time LMPs, provide compensation to resources for having to supply reserve rather than energy in order to meet Regulation and Synchronized Reserve requirements. The market clearing price

\textsuperscript{12}Imbalance energy refers to the energy amounts needed to resolve any differences between the load and generation amounts cleared in the forward DAEM subject to nodal energy balance constraints and the load and generation amounts realized in real time.

\textsuperscript{13}As will be clarified below, this permits the calculation of an “opportunity cost” for this resource reflecting its lost opportunity to offer energy instead of reserve.
($/MWh) determined in the DASRM for Operating Reserve for each hour of day D is set equal to the “merit-order price” of the highest-cost DAEM resource necessary to meet any remaining Operating Reserve requirement. This merit-order price is equal to the reserve price offered by this resource in the DASRM plus the opportunity cost incurred by this resource in the DAEM (measured in terms of DAEM LMPs) as a result of being backed down in the DAEM in order to meet Operating Reserve requirements.

At the beginning of hour 16 on day D-1, PJM posts a first resource commitment schedule for the operating day D that includes the results of the energy and reserve market clearing analysis for day D conducted up through hour 16 on day D-1. From hour 16 up to hour 18 on day D-1, PJM opens a *Balancing Market Offer Period* during which market participants can submit revised offers for energy and reserve resources not cleared in this first resource commitment.

The Balancing Market Offer Period closes at the beginning of hour 18 on day D-1. PJM then performs a second resource commitment based on revised offers, updated resource availability information, updated PJM load forecast information, and load forecast deviations. The focus of this second resource commitment is reliability and the objective is to minimize start-up and no-load costs for any additional resources that are committed. From hour 18 through the end of day D-1, PJM then performs as many additional resource commitment runs as deemed necessary based on updated PJM load forecasts and updated resource availability information. Resource commitment and dispatch schedules for each operating hour on day D are updated accordingly. Subsequently, PJM continues to make adjustments in its co-optimized energy and reserve dispatch schedules for each operating hour H of day D up until the start of hour H. PJM also makes intra-H adjustments in Regulation and Synchronized Reserve dispatch levels as needed to ensure system reliability.

PJM allocates its costs of Regulation and Synchronized Reserve procurement to its LSEs on the basis of their relative loads [39]. More precisely, the process by which PJM allocates its costs for Regulation procurement is as follows. (PJM’s cost allocation for Synchronized Reserve procurement is entirely analogous.) For each operating hour H, each LSE participating in PJM has a Regulation Obligation equal to its real-time load ratio share in its Regulation Zone multiplied by the total Regulation requirement for this Regulation Zone. This Regulation Obligation can be satisfied from an LSE’s own resources capable of providing Regulation, by bilateral arrangements with other market participants capable of providing Regulation, and/or by purchases of Regulation from the PJM Regulation Market. Each PJM LSE, or other Regulation buyer, is charged at the locational hourly Regulation Market Clearing Price (RMCP) applicable for their Regulation Zone for the amount of Regulation purchased to meet their hourly Regulation Obligation. In addition, net purchasers of Regulation in an hour are also charged a proportionate share of any opportunity cost credits paid to regulating generators for unrecovered costs over and above their RMCP payments.

Each LSE participating in PJM also incurs a Day-Ahead Scheduling Reserve Obligation equal to its load ratio share within PJM times the total amount of additional Operating Reserve required by PJM. This obligation can be satisfied from an LSE’s own resources capable of providing Operating Reserve, by bilateral arrangements with other market participants capable of providing Operating Reserve, and/or by purchases of Operating Reserve from the DASRM [39].
3.4 ISO-NE Operating Reserve Acquisition

The ISO-NE supports decoupled co-optimizations for the procurement of energy and OR [19, 20, 21, 22, 23, 25, 26]. Long-term forward commitments and prices for OR are secured through a Forward Reserve Market (FRM) that takes into account system-wide and locational reserve requirements. Day-ahead energy prices (LMPs) together with energy generation commitments and scheduled dispatch levels are determined through a Day-Ahead Market (DAM) similar in its design and operation to PJM’s DAEM. Real-time prices and dispatch levels for energy are determined through a Real-Time Market (RTM) that includes system-wide and locational reserve requirements. A separate optimization is conducted to activate and price OR in real time that takes into account the opportunity cost of providing OR rather than energy, calculated in part on the basis of RTM LMP outcomes.

Because the FRM is unique to ISO-NE, its operation will now be explained in more detail. As depicted in Figure 3, twice a year the ISO-NE manages an FRM for the acquisition of forward commitments to OR for delivery in real time. A needed input for the FRM are reserve requirements.

The ISO-NE determines Locational Reserve Requirements for each of its three OR products for three ISO-NE interfaces (import-constrained areas), and determines prices for each of these three OR products at four pricing locations (called Reserve Zones). The Locational Reserve Requirement for an interface is the amount of 30-minute contingency response (given the available transfer capability for the interface) that must be physically located within the interface to ensure recovery from the loss of a generator or a second line in the local area. The ISO-NE also determines System-Wide Reserve Requirements. The Ten-Minute Reserve Requirement (including TMSR and TMNSR) equals the largest First Contingency in the system. The Total Thirty-Minute Operating Reserve Requirement (including TMSR, TMNSR, and TMOR) is equal to the sum of 100% of the largest First Contingency and 50% of the largest Second Contingency in the system.

The Forward Reserve Market includes a Forward Reserve Auction to acquire, in advance, capability to supply required OR to meet the reserve requirements in each Reserve Zone in real time. Market Participants participate in the Forward Reserve Auction by submitting supply offers for their reserve resources at prices at least as high as the current Forward Reserve Threshold Price. The ISO-NE changes the Forward Reserve Threshold Price each month, based on heat rates and fuel price indices.

A market participant whose offers have cleared in the Forward Reserve Auction receives a Forward Reserve Obligation for each Reserve Zone equal to the amount of that market participant’s Forward Reserve Auction Offers that cleared in the auction. A market participant’s Forward Reserve Obligation also reflects any bilateral transactions for forward reserve that the market participant may have concluded. To meet their Forward Reserve Obligations, market participants must assign Forward Reserve to their Forward Reserve Resources on a daily basis at any time prior to start of an operating day such that the aggregate assignments are greater than or equal to their Forward Reserve Obligations for this operating day. The ISO-NE allocates its procurement

---

14As indicated in Table 2, the ISO-NE divides its Operating Reserve (OR) into three categories: Ten-Minute Spinning Reserve (TMSR); Ten-Minute Non-Spinning Reserve (TMNSR); and Thirty-Minute Operating Reserve (TMOR).
Determination of Forward Reserve Requirements

Submittal of Forward Reserve Auction Offers

Forward Reserve Auction Clearing

Internal Bilateral Transactions between Participants

Assignment by Participants of Specific Resource to Meet Forward Reserve Obligations

Submission of Bid Data / Self-Schedule for Units Assigned a Forward Reserve Obligation

Determination of Forward Reserve MW Delivered to the Real-Time Market:
- determine Forward Reserve Failure-to-Reserve Penalty
- determine Forward Reserve Failure-to-Activate Penalty

Market Settlement / Billing

Figure 3. ISO-NE Locational Forward Reserve Market, High-Level Process Flow
costs in the Forward Reserve Market to market participants on the basis of their Real-Time Load Obligations within their applicable Load Zones.

For completeness, it should be noted that the ISO-NE also supports a decoupled co-optimization for the real-time procurement of energy and Regulation [25, Section 3] [26, Section III]. This decoupled co-optimization is briefly summarized as follows. The ISO-NE determines and posts a Regulation requirement for each hour of each operating day on the basis of historical control performance. Generation owners wishing to sell Regulation service are required to submit supply offers and operating characteristics to the ISO-NE by hour 18 on the day prior to the operating day. The ISO-NE uses these offers to calculate, for each operating hour, a real-time Regulation Clearing Price (RCP) together with scheduled Regulation assignments to satisfy its Regulation requirements. The RCP is used to determine the Time-on-Regulation Credits and Regulation Service Credits awarded to providers of Regulation.\footnote{As explained more carefully in [25, Section 3], Time-on-Regulation Credit for any hour is calculated as Time-on-Regulation Megawatts (capacity availability) multiplied by \( \max(\text{RCP, Regulation offer price}) \). Regulation Service Megawatts for any hour are determined as the sum of the absolute value of positive and negative Regulation adjustments requested by the ISO for that hour. Regulation Service Credit is then calculated as \( \max(\text{RCP, Regulation offer price}) \) multiplied by Regulation Service Megawatts multiplied by a “capacity-to-service ratio”.} ISO-scheduled providers of Regulation also receive compensation for Regulation Opportunity Cost. The unit-specific Regulation Opportunity Cost is calculated ex-post for each unit that the ISO schedules to provide Regulation during an operating hour, using the final real-time LMPs calculated for use in the real-time energy market settlements for that hour.

Each market participant has an hourly Regulation Obligation equal to its pro rata share of the New England Control Area Regulation requirements for the hour, based on the market participant’s total real-time load obligation in the New England Control Area for the hour. A market participant may satisfy its Regulation Obligation by self-scheduling its own resources capable of performing Regulation service, by internal bilateral transactions with other market participants, or by purchases from the Regulation Market.

As detailed in [26, Section 4], each market participant that purchases Regulation from the Regulation Market is charged a pro-rata share of the Time-on-Regulation Credits and Regulation Service Credits based on its Adjusted Regulation Obligation, i.e., its Regulation Obligation adjusted downward by the amount of its net Regulation purchases through internal bilateral transactions (if any). In addition, defining Net Regulation Purchase as Adjusted Regulation Obligation minus self-scheduled Regulation, each market participant with a positive Net Regulation Purchase is charged a pro-rata share of any Regulation Opportunity Cost.

### 3.5 ERCOT Operating Reserve Acquisition

ERCOT supports the integrated co-optimization of energy, Regulation Up/Down, Responsive Reserve, and Non-Spinning Reserve in a \textit{Day-Ahead Market (DAM)} \cite{9, 10, 12, 13}. The operation of ERCOT’s DAM is similar to the operation of PJM’s DAEM except that ERCOT’s DAM permits supply offers for reserve as well as demand bids and supply offers for energy.
For each operating day D, ERCOT conducts DAM operations as follows. At least one day prior to day D, ERCOT analyzes the expected load conditions for day D and develops an Ancillary Service (A/S) Plan that identifies the reserve requirements it deems to be necessary for each hour of day D. ERCOT assigns part of these reserve requirements, by service and by hour, to each LSE based on its Load Ratio Share and then aggregates these requirements, by service and by hour, to the level of Qualified Scheduling Entities (QSEs). The resulting reserve requirement for each QSE, by service and by hour, is called its Ancillary Service (A/S) Obligation.

From hour 00 of day D-1 through the close of the DAM at hour 10 of day D-1, each QSE can submit energy demand bids, energy supply offers, and supply offers for Regulation Up/Down, Responsive Reserve, and Non-Spinning Reserve into the DAM for each hour of day D. ERCOT sets the hourly requirements (demands) for reserve for each hour of day D based on its A/S Plan for day D. At hour 10 of day D-1, ERCOT starts the DAM clearing process. This process results in hourly prices and hourly scheduled commitments and dispatch levels for energy and reserve for each hour of day D. The costs of the reserve procured by ERCOT through the DAM on day D-1 are allocated to the QSEs on the basis of their A/S Obligations.

ERCOT load resources can participate in the provision of reserve through the DAM. A QSE can submit load resource-specific supply offers for Regulation Up/Down, Responsive Reserve, and Non-Spinning Reserve, and it can offer the same load resource capacity for any or all of these reserve products simultaneously.

For each operating hour H of day D, an Adjustment Period takes place starting from hour 18 of day D-1 up to the start of hour H [11]. During this Adjustment Period each QSE has an opportunity to adjust its resource commitments for hour H, including its DAM supply offers for energy and reserve, as more accurate information becomes available. QSEs can also submit supply offers for Replacement Reserve during the Adjustment Period [9]. Replacement Reserve is resources capable of providing additional balancing energy services to ERCOT to handle capacity insufficiency, either because they are expected to be off-line during the required period or because they are load that is available for interruption during the required period. ERCOT attempts to procure any needed Replacement Reserve at minimum feasible cost, based on lowest offered supply prices.

During the Adjustment Period leading up to an operating hour H on day D, ERCOT continues to evaluate system sufficiency and security by use of Hour-Ahead Reliability Unit Commitment (HARUC) processes. If ERCOT determines that more reserve is needed for hour H than is currently provided for, it notifies each QSE of its increased A/S Obligation. ERCOT allocates additional A/S Obligations to QSEs using the same percentages as the original day-ahead allocation of A/S Obligations.

If the current QSE self-arranged reserve quantities are insufficient to meet these increased A/S Obligations for hour H, ERCOT can notify QSEs at least two hours in advance of hour H of its intention to hold a Supplemental Ancillary Services Market (SASM) in order to facilitate the procurement of the needed additional A/S Obligations. Each QSE can submit to the SASM additional self-arranged reserve quantities limited to its additional A/S Obligation. ERCOT then

---

16In the ERCOT system, only QSEs are allowed to bid or offer energy and ancillary services in the Day-Ahead Market.
posts the reserve amounts cleared in the SASM together with the market clearing reserve prices. The SASM price for each type of reserve is the shadow price for the corresponding inequality constraint indicating the minimum required amount for that type of reserve.

3.6 MISO Operating Reserve Acquisition

MISO supports the integrated co-optimization of energy and Operating Reserve (OR) in both a Day-Ahead Market (DAM) and a Real-Time Market (RTM) [28, 29, 30, 31]. The operation of MISO’s DAM and RTM is similar to the operation of PJM’s DAEM and RTEM except that MISO’s DAM and RTM permit supply offers for OR.

For each operating day D, the MISO DAM is managed as follows. Between hour 00 and hour 11 of day D-1, market participants can submit supply offers for energy and OR (as well as demand bids for energy) into the DAM for each of the twenty-four hours of day D. Stored Energy Resources (SERs) are registered as Regulation Qualified Resources for DAM participation, and SERs can only submit supply offers for Regulation. The MISO constructs and uses market-wide and reserve zone OR demand curves for the DAM to ensure that energy and OR are priced in the DAM to reflect conditions when OR becomes scarce. After the DAM closes at hour 11, the DAM is cleared to produce a 24-hour schedule of hourly energy and OR commitment and dispatch levels, hourly energy price (LMP) levels, and hourly OR price levels for day D. This DAM schedule is publicly posted at hour 15.

During hour 16 of day D-1, the MISO conducts a Reliability Assessment Commitment (RAC) process to ensure there is enough capacity committed on day D-1 to cover the forecasted system load and OR requirements for day D. Both commitment and decommitment decisions are made during these RAC processes. Resources committed or decommitted during any RAC process are subject to make whole payments. The RAC process is executed several more times during the remainder of day D-1. During the RAC, available SERs not previously selected in the DAM or prior RAC processes are also considered based on their Regulating Reserve Offers in the DAM.

An RTM permitting co-optimization of imbalance energy and OR is run in parallel with the DAM and RAC processes. The RTM is cleared every five minutes to produce price and dispatch levels for both imbalance energy and OR.

Demand response resources (DRRs) are permitted to participate in the DAM and the RTM [28, 29]. DRR are divided into two categories, as follows.

- Type-1 DRRs: These DRRs may be qualified to supply energy at fixed target MW reduction (when committed), or to provide Contingency Reserve (when not committed), subject to shut-down and hourly curtailment costs, minimum interruption/non-interruption times, and exclusivity of energy or reserve dispatch.
- Type-2 DRRs: These DRRs may be qualified to provide energy and OR in the same manner as generation resources, with load treated as negative generation.
The allocation of costs for the procurement of OR is based on the obligations incurred by MISO market participants through their regional loads and exports. MISO provides percent-of-load values for each Reserve Zone that represent the percentage of the MISO Balancing Authority Area Load Forecast that resides within each Reserve Zone (the sum of all Reserve Zone percentages is 100%). Market participants can use this Load Forecast and Reserve Zone percentage data to estimate their OR obligation on both a market-wide and zonal basis. A market participant has the option of self-scheduling OR from qualified resources located anywhere within the MISO Balancing Authority Area to meet its obligations, or it can purchase these obligations directly from the energy and OR markets.

3.7 NYISO Operating Reserve Acquisition

NYISO manages a Day-Ahead Market (DAM) and a Real-Time Market (RTM) that permit the integrated co-optimization of energy, Regulation, and Operating Reserve [34, 35, 17]. The operation of these markets is similar to the operation of the DAM and RTM for MISO. Here we concentrate on the operation of the NYISO DAM and RT markets for the procurement of Regulation. A discussion for Operating Reserve would be similar.

Regulation can be offered into the DAM and/or the RTM by generating units (including Limited Energy Storage Resources) and demand-side resources whose energy production and/or energy demand can be raised or lowered by automatic generation control (AGC) to follow moment-by-moment changes in load. For each hour H of an operating day D, the NYISO establishes a Regulation demand curve that applies to both the DAM and the RTM for hour H.

On each day D-1, the DAM simultaneously produces a solution for hourly load, energy, Regulation, and other forms of reserve for each hour H of operating day D. In particular, the DAM produces a market clearing price for Regulation for each location and each hour H of day D. The DAM Regulation clearing price for hour H for a particular location equals the locational shadow price for the NYISO’s Regulation requirement constraint for hour H in this location. This locational shadow price takes into account the day-ahead Regulation supply offer of the marginal resource selected to provide Regulation (or the applicable price on the Regulation demand curve during shortage conditions), plus any margins on the sale of energy or Operating Reserve in the DAM that the resource would forego if scheduling it to provide additional Regulation would lead to it being scheduled to provide less energy or less Operating Reserve.

On each operating day D, the RTM simultaneously produces a solution for load, imbalance energy, regulation, and other reserve over a real-time dispatch (RTD) interval (e.g., sixty minutes). RTD runs normally occur every five minutes during day D. Consequently, RTM integrated co-optimization in the NYISO is conducted on a “rolling horizon” basis, with re-optimizations occurring every 5 minutes for planning horizons (RTD intervals) on the order of an hour.

The NYISO calculates an RTM clearing price for Regulation for each location during each RTD interval on each operating day D. The RTM Regulation clearing price for each location during each RTD interval equals the real-time locational shadow price for the NYISO’s Regulation re-
requirement constraint for that location during that RTD interval. This locational shadow price takes into account the real-time Regulation supply offer of the marginal resource selected to provide Regulation (or the applicable price on the Regulation demand curve during shortage conditions), plus any margins on the sale of energy or Operating Reserve in the RTM that the resource would forego if scheduling it to provide additional Regulation would lead to it being scheduled to provide less energy or less Operating Reserve.

All LSEs operating within the footprint of the NYISO Open Access Transmission Tariff (NYISO OATT) pay service charges to participate in the NYISO DAM and RTM. The charges for any hour of service are allocated to each LSE in proportion to its load ratio share for that hour.

### 3.8 CAISO Operating Reserve Acquisition

CAISO manages a Day-Ahead Market (DAM) and a Real-Time Market (RTM) for the integrated co-optimization of energy, Regulation Up/Down, Spinning Reserve, and Non-Spinning Reserve [3, 4, 5]. The operation of CAISO’s DAM and RTM is similar to the operation of the DAM and RTM in both MISO and NYISO.

Seven days before each operating day D, the DAM is opened and ready to accept virtual and physical bid/offer information from entities called Scheduling Coordinators (SC). Two days before day D, CAISO produces a forecast of CAISO demand. By hour 18 two days ahead of day D, CAISO publishes forecasted reserve requirements and regional constraints by Ancillary Service Region.

Any SC wishing to participate in the DAM for operating day D must submit its bids/offers prior to hour 10 on day D-1. These bids/offers include energy demand bids, energy supply offers, and supply offers for Regulation Up/Down, Spinning Reserve, and Non-Spinning Reserve. CAISO sets hourly reserve requirements for each of the twenty-four hours of day D based on its forecasted reserve requirements and regional constraints.

After the close of the DAM at hour 10 on day D-1, CAISO undertakes an Integrated Forward Market (IFM) process to determine the day-ahead schedule for energy prices (LMPs), energy commitment and dispatch levels, reserve prices, and reserve commitment and dispatch levels. After the completion of the IFM, CAISO carries out multi-interval real-time optimizations to minimize the cost of dispatching imbalance energy and procuring additional needed reserve, subject to resource and network constraints. The Hour-Ahead Scheduling Process (HASP) is included in a special hourly run of real-time unit commitment (RTUC). Reserve procurement in the HASP is done through an optimization process that is based on repeatedly updated system conditions. After the HASP closes for a particular operating hour H, the bids/offers for hour H are validated and a Market Power Mitigation and Reliability Requirement Determination (MPM-RRD) process is performed. Real-time dispatch levels and settlement prices for imbalance energy and reserve are then determined in the RTM for hour H.

The reserve procurement cost allocation for all reserve products is hourly, system-wide, and

31
across IFM, HASP, and RTMs. The cost of procuring reserve is viewed by CAISO as being on behalf of demand and is therefore allocated to demand using a system-wide user rate. The user rate for each form of reserve is the average cost of procuring this form of reserve in both the DAM and RTM for the whole CAISO system.

### 3.9 SPP Operating Reserve Acquisition (Planned)

The SPP is set to launch its *New Integrated Marketplace* on March 1, 2014 [47, 46, 48, 1]. The New Integrated Marketplace will support the integrated co-optimization of energy, Regulation Up/Down, Spinning Reserve, and Supplemental Reserve in both a *Day-Ahead Market (DAM)* and a *Real-Time Balancing Market (RTBM)*.

The planned operations of the DAM and RTBM closely resemble the operations of the day-ahead and real-time market processes in MISO, NYISO, and CAISO, as outlined in sections 3.6, 3.7, and 3.8. Consequently, these planned operations are only briefly summarized here.

On the morning of each day D-1, SPP market participants will be able to submit energy demand bids, energy supply offers, and reserve supply offers into a DAM for each hour of day D. The SPP will set the hourly reserve requirements for each of the twenty-four hours of day D. The DAM will then be cleared to produce a 24-hour schedule of hourly price, commitment, and dispatch levels for energy and reserve for day D.

After the close of the DAM, a *Reliability Unit Commitment (RUC)* process will be conducted to ensure there is enough capacity committed for day D to cover the forecasted system load and reserve requirements for day D. The RUC process will be executed several times (every four hours at a minimum) during the remainder of day D-1. Both commitment and decommitment decisions will be made during these RUC processes. Resources committed or decommitted during any RUC process will be subject to make whole payments. An RTBM permitting co-optimization of imbalance energy and reserve will be run in parallel with the DAM and RUC processes. The RTBM will be cleared every five minutes to produce dispatch and price levels for both imbalance energy and reserve.

The method for allocating the costs of procuring reserve from DAM, RUC, and RTBM operations will be identical, based on obligations incurred through regional loads and exports.
4 Direction for Future Research

The regulation and spinning reserve markets reviewed in this paper assume that resources bidding into the reserve markets are ramp-rate constrained. Such resources may have additional capacity to offer but are prohibited from doing so by the ramp rate limitation specified in the market rules.\textsuperscript{17} Thus, specifying the reserve capacity required is the equivalent of specifying the required ramp rate. If 50 MW of regulation reserve is acquired, and the nominated capacity must be reached within 5 minutes, then this means that the available ramp-rate would be 10 MW/minute.

Some types of EES devices have smaller capacities but extremely fast ramp rates and greater response accuracy as compared with conventional generators. These resources are capacity constrained, not ramp rate constrained. Current market rules assign the same capacity price to this and all other reserve resources, thereby providing no additional compensation for speed and accuracy of response.\textsuperscript{18} An EES device capable of offering 10 MW of regulation reserve, and having a 1000 MW/min ramp rate, would be treated by the market as if it had a 2 MW/min ramp rate, and be compensated the same as any other resource offering 10 MW of regulation reserve.

This approach may undermine a main advantage of fast response resources that may result in higher cost of producing electricity to meet demand. If system operators could take the fast ramp rate into account in required reserve calculations, it could reduce the amount of capacity required for regulation reserves. Furthermore, splitting the AGC signal would permit faster responsive resources to receive a faster signal and slower devices a slower signal. Thus, conventional generators might be enabled to operate with less headroom thereby reducing the cost of energy.

Market reform efforts across the U.S. are focused on reducing and/or removing the constraints to greater reserves market contribution by new resources, including demand curtailment offers. New market designs are needed that define market products in terms of the services required by the grid rather than on the inherent characteristics of the technologies used to supply the services. Thus, new approaches to definition of reserves might move away from the arbitrary definitions of reserve categories common in current market designs, most of which reference the response characteristics of the technologies thought best able to provide that category of reserves. Instead, a generalized reserve product might be defined such that it references the important parameters of response time and location, ramp rate, and duration of service delivery. This type of product specification leaves the field open to supply offers from any resource, including demand curtailment. It could involve splitting the AGC signal so that fast responsive resources receive one signal and slower responders receive yet another.

Future research could investigate the feasibility of collapsing the various commonly-used operating reserve categories into a single reserve market product as an alternative to having separate market products for regulation reserve, spinning reserve, quick-start reserve, etc. Using the elapsed time after a dispatch signal (commanding a change in energy production) has been sent may be a

\textsuperscript{17}The market’s ramp rate limitation arose from the technological limitation of the resources considered capable of providing ancillary services at the time that the market rule specifications were adopted.

\textsuperscript{18}This is the result of the application of the market principle of paying a single price to all cleared offers rather than paying cleared offerors as offered.
more efficient way to specify the reserves required for reliable operation. Resources would then submit their offers for change in energy production by a certain point in time. The market would select the mix of resources from those offered that satisfies the reserves specified and minimizes cost.

Another way this optimization could be done is to convert the grid operating reserve requirements into the frequency domain. The ramp rate of the various assets could be converted into a “frequency response.” The market would use this information, together with the price offered by the resources, to find the optimal mix of resources to meet the grids operating reserve requirements at least cost.

This is captured in Figure 4. The figure to the left depicts the operating reserve requirements in the frequency domain. The figure on the right shows the optimum selection of assets based on their frequency domain response and cost curves. This figure is a simplification. In reality, more than one asset would be called upon to meet the requirements in each frequency bin.

This approach of decoupling ramp-rate and capacity requirements in other words, specifying what operating reserves the grid needs for reliability, and allowing resources to offer what they are capable in terms of both ramp-rate and capacity – should lower the cost of reserves and improve the performance of the grid by better aligning the services provided with the technical requirements.

![Figure 4. Operating reserve requirements in the frequency domain.](image-url)
Operating reserve is an essential component of power systems. It enables system operators to keep system frequency within defined limits as part of normal operations, allowing generation to be adjusted to match unexpected changes in load and variable generation. It is also required in the event of a contingency to arrest sudden changes in frequency and to bring the frequency back to a more acceptable level.

For vertically-integrated utilities, providing operating reserve is simply part of the cost of providing electricity. However, in ISO/RTO-managed energy regions, the cost of providing operating reserve is not recovered as part of the price of electricity. Therefore, procedures must be in place that allow resources to be compensated for the supply of operating reserve.

The question posed at the outset of this survey is whether the operating reserve markets of the energy regions offer the incentives needed for the efficient, reliable operation of the grid even as the resource mix on the grid and resource technologies change. In our view, it is unlikely that these markets have all of the attributes necessary to promote efficiency and reliability in the face of change.

Currently, none of the seven U.S. ISO/RTO-managed energy regions sets hourly targets for primary frequency control, and suppliers of primary frequency control do not receive any compensation for this service. Only two of the seven regions make the ability to provide primary frequency control a criterion for participating in their operating reserve markets. As the fraction of energy from variable renewable generation increases, there will be a corresponding decrease in the fraction of energy provided by generators capable of providing primary frequency control response. In the future, the recognition of primary frequency control as an essential reserve product, compensated through market mechanisms, might become necessary.

Only three of the seven energy regions currently recognize regulation up and regulation down as separate reserve products. At times of low load, regulation up may be plentiful and regulation down may be in short supply. Having separate markets for regulation up and regulation down better reflects system conditions such as these, and therefore might be more efficient than a single market for regulation.

Finally, current rules across the seven energy regions allow storage and demand response resources to participate in operating reserve markets to varying degrees. In order to allow storage and/or demand response resources to participate in an unbiased manner, some of the energy regions have made modifications to their market rules while others have made exceptions to them. One aspect of these modifications involves rules on whether and to what extent a resource can participate in an operating reserve market; these rules define the minimum capacity and minimum operating duration of a resource. Another aspect deals with how resources are compensated for their market participation. In general, reserve resources are compensated based on capacity rather than actual performance, although FERC Order 755 requires this to be changed [16].

As discussed in previous sections, the seven U.S. ISO/RTO-managed energy regions have
adopted the same basic market design for energy trading: the two-settlement system recommended by [15] that involves the parallel operation of day-ahead and real-time energy markets. On the other hand, each of these regions currently maintains its own idiosyncratic definitions and acquisition practices for operating reserve products.

In our view, four key issues arise for future restructuring efforts in these energy regions. First, how should operating reserve and energy products be defined? Second, should these product definitions be standardized across the energy regions? Third, to what extent is it both desirable and practical to handle the procurement and pricing of these products through markets? Fourth, to what extent is it both desirable and practical to standardize the forms of these markets across the energy regions?

It is our hope that the current survey of operating reserve practices in the seven U.S. ISO/RTO-managed electric energy regions will facilitate the thoughtful consideration of these and other critical issues in future restructuring efforts.
References


[9] ERCOT. Ancillary service market transactions in the day-ahead and real time adjustment period, January 2011.


[36] NYISO. NYISO market services tariff, MST 15 rate schedules, November 2011.


38


This page intentionally left blank.