Managing Uncertain Intermittent Resource Output

in the Unit Commitment and Dispatch

Scott Harvey

6-17-2013

Rutgers Center for Research in Regulated Industries
26th Annual Western Conference

TABLE OF CONTENTS

Managing Uncertain Intermittent Resource Output in the Unit Commitment and Dispatch

Scott Harvey[1]

6-17-2013

I. Introduction

A fundamental characteristic of intermittent or variable energy resources is that their energy output capability varies from interval to interval with variations in their fuel input (e.g. variations in wind for wind turbines, variations in sunlight for solar generation resources). In order to balance load and generation on a short-term basis, the system operator, whether it is an ISO, RTO or vertically integrated utility, must dispatch other resources to compensate for these variations in intermittent resource output.

A second important characteristic of intermittent energy resources is that the variations of their energy output capability are not independent across the individual generating resources but tend to be correlated, in some circumstances highly correlated. This correlation has the consequence that the interval to interval or hour to hour variation in the overall output capability of intermittent resources can be material relative to their average output.[2]

A third important characteristic of the output capability of intermittent resources is that output varies in multiple time frames, there are longer term variations in expected output over the year and hours of the day (solar output capability will be higher in the summer and lower in the winter, will rise when the sun comes up and fall as it goes down, wind

---

[1] The author is or has been a consulting on electricity market design, transmission pricing and/or market power for the entities listed in endnote A. The views presented here are not necessarily attributable to any of those entities and any errors are solely the responsibility of the author. This paper has benefitted from a discussion of these issues over the past several years with Dhiman Chatterjee, Navid Nivad, Ryan Sutton of the Midwest ISO, Paul Gribik (now with PG&E but at the Midwest ISO during the initial development of these ideas), Lin Xu and Mark Rothleder of the California ISO, and my colleagues on the California ISO Market Surveillance Committee, Ben Hobbs, James Bushnell and Shmuel Oren as well as participants in a variety of stakeholder and industry meetings. The views expressed in this paper, however, reflect my way of looking at these issues and do not necessarily reflect the views of any of these individuals, nor do they necessarily reflect the views of the Midwest ISO, the California ISO or the collective views of the California ISO Market Surveillance Committee.

[2] For example, if a region had 4000 2.5 megawatt wind turbines with an average output of .75 megawatts (30% of capacity) whose 5 minute output was independently distributed, the total output of these resources would be tightly centered around 3000 megawatts and it would not be difficult for a system operator to manage their output variations. However, if the energy output capability of many of those 4000 turbines is correlated, the variations in total intermittent output that would need to be managed would be much larger. It is the correlation in variations in output over different resources, not the variations themselves, that is the key problem in using intermittent resources to reliably meet load.

generation capability will be lower in the summer and will rise at night and fall in the morning), and there is also a substantial unpredictable short-term intra hour variation in output capability.

This short and long term variability of intermittent resource output gives rise to two distinct reliability issues for balancing authority areas. The first issue is to determine the amount of non-intermittent capacity that will be needed to reliably meet load given the longer term variations in intermittent resource output. Resolving this first issue requires that the balancing authority area or utility take account of the variability of its load and intermittent resource output and the correlations between them to determine the amount of non-intermittent capacity that is needed to reliability meet load over the year. Because of the correlations between load and intermittent resource output (solar output is lower at night but so is load, solar output is lower on a cloudy summer day but so is load, wind output is lower on a hot summer day but load is high), this problem can be thought of as reducing to analyzing the probability distribution of net load, load less intermittent resource output. This is a complex question, particularly when there is limited historical data that can be used to analyze correlations between gross load and the output of various kinds of intermittent resources, but is not the focus of this paper.

The second issue is to determine the amount, and type, of non-intermittent resource capacity that needs to be on-line and available to balance load and generation in the time frame of the 5-minute economic dispatch. Maintaining reliability in this time frame requires both that the balancing authority area have enough capacity on line that can be dispatched up to replace reductions in intermittent resource output relative to the expected level and enough capacity that can be dispatched down to accommodate increases in intermittent resource output relative to the expected level.

This second issue must in turn be addressed in three time frames. The first time frame is the time frame of short-term unit commitment and scheduling decisions in which additional resources can be brought on line to meet load, units decommitted, pump storage units committed to pump or generate, and net interchange adjusted up or down.

These evaluations generally need to be made 30 minutes or more prior to the beginning of the operating interval. The second time frame is the time frame of the 5-minute economic dispatch in which the unit commitment is fixed and the level of net interchange is fixed except for dynamically scheduled resources. The third time frame is the 4 to 6 second time frame in which regulating resources are instructed to balance very short-term variations in load and generation.

The focus of this paper is on balancing load and generation in the time frame of the 5 minute economic dispatch. However, section two will discuss the trade off between balancing load and generation to the extent feasible in the time frame of the economic dispatch rather than in the 4 to 6 second time of regulation, and section three includes a background discussion of the recent evolution of the short-term unit commitment process in ISOs and RTOs.

Balancing authority area operators use three mechanisms to balance load and generation, the economic dispatch of generation on a 5, 10 or sometimes 15 minute time frame, the instruction of resources on regulation on a 4 to 6 second basis, and the activation of reserves. Section II discusses the potential use of regulation and reserves to balance the output of intermittent resource output. Section III discusses how changes in the intra-day unit commitment to better account for ramp capability can aid in balancing the variations in the output of intermittent resources. Section IV discusses how changes in the real-time dispatch to better account for ramp capability can be used to balance the variations in the output of intermittent resources.

II. Regulation, Reserves and Shortage Pricing

A. Regulation

Regulating capacity is used throughout North America to balance load and generation on a short-term basis, typically 4 to 6 seconds. Hence, regulating capacity could also be used to balance the longer term differences between load and generation resulting from variations in intermittent resource output. While this approach could be used to balance load and generation for the variations in intermittent resource output, it has four features that make it an unattractive alternative for balancing load and generation for variations in intermittent resource output in the time frame of the economic dispatch.

First, unlike the 5 or 10 minute dispatch which is based on security constrained least cost dispatch, regulation instructions generally do not take resource costs into account. Individual system operators, whether RTOs or transmission owning utilities, use a variety of rules to select which resources are sent regulation instructions, but these rules are generally based on criteria such as moving all resources in proportion to their ramp rate, or moving the fastest resources first. Hence, if a balancing authority area chose to use regulation to balance variations in intermittent resource, it would forgo balancing these variations at least incremental cost,[3] which would raise the overall cost of meeting load, relative to an approach based on economic dispatch. The larger the variations in net load being met with regulation instructions rather than economics dispatch, the larger the cost impact.

A long run approach to balancing generation and load in the 4 to 6 second time frame that might avoid this disadvantage would be to take account of incremental resource costs in providing instructions to regulating resources. In addition to requiring changes in existing regulation software, which is already complex, solving the optimization problem in the time frame of regulation instructions would be very challenging to say the least.[4]

A second disadvantage of using regulation for this purpose is that increases in the regulation requirement will generally not lead to a corresponding increase in the total available ramp capability available to balance load and generation in the 5 minute time frame. Hence, achieving a given increase in total ramp capability solely through an increase in the regulation requirement would require increasing the regulation requirement by much more than the increase in ramp capability that is sought. The

[3] Incremental cost in this time frame generally includes fuel cost, incremental operation and maintenance costs, and emissions costs but can also include a variety of opportunity costs and sometimes external costs in addition to emissions costs.

[4] A balancing authority area could of course address this by initializing its AGC software further in advance of the time the instructions are sent, but this would increase the disconnect between the regulation instruction and the actual load and generation balance and reduce the effectiveness of the regulation in maintaining ACE.

reason for this pattern is that a need to balance variations in intermittent resource in a 5 minute time frame requires an increase in the total available 5 minute ramp capability, relative to the 5 minute ramp capability that would otherwise be available. To the extent that there is 5 minute ramp capability available from on dispatch resources, the first order effect of an increase in the regulation requirement would not be an increase in the available 5 minute ramp capability but simply to shift some of the available 5 minute ramp capability from on dispatch resources to regulating resources. This would actually increase the frequency and magnitude of price spikes, by reducing the amount of ramp capability available in the economic dispatch, and raise the cost of meeting load, by replacing economic dispatch with regulation instructions.

Suppose for example, that a balancing authority area normally carries 400 megawatts of regulation (i.e. 400 megawatts of generation able to regulate upwards 400 megawatts in 5 minutes) and wanted to have an additional 200 megawatts of 5-minute ramp capability, beyond the 300 megawatts of ramp capability available from on-dispatch resources not originally scheduled to provide regulation. If the balancing authority area simply increased its regulation requirement by 200 megawatts, that would generally simply shift 200 megawatts of ramp capability from on-dispatch generation to resources scheduled to provide regulation, without increasing the total amount of 5 minute ramp capability available. Hence, achieving the intended goal could require either increasing the regulation requirement from 400 megawatts to 900 megawatts, or to both increase the regulation requirement and commit and dispatch additional rampable resources to maintain the amount of on dispatch ramp capability. Hence, increasing the regulation requirement does not necessarily address the need for additional ramp capability so would need to be combined with some other design changes to maintain on-dispatch ramp capability when the regulation requirement is increased.

Third, regulating requirements and capabilities are typically based on 5 minute ramping capability. Hence, 400 megawatts of regulation would be carried on capacity capable of ramping up 400 megawatts in a five-minute period. This 1-1 relationship between total undispatched capacity and ramp capability that characterizes regulating capacity may not be the ideal relationship for balancing variations in intermittent output. Although the nature of the ramping capability that high levels of intermittent resource output will ultimately require balancing authority areas to maintain is not yet fully understood and may not be well defined until there is more experience with actually managing systems with high levels of intermittent resource output, there are indications that the total ramp capability needed over a 15-30 minute period will be greater than the total ramp capability needed over a five minute period.

If this turns out to be the case, meeting these ramping requirements solely through 5 minute regulating capability would likely be unnecessarily high cost, entailing higher opportunity costs to maintain the needed ramp capability in real-time and likely also entailing higher unit commitment costs, then use of a mixture of on dispatch capacity able to provide ramp capability over a 5, 10, 15 minute or longer time frame. Moreover, use of regulation to balance serially correlated disturbances in intermittent resource output would hinder the use of fast, but energy limited resources such as flywheels and batteries to provide regulation because they would recurrently be drained of energy and pinned by sustained deviations in intermittent resource output. This could perhaps be avoided by introducing more complex software that would hold them back at times, but not only does this software not exist at present but it is not at all clear it would make sense to develop because it would in effect be combining dispatch with regulation.

Fourth, if a substantial additional amount of regulation were scheduled in order to balance variations in intermittent resource output, this would necessarily at the margin entail scheduling resources that incur higher costs in performing regulation and are less able to follow regulation instructions. If the incremental resources were only sent instructions when they are needed in addition to the current level of regulating resources the cost impact of the higher cost resources would only be felt in the periods in which the additional regulation was used. However, most ISOs and RTOs do not have regulation software that sends instructions to the resources with the lowest cost of moving. As a result, adding high cost resources to the pool of resources providing regulation would typically mean that low cost resources for providing regulation would be moved less than would otherwise be the case, replaced with the movement of higher cost resources. [5] This outcome could be avoided by designing and implementing regulation algorithms that prioritized instructions based on the each resources bid for its cost of providing regulation, i.e. movement but that software does not exist.

The bottom line is that 1) extensive changes in regulation software would be needed to avoid increasing the cost of meeting load, 2) some process would be needed to commit and dispatch additional capacity so that the amount of dispatchable capacity was increased, and 3) another process would be needed to commit capacity need to maintain ramp capability over a sequence of intervals.

B. Reserves

Another approach to balancing variations in intermittent resource output would be to dispatch reserves to compensate for reductions in intermittent resource output if there is insufficient ramp available on on-dispatch generation. Importantly, the concept would be to only use the ramp capability available on capacity providing reserves to balance reductions in intermittent resource output, while maintaining the target level of reserve capacity.

Suppose, for example, that a balancing authority area had 1000 megawatts of ten-minute reserves, 800 megawatts of intermittent resource output and 900 megawatts of undispatched capacity in addition to the reserves. This system has sufficient capacity on line to meet load and maintain the target level of reserves even if intermittent resource output falls to zero. However, if the amount of the 900 megawatts of undispatched capacity that could be dispatched up in a single 5 minute dispatch interval was only 300 megawatts, while the drop in intermittent output over a 5 minute period could be as large as 500 megawatts, then there would be a potential for the balancing authority area to be unable to balance load and generation in the real-time dispatch when there are large changes in intermittent resource output.

[5] Some ISOs and RTOs will be shifting under Order 755 to prioritizing regulation instructions to move the fastest units first. In some cases the fastest units may also be the lowest cost resources but this may not always be the case.

In such a situation, one approach to balancing load and generation in response to such a large change in intermittent resource output would be to make use of some of the ramp capability of the 10 minute reserves to balance load and generation when there are larger decreases in intermittent resource output than can be balanced with the ramp capability of on-dispatch generation. The reserves used to balance load and generation would then be restored in subsequent intervals as the on dispatch generation continued to ramp up and capacity constrained generation providing reserves could be ramped down to restore the level of 10 minute reserves.

There are a few limitations of relying on reserves to balance load and generation in this manner. First, if a large generating unit or transmission line tripped off line at the same time that there was a large drop in intermittent resource output, the balancing authority area might not have enough ramp capability to recover from the outage and balance the change in intermittent resource output within the required time frame.

Second, if off-line reserves were used to provide ramp capability in this manner, minimum down times might prevent the resources from being turned off when intermittent resources output recovers, requiring that other resources be dispatched down to accommodate the intermittent resource output. Because intermittent resource output can go up and down more than once over the day, reserves provided by off-line units with minimum down times would likely be too inflexible to provide a good source of ramp capability to balance routine variations in intermittent resource output. This limitation could be addressed by only using the ramp capability of on line spinning reserves to balance variations in intermittent resource output.

Third, the ramp capability, as opposed to simply capacity, of on-line spinning reserves may provide support to the stability of the interconnected grid in response to events external to the control area carrying the reserves. This stabilizing role might be comprised by temporary, but frequent, reductions in on-line spin in order to balance reductions in intermittent resource output.

Fourth, while reserves can be used to balance sudden decreases in intermittent resource output, they are not well suited to balancing sudden increases in intermittent resource output. Hence, another mechanism would be needed to balance increases in intermittent resource output.

There are, however, other mechanisms that can be used to balance large increases in wind output, at least from a reliability perspective. The Alberta Electric System Operator uses spinning reserves to balance reductions in intermittent resource output and manages increases in intermittent resource output through a rule that limits increases in intermittent resource output to the ramp available on dispatchable resources.[6]

Other ISOs and RTOs such as ERCOT, MISO, New York ISO and PJM have implemented economic dispatch of wind generation over the past few years, [7] and the California ISO and ISO New England are in the process of doing so.[8] The application of economic dispatch to intermittent resource output allows the ISOs and RTOs to dispatch intermittent resource output down when there is not enough ramp capability to absorb the full potential output of the intermittent resources. In some ISOs and RTOs introduction of wind dispatch has been accompanied by reductions in the bid floor so that the bid floor would be lower than the tax and other subsidies provided to wind generation.[9]

6 See ISO rules section 304.3 and Kris Aksomitis, Alberta Electric System Operator, "Short-Term Wind Integration," Recommendation Paper, September 23, 2010.

7 ERCOT, see "Operational Requirements for Managing Wind Generation." More than 50% of MISO wind generation was on dispatch by 9/1/2012 and most must be dispatchable by June 1, 2013, see David Patton, IMM Quarterly Report, summer 2012, September 2010 p. 50. New York ISO, see Technical Bulletin 154, revised 2/29/2012. PJM see Manual 12 and PJM, "Wind –Specific Requirements, August 10, 2012.

8 ISO New England filed its wind dispatch rules at FEERC on September 26, 2012 with implementation planned for 2014. The California ISO plans to introduce economic dispatch of wind generation in combination with introduction of 15 minute scheduling and changes to the PIRP program. See California ISO, "FERC Order 764 Compliance 15-Minute Scheduling and Settlement, Draft Final Proposal, March 26, 2012 section 7; California ISO, FERC Order 764 Compliance 15-Minute Scheduling and Settlement, Addendum to Draft Final Proposal, April 24, 2013 section 7.

9 PJM allowed supply to be offered at negative prices beginning on June 1, 2009, see Monitoring Analytics, State of the Market Report for PJM, Volume 2, 2011 march 15, 012 p. 174. The California ISO plans to reduce its bid floor in two steps, from $30 per megawatt hour to -$150 per megawatt hour and then to -$300 per megawatt hour, see California ISO, Renewable Integration Market Vision and Roadmap, October 10, 2011 p. 7. The first reduction in the California ISO bid floor was originally planned for implementation in fall 2012 but was pushed back until Fall 2013 because of delays in associated changes in the bid cost recovery design. ISO New England currently has a bid floor of zero but is in the process of reducing it, see Aleks Mitreski, ISO New England, Hourly Offers and Intraday Reoffers, Markets Committee, October 10, 20112 p. 30; The New York ISO's bid floor is -$1,000 per megawatt

Most ISOs and RTOs also have rules allowing the output of intermittent and other producers to be curtailed down to the resource's physical minimum operating point when prices fall to the bid floor and generation still exceeds load, either in aggregate or within a generation pocket.[10] These curtailment mechanisms are in general not automated, particularly when applied to generator pockets. Hence, reliance on these curtailment mechanisms to manage variations in intermittent output on an ongoing gasis can be problematic from an operational perspective.

Overall, system operators, and particularly ISOs and RTOs, are better able to manage the reliability impacts of increases in potential intermittent resource output through the application of ramp rules such as those used by Alberta, the application of economic dispatch, and in the last resort through curtailment of intermittent generation, than to manage sudden decreases in the output of these resources.

Nevertheless there is an economic cost associated with extreme negative prices due to lack of sufficient capability to accommodate increases in intermittent resource output, and the associated increase in production cost (dispatching negatively priced generation down raises total production cost). The economic consequences of these negative prices may fall on the intermittent resource owner or on those who have contracted with the owner, and will also fall on other resources that cannot reduce their output to zero when prices fall to extremely negative levels. While these extreme negative prices send the correct price signal in the very short-run, the long-run impact of many ramp driven negative price spikes may be costly for consumers. In addition, the negative prices send a short-term signal for decommitment that may not be consistent with reliability needs, because minimum down times can make decommitted generation unavailable when

hour, MISO -$500 per megawatt hour, and ERCOT -$250 per megawatt hour and no reductions are under consideration.

[10] Alberta, ISO rules section 202.5 and Alberta Electric System Operator, Supply Surplus Discussion Paper, December 2, 2010; California ISO, Renewable Integration: Market and Product Review, Fourth Revised Straw Proposal, August 22, 2011 p. 12, ERCOT, Operational Requirements for Managing Wind Generation, July 19, 2010; MISO Tariff Module C, Section 40.2.21; PJM Wind Specific Requirements, August 10, 2010 and Manual 12 attachment B; ISO New England, see Aleks Mitreski, ISO New Engalnd, Hourly Offers and Intraday Reoffers, Markets Committee, October 10, 20112 pp. 27-29; New York ISO Technical Bulletin 154 revised February 29, 2012

intermittent resource output falls. Hence, dispatch methods that allow variations in intermittent resource output to be accommodated with less extreme variations in prices and avoiding decommittment of generation potentially needed to maintain reliability in future intervals would be beneficial if they can be implemented in a cost effective manner.

C. Shortage Pricing

While regulation and the ramp available on spinning reserves should not be the primary means of balancing the output of intermittent resources, it does make economic sense for them to be used to balance occasional extreme events. This should be combined with appropriate shortage pricing, so that events that cause regulation or the ramp available on spinning reserves to be used to balance intermittent resource output when there is not enough ramp capability available on on-dispatch resources sends an appropriate price signal.

Several ISOs and RTOs currently employ real-time ancillary service pricing that sends a price signal when there is inadequate ramp capability. Both the New York ISO, Midwest ISO, and ISO New England optimize ancillary services in the real-time dispatch and have explicit penalty prices that impact both energy and ancillary service prices when ramp capability on spinning reserves or regulation is used to balance load and generation in the economic dispatch.

The Midwest ISO releases 150 megawatts of spinning reserves at a penalty price of $65 per megawatt and releases additional spinning reserve at a penalty price of $98 per megawatt. The New York ISO on the other hand, releasing up to 25 megawatts of regulation at a penalty price of $80 per megawatt, up to 55 megawatts at a penalty price of $180 per megawatt, and additional regulation at a penalty price of $400 per megawatt. The New York ISO will also shift spinning reserves from east to west at a penalty price of $25, release western 10 minute reserves at a price of $450 per megawatt, and eastern 10 minute reserves at a price of $500 per megawatt.

The California ISO releases regulation and spinning reserves in its look-ahead commitment program (RTPD or RTUC) but not in its real-time dispatch, This design can produce ramp shortages in the real-time dispatch that can cause real-time prices to spike to $1,000 per megawatt hour, the penalty price for a load balance violation in the real-time dispatch. These price spikes are at least in part a result of the inconsistency in the California ISO shortage pricing design in which spin could be released in RTPD at a shadow price of $100 per megawatt hour, but if the need for the ramp capability is not foreseen in RTPD, the capacity scheduled to provide spinning reserves in RTPD is not available to provide ramp capability at any price in RTD. Moreover, because the location of spinning reserves is determined in RTPD, RTD is not able to shift spinning reserves across resources to make additional ramp capability free within constrained regions in real-time.

PJM has implemented real-time reserve shortage pricing on October 2, 2012, with initial shortage values of $250 per megawatt (rising over a couple of years to $850 per megawatt) for both synchronized and non-synchronized reserves. PJM applies these penalty price two 2 zones, the RTO zone (PJM as a whole) and the MAC Dominion sub-zone.

ISO New England has shortage prices of $50 for spinning reserves and a shortage price of $850 for 10 minute reserves.[11]

D. Time Interval Pricing

Another element of real-time pricing designs that is relevant to the supply of ramp capability to accommodate variations in intermittent resource output is the time interval used for settlements. A number of ISOs and RTOs settle both load and generation based on time weighted average hourly prices. This is the simplest pricing system to implement and is sometimes viewed as avoiding the need for additional revenue quality metering. This time weighted pricing system is used by the Midwest ISO, PJM, ISO New England

[11] ISO New England also has shortage prices for 30 minute operating reserves but shortages of these reserves reflect a lack of capacity, not a shortage of ramping capability.

and SPP.[12] The New York ISO on the other hand, settles generators based on 5 minute interval prices,[13] the California ISO settles payments to generators based on 10 minute interval prices,[14] and ERCOT settles payments to generators based on 15 minute interval prices.

A key problem with time weighted average hourly pricing systems is that they do not provide generators or dispatchable power consumers with efficient incentives to respond to high or low prices at the point in time when they are high or low and can even incent perverse output responses at the end of an hour, i.e. increases in output when prices were high early in the hour but are low late in the hour.

In addition, time weighted hourly pricing systems do not pay as much to generators responding to dispatch instructions by increasing or decreasing their output as do output weighted pricing systems so provide less incentive for generators to increase their ramp rate or perhaps even to participate in the economic dispatch. Table 1 illustrates the differential revenues for a generator varying its output over three periods with varying prices, under time weighted and interval or output weighted pricing. Column A shows the output in each of the three periods, for two generators with the same average output over the hour. Columns B and C show the prices and revenues under interval or output weighted pricing, while columns D and E show the prices and revenues under time weighted pricing. The example shows that the revenues of the generator that does not vary its output are the same under both pricing systems and the same as the generator that increased its output under time weighted pricing, but the generator that follows dispatch instructions up and down has higher revenues for the same average output under interval or output weighted pricing.

[12] SPP will be moving to five minute pricing in conjunction with implementation of its day-ahead market, see Richard Dillon, SPP, Integrated Marketplace, June 22, 2012 p. 7

[13] The New York ISO has hourly settlement prices but it effectively settles generators based on their five minute output as the hourly price for each generator is computed as the output weighted average, where the output weights are that generator's output in each five minute interval.

[14] The California ISO is considering moving to 5 minute pricing for internal generation in conjunction with its implementation of 15 minute interchange scheduling, see California ISO, FERC Order 764 Compliance 15-Minute Scheduling and Settlement, Straw Proposal, October 23, 2012 p. 14.

Table 1: Illustration of Time Weighted Pricing

Output	Interval Price	Revenue	Time Weighted Price	Revenue
A	**B**	**C**	**D**	**E**
Generator Following Dispatch Instructions				
80	10	266.6667	20	533.3333
100	20	666.6667	20	666.6667
120	30	1200	20	800
Total		2133.333		2000
Generator with Fixed Output				
100	10	333.3333	20	666.6667
100	20	666.6667	20	666.6667
100	30	1000	20	666.6667
Total		2000		2000

While shortage pricing and output weighted or interval pricing do not directly contribute to accommodating increasing levels of intermittent resource output, they contribute to this goal by more accurately reflecting the economic value of faster ramping generating resources and increased ramp capability.

III. Intra Day Unit Commitment, Interchange Scheduling, and Ramp Capability

A. Introduction

Another potential approach to ensuring the availability of sufficient ramp capability to efficiently balance variations in intermittent resource output is through adjustments to the commitment of resources during the operating day, and by taking ramp capability into account in adjusting the level of net interchange. This approach requires that the ISO or

RTO look forward in time and anticipate the need for additional ramp capability in making these decisions. Subsection B discusses the introduction of look-ahead unit commitment programs in the RTOs and the limitations these programs have as a mechanism for managing the variability of intermittent resource output.

Subsection C then turns to a discussion of the design changes that the California introduced in late 2011 in their look-ahead commitment program in order to reduce the frequency and cost of ramp capability driven price spikes. This subsection also discusses the limitations with this design that the California ISO has encountered.

Finally, subsection D discusses how 15 minute scheduling can either help manage the variability of intermittent resource output or contribute to it, depending on the process used to schedule interchange.

B. Intra-day Unit Commitment Mechanisms

Most US ISOs and RTOs have implemented some kind of process for the ISO or RTO to commit non-quick start generating capacity during the operating day to supplement self-commitment decisions by market participants. These ISO or RTO commitment decisions may be in response to load forecast error, to changes in net interchange, to generation or transmission outages or to changes in intermittent resource output. The criteria for commitment include both maintaining reliability and economic commitments based on production cost minimization.

The ability of ISOs and RTOs to commit generation, rather than relying entirely on market participant self-commitment decisions, can be beneficial because in committing resources the ISO or RTO can 1) take account of transmission congestion that is projected to exist at the time the resources would come on line; 2) take account of other unit commitment decisions; and 3) potentially take advantage of more up to date forecasts for net load. Resources committed by ISOs or RTOs using these intra-day commitment process receive a generation cost guarantee. Resources can also be self-committed by market participants in most ISO and RTO markets, but such self-committed resources do not receive a generator cost guarantee covering their start-up and minimum load costs.

The earliest version of these intra-day unit commitment processes, the New York ISO's Balancing Market Evaluation, often referred to as BME, evaluated the economic and reliability needs of the New York market. BME was used to schedule inter-change and commit 30-minute gas turbines, and was run once an hour and looking forward 3 hours in hour-long blocks. The New York ISO replaced its hourly Balancing Market Evaluation with a forward evaluation mechanism called RTC in 2005. RTC economically evaluates the commitment of resources that can come on line in 30 minutes or less (which includes gas turbines, combined cycles and pumped storage resources), and schedules net interchange. A significant innovation in the design of RTC relative to the New York ISO's Balancing Market Evaluation is that RTC runs every 15 minutes and evaluates operating conditions in 15 minute, rather than hour long segments.

The use of 15-minute time segments in intra-day unit commitment mechanisms has a number of advantages in terms of more efficient unit commitment and pricing, some of which are relevant to balancing variations in intermittent resource output. First, by evaluating load and resources in 15-minute increments, resources that are off-line can be committed closer to the period of time in which their capacity is needed, rather than at the start of the hour. Starting units needed to meet load near the end of the hour later in the hour reduces uplift costs because the units are less likely to be out of merit, and thereby also avoids distorting prices at the beginning of the hour.

Second, the use of 4 periods to evaluate either hourly imports or resources with hour long minimum run-times, results in a more accurate economic evaluation because the four periods allow the peak load to differ from the average load, enabling the peak load used to determine commitments for reliability to differ from the average load that would determine economic commitments.

Third, deferring unit commitment decisions until closer to the time resources are projected to be needed reduces the frequency with which units are committed uneconomically due to load forecast error or changes in intermittent output. Fourth, the use of 15-minute time increments has the potential to identify ramping constraints that would not be apparent using hour-long blocks. In practice, however, most ramping constraints only show up in the 5 minute time frame so would not identified in these forward looking evaluations, even if the changes in supply and demand were anticipated.

Another limitation of these programs in managing variations in intermittent resource output is that it typically takes 10-15 minutes for them to solve, then another 30 minutes or so to start resources other than quick start gas turbines. Moreover, the programs only initialize every 15 minutes, so a change in intermittent output could occur nearly 15 minutes before the program initializes for its next solution. The combined effect of these lags is that reaction time of these programs to changes in intermittent resource output is too long for them to be used to directly manage short-term variations in intermittent resource output which often cannot be predicted 30-55 minutes in advance.

The California ISO implemented a look-ahead unit commitment process similar to RTC in April 2009 as part of its MRTU market design. Like RTC, this program (referred to sometimes as Real-Time Unit Commitment (RTUC) and sometimes as Real-Time Pre-dispatch (RTPD) runs every 15 minutes and evaluates interchange schedules and unit commitment decisions in 15 minute time segments.

Midwest ISO operators initially used ad hoc processes to commit units during the operating day following the Midwest ISO start-up in 2005. [15]The Midwest ISO implemented a look-ahead commitment program similar to that used by the California ISO and New York ISO on April 1, 2012.[16]

---

[15] Units committed by the Midwest ISO operators to maintain ramp capability or for other reasons are made whole for any start-up and minimum load costs that are not recovered in their energy and ancillary service revenues through a generator cost guarantee. The associated uplift costs are referred to as RSG (revenue sufficiency guarantee) in the Midwest ISO.

[16] See Potomac Economics, 2011 State of the Market Report for the MISO Electricity Markets, June 2012 p. 64.

PJM has recently implemented a look-ahead program called "Intermediate Security Constrained Economic Dispatch" (IT SCED), which looks forward 2 hours in 2 15 minute increments and 2 45 minute increments to aid in committing generation, as well as to apply the three pivotal supplier test and identify reserve shortages, among other things.[17]

The intra-day unit commitment (and decommittment) programs of these ISOs and RTOs are similar in that they evaluate unit commitment decisions on a production cost minimization basis, taking account of the load forecast, incremental energy costs, start-up costs, minimum load blocks, and minimum run times.

While these look-ahead programs provide a platform that could potentially be used to aid managing the variability of intermittent resource output over the operating day, the bottom line is that the conventional implementation of these look-ahead unit commitment programs is not very useful for managing variations in intermittent resource output. There are two key reasons for this conclusion. First, the time lag between when a change in intermittent resource would need to be projected and the time additional capacity could come on line, is too long to help with managing many intra-day variations in intermittent resource output.

Second, even the shorter 15-minute time intervals used in the newer forward evaluation designs do not do a very good job of identifying upcoming ramp constraints. While if the programs foresaw a large change in intermittent resource output, they likely would be able to identify any potential capacity shortage this would produce, they likely would not identify any problems the change in output would create due to ramp constraints in the time frame of the 5 minute economic dispatch. This is because the 15 minutes of ramp available over the interval would likely be larger than the change in net load over the 15 minute period. Moreover, even a sustained ramp constraint lasting 15 or 20 minutes might not be foreseen in these 15 minute evaluations, because it would likely fall over

[17] See PJM, Manual 11, "Energy & Ancillary Services Market Operations," November 29, 2012 pp. 33-34.

more than one 15-minute evaluation period, and the ramp constraint would not be evident with 30 minutes of ramping capability available to manage it.

C. Intra-day Unit Commitment for Ramp Capability

The next evolution of intra-day unit commitment programs has been to try to adapt them for use in managing the variability of intermittent resource output, rather than simply committing generation to meet peak load at least cost. Thus, some ISOs and RTOs are attempting to use these programs as a platform to ensure that enough ramp capability is on-line to accommodate changes in intermittent resource output, not just to ensure that enough total capacity is on-line to meet load. Achieving this goal requires evaluating whether there is enough ramp capability on line to accommodate potential upward and downward ramp capability requirements associated with unpredictable changes in net load.

The Midwest ISO has attempted to manage this in an ad hoc manner by committing resources to maintain "headroom" above the capacity needed to meet forecasted net load.

California implemented a change in its look-ahead unit commitment program (RTPD) in December 2011, introducing an upward ramp capability target, referred to as the "flexible ramping constraint."[18] The core concept of this design is to add to the objective function of the program a target for maintaining ramp capability above that needed to meet the forecasted change in net load, so that an additional margin would be available to meet unforeseen variations in net load, such as those due to unexpected changes in intermittent resource output. The key choice variable the program can use to meet this target is to commit additional generation, but it also dispatches generation out of merit in order to maintain additional ramp capability and the ramp capability target can also be met by varying hourly interchange schedules (a higher level of imports backs down internal generation, creating more upward ramp capability).

[18] See California ISO October 7, 2011 filing in docket ER12-5-000., California ISO, "Opportunity Cost of Flexible Ramping Constraint, Draft Final Proposal," July 20, 2011, and 137 FERC ¶61,191, December 12, 2011 . At present, the adequacy of downward ramp capability is not considered in RTPD.

While this design has had some success in reducing the frequency of real-time load balance violations and the associated price spikes, its performance has illustrated the complexity of successfully implementing such a look-ahead commitment program to manage ramp constraints. The California ISO Department of Market Monitoring has found that the commitment of capacity based on the ramp constraint has been relatively high cost in terms of load payments.[19] Data on the incremental production cost changes in the cost of meeting load resulting from the design have apparently not been studied.[20]Moreover, the program has been less effective than hoped in avoiding energy balance violations due to ramp constraints.[21]

Some of the reasons for the limited success and high cost of this design are likely:

- the ramp capability calculated to be available in RTPD through out of merit dispatch is not maintained in the real-time dispatch;
- the evaluation of ramp capability in RTPD does not account for transmission constraints;
- the California ISO is still improving the criteria used to set the ramp capability targets;[22]
- the California ISO is using a very high penalty factor for a failure to meet the ramp capability target.

[19] See, for example, California ISO, Department of Market Monitoring, Q3 2012 Report on Market Issues and Performance, November 13, 2012 pp. 41-44; and 2012 Annual Report on Market Issues and Performance, April 2013 p. 86. Generators evaluated as providing ramp capability in the RTPD evaluation are paid the shadow price of incremental ramp capability in RTPD schedules (determined by opportunity costs or by a penalty value if the target amount of ramp cannot be scheduled). Because the California ISO has not implemented a real-time ramp capability based dispatch, the ramp capability modeled in RTPD is not necessarily available in real-time as discussed below. If any calculations of the net production cost impact of this constraint have been carried out they have not been made public.

[20] It would be useful to evaluate the economics of the incremental ramp capability on a production costs basis, but evaluating the impact of changes in the unit commitment would require running the program with a different ramp capability penalty value to identify instances in which the high penalty value lead to the commitment of additional generation, then simulating the economic dispatch with the different unit commitment.

[21] See California ISO, Department of Market Monitoring, Q2 Report on Market Issues and Performance, August 14, 2012 p. 35-38, Q3 Report on Market Issues and Performance, November 13, 2012 pp. 41-44; and 2012 Annual Report on Market Issues and Performance, April 2013 p. 79-83.

[22] The California ISO has been adjusting the ramp capability based target over time as it has gained experience with this design but is not yet able to set targets based on time of day, projected change in load, and the level of intermittent output (if intermittent output is high, substantial upward ramp capability is much more likely to be needed than if intermittent output is low). California ISO, Department of Market Monitoring, Q3 2012 Report on Market Issues and Performance, November 13, 2012 p. 41. 2012 Annual Report on Market Issues and Performance, April 2013 p. 85.

The failure to actually dispatch the system in real-time the way RTPD assumes in evaluating ramp capability means that the system may have less, potentially a lot less, ramp capability available in real-time than calculated in the RTPD evaluation. When RTPD calculates the ramp capability that will be available in future periods and compares those quantities to the ramp capability target, RTPD may “create” ramp capability for the next period by dispatching generation down out of merit in the current period. The “dispatch” in RTPD is used only for evaluating unit commitment and scheduling decisions, however, it has no impact on the actual real-time dispatch.

Because the ramp capability target used in RTPD is at present not included in the objective function for the real-time dispatch, generation will not be dispatched down out of merit in this manner to maintain ramp capability in real-time, so no ramp capability “created” by out-of-market dispatch in RTPD will actually be available in real-time. Hence, any time generation is dispatched out of merit in RTPD in order to meet the ramp capability target, less ramp capability will be available in real-time than projected in RTPD. Data compiled by the California ISO relating to ramping capacity shadow prices is consistent with less ramp capability being available in real-time than calculated in RTPD due to this consideration but the magnitude of the effect is not clear. [23] Similarly, data compiled by the California ISO Department of Market Monitoring shows that the ramping constraint bound in far more intervals than the number of intervals in which there was a procurement shortfall, so the presence of a binding constraint and non-zero shadow prices generally meant that generation was being dispatched down out of merit in RTPD in order to meet the ramp capability target.[24]

Another factor limiting the effectiveness of this design has been the failure of the ramp capability target to account for potential transmission constraints. This failure to take account of transmission constraints has meant that RTPD may commit units to provide

23 See Lin Xu and Don Tretheway, California ISO, “Flexible Ramping Product,” Market Surveillance Committee Meeting October 19, 2012, pp. 9-12.

24 See California ISO, Department of Market Monitoring, 2012 Annual Report on Market Issues and Performance, April 2013 p. 86.

ramp capability or count on too much ramp capability in locations at which it has a low opportunity cost, but also limited value in providing ramp capability because it is on the wrong side of important transmission constraints.[25]

The third factor limiting the effectiveness of this design has been the need to gain experience in setting the appropriate ramp capability target. The potential need for ramp is hard to analyze and there will be learning over time in assessing the amount of ramp capability that is likely to be needed at different times of day and year, under different system conditions, differing load levels, and different projected levels of intermittent resource output (the higher the projected level of intermittent resource output, the more upward ramp capability could be needed to replace declines in output and the less downward ramp capability could be needed to accommodate further increases in intermittent resource output and the converse when projected intermittent resource output is low).

Unfortunately the accuracy of the target is critical to the cost and operational effectiveness of a ramp capability based commitment. If the target is set too high when the ramp capability is less likely to be used, the excess commitment costs incurred when the ramp capability turns out after the fact to not be needed will swamp the economic benefits realized when the incremental ramp capability is used. Conversely, if the target is set too low when it is likely to be used, too few benefits will be realized to compensate for the costs when the capability is not used.

The difficulties the California ISO has encountered in choosing ramp capability targets that perform well over time suggest that it may be necessary to develop a model of the need for ramp that allow the target to modified from day to day and perhaps hour to hour as conditions change rather than relying on discrete changes in the target.

[25] California ISO, Department of Market Monitoring, Q3 2012 Report on Market Issues and Performance, November 13, 2012 pp. 44-45; 2012 Annual Report on Market Issues and Performance, April 2013 pp. 88-89. The California ISO has not yet begun analysis that would support the application of regional ramp capability targets.

The high cost of the flexible ramping constraint as currently implemented by the California ISO reflects in part the $250 per megawatt penalty value used to schedule ramp. This is an extremely high value. For ramp capability with a shadow price of $250 per megawatt to be economic at the margin, incremental rampable capacity would need to dispatched with a high probability and for large production costs savings. While the fact that the California ISO does not release spinning reserves or regulation in the real-time dispatch and uses a $1000 per megawatt penalty for power balance violations implies that avoiding these violation has a much higher value than in other RTOs, this is still a very high penalty value and contributes to the high costs of the ramp capability procured.

The experience of the California ISO shows that the cost effective and operationally effective incorporation of ramp constraints in forward unit commitment evaluations is complex and may need to be accompanied by additional design changes such as real-time ramp capability based dispatch and locational ramp targets in order to achieve the intended goals in a cost effective manner.

While the California ISO's flexible ramping constraint in RTPD has not performed as well as was hoped, the 1.5 years of operational experience can be analyzed to provide valuable insights both into improving the design of the constraint and understanding the factors that will determine the success of a ramp capability based dispatch. In particular, it is possible to utilize historic data to analyze the extent to which the performance problems are due either to the failure to enforce the constraint in the real-time dispatch, and hence would be resolved by implementation of a ramp capability based dispatch in real-time, are due to the failure to take account of transmission congestion in scheduling ramp, or are due to other factors, such as the complexity of setting a cost effective target for ramp capability, or the inability to commit additional capacity.

The California ISO could identify all instances of ramp constrained price spikes above a specified threshold, such as $500, and determine: 1) how often was less than the target amount of capacity procured because of its high cost (more than $250 per megawatt), suggesting the need for changes in HASP (to schedule more interchange) or the day-ahead market (to commit more non-quick start generation); 2) how much of the ramp capability identified in RTPD was produced by out-of-merit dispatch in RTPD and hence not available in real-time because the capacity was already dispatched to its upper limit, but would be available in real-time had a real-time ramp capability dispatch been in operation; 3) how much of the ramp capability identified in RTPD was not be dispatched in real-time because of transmission constraints, 4) how often was the target apparently simply too low, and 5) does the data suggest another factor limiting the effectiveness of the design.

This analysis could lead to near term and longer term improvements in the cost effectiveness of the ramp capability commitment. If the price spikes are associated with much less ramp capability actually being available in real-time than in RTPD because of the failure to dispatch generation out of merit in real-time, this would demonstrate the importance of moving forward to implement a ramp capability based dispatch. Alternatively, if the price spikes tend to be associated with a large proportion of the ramp capability in RTPD being undispatched due to congestion, then it would be important to introduce some locational targets into the RTPD procurement. Other findings could lead to other diagnoses.

While the California ISO's implementation of ramp capability evaluation in its look-ahead commitment program is limited to considering the supply of up ramp, this kind of design can also be used to maintain down ramp capability. While down ramp capability might at times be managed be decommitting units, a more important contribution to cost effectively maintaining down ramp capability could be in the scheduling of net interchange by ISO's such as the California ISO, New York ISO and Ontario IESO, that schedule net interchange based on bids and offers in their forward evaluations.

Scheduling additional imports during low load hours will tend to back more units down to their minimums, reducing both the downward ramp available in a dispatch interval and the total amount of downward dispatch capacity available. The traditional formulation of the programs used to economically evaluate interchange schedules treats the load forecast as a given, rather than as an estimate subject to error. Hence, they can find it economic to schedule imports up to the point at which every resource is dispatched down just barely above its minimum load point, in order to save pennies. When the interchange scheduling programs set imports at levels that back dispatchable resources down to their economic minimums based on the load forecast, however, then whenever the load forecast turns out to be too high, the system will run out of downward ramp, causing a downward price spike, and perhaps causing a power balance violation in California or a reduction in downward regulating capacity in New York.[26] The potential for deterministic interchange scheduling programs to create these kind of problems will increase as intermittent output increases if the expected net load is treated as a given in scheduling net interchange.

Adding a downward ramp capability target with a small penalty value to these interchange scheduling programs, would deter the look-ahead scheduling software from scheduling imports that would reduce ramp capability below the target level unless there were enough cost savings to warrant the risk of running out of downward ramp capability.

Scheduling too high a level of imports during low load hours is less of a concern for ISOs and RTOs such as PJM, MISO, SPP, and ISO New England that rely largely on price taking imports. Importers exposed to downward price risk would be cautious about scheduling too many imports during hours in which the system would potentially run out of downward ramp capability or dispatch down resources with negative offer prices to balance load and generation.[27] However, in these systems, the individual importers have less visibility into system conditions than the ISO or RTO so may unintentionally schedule imports that pin generation resources at their minimum, exposing the system to load balance violations and downward price spikes.

D. 15 Minute Interchange Scheduling

---

[26] In New York and Ontario import suppliers are insulated from the consequences of these downward price spikes by import offer guarantees. In California, import suppliers sell at the HASP price so are not impacted by downward price spikes in real-time.

[27] Until PJM allowed negative prices in 2009, however, imports were not exposed to the potential for negative prices, requiring PJM to manage the level of imports by curtailing transactions in checkout.

Traditional hourly interchange scheduling processes contribute to problems in balancing variations in intermittent output in two ways. First, because hourly interchange schedules are adjusted over the 10 or 20 minutes at the top of the hour, less ramping capability is available during this period to compensate for variations in intermittent resource output in the same direction, i.e. if decreased import schedules are accompanied by decreased intermittent resource output or if increased import schedules are accompanied by increased intermittent resource output.

Second, the longer time frame of hourly interchange schedules provides essentially no flexibility for interchange to be adjusted to compensate for changes in intermittent resource output. We noted above that the California ISO's implementation of a ramp capability target in its look ahead scheduling process (RTPD) really only has one choice variable to adjust in the relevant time frame in order to maintain the target level of ramp capability, which is the commitment of quick starting units. If interchange could be adjusted on a shorter term basis than the current hourly time frame, this would enable ISOs and RTOs to use changes in net interchange as a second tool for adjusting the level of net interchange in order to maintain a target level of ramp capability.

Some ISOs and RTOs, such as the Midwest ISO and PJM, have historically allowed interchange schedules to be adjusted by market participants every 15 minutes, roughly 20 minutes prior to the start of the period. This design offers the potential to spread the ramping of interchange schedules over the hour, reducing the potential for severe ramp constraints during the top of the hour ramping periods.

However, because in the MISO and PJM markets changes in net interchange are determined by market participants, rather than by the ISO or RTO, the ISOs and RTOs cannot use short-term changes in net interchange as a tool for maintaining a target level of ramp capability. On the contrary, in these designs 15 minute scheduling has the potential to contribute to ramp constraints and require the commitment of generation to supply additional ramp capability because of the potential for sudden unpredictable changes in interchange schedules every 15 minutes. This has been a problem for a number of years in the Midwest ISO. [28]

In the last few years the New York ISO has introduced a different type of 15 minute scheduling that does not contribute to ramp problems (for the New York ISO) on its Hydro Quebec and PJM interfaces. Under the New York design, market participants submit bids and offers that are fixed for the hour and the New York ISO adjusts interchange schedules based on these bids and offers every 15 minutes. Thus, it is the New York ISO, rather than the market participant, that initiates changes in interchange schedules during the hour. This design can potentially help the New York ISO balance variations in the output of intermittent resources by adjusting net interchange in response to actual or projected changes in intermittent resource output. Although the frequency of adjustment is not fast or frequent enough to allow the New York ISO to use interchange adjustments to directly response to changes in intermittent resource outout, the design would allow the New York ISO to adjust interchange in increments every 15 minutes to maintain a target level of ramp capability[29] and would also reduce the likelihood of top of the hour ramp constraints due to large changes in net interchange.

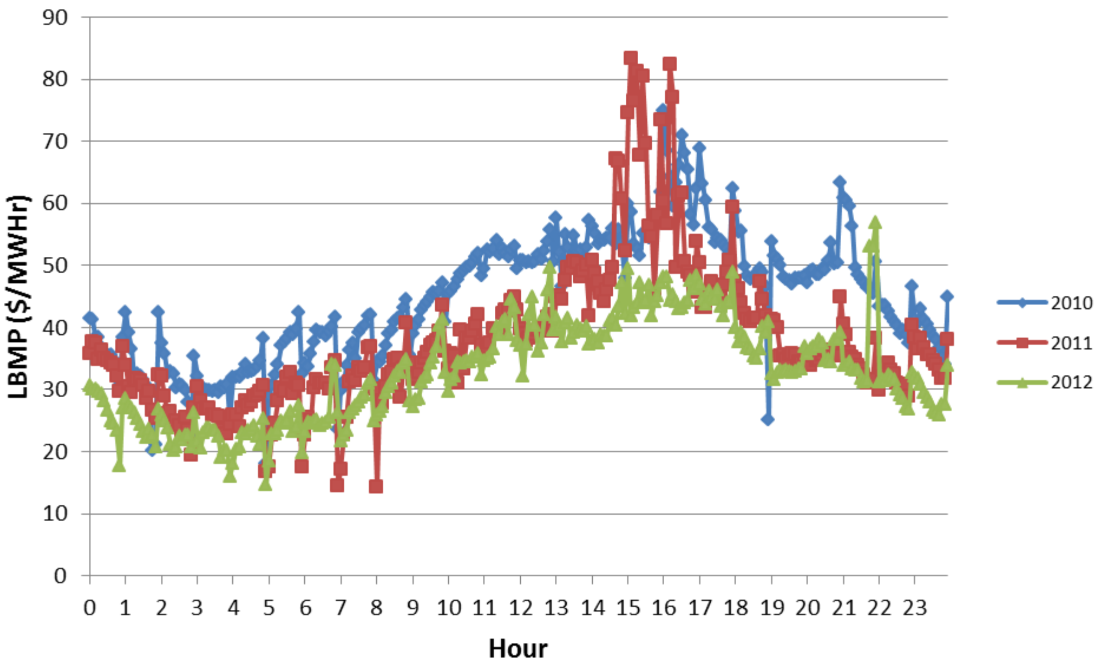
Figure 2 shows the pattern of intra-hour price volatility in New York. It can be seen that during the morning ramp up when imports are increasing, the real-time price plunges at the top of the hour, then rises throughout each hour before plunging at the beginning of the next hour. There is a somewhat less pronounced pattern of price spikes, followed by gradual declines at the end of the day when the level of imports is being reduced. The data in the figure suggests that the introduction of 15 minute scheduling in 2011 with Hydro Quebec and in 2012 with PJM may have somewhat reduced the intra-hour ramp cycle in New York, but perhaps because most interchange is still scheduled hourly, the impact has not been dramatic.

---

[28] See, for example, the discussion in Potomac Economics, 2007 State of the Market Report for the Midwest ISO, pp. 122-123. 2008 State of the Market Report for the Midwest ISO, p. 141, 2006 State of the Market Report, the Midwest ISO July 2007, pp. 120-123. This volatility in import schedules is in part due to the MISO pricing interchange on an hourly basis while allowing schedule changes on a 15 minute basis, but that is not the only source of the volatility in interchange schedules.

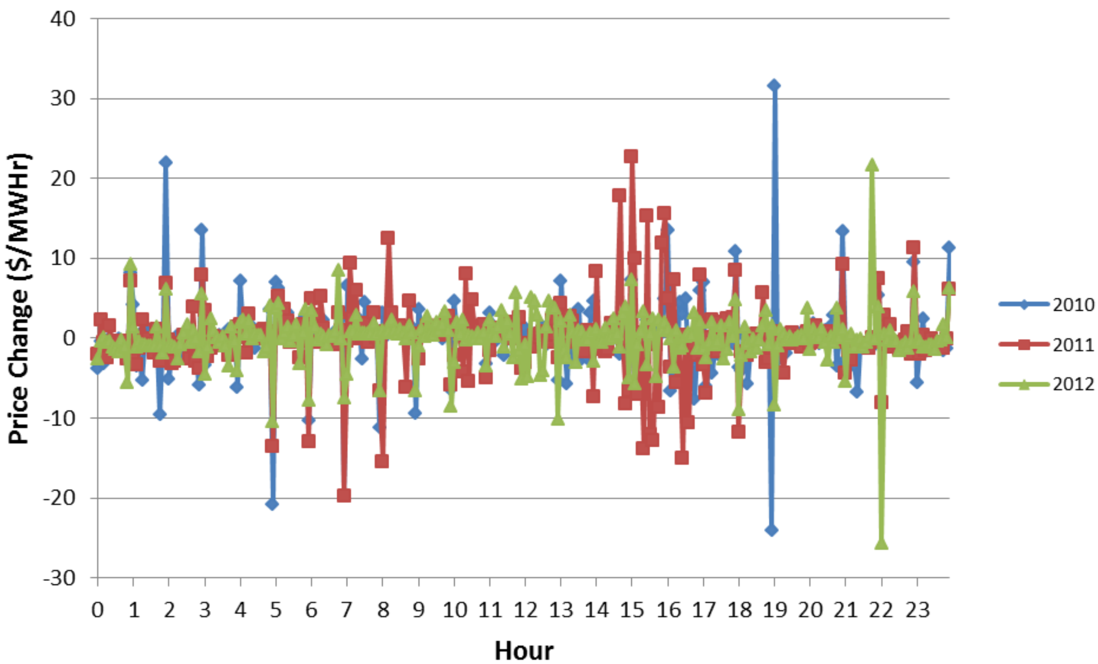
[29] At present, however, the New York ISO is not developing such a design.

Figure 2 NYISO Real-Time Interval Level Price Averages for August (Ontario Proxy Bus)



The data is perhaps somewhat easier to evaluate in Figure 3 which shows the change in the real-time price from interval to interval. The data portrayed in Figure 3 also suggests that the reduction in volatility between 2010 and 2012 has not been very dramatic.

Figure 3: NYISO Average Real-Time Price Change by Interval for August (Ontario Proxy Bus)



A key limitation of the New York ISO design for 15 minute scheduling from the standpoint of balancing variations in intermittent resource output, in addition to the time lags, is that the design only helps the ISO or RTO doing the scheduling, not the balancing authority area sourcing or sinking the other end of the interchange transaction. Under this design the New York ISO is adjusting interchange every 15 minutes based on the bids and offers submitted to the New York ISO by market participants. These changes can be used to help the New York ISO balance variations in intermittent resource output or to maintain a target level of ramp capability. However, since the changes are determined by the New York ISO, the more frequent schedule changes do not help Hydro Quebec or PJM maintain ramp capability. In fact, the schedule change that is good for the New York ISO may exacerbate ramp problems in PJM.[30]

[30] The schedule changes are less likely to cause ramp problems for Hydro Quebec because the Hydro Quebec system is largely hydro with substantial ramping capability in a normal operating range and the bids and offers submitted at the Hydro Quebec proxy bus are presumably made taking into account the ramp available on the system resources.

In the U.S., FERC's Order 764 now requires jurisdictional utilities to allow 15 minute interchange scheduling. This shorter time frame for scheduling interchange will facilitate scheduling of interchange supported by intermittent resources located outside ISOs and RTOs, which do not have access to a real-time spot market. The intermittent resource producers may at present be selling their output using hourly transmission schedules, exposing them to imbalance charges in the source balancing authority area on the difference between their hourly schedule and their actual real-time output, and imposing the need to balance these deviations on the balancing authority area in which they are located.

Order 764 will be much less significant for intermittent resources located in ISOs and RTOs, as these resources already have access to a real-time spot market; hence there is no need for them to make use of 15 minute scheduling of interchange.

The introduction of 15 minute scheduling by non-ISO and RTO balancing authority areas will not, however, contribute to reducing balancing issues associated with intermittent resource output but will simply shift the balancing issues from the source balancing authority area to the sink balancing authority area. This will make it even more important for ISOs and RTOs likely to be impacted by increased imports supported by intermittent resources to have sufficient ramp capability to manage high levels of unpredictable changes both in their internal net load and in net interchange. In practice, this is likely to primarily be an issue that the California ISO will need to address and to a lesser extent the MISO and Southwest Power Pool.

While the form of 15 minute scheduling implemented by the New York ISO with PJM and Hydro Quebec enables the New York ISO to evaluate interchange bids and offers closer in time to the operating interval, the bids and offers it uses in this evaluation must be submitted 75 minutes prior to the operating hour and may be far out of line with conditions in the other RTO by the end of the operating hour. As a result, this type of 15 minute scheduling only allows interchange schedules to respond to changed market conditions in one of the involved ISOs, in this case the New York ISO.

One way to enable interchange schedules to respond more quickly to changed market conditions in either the source or sink ISO or RTO would be for the ISOs and RTOs to jointly adjust interchange schedules based on ISO and RTO projections of market conditions rather than based on market participant bids.

The New York ISO and ISO New England have filed with FERC to implement such a coordinated interchange scheduling system between the regions,[31] and FERC has accepted the changes to be effective when the coordinated interchange process is implemented.[32] Under this design, market participants seeking to schedule interchange would submit bids reflecting the difference in projected prices at which they would be willing to have their transaction scheduled. Market participants could thereby choose to submit interchange schedules that would only be scheduled if the price difference were projected to exceed their target, but under the current design there would be no production cost guarantee for the scheduling market participant if the real-time price difference turned out to be lower than projected.

The ISOs and RTOs would adjust interchange every 15 minutes based on their price projections. While this design is projected for implementation between the New York ISO and ISO New England in 2015, the New York ISO has been working with ISO New England to implement coordinated interchange scheduling for more than a decade [33] and implementation is still at least two years away. The long delay in implementing coordinated interchange scheduling reflects in part the complexity in developing processes to optimize interchange between two distinct system operators and the need for ISO New England to implement a look-ahead process that can be used in the scheduling process.[34]

---

31 New York ISO filing Dec 28, 2011 Docket ER12-701-000

32 See 139 FERC ¶61,048 April 19, 2012.

33 Scott Harvey, "Inter-ISO Dispatch Proposal," NYISO Market Structures Working Group, January 14, 2003 and Scott Harvey, "Inter-ISO Dispatch Proposal," NEPOOL Markets Committee, February 11, 2003.

34 Another possible approach to implementing 15 minute scheduling would be to allow market participants to change their offer prices every 15 minutes. This approach is not being pursued at present. Some of the issues with this approach are allowing market participants to change their bids and offers after the affected ISOs and RTOs had determined the commitment for units with 30 minute start times and hour long import schedules could increase rather than decrease price volatility and could create incentives for opportunistic changes in bids and offers.

The New York ISO and PJM more recently began working on implementing coordinated interchange schedules and have been moving forward at a fairly rapid pace, with a time frame for implementation in 2014 (part of the reason for the more rapid implementation than with ISO New England is that PJM has already implemented a look-ahead process that can be used for interchange scheduling, ITSCED).[35] Under the proposed design, CTS bids and offers would be submitted along with conventional economic bids 75 minutes prior to the beginning of the operating hour, but unlike the current design both CTS and economic bids could change in bid price and megawatt quantity every 15 minutes.[36] In recent month, the New York ISO and PJM have been studying the convergence between these look-ahead projections and real-time prices.[37]

The development of interchange scheduling processes that will jointly determine the level of net interchange has been motivated by a goal of reducing the overall cost of meeting load by better converging the incremental cost of power between adjacent ISOs and RTOs. However, these coordinated interchange scheduling processes would also address the limitations of 15 minute interchange schedules determined by only one of the impacted balancing authority areas as a balancing mechanism for ISOs and RTOs seeking to manage variations in intermittent resource output. Coordinated interchange scheduling would enable changes in interchange to be managed in a manner that helps balancing, or at least does not exacerbate variations in intermittent output, in both the source and sink balancing authority area.

Coordinated interchange may also be helpful in managing interchange scheduling issues that may emerge as the level of intermittent resource capability rises. Increases in intermittent resource output tend to naturally create the unloaded capacity needed to meet a subsequent drop in intermittent output. However, if the increases and decreases in

---

35 See PJM and New York ISO, "Coordinated Transaction Scheduling (CTS) between NYISO & PJM – Proposal Kickoff," Joint NYISO PJM Meeting, November 28, 2012.

36 See PJM and New York ISO, PJM and New York ISO, Coordinated Transaction Scheduling (CTS) between NYISO & PJM – Third Joint Meeting, April 2, 2013; "Coordinated Transaction Scheduling (CTS) between NYISO & PJM – Proposal Kickoff," Joint NYISO PJM Meeting, November 28, 2012. p. 10.

37 See PJM and New York ISO, Coordinated Transaction Scheduling (CTS) between NYISO & PJM – Third Joint Meeting, April 2, 2013.

intermittent resource output are large, there is a potential for the lower prices when intermittent resource output is high to cause reductions in net imports, which will erode the capacity needed to meet load when intermittent resource output declines. The long time lines of the current processes for market participant driven interchange adjustments may therefore not be well suited to market based interchange scheduling systems at higher levels of intermittent resource output. Coordinated interchange processes may at least be part of the solution to managing these situations.

Finally, a further step in using interchange to adjust to changes in intermittent resource output would be to move to adjusting interchange on a five-minute basis. This approach is under development in three regions. First, the New York ISO and Hydro Quebec have considered eventually implementing 5-minute dispatch, using the DC interconnection to utilize the capabilities of Hydro Quebec's hydro system in managing variations in intermittent output and New York, while providing Hydro Quebec with a low cost source of power to back down hydro generation.

Second, the MISO has been working with Manitoba Hydro on implementing five minute dispatch between the MISO and Manitoba Hydro system, which is linked to the MISO by a DC interconnection and is a largely hydro system, similarly offering the potential for benefits to the MISO in managing intermittent output and to Manitoba Hydro as a low cost source of power. Third, the California ISO's EIM would introduce 5 minute scheduling between the California ISO and participating adjacent balancing authority areas. [38]

The bottom line is that price based interchange scheduling coordinated by the ISO or RTO can contribute to managing the variability of intermittent resources in three ways. First, it has the potential to reduce the amount of ramp capability needed to accommodate large changes in net interchange at the top of the hour, reducing the impact of variations in intermittent output during this period. Second, it has the potential to enable ISOs and RTOs to adjust net interchange, with a lag, to changes in intermittent output, perhaps somewhat cushioning the effect of large changes in intermittent output. Third, it could be

[38] See, California ISO, Energy Imbalance Market, Revised Straw Proposal, May 30, 2013.

combined with a look-ahead ramp capability target to maintain a cushion of ramp capability to respond to unexpected variations in intermittent resource output. However, it is complex to implement and no ISOs and RTOs have yet implemented such a design, much less implemented a successful design.

IV. Dispatch Optimization

A. Introduction

The next time frame in which variations in intermittent output can be managed is the time frame of the economic dispatch, generally based on 5 or 10-minute dispatch instructions. The initial development of economic dispatch was based on single period optimization, in which the value of the objective function of the dispatch depends only on the cost of reliably meeting load in the current interval. This structure meant that the dispatch sometimes did not position units the way that would be lowest cost for meeting large changes predictable changes in net load in subsequent intervals, such as would occur during the morning load pickup or the evening load drop off.

Operators have used and continue to use a variety of ad hoc mechanisms to manage ramp in these circumstances. One method is to dispatch slow moving units up out of merit prior to the beginning of the rapid rise in load, getting them to a higher output, and perhaps also taking fast moving units down out of merit in preparation for the upcoming rise in load, while letting the remaining units be managed by the economic dispatch. All of the units would then be moved up when the load pickup occurs.

This ad hoc approach can work reasonably well for managing morning and evening load changes in a cost based dispatch, because the operators will be able to observe through trial and error over time which units make good choices for being moved out of merit in this manner without creating other problems This kind of ad hoc approach will work less well in a market based dispatch in which taking units out of market may have financial impacts and if predictable, may affect how resources bid.

Another method used to maintain additional ramp capability is to bias the load forecast in the economic dispatch so that on dispatch generation is dispatched to a higher level than needed to meet actual load, with regulating resources being backed down to balance load and generation. If the regulating resources are fast moving, capacity constrained resources, this can create additional ramp capability when load begins to rise rapidly. However, if the economic dispatch sets price, this approach results in artificially high prices when the load forecast is biased up and can even create price spikes when none would otherwise have occurred. Moreover, it may not accomplish much in providing additional ramp capability if the regulating resources are ramp constrained, rather than capacity constrained.

B. Multiple Interval Dispatch Optimization

Given the limitations of these ad hoc approaches, particularly in market based dispatches, some ISOs and RTOs have begun to take another approach to managing such predictable future variations in net load. This approach has been to avoid the need for ad hoc operator actions by factoring these changes in future load into the economic dispatch, reducing the need for the operators to manage these variations through ad hoc actions.

The Ontario IESO was the first ISO to implement a form of multi-interval optimization that could take account of projected future changes in load in its real-time dispatch in 2004. The New York ISO implemented multi-interval optimization in its real-time dispatch in 2005 as did the California ISO in April 2009. [39] These forward looking dispatch tools can potentially avoid or reduce the extent of real-time ramp capability driven price spikes, but only to the extent that the price spikes can be foreseen. This forward looking dispatch capability is most useful for managing predictable changes in load (such as the morning ramp up or evening ramp down) or predictable changes in supply (such as pump storage units going on or off line or large changes in scheduled net interchange.

[39] The market monitor for the Midwest ISO has recommended for several years that the Midwest ISO implement such a multi-interval optimization but this has not yet been done. See Potomac Economics 2011 State of the Market Report for the MISO Electricity Markets, June 2012, p. 64; 2010 State of the Market Report for the MISO Electricity Markets, June 2011, p. xxvi; 2009 State of the Market Report for the Midwest ISO, p. xxiii.

It needs to be kept in mind that the multiple interval dispatch optimization extends the solution time for the real-time dispatch software, which increases the time lag between the time the dispatch software is initialized based on system conditions and the time dispatch instructions can be send to resources. ERCOT does not utilize multiple interval optimization nor does it jointly optimize energy and ancillary service schedules, which reduces the solution time for its dispatch software.

While optimizing the dispatch over time to meet expected variations in net load can reduce the likelihood of small unpredictable variations creating a price spike, this form of multi-interval optimization is likely to be of limited value in balancing the output of intermittent resources because the predictable changes may be swamped by unpredictable changes. This has led some ISOs and RTOs with substantial, and growing, intermittent resource output to evaluate alternative dispatch concepts that go beyond conventional multi-interval optimization in order to better manage the reliability challenges posed by high levels of intermittent resource output. This is the topic of the next section.

C. Ramp Capability Based Dispatch

1. Ramp Capability Based Dispatch Optimization

With the increased importance of intermittent generation some U.S. ISOs and RTOs have begun to examine the development of dispatch concepts that would better position the transmission system to respond to unpredictable changes in net load (load net of intermittent resource output). In particular, the Midwest ISO and California ISO have been developing the concept of a ramp capability based dispatch which would not only seek to minimize the cost of meeting load in the current dispatch interval, but would also dispatch the system so as to preserve incremental ramp capability for use in responding to unexpected changes in net load in subsequent intervals.[40]

[40] See Midwest ISO, Ramp Capability in MISO Markets, Stakeholder 5th Technical Workshop, April 14, 2012; Lin Xu and Donald Tretheway, California ISO, "Flexible Ramping Products, Revised Draft Final Proposal," August 9, 2012.

The ramp capacity available on an electric system at any point in time depends on the ramp rates of on-line units at their current operating point and by the upper or lower limits of the resources. Resources that have been dispatched to their upper limit have no upward ramp capability, regardless of their nominal ramp rate. Similarly, resources that have been dispatched down to their lower limit have no available downward ramp capability, regardless of their nominal ramp rate. Hence, additional ramp capability can be preserved for future intervals by dispatching capacity limited resources out of merit either below their upper limit (to maintain up ramp capability) or above their lower limit (to maintain down ramp capability).

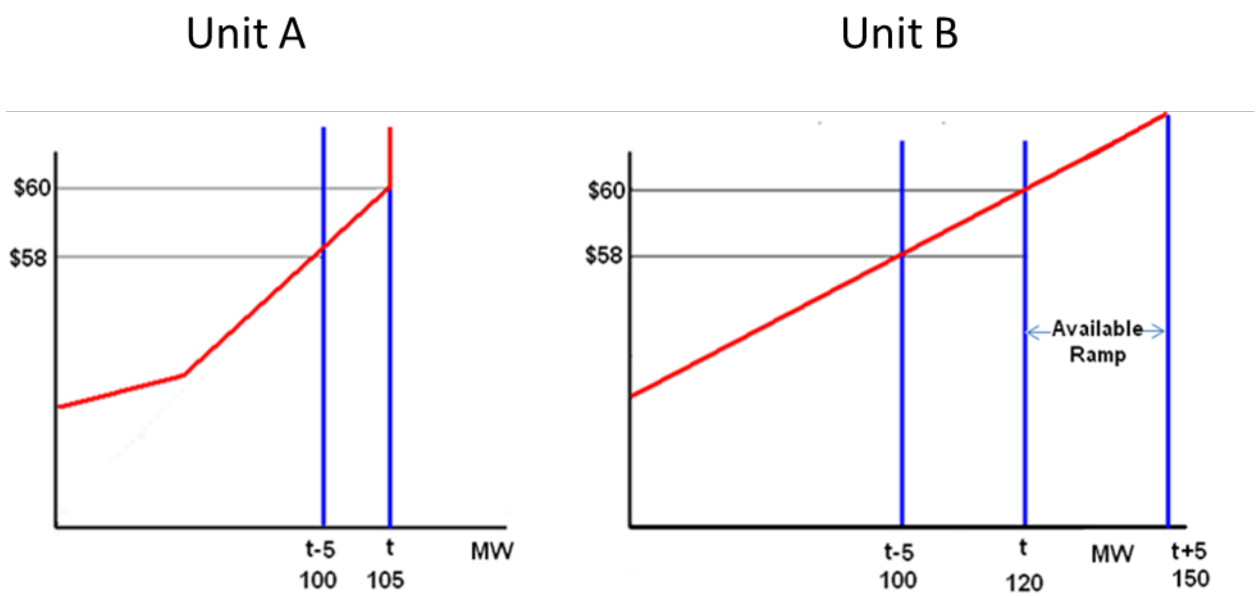
If resources are dispatched down out of merit below their upper limit so that they have ramp capability available for the next dispatch interval that would not be available in a pure economic dispatch, their output must be replaced with the output of other higher cost generation whose ramp capability is not capacity constrained.[41] This out of merit dispatch would raise real-time prices and raise the cost of meeting load in the current interval, because low cost generation would be replaced with higher cost generation. This substitution could be economic on a production cost basis if the availability of the additional ramp in the next interval would either avoid the need to dispatch of generation with very high offer prices, avoid potential power balance violations with high penalty costs or avoid shortage of spinning reserves or regulation in future intervals.

This substitution is illustrated in a very simple two unit context in Figures 4 and 5. In Figure 4 unit A has a 1 megawatt per minute ramp rate and is dispatched to its upper limit (105 megawatts) in the interval ending at t, with its incremental cost of $60 per megawatt hour equal to the price at its location. Because unit A is dispatched to its upper limit in the interval ending at t, it has no upward ramp capability available for the dispatch interval ending at t+5. When unit A is dispatched to its upper limit its upward ramp capability is capacity constrained.

[41] These would be generators whose undispatched capacity exceeds their ramp capability.

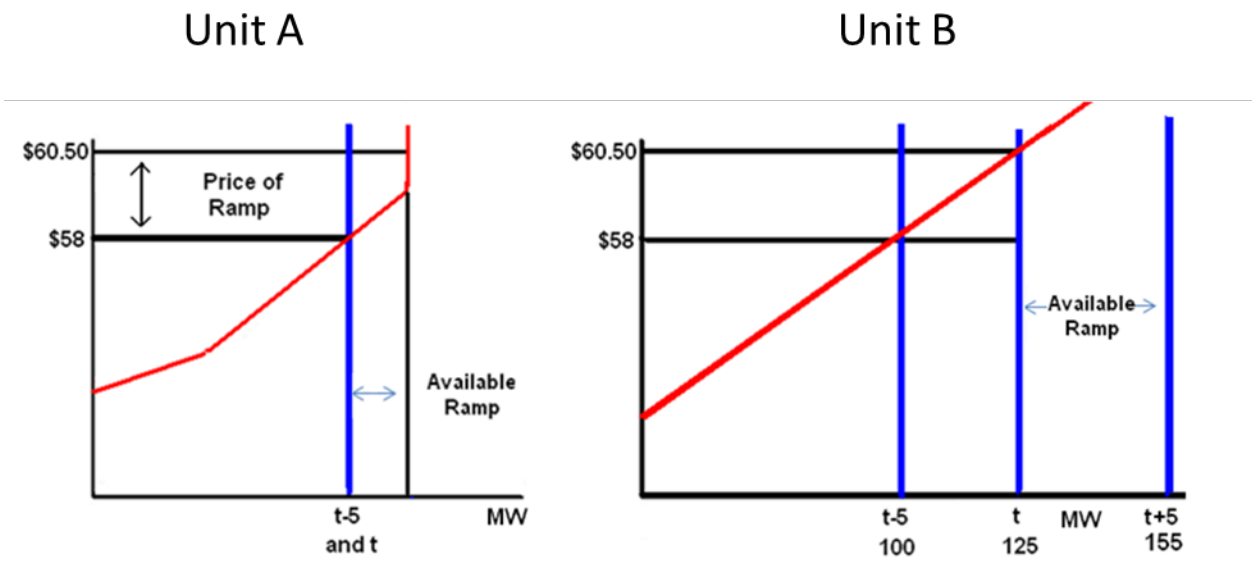
Unit B is assumed to have a ramp rate of 6 megawatts per minute and is dispatched up 20 megawatts in the interval ending at t, with its incremental offer of $60, also equal to the price at its location. Because the upper limit of unit B is 200 megawatts, its ramp capability for the interval ending at t+5 is constrained by its ramp rate (30 megawatts over 5 minutes), not its capacity.

Figure 4



If the system operator wanted to maintain additional ramping capability for the period ending at t+5, it could do so by not dispatching unit A up to its upper limit in the interval ending at t, and instead meeting that load by dispatching unit B higher on its offer curve (to $60.5 per megawatt hour) as shown in Figure 5. As a result of this out of merit dispatch, the cost of meeting load would rise, because generation costing $60 or less per megawatt hour has been replaced with generation costing up to $60.5 per megawatt hour, and in market based systems the real-time price would rise to $60.5 per megawatt hour. However, there would be 35 megawatts of upward ramp available for the interval ending at t+5, instead of only 30 megawatts.

Figure 5



Conversely, additional downward ramp capability can be created by dispatching generation that is at its lower limit (and that hence could not be dispatched any lower) up out of merit above its lower limit, and dispatching other generation that is lower cost and operating above its lower limit down to balance load and generation, raising the total cost of meeting load, lowering the cost of meeting incremental load and lowering the real-time energy price,[42] and maintaining additional downward ramp capability.

There can also be a potential to create additional ramp capability on resources that are not ramp constrained, if the resources have ramp rates that vary materially based on the output range to which they are dispatched. If the economic dispatch would place units in an operating range in which they have a relatively low ramp rate, dispatching them out of merit to an operating point at which they have a higher ramp rate could also raise the available ramp capability. This is most commonly the case for resources that have very low upward ramp rates when dispatched down to or near their minimum load, but have a higher ramp rate if dispatched to a somewhat higher output. In this case, the units on which the additional ramp is created would be dispatched up out of merit, and other units

[42] The imposition of the ramp capability target causes the price of power to fall because when capacity constrained generation is dispatched up out-of-merit above its minimum load point to provide additional ramping capability, ramp constrained generation is dispatched down to balance load and generation. Hence, incremental load is met with lower cost generation than would be the case absent the ramp constraint.

would be dispatched down, so the real-time price would fall, but the total cost of meeting load in that interval would rise.

The real-time dispatch can therefore be used to create additional ramp for the next interval, at a cost. This ramp capability would then available for use in the next interval to meet expected or unexpected variations in net load. This dispatch can be cost effective on a production cost basis if the expected reduction in the cost of meeting load in the next interval is greater than the increase in the cost of meeting load in the current interval.

Turning this concept for a ramp capability dispatch into a cost effective operating design requires addressing and resolving three key empirical questions.

- How much ramp should the dispatch try to create in each interval for potential use in the next interval?
- When should ramp capability be released to reduce the cost of meeting load in the current interval rather than maintained to potentially reduce costs in future intervals?
- Which undispatched capacity should be counted towards the ramp capability target?

These design issues are the subject of the next subsection.

2. Key Design Issues

Within the general framework described above there are three key design questions that must be resolved in order to implement such a dispatch concept in the real-time dispatch in a cost effective manner. Each is discussed below.

- Ramp capability target

The selection of the ramp capability target is extremely important both from the standpoint of achieving the intended benefits and keeping costs low. If the ramp capability target is set too low, there will be times when additional ramp capability is needed to manage increases in net load and could have been created at low cost. Scheduling too little ramp thus risks forgoing potential benefits from reducing production costs when the system becomes ramp constrained in the subsequent interval. Conversely, scheduling too much ramp capability relative to the likely need can lead to significant increases in the cost of meeting load when the ramp capability is not needed to compensate for variations in net load or other changes in system conditions, reducing or eliminating the potential net benefits from the overall design.

Optimizing the scheduling of ramp capability is not straightforward, because the value of increased ramp capability for the next interval is uncertain at the time the dispatch is solved. This is similar to the scheduling issue the California ISO has been working in its ramp capability commitment design over the past year, how does one set the right target for ramp capability. It will be essential for ISOs and RTOs using this kind of dispatch to carefully analyze patterns in the demand for ramp capability both over the day, over the year, and as a result of factors such as the level of intermittent resource output so that they can choose cost effective targets.[43]

The setting of the ramp capability dispatch target actually has two components. The first component is a megawatt quantity that is used to define the target amount of ramp capability. The second component is a shortage price that caps the amount of additional costs that would be incurred in the dispatch in order to create additional ramp.

My view is that selection of the ramp capability target is both important and complex. In order to achieve the potential benefits the target needs to be account for differences over the day, time of year, load levels and the level of intermittent output (if intermittent output is currently high, large increases are not likely so more ramp up capability is needed than ramp down capability and the converse applies if intermittent resource

[43] A high level of upward ramp capability is more likely to be valuable when the level of intermittent output is high (and hence could fall a lo), than when it is low (and hence could only fall a little).

output is low). The complexity of these ramp capability targets can be reduced somewhat in a multi-interval optimization which accounts for the expected load in the next interval.

The second component of the target, the selection of the shortage price, is actually the same as the question of when ramp capability should be released, the question we now turn to.

- Reserving or Using Ramp Capability

While the questions of what cost cap should be applied to the target for maintaining ramp capacity and when ramp capability should be released to reduce price spikes may seem like two distinct questions, there is actually only a single choice, to either hold back ramp capability for use in a future period or to use it in the current period. Thus, the question of when is the cost of ramp so high in the current interval that the ISO should not attempt to create additional ramp capability for future intervals is the same as the question of when is the value of ramp capability so high in the current period that some or all of the available ramp should be used to minimize the cost of meeting load in the current period.

There is more than one way to make this tradeoff but it is argued here that the best way is to set an appropriate penalty price for the value of ramp in future intervals and let the optimization solve for the amount of ramp capability that is maintained. This is the approach the MISO is planning to take if it implements a ramp capability based dispatch and the approach the California ISO has been tending toward. [44] A high shadow price of ramp capability in the current interval not only signifies that it would be expensive on a production cost basis to maintain additional ramp capability, but it indicates that the system is ramp constrained in the current interval.

The shortage or penalty value for ramp capability does not have to be a single value. Although the early analyses by the MISO and the California ISO's implementation of a ramp capability target in RTPD have been based on a single penalty value, this is not an

[44] See Midwest ISO, Ramp Capability in MISO Markets, Stakeholder 5th Technical Workshop, April 14, 2012.

inherent feature of the design. A ramp capability based dispatch could apply a lower penalty value an initial reduction in the availability of ramp capability for the next interval, then a higher penalty value (or values) for reductions below other lower thresholds. This kind of change in the basic design might be found to be cost effective after the basic design has been in operation for a while and its performance has been studied.

- Which ramp capability should be counted?

This issue has two primary dimensions. The first is spatial. At what locations should available ramp capability count towards the target. The need for a locational element to the procurement of ramp capability has been established by the issues the California ISO has experienced with a single system wide target for ramp capability in RTPD, which has frequently been met with undispatched capacity in PG&E's service territory in Northern California, which has a low opportunity costs precisely because it cannot be used to meet load, or variations in net load, in southern California.[45]

The MISO has been analyzing the congestion patterns associated with ramp constraint driven price spikes will be addressing this issue at least in part in the next iteration of the design.

The second dimension of what capacity counts towards the target is at what offer price the cost of energy associated with the ramp capability so high, that it has little value in terms of reducing the cost of meeting when the system becomes constrained.. At a theoretical level, the value of incremental ramp capability in reducing the cost of meeting load decreases with increases in the offer price, so the ideal design, if implementation complexity were not an issue would be a design which applied a continuously decreasing value to incremental ramp capability based on its offer price. The California ISO

[45] California ISO, Department of Market Monitoring, Q3 2012 Report on Market Issues and Performance, November 13, 2012 pp. 44-45; 2012 Annual Report on Market Issues and Performance, April 2013 pp. 88-89

considered this general kind of approach at one stage in the development of its ramp capability product.[46]

Implementation complexity is an issue, however, therefore alternative approaches which simply do not count ramp capability offered at prices above a threshold related to the typical level prices reach during price spikes, might provide an imperfect but more workable approach.[47]

3. Empirical Application

Empirical examination of the application of this concept to historical Midwest ISO dispatch data and price spike events has shown that the application of this dispatch concept can be cost effective on a production cost basis given current price spike frequencies and current Midwest ISO penalty values for spinning reserves shortages.[48] This examination also showed that the benefits of such a ramp capability based dispatch are sensitive to the penalty value used to value incremental ramp capability, the criteria used to set the target level of ramp capability,[49] as well as the frequency with which the ISO is unable to meet the load balance constraint with its real-time dispatch, and the value assigned to being able to avoid these situations.[50]

Further analysis will be needed by the Midwest ISO to assess whether a ramp capability based dispatch needs to be implemented within a multi-interval framework in order to provide large enough cost reductions to be beneficial in its market. The California ISO

---

[46] See, for example, Lin Xu and Donald Tretheway, California ISO, Flexible Ramping Products, Draft Final Proposal, April 9, 2012 section 2.1.4.

[47] It is important that putting in such a ceiling does not create any incentive to raise offer prices, as suppliers face the same competition in the energy market with or without a ramp capability product. The issue is simply that of paying the same price for ramp capability which provides lower benefits, and potentially incenting high cost resources to incur costs to increase their ramp rate, when that added ramp rate has limited value.

[48] See Midwest ISO, Ramp Capability in MISO Markets, Stakeholder 5th Technical Workshop, April 14, 2012. pp. 54-61. The MISO is in the process of completing another round of more complex simulations but the results have not yet been publicly posted.

[49] See Midwest ISO, Ramp Capability in MISO Markets, Stakeholder 5th Technical Workshop, April 14, 2012. pp. 54-61

[50] The California ISO, for example, assigns a $1000 value to being able to avoid violating the load balance constraint, which increases the benefits to implementing such a dispatch methodology.

would implement a ramp capability based product within its existing multi-interval optimized real-time dispatch.

The Midwest ISO's initial analysis was restricted to Midwest ISO wide price spikes with no congestion, which is a small proportion of all price spikes. The California ISO's experience with its ramp capability based commitment has also shown the importance of taking account of transmission constraints in scheduling ramp capability in the dispatch. The development of a ramp capability based dispatch tool that would maintain incremental ramp capability in the appropriate locations will be complex and these implementation issues are only beginning to be addressed. Hence, the implementation of a ramp capability based dispatch is still a concept under development with an uncertain implementation time frame in both the Midwest ISO and the California ISO.

The application of the ramp capability dispatch concept is illustrated in Table 6, drawn from a simulation of the operation of a ramp capability based dispatch in the Midwest ISO footprint. Table 6 portrays the operation of a ramp capability based dispatch with a $10 penalty value and no ramp target.

In this simulation, the production cost of additional ramp capability exceeded the $10 penalty value in the first two intervals, so no additional ramp was created and there was no change in production cost. Then in the interval beginning 1:10 it was economic to redispatch the system to create 4.1 megawatts of additional ramp at total production cost of $6.63 per hour. The generation dispatched down out of merit to create the additional ramp was replaced with slightly more expensive generation, raising LMP prices by $.22 per megawatt hour at the location at which generation was dispatched up.[51]

In the dispatch for the interval beginning 1:10 it was economic to create an additional 18.3 megawatts of ramp at a total production cost of $64 per hour, raising the LMP price by $1.83 per megawatt hour. In the following dispatch interval a total of 48.4 megawatts

[51] Because the cost of marginal losses varies by location, the change in prices would have varied slightly around this $.22 per megawatt hour figure from location to location.

of additional ramp were created at a total production cost of $142.39 per megawatt hour and raising the LMP price by $3.76 at the location at which generation was dispatched up to replace the output of the generation dispatched down to create additional ramp.

Table 6[52]

Spike on January 26, 2011
$10 Cap, No Ramp Capacity Cap

[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]	[K]	[L]
Interval	LMP	Change in LMP[1]	Change in Production Cost[2]	Spin Shortage	Change in Available Spin	Change in Spin Shortage Costs[3]	Up Ramp Left in Original Dispatch	Up Ramp Used in Original Dispatch	Original Total Up Ramp [H] + [I] - [E]	Additional Ramp in New Strategy	Total Up Ramp in New Strategy [J] + [K]
1:00	44.2	0.00	0.00	0	0	0	80	221.1	301.1		301.1
1:05	41.42	0.22	6.63	0	0	0	132.7	134.5	267.2		267.2
1:10	36.6	1.83	64.09	0	0	0	150.6	107.6	258.2	4.1	262.3
1:15	35.68	3.76	142.39	0	0	0	237.9	90.6	328.5	18.3	346.8
1:20	47.28	1.61	3.12	0	0	0	39.9	334.7	374.6	48.4	423
1:25	191.75	-128.04	-1885.08	156.9	15	1202.7	2.7	403.8	249.6	13.1	262.7

Total Production Cost	-1668.85
Total Production Cost and Shortage	-2871.55
Production Cost Benefit in Spike Interval	1885.08
Total Production Cost Benefit in Spike Interval Including Shortage	3087.78
Average Production Cost in Non-Spike Intervals	43.25
Ratio without Shortage Benefits	43.59
Ratio with Shortage Benefits	71.40

In the interval beginning at 1:20 the LMP price rose by more than $10 in the original dispatch, meaning that all of the resources dispatched down in the prior interval would have an opportunity cost in excess of $10 per megawatt in this interval. The amount of additional ramp created in the dispatch was only 13.1 megawatts, at total production cost of only $3.12 per megawatt hour and raising the LMP price by $1.61.

The additional 13.1 megawatts of ramp capability maintained in the interval beginning at 1:20 was used in the interval beginning at 1:25 when the price originally rose by almost $150 to $191.75 per megawatt hour. With the 13.1 megawatts of additional ramp capability available, the price rose to only around $63 per megawatt hour in the simulation. In addition to reducing prices, the additional ramp would have reduced

52 Source Midwest ISO, Ramp Capability in MISO Markets, Stakeholder 5th Technical Workshop, April 14, 2012. P. 56.

production costs in that interval by $1885 per hour. In addition, the additional ramp capability would have permitted a reduction of 15 megawatts in the shortage of spinning reserves. At current MISO shortage values for spinning reserve, this reduction would have been valued at $1202.7 per hour, so the total production and shortage costs savings would have been more than $3000 at an hourly rate.[53]

[53] The period analyzed in the simulation was prior to the time that the MISO implemented shortage pricing for spinning reserves in its real-time dispatch so the reserve shortage was not reflected in energy prices in the original dispatch.

Table 7 shows a redispatch of the same time period using a penalty value of $20 per megawatt of ramp. It is noteworthy that although there is a larger reduction in both production costs and shortage costs during the price spike interval, there is a net increase in production costs over the period modeled and a much small net reduction in the sum of production costs and shortage costs than with the lower penalty value. This example illustrates the importance of carefully choosing the ramp capability shortage value and the potential for a shortage value that is too high to diminish the efficiency benefits from implementing this design in the real-time dispatch.

Table 7[54]

but not at $20.

Spike on January 26, 2011

$20 Cap, No Ramp Capacity Cap

[A] Interval	[B] LMP	[C] Change in LMP[1]	[D] Change in Production Cost[2]	[E] Spin Shortage	[F] Change in Available Spin	[G] Change in Spin Shortage Costs[3]	[H] Up Ramp Left in Original Dispatch	[I] Up Ramp Used in Original Dispatch	[J] Original Total Up Ramp [H] + [I] - [E]	[K] Additional Ramp in New Strategy	[L] Total Up Ramp in New Strategy [J] + [K]
1:00	44.2	0.77	139.11	0	0	0	80	221.1	301.1		301.1
1:05	41.42	0.22	560.50	0	0	0	132.7	134.5	267.2	26	293.2
1:10	36.6	1.83	735.61	0	0	0	150.6	107.6	258.2	74.4	332.6
1:15	35.68	3.76	1182.17	0	0	0	237.9	90.6	328.5	86.7	415.2
1:20	47.28	1.61	54.46	0	0	0	39.9	334.7	374.6	129	503.6
1:25	191.75	-128.04	-2282.29	156.9	16.9	1326.2	2.7	403.8	249.6	20.1	269.7

Total Production Cost	389.56
Total Production Cost and Shortage	-936.64
Production Cost Benefit in Spike Interval	2282.29
Total Production Cost Benefit in Spike Interval Including Shortage	3608.49
Average Production Cost in Non-Spike Intervals	534.37
Ratio without Shortage Benefits	4.27
Ratio with Shortage Benefits	6.75

The discussion of ramp capability dispatch has focused on managing the ramp capability available to meet load in the next dispatch interval. The same basic design could be used by ISOs and RTOs utilizing multi-interval optimization to maintain a target level of ramp

54 Source See Midwest ISO, Ramp Capability in MISO Markets, Stakeholder 5th Technical Workshop, April 14, 2012. p. 57.

capability to be available over a multi-interval period. Such a design might to enable the system operator to better meet rapid changes in net load over a series of intervals. Whether such an elaboration in the design would be beneficial for some ISOs and RTOs cannot reasonably be assessed until there is operating experience with the basic design discussed above.

4. Ramp Capability Based Dispatch –Other Implementation Complexities

The preceding sections of discussed the operational issues associated with implementation of a ramp capability based dispatch. There are also a few economic issues to be addressed, in particular, how to compensate generators providing ramp capability, how to allocate the cost of the payments for ramp capability and whether and how ramp capability targets should be integrated in the day-ahead market. These topics are discussed briefly below.

Pricing.

The issue of pricing of ramp capability arises in two dimensions. First, when a generator in dispatched down out of merit in the current interval in order to provide additional ramp capability in future intervals, the generator does not receive any economic benefit from this out-of-merit dispatch and forgoes the additional margin it would have generated had it operated at its upper limit. Similarly, a generator dispatched up out of merit in order to provide additional downward ramp capability would be paid an LMP price that would be less than the cost of its incremental output, incurring losses at the margin. Second, there is a potential to use the settlements for ramp capability to provide an economic incentive for resources that are not capacity constrained in the amount of ramp they provide to provide more ramp capability by taking actions that increase their ramp rate.[55]

[55] The increase in ramp rates could be a result of both changes in physical capability or a change in the economic tradeoff between higher maintenance costs and the profits from offering a higher ramp rate for use in the economic dispatch.

There are four general approaches that can be taken to providing compensation to generators for providing ramp capability. The first approach would be to simply apply a bid production cost guarantee to the costs of resources dispatched up out of merit to provide downward ramp capability and to pay resources dispatched down out-of-merit the opportunity cost of their forgone output based on their offers. This approach would make resources whole for being dispatched to provide additional ramp capability. However, since resources that are ramp rate constrained in the amount of ramp they provide would receive no compensation under this design, this pricing system would not provide any additional incentive for resources to take actions to increase their ramp rate. Effectively, high cost providers of additional ramp capability would be paid for their high costs, but no incentive would be provide for more supply from low cost sources of incremental ramp capability.

The second approach would be to pay all resources providing ramp capability the market price of ramp capability, with the market price determined by either the opportunity cost of the marginal resource, or the ramp capability penalty value if less than the target amount of ramp capability is available. This is the approach the California ISO currently applies to compensating generators identified as providing ramp in RTPD and is the approach the Midwest ISO envisions applying to compensating resources providing ramp capability when it implements its ramp capability based dispatch. Like the first approach, this approach makes all resources whole for the costs they incur when they are dispatched out-of-merit to provide ramp capability. This design also provides compensation to resources that provide ramp capability but do not incur any out of merit costs. By providing compensation to resources that are constrained in the amount of ramp they provide by their ramp rate, this design provides incentives for these resources to incur costs in order to improve their ramp rates in the longer term or to choose a different trade off between higher maintenance costs and offering a higher ramp rate for use in the economic dispatch.

This second pricing system also has the property that if there is generally enough ramp capability without incurring out-of-merit dispatch costs, the payments for ramp capability will be low. If, however, the system is frequently short of ramp capability unless resources are dispatched out-of-merit, then the payments for ramp capability will be higher.

This is the compensation design that the California ISO currently uses to compensate resources identified as providing ramp capability in RTPD, which as discussed in section IIIC above has resulted in fairly high prices for ramp capability and high overall costs of procuring additional ramp capability. It is critical in considering the application of this pricing system to a ramp capability dispatch to keep in mind that in the current California ISO ramping capability design RTPD calculates the amount of ramp capability available and the opportunity costs of this ramp capability based on a hypothetical dispatch that does not reflect how the system is actually dispatched. Hence as discussed above, it overstates the ramp capability that will be available in real-time, by an unknown amount.

Because RTPD identifies more ramp capability than will actually be available in real-time, the California ISO must set a higher ramp capability target in order to get the amount of ramp capability it needs and likely also needs to set a higher penalty price (to ensure that the RTPD takes actions such as committing units that will actually increase the available ramp capability). This feature of the current design directly increases the cost of ramp capability because the California ISO pays for all of the ramp capability identified in RTPD, while receiving a lesser amount of ramp capability in the real-time dispatch.

The indirect effects of this design on the cost of ramp capability are potentially even larger. The need to inflate the ramp target in order to get the desired amount of ramp capability also inflates the opportunity cost of the marginal resource, raising the clearing price, and increases the potential for a shortfall relative to the target, resulting in prices being set more often that would otherwise be the case by the penalty price. Hence, the prices and costs of ramp capability in the California ISO design may not provide much insight into the actual cost of a real-time ramp capability dispatch in which the target is set based on the ramp capability that is actually available in real-time.

If the high cost and relatively poor performance of the California ISO's ramp capability commitment has been due to the lack of a ramp capability dispatch in real-time, the cost to load of the payments to generation for a real-time ramp capability dispatch could be negative, i.e. load payments would decline relative to the status quo with implementation of a ramp capability based dispatch. This could be the case because the California ISO is already paying for ramp capability in RTPD, and may be paying for more than it would need if it had a real-time ramp capability based dispatch that would enable it to get the ramp capability it is already paying for. If the amount of ramp capability the California ISO is paying for but not getting in the current design is significant, the California ISO would be able to reduce its ramp capability target if it implemented a ramp capability based dispatch, both paying for less ramp capability, and paying a lower price based on lower marginal opportunity costs and lower frequency of insufficient available ramp capability.

A third approach to pricing would be to employ a market based pricing sytem as under the second approach but to also allow suppliers to submit capacity bids for supplying ramp capability. Like the second approach, this pricing system would ensure that suppliers of ramp capability are made whole for being dispatched out of merit and would provide additional incentive for resources to incur costs to improve their ramp capability.

Extensive discussion of this pricing approach in the California ISO stakeholder approach failed to identify any costs that would be reflected in such a real-time capacity bid. Resources would in fact be dispatched up using their ramp capability to meet load whether or they were paid for providing ramp capability, just as they are today. The only apparent impact of high capacity bids would be to inflate the price of ramp capability by potentially causing resources providing ramp capability to not be counted against the ramp capability target.

A fourth approach to pricing would be to provide compensation for ramp capability in the resource adequacy design, rather than in the energy market. There are several fundamental limitations of an approach based on compensation through the resource adequacy system, whether a capacity market or other mechanism. First, the system operator does not need all resources to have ramping capability, it just needs some of them to be able to provide ramp capability. Hence, a broad requirement that all capacity market resources provide ramp capability is not necessary, would restrict the supply of capacity and needlessly raise the cost of incremental capacity needed to meet peak load. Second, the cost of resources providing ramp capability depends not only on their ramp rate but also their economics relative to the market price and whether they are even economic to commit. It would be very difficult to accurately evaluate these values in a forward capacity market or resource adequacy procurement process

Cost Allocation

A second financial issue in implementing a ramp capability dispatch is the allocation of the costs associated with the ramp capability dispatch which under at least the first three pricing designs will require at least some payments in excess of the energy price and under the fourth approach would likely entail additional capacity market payments. The California ISO has discussed this issue at length with its stakeholders.[56] Cost allocation issues are complicated by the variety of contracts relating to intermittent resources, and the variety of entities that may bear price risk under the contract.

The direct benefits of reduced frequency and depth of negative price spikes will benefit non-dispatchable resources such as nuclear units and intermittent resources whose increase in output is causing the downward ramp constraint to bind, and to a lesser extent, resources with large minimum load blocks and high cycling costs. The direct benefits of reduced frequency and height of upward price spikes will mostly flow to power consumers, but would also flow to resources settling deviations against forward schedules during the price spikes. This could include intermittent resources whose reduced real-time output causes the price spike, if their output is settled against a forward schedule.

[56] See Lin Xu and Donald Tretheway, California ISO, “Flexible Ramping Products, Draft Final Proposal, April 9, 2012 section 5; Offer of Settlement, Docket No. ER12-50-000 July 27, 2012;Lin Xu and Donald Tretheway, California ISO, Second Revised Draft Final Proposal, October 24, 2012 section 6.

The actual short-run incidence could be different, particularly for intermittent resources, because of existing contracts that may shift the impact of low or high prices to the load serving entity that has contracted for the output of the intermittent resource. Similarly, the long run impact of these changes is likely to shift more to consumers through reductions in capacity costs, whether determined in a procurement process or a capacity market.

Day-Ahead Market

A third economic design choice is how to integrate these real-time dispatch designs with the day-ahead market. These design choices are not discussed in this paper. The level of ramp capability available in real-time can be impacted to degree by the resources committed in the day-ahead market. Resources with a lower minimum load in proportion to their upper limit and faster ramp rates would provide more flexibility in real-time and could be favored. Another choice variable that may be more important is the level of interchange scheduled in the day-ahead market, as hourly interchange schedules provide no flexibility relative to generation resources that can be dispatched up or down. Hence, there could be a trade off that would favor internal dispatchable resources over imports for meeting load in the off-peak hours, with the dispatchable resources providing more ability to accommodate increases in intermittent resource output. Conversely, there will also be a need to ensure that there is enough rampable capacity to allow load to be met if intermittent resource output is lower than expected, which could favor scheduling more imports in high load hours to back down internal resources to provide more available upward ramp capability.

Approaches to integrating a real-time ramp capability dispatch into the day-ahead market have been discussed by the California and MISO but are less far along than the discussion of the real-time design. It is not even clear that there is any need to make changes in the day-ahead market when a real-time design is initially implemented.[57]

D. Other Approaches

The discussion in section IV C above has focused on one approach to implementing a ramp capability based dispatch. This is not the only possible approach. Three alternative approaches are briefly outlined and discussed below.

Scenario Based Approach

Some ISOs and RTOs currently run multiple cases of their real-time dispatch or look-ahead commitment programs with the operators choosing which case to use for the actual unit commitment and dispatch. One can envision an alternative approach to managing ramp capability that would run multiple cases with differing levels of intermittent resource output at various locations in future intervals. The operators could then choose which commitment or dispatch case to use for operations.

Some of the potential limitations of this approach are 1) the potential burden on the operators of both setting up these cases and selecting the appropriate case for operations in real-time, 2) the case selected for operations would generally be optimal for a particular scenario, not the best option over a range of scenarios.

This kind of approach might be useful, however, as part of another design to provide a check on whether particular solutions would be able to manage particular scenarios in future intervals.

Surprise Based Ramp Release Rules

[57] George Angelidis, California ISO, "Integrated Day-Ahead Market," September 28, 2012; Lin Xu and Donald Tretheway, California ISO, Second Revised Draft Final Proposal, October 24, 2012, section 2.5; Lin Xu and Donald Tretheway, California ISO, "Flexible Ramping Products, Revised Draft Final Proposal," August 9, 2012, section 2.5.

The ramp capability based dispatch design discussed in section IV releases ramp capability for use in the current period based on a penalty factor for the ramp capability target. This is not the only kind of rule that could be used for this purpose. An alternative approach that was considered at length by the California ISO would be to release ramp capability to meet surprises, i.e. when the real-time net load differs from that projected in the look-ahead scheduling process (RTPD in the case of the California ISO).[58]

This approach has two limitations. First, defining a "surprise" will be complex and introduces the likelihood of unintended consequences. While a defining a surprise may seem straight forward in terms of a difference between the net load or dispatch that was projected in the forward evaluation and that seen in the real-time dispatch, these differences will arise both because of "surprises" occurring in the current interval, and those that occurred prior to the current interval. Depending on the timing employed by a particular ISO or RTO the forward evaluation may have initialized between 20 and 45 minutes prior to the current dispatch interval. Differences between the net load in the two solutions arises both from surprises in the current interval and surprises that occurred in prior intervals after the forward evaluation initialized. It would not be necessary or cost effective to release ramp capability to deal with surprises that have already occurred and been met in the economic dispatch. Defining a "surprise" and the megawatt amount of the surprise is therefore potentially complex, very complex.

Second, even if a current period "surprise" could be readily defined, not all surprises have an equal cost impact and it would not be optimal to release scarce ramp to deal with a "surprise" that had only a minor cost impact. The penalty price approach provides a better outcome, releasing ramp capability when it has a high value in the current interval and maintaining it when it has a low value, independent of whether the "surprise" relative to some baseline is big or small.

[58] See Lin Xu and Donald Tretheway, Flexible Ramping Products, Draft Final Proposal, April 9, 2012; Lin Xu and Donald Tretheway, Flexible Ramping Products, Third Revised Straw Proposal, March 6, 2011.

Stochastic Optimization

It is also possible to imagine optimizing the dispatch to account for ramp capability using stochastic optimization methods, rather than the targets, penalty factors and eligibility rules of the approach described in Section IVC. Beibei Wang and Benjamin Hobbs compare the performance of the two approaches and show the potential for stochastic optimization to achieve a more optimal solution if it includes a better representation of the relevant outcomes.[59]

Another way of looking at the “deterministic” and stochastic optimization approaches is in terms of a tradeoff between detailed forward analysis of generic probabilities embodied in targets, penalty factors and ramp rules, which will generally not be fully optimal because the choices will not be optimized for the exact real-time state of the system, and real-time stochastic optimization based on the current real-time state of the system, which will also generally not be fully optimal because it will be optimized only over the cases analyzed, which will be limited by performance considerations in real-time.

To the extent that the number of possible future states of the system constrains which can be probabilistically evaluated in real-time is limited, there will likely be a trade off between not analyzing particular outcomes at all in a stochastic optimization framework or analyzing them in advance and applying the kind of generic dispatch rules described in section IVC, with key states that can be evaluated in detail within the relevant solution time frame solved using a stochastic optimization framework.

V. Conclusions

The commitment of capacity and scheduling of interchange to provide sufficient ramp capability is important in managing ramp capability but may not be sufficiently effective operationally or sufficiently cost effective to use as the sole mechanism for managing

[59] Beibei Wang and Benjamin Hobbs, “Flexiramp Market Design for Real-Time Operations: Can it Approach the Stochastic Optimization Ideal?” IEEE Summer PES meeting, April 14, 2013

ramp capability in the future. Combining forward commitment and interchange scheduling mechanisms that account for ramp capability with a ramp capability based dispatch has the potential to provide additional ramp capability at a lower cost than simply committing additional capacity or adjusting net interchange and could improve the performance and economics of changes in unit commitment and interchange scheduling that are intended to adjust the available ramp capability.

The operational and cost effectiveness of a ramp capability dispatch will depend on the accuracy with which the ramp capability targets, penalty values and locational requirements can be determined. As the level of intermittent output increases, some regions may have to utilize ramp capability based commitment and interchange scheduling evaluation mechanisms and a ramp capability based dispatch may be essential to ensure that those evaluation mechanisms operate in a cost effective manner.

The operational experience of the California ISO with a ramp capability based commitment provides a valuable source of data for analyzing the value of such a design.

References

Angelidis, George, California ISO, “Integrated Day-Ahead Market,” September 28, 2012.

MISO, “Ramp Capability in MISO Markets,” Stakeholder 5th Technical Workshop, April 14, 2012;

Navid, Nivad, Gary Rosenwald, Dhiman Chatterjee, “Ramp Capability for Load Following in the MISO Markets, July 15, 2011;

Navid, Nivad, “Ramp Capability for Load Following in the Midwest ISO Markets, march 4, 2011;

Navid, Nivad , Gary Rosenwald, “Market Solutions for Managing Ramp Flexibility with High Penetration of Renewable Resource,” IEEE Transactions on Sustainable Energy Vol 3 No. 4 October 2012;

Xu, Lin and Donald Tretheway, California ISO, Second Revised Draft Final Proposal, October 24, 2012.

Xu, Lin and Don Tretheway, California ISO, “Flexible Ramping Product,” Market Surveillance Committee Meeting October 19, 2012,

Xu, Lin and Donald Tretheway, California ISO, “Flexible Ramping Products, Revised Draft Final Proposal,” August 9, 2012.

Xu, Lin, and Donald Tretheway, California ISO, Flexible Ramping Product, Supplemental: Foundational Approach,” July 11, 2012.

Xu, Lin and Donald Tretheway, California ISO, Technical Workshop on Flexible Ramping Products, May 29, 2012;

Xu, Lin and Donald Tretheway, California ISO, Flexible Ramping Products, Draft Final Proposal, April 9, 2012 .

Xu, Lin and Donald Tretheway, California ISO, Flexible Ramping Products, Third Revised Straw Proposal, March 6, 2011

Wang, Beibei and Benjamin Hobbs, “Flexiramp Market Design for Real-time Operations, Can It Approach the Stochastic Optimization Ideal?” IEEE Summer PES meeting, April 14, 2013.

End Note A

The author is or has been a consultant on electricity market design, transmission pricing and/or, market power for Allegheny Energy Global Markets; American Electric Power Service; American National Power; Aquila; Avista Corp; California ISO; Calpine Corporation; Centerpoint Energy; Commonwealth Edison; Competitive Power Ventures, Conectiv Energy, Constellation Power Source; Coral Power; Dayton Power and Light, Duke Energy, Dynegy; Edison Electric Institute; Edison Mission; ERCOT, Exelon Generation; General Electric Capital; GPU; GPU Power Net Pty Ltd; GWF Energy; Independent Energy Producers Association; ISO New England; Koch Energy Trading; Longview Power; Merrill Lynch Capital Services; Midwest ISO; Morgan Stanley Capital Group; New England Power; New York Energy Association; New York ISO; New York Power Pool; Ontario IMO/IESO; PJM; PJM Supporting Companies; PP&L; Progress Energy, Public Service Co of New Mexico; Reliant Energy; San Diego Gas & Electric; Sempra Energy; Mirant/Southern Energy; Texas Utilities; TransAlta Energy Marketing, TransCanada Energy; Transpower of New Zealand Ltd; Tucson Electric Power; Westbrook Power; Williams Energy Group; and Wisconsin Electric Power Company.

The views presented here are not necessarily attributable to any of those mentioned, and any errors are solely the responsibility of the author.