Current and Future Ancillary Services

Ten Important and Unique Considerations

Erik Ela

December 2024

Product ID: 3002031421

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

THIS DOCUMENT WAS PREPARED BY THE ORGANIZATION(S) NAMED BELOW AS AN ACCOUNT OF WORK SPONSORED OR COSPONSORED BY THE ELECTRIC POWER RESEARCH INSTITUTE, INC. (EPRI). NEITHER EPRI, ANY MEMBER OF EPRI, ANY COSPONSOR, THE ORGANIZATION(S) BELOW, NOR ANY PERSON ACTING ON BEHALF OF ANY OF THEM:

(A) MAKES ANY WARRANTY OR REPRESENTATION WHATSOEVER, EXPRESS OR IMPLIED, (I) WITH RESPECT TO THE USE OF ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT, INCLUDING MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, OR (II) THAT SUCH USE DOES NOT INFRINGE ON OR INTERFERE WITH PRIVATELY OWNED RIGHTS, INCLUDING ANY PARTY'S INTELLECTUAL PROPERTY, OR (III) THAT THIS DOCUMENT IS SUITABLE TO ANY PARTICULAR USER'S CIRCUMSTANCE; OR

(B) ASSUMES RESPONSIBILITY FOR ANY DAMAGES OR OTHER LIABILITY WHATSOEVER (INCLUDING ANY CONSEQUENTIAL DAMAGES, EVEN IF EPRI OR ANY EPRI REPRESENTATIVE HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES) RESULTING FROM YOUR SELECTION OR USE OF THIS DOCUMENT OR ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT.

REFERENCE HEREIN TO ANY SPECIFIC COMMERCIAL PRODUCT, PROCESS, OR SERVICE BY ITS TRADE NAME, TRADEMARK, MANUFACTURER, OR OTHERWISE, DOES NOT NECESSARILY CONSTITUTE OR IMPLY ITS ENDORSEMENT, RECOMMENDATION, OR FAVORING BY EPRI.

EPRI PREPARED THIS REPORT.

NOTE

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

Together…Shaping the Future of Energy®

Acknowledgments

EPRI prepared this report.

Principal Investigator
E. Ela

This publication is a corporate document that should be cited in the literature in the following manner:
Current and Future Ancillary Services: Ten Important and Unique Considerations. EPRI, Palo Alto, CA: 2025. 3002031421.

Abstract

Ancillary Service Markets are an important part of maintaining reliability and achieving economic efficient solutions in electricity market regions. With increased variability and uncertainty, increased levels of inverter-based resources, and other changes to the power system, ancillary services and the markets that run these services are evolving. Many regions are exploring new ancillary service products or improving the ways in which they set the needs for those services, including which resources can provide them. This review explores 10 ideas related to ancillary services and ancillary service markets, some which are counterintuitive, others that are not obvious, to support market operators, balancing authorities, and other stakeholders on making improvements.

Keywords

Ancillary services

Electricity markets

Frequency response

Independent system operators

Reserve demand curves

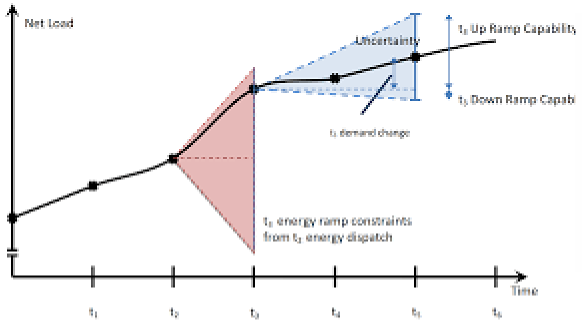
Variable energy resources

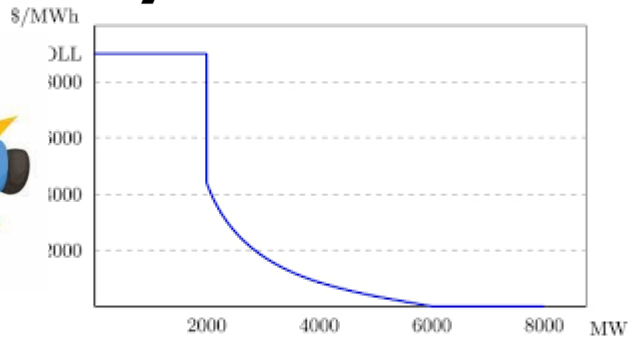
Introduction

In this document, EPRI provides ten insights on ancillary services (with an additional bonus insight) related to how to think about the different types that are in existence or that may be in the future. Many of these insights may be counterintuitive or less obvious. In each insight, one or more slides are shared that provides examples that illustrate the insight. Some simple quantitative examples while others are showing qualitatively how to consider the insight in practice. A summary of the slide is seen using a mouse-over to the upper left side of the slide to read how to understand further. Copy of the response to technical comments by FERC following a technical conference (AD21-10) are also included in the appendix.

Ancillary Service Insights Table of Contents

Operational Flexibility and Reliability in Markets

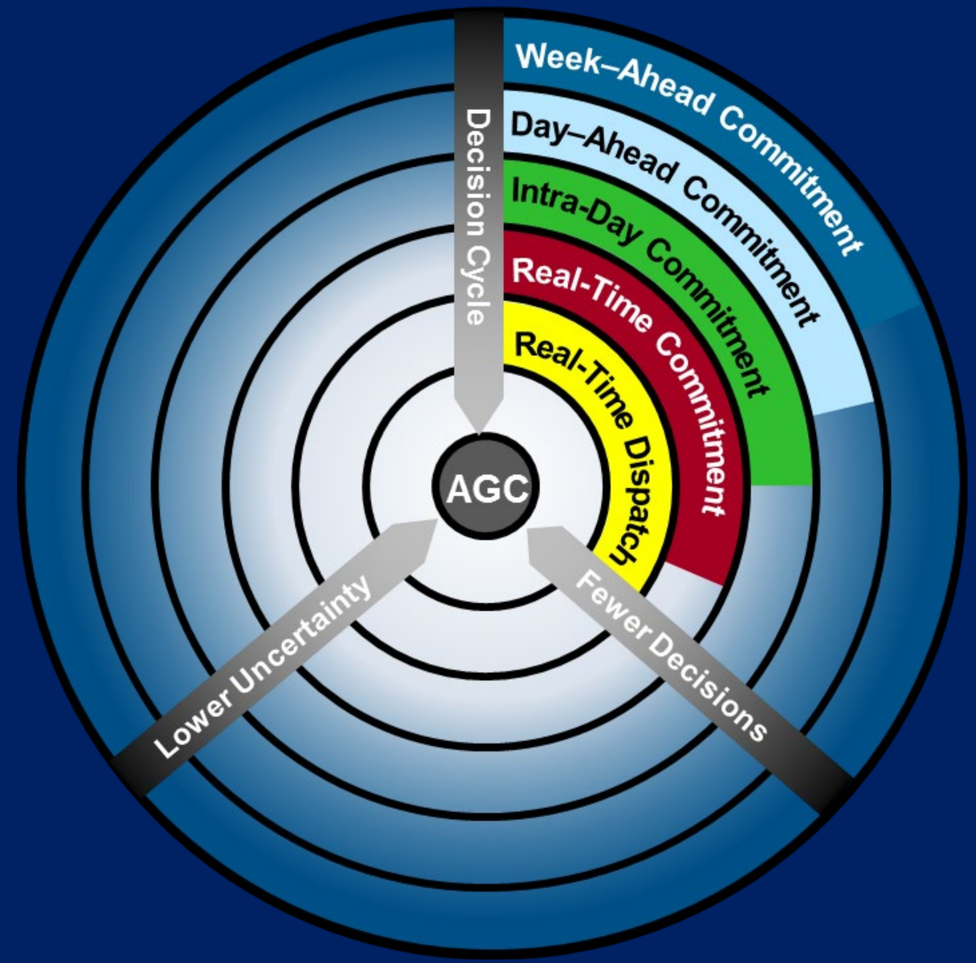




- Forecasting dynamic reserve is similar to forecasting load or VER output, and can be used to improve reliability and economic efficiency
- Extended downward sloping ORDCs beyond the minimum contingency reserve requirement can be beneficial if designed adequately; so can flexibility reserve;
  - **These two flexibility mechanisms may be identical under different names; however, the current implementation of both may lead to different outcomes**
- Reserve does not only improve reliability, it can also reduce operating costs
- Designing and introducing products may demonstrate benefits in studies; however, the design (including stakeholder debate), implementation, and administration can also be costly

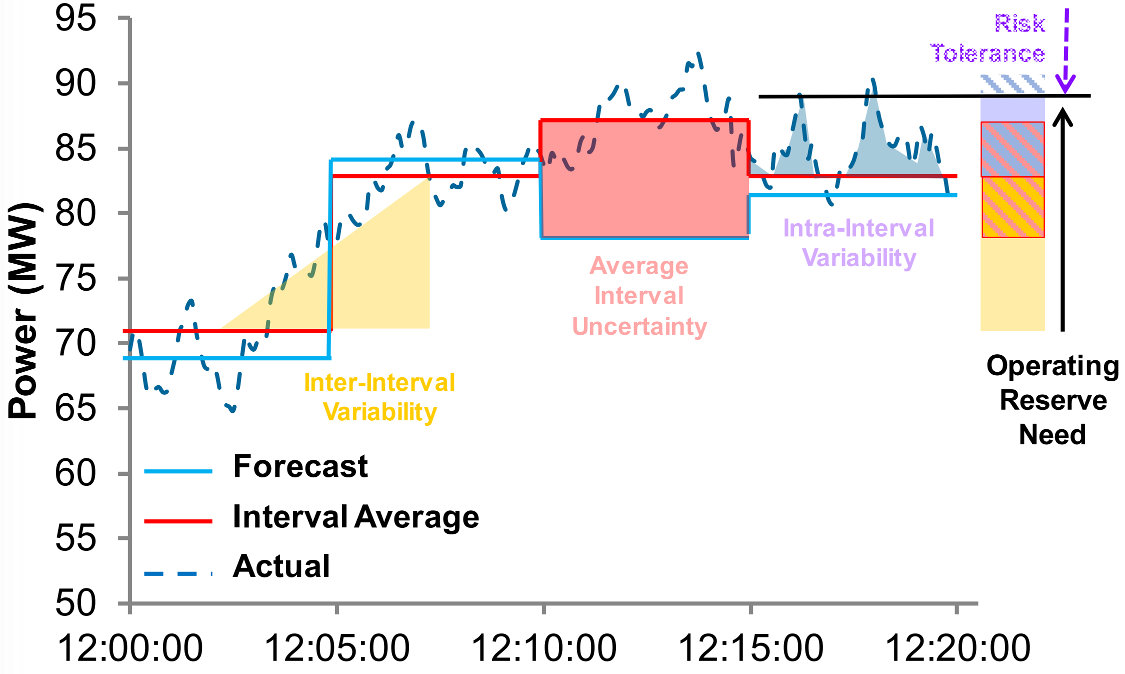
Operational Flexibility and Reliability in Markets

- Reserve products may be better defined by the decision processes compared to response times (e.g., day-ahead reserve released in hour-ahead), this can align the reserve holding to the decisions that can be made according to the scheduling process
- The flexibility response time need (e.g., 10-min vs. 30-min vs. hour ramp product) is a continuum, but the most critical time horizon can be determined by 1) the largest need or inflection point 2) where lack of associated flexibility may exist
- Reserve needs can be met implicitly through advanced scheduling. The residual need thus depends on how you schedule, not just how much variability and uncertainty there is



Operational Flexibility and Reliability in Markets

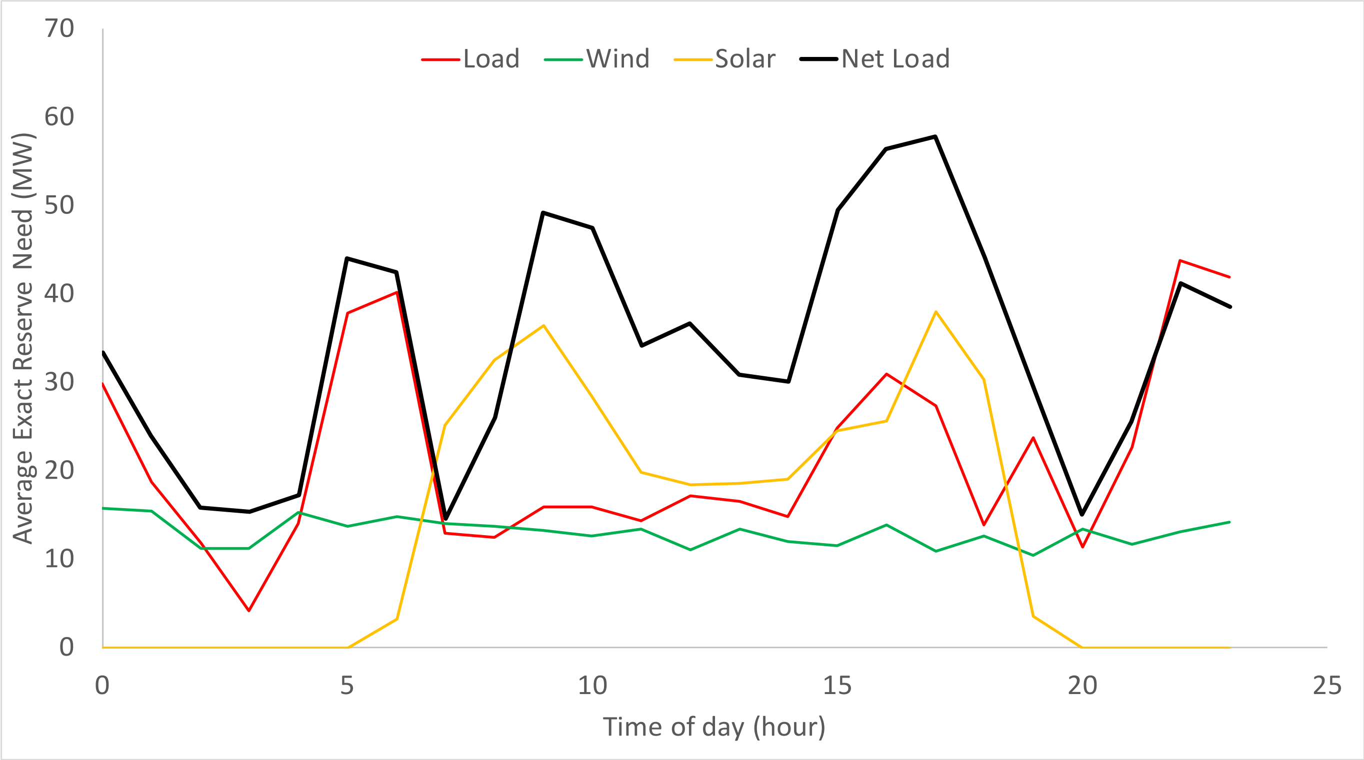
- Implicit ways of meeting operating reserve such as multi-period dispatch and shorter market intervals, have efficiency gains but may impact incentives
  - Multi-period real-time markets in particular due to single binding interval settlement
- Frequency response ancillary service markets may provide benefits in regions that foresee difficulty in meeting in BA's frequency response obligation
- Because ISO DA forecasts are only used in the RUC processes, there may be limited benefits from more advanced forecasting techniques in this time frame
- The price of ancillary services under zero-fuel cost futures is complex and unclear. The interval lost opportunity cost when energy price is zero may also be zero.
  - Non-convex pricing, inter-temporal opportunity costs, and shortage pricing design can all impact ancillary service prices in the future



Ancillary Service Insights Illustrations and Details

#1 Reserve Forecasting

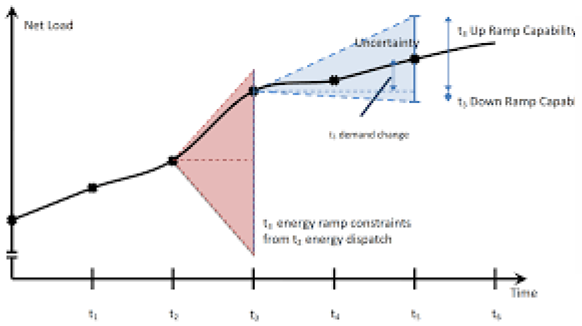
- Reserve needs change across time
  - E.g., Different night from day
- Reserve needs change through time
  - The need for hour 10 a day ahead is different from the need for hour 10 an hour ahead
- Certain factors can help us understand the needs
  - Uncertainty and variability anticipation
- Unequal risk between over- and under-forecasting
  - Forecasting expected value may not be appropriate

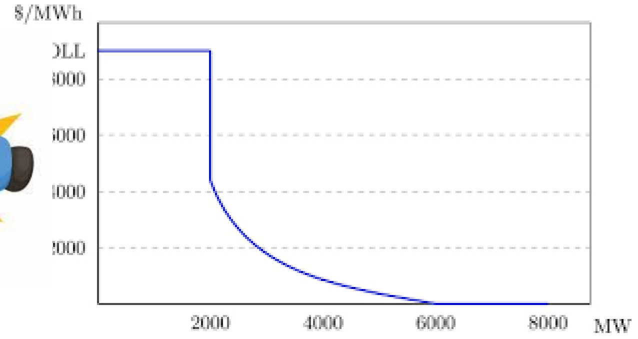


Historically, ancillary services were based on conditions that are generally consistent by time of day and by season. For example, the probability of a contingency occurring was generally consistent regardless of system conditions. However, with reserve that are based on the variability and uncertainty, and with contingencies being more impacted by weather conditions, the likelihood of needing reserve and the amount of reserve that can keep a system at a consistent risk level will change based on existing and anticipated conditions. The reserve need will change by time, by season, and by time ahead. Reserve needs during one hour are different from another hour, and vary by season and vary between day-ahead and real-time for the same hour. As a simple example, reserve for a drop in solar production is not needed during the night. Reserve can and should be forecasted, just as it is for load and renewable production. There is a lot of information that operators can use to help determine when greater levels are needed, forecast that reserve and use that as requirements for the ancillary service markets. Given that being short on reserve has a larger impact than being long, the way that the reserve is forecast may be different than other forecasted values (e.g., a 90th exceedance forecast may be preferred over forecast of expected value).

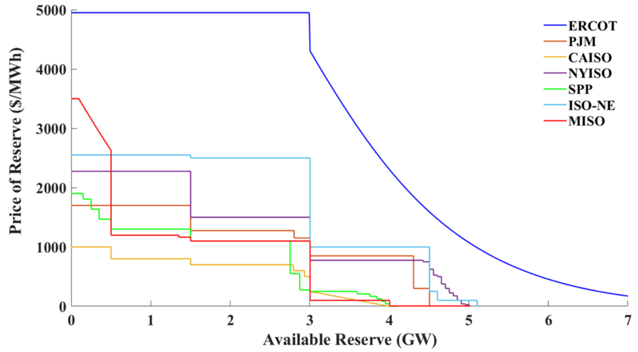
#2: Extended ORDC and Separate Flex Product

Same in Principle, different in implementation





Design Feature	Extended ORDC	Separate Flex Product
Requirement setting	Uncertainty between scheduling processes	Variability + ~10-min uncertainty
Price setting	VOLL*LOLP	Can be arbitrary
Steps	Multiple, approximate non-linear curve	Single, up to 5
Eligibility	10-min reserve	Varies, may be diff from ctgc reserve
Pricing Hierarchy	All	May not cascade
Market Processes	All	Sometimes RT only
Offer	Offer rules for ctgc reserve	Usually zero cost offers



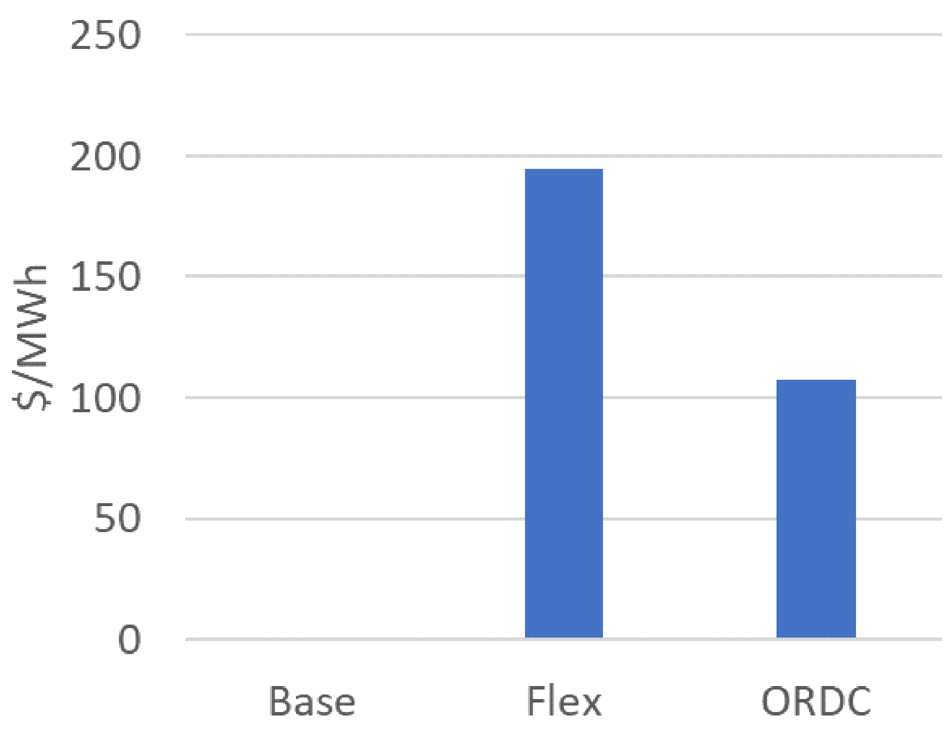
Mehrtash, Hobbs, & Ela, "Reserve and Energy Scarcity Pricing in United States Power Markets: A Comparative Review of Principles and Practices" *Renewable and Sustainable Energy Reviews*, vol. 183 (2023).

https://www.epri.com/research/products/000000003002025037

The two most popular methods for getting additional capacity for flexibility to mitigate the errors associated with uncertainty of the net load are by introducing a new "flexibility" product that is used regularly between economic dispatch intervals to provide additional capacity than would otherwise be scheduled, and to apply long extended operating reserve demand curve to the existing reserve products, such as those for contingency reserve. Further investigation of these two methods have shown that the methods themselves can essentially be designed to have a near identical result, but that in practice they have been designed differently. ORDC have been designed to support long-term revenue adequacy with flexibility products being designed for more of a short-term flexibility incentive. ORDC have to follow the product design of the existing product where flexibility products, which can have a demand curve of their own, can be designed from scratch.

#2: Extended ORDC and Separate Flex Product

Same in Principle, different in implementation



Cost per MWh of shortage reductions

ORDC for Existing Contingency Reserve	Separate Flexibility Product
► To which existing product(s) applied ► How price/quantity is calculated, including loss of load probability calculations if applicable ► How demand curve prices set ► Regulation (other reserve) adjusted ► Different curves (price and quantity) day-ahead versus real-time ► Number of steps to approximate curve ► Deliverability and locational curves ► Convex Hull / ELMP features	► Requirement setting ► Single requirement vs. multi-step ► Demand curve or shortage price setting ► Response time of product ► Online only versus offline allowed ► Which scheduling processes product held, and any changes between them ► Availability bids versus LOC only ► Cascading price effect with other products ► Post-deployment deliverability constraints ► Convex Hull / ELMP features

https://www.epri.com/research/products/000000003002028054

Analysis showed some of the benefits of both mechanisms for avoiding shortages of minimum reserve requirements and load shedding at a cost that is far lower than the values assigned to those reserve (i.e., as penalty values for violations). Again, it can depend on the design so it is important to design each mechanism carefully to get the desired result. Overall, the sloped demand curve (which can also be used for a flexibility product) provides some value by raising prices as the system gets closer to critical conditions, but does not have to actually have a shortage condition before they do provide the incentives. Results have shown value for either mechanism if designed adequately. Some of the design components can be seen on the table on the right.

#3 Reserve for Cost Reduction

	Variable Fuel Cost	Min Gen	Capacity	Started in RT?
G1	20$/MWh	25 MW	100 MW	No
G2	30$/MWh	25 MW	100 MW	No
G3	80$/MWh	5 MW	100 MW	Yes

Load / Spin Reserve	DA	RT
Scenario 1	100 / 0	125 / 0

Reserve supports reliability and having that reserve capacity typically has reliability as its primary purpose. However, reserve capacity, if used efficiently, can also reduce production costs. Reserve can in some cases both improve reliability and reduce costs. We will show a simple, exaggerated example to illustrate the concept of lowering costs using reserve that is released. In the example, three generators exist, with one resource being expensive but able to start quickly in real-time. The two cheaper resources have long-start-times and notification times meaning they have to be started in day-ahead. The day-ahead has a load of 100 MWh, but in real-time in increases to 125 MWh. The operator has some information that the load can increase in real-time but the forecast does show it.

#3 Reserve for Cost Reduction

	Variable Fuel Cost	Min Gen	Capacity	Started in RT?
G1	20$/MWh	25 MW	100 MW	No
G2	30$/MWh	25 MW	100 MW	No
G3	80$/MWh	5 MW	100 MW	Yes

Load / Spin Reserve	DA	RT
Scenario 1	100 / 0	125 / 0

Scenario 1	DA	RT
G1	100	100
G2	0	0
G3	0	25
Cost ($)	2,000	4,000

Without reserve scheduled day-ahead, we have to use G3, the expensive generator to accommodate the 25 MWh of increased load in real-time. This leads to an overall production cost $4,000 in real-time.

#3 Reserve for Cost Reduction

	Variable Fuel Cost	Min Gen	Capacity	Started in RT?
G1	20$/MWh	25 MW	100 MW	No
G2	30$/MWh	25 MW	100 MW	No
G3	80$/MWh	5 MW	100 MW	Yes

Load / Spin Reserve	DA	RT
Scenario 1	100 / 0	125 / 0

Scenario 1	DA	RT
G1	100	100
G2	0	0
G3	0	25
Cost ($)	2,000	4,000

Scenario 2	I1	I2
G1 (Sched/Spin)	75/25	100
G2 (Sched/Spin)	25/0	25
G3 (Sched/Spin)	0/0	0
Cost ($)	2,250	2,750

Not only does smart reserve requirements improve reliability, it also can reduce costs

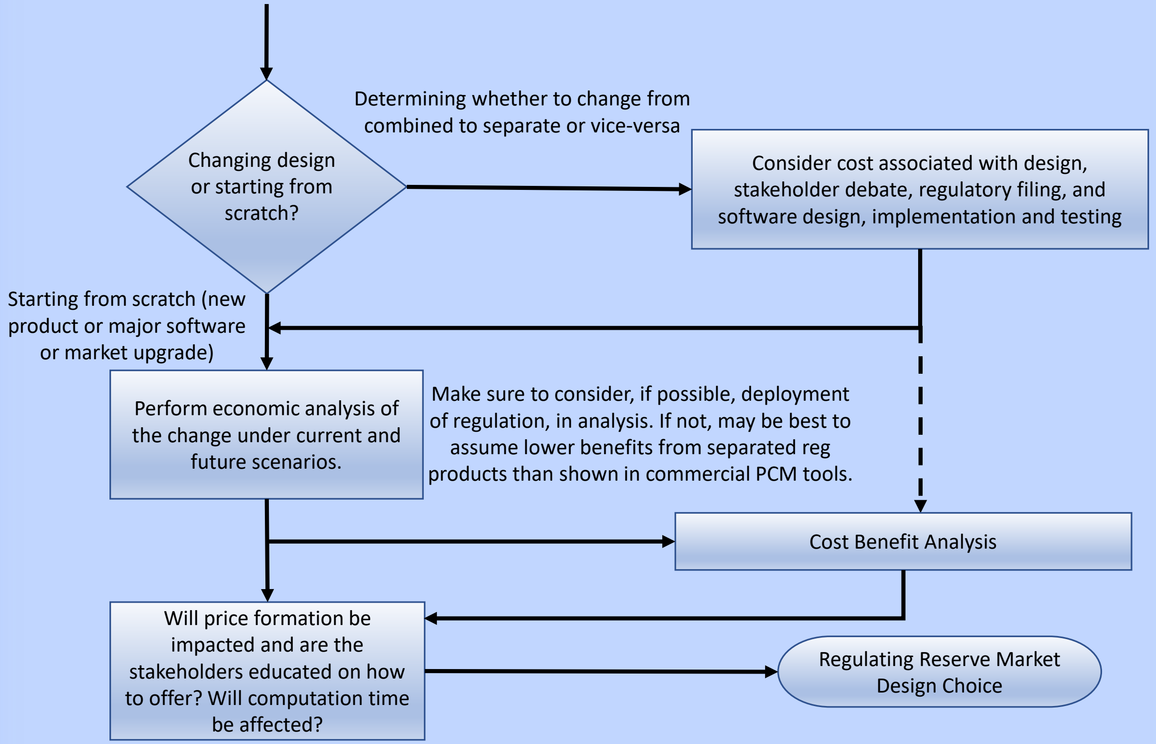
https://www.epri.com/research/products/000000003002012346

In the second scenario the operator meets the day-ahead load but also ensures 25 MW of reserve in case of the increase. Now G2 is started in day-ahead and G1 is able to provide reserve. When the 25 MWh increase happens in real-time G1 is able to increase production. The overall production cost is reduced significantly down to $2,750 because of the reserve scheduled. Obviously, the costs would have been increased if the 25 MWh increase in real-time did not happen but the reserve was scheduled. This shows that intelligence on the condition happening is important for reserve forecasting so that both costs and reliability can be improved. In a study with Hawaiian Electric Company, it was found that both reliability was improved and costs were reduced using these types of reserve requirements for their system.

#4: Design of products can introduce benefits, but design, implementation and admin costs may be costly

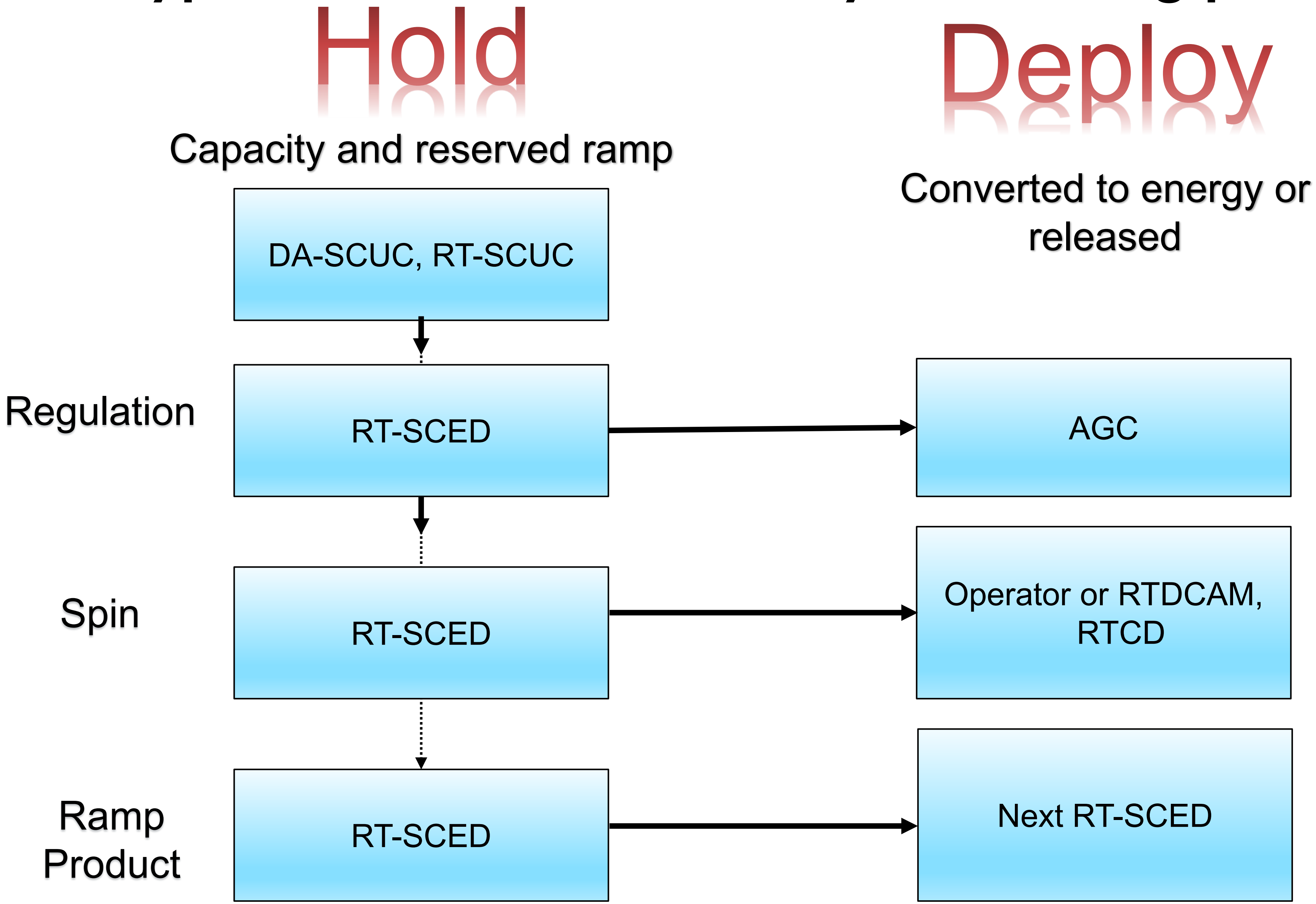
Combined bi-directional regulating reserve product	Separate bifurcated directional regulating reserve up and regulating reserve down products
► **Ability to accurately reflect lost opportunity costs and product prices directly in the market software** ► **Likely lower cost differences from regulation deployment** ► **Capture net zero energy neutrality** ► **Fewer variables**	► Economic efficiency benefits for the economic dispatch of resources to meet energy and regulating reserve up and regulating reserve down ► Ability for different product requirements when the magnitude of upward and downward need are ► May be easier to find solutions

https://www.epri.com/research/products/000000003002027004



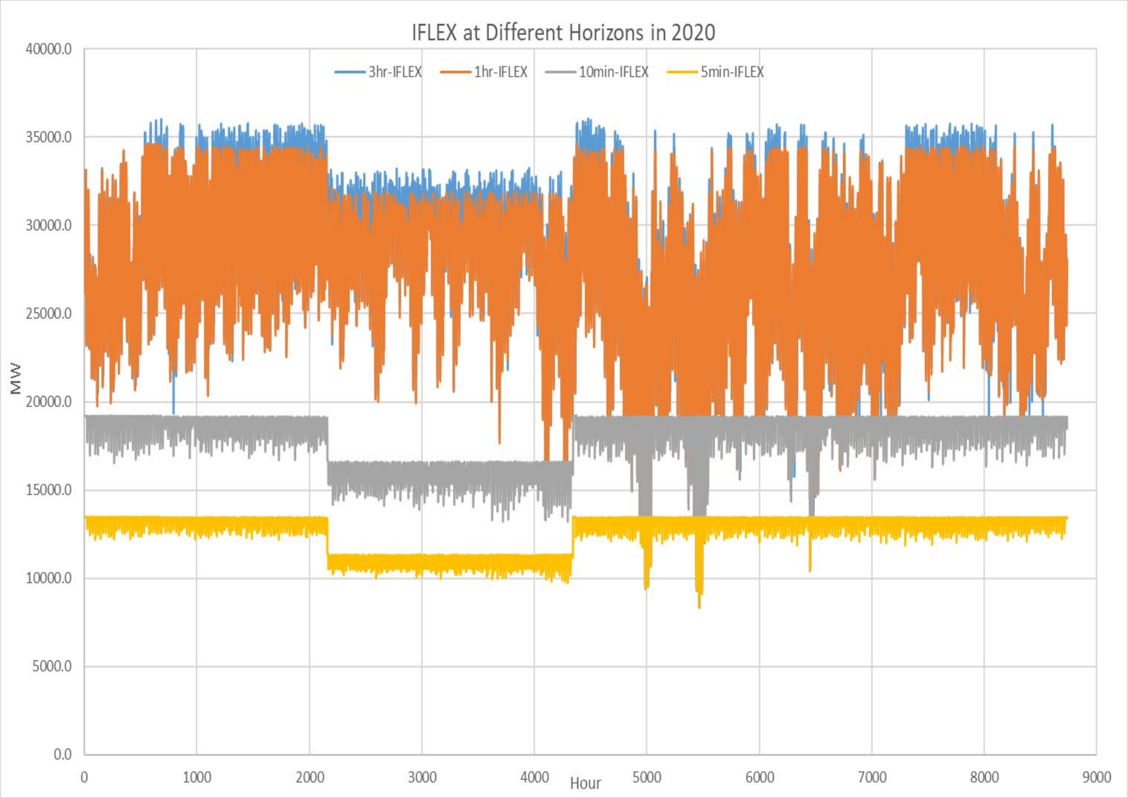
It is easy to show when new ideas, algorithms, and market designs can show theoretical economic efficiency gains through simulations or simple examples. It may lead market designers or stakeholders to pursue the changes immediately. However, it is important to include the various soft costs associated with making a change to market design. This includes not only the software changes, but stakeholder time to deliberate, stakeholder time to adapt to the change, education time, any downstream market changes such as settlement software, legal costs associated with filing and regulatory approval, computational changes, and many other costs. These may make it so changes with positive, yet small relative benefits, may not necessary be worth pursuing, without considering first. An example is changes to regulation market to move from a combined regulation product to a separate product. EPRI research found benefits to this change in market design, but that the benefits were much lower than what are seen on the surface due to testing software limitations, as well as the changes to market participant behavior and education, and other costs associated with time.

#5 Reserve types should be defined by scheduling process



Often reserve products are defined by their response time requirement. For example, reserve is set to "20-minute reserve" with a response time to be within 20 minutes. While these values are useful for transparency, we find that they may sometimes be arbitrarily decided. It is better to define the reserve product with the scheduling process for which the capacity is held and the scheduling process where the reserve is released, and if needed, converted into an energy injection/ withdrawal. This makes it more clear when and which decision process can utilize the capacity. It also can help make clear the requirements necessary, as the requirements are directly dependent on the uncertainty and variability occurring between those two scheduling processes.

#6: Choosing horizon based on largest need and largest lack of supply



Year	Items	Horizon			
		3 hour	1 hour	10 minute	5 minute
2016	Annual Average IFLEX (MW)	27,292	27,136	17,995	10,858
	Percentage Avg IFLEX to 3hr-IFLEX	-	99%	66%	40%
	Percentage of NL ramp to 3hr-NL Ramp	-	44.1%	10.2%	7.0%
	Percentage of NL Ramp +CR to 3hr-NL Ramp + CR	-	50.7%	24.9%	N/A
2020	Annual Average IFLEX (MW)	27,932	27,652	17,894	10,753
	Percentage Avg IFLEX to 3hr-IFLEX	-	99%	64%	39%
	Percentage of NL ramp to 3hr-NL Ramp	-	47.7%	10.9%	7.5%
	Percentage of NL Ramp +CR to 3hr-NL Ramp + CR		50.3%	20.3%	N/A

Flexibility Assessment for the California ISO: Evaluating Flexibility Needs and System-wide Feasible Installed Flexible Capacity. EPRI, Palo Alto, CA: 2018. 3002013725. Available: https://www.epri.com/research/products/000000003002013725

Almost the same amount of flexibility but double the need!

While Insight #5 described the need to define reserve products by the scheduling process by which they are held and released, operators and planners may be wanting to try to include some signals that are specifically at a type of horizon of flexibility. For example, the ISO or BA may be wanting to increase the amount of flexibility on their system but are unsure how to express it. The ability to move in 5 minutes? 15 minutes? One hour? Several hours? The first evaluation is where the greatest need for flexibility might be. For example, if net load changes quickly in X hours, but rarely increases in greater magnitude in Y hours. However, the second evaluation is just as important. And that is how much flexibility do you currently have that needs a signal to achieve more. An example of this is if most of your resources provide ramping within an hour, but there isn't much more ramping upward that can happen within two hours. The results on this chart show that while the needs were greater per hour at a one hour horizon, the amount of available flexibility was almost the same in MW between one and three hours. This meant there was a larger gap in the available flexibility to meet the 3-hour ramp need than the gap to meet the one-hour ramp need, even though the one-hour ramp need appeared to be more significant.

EPRI

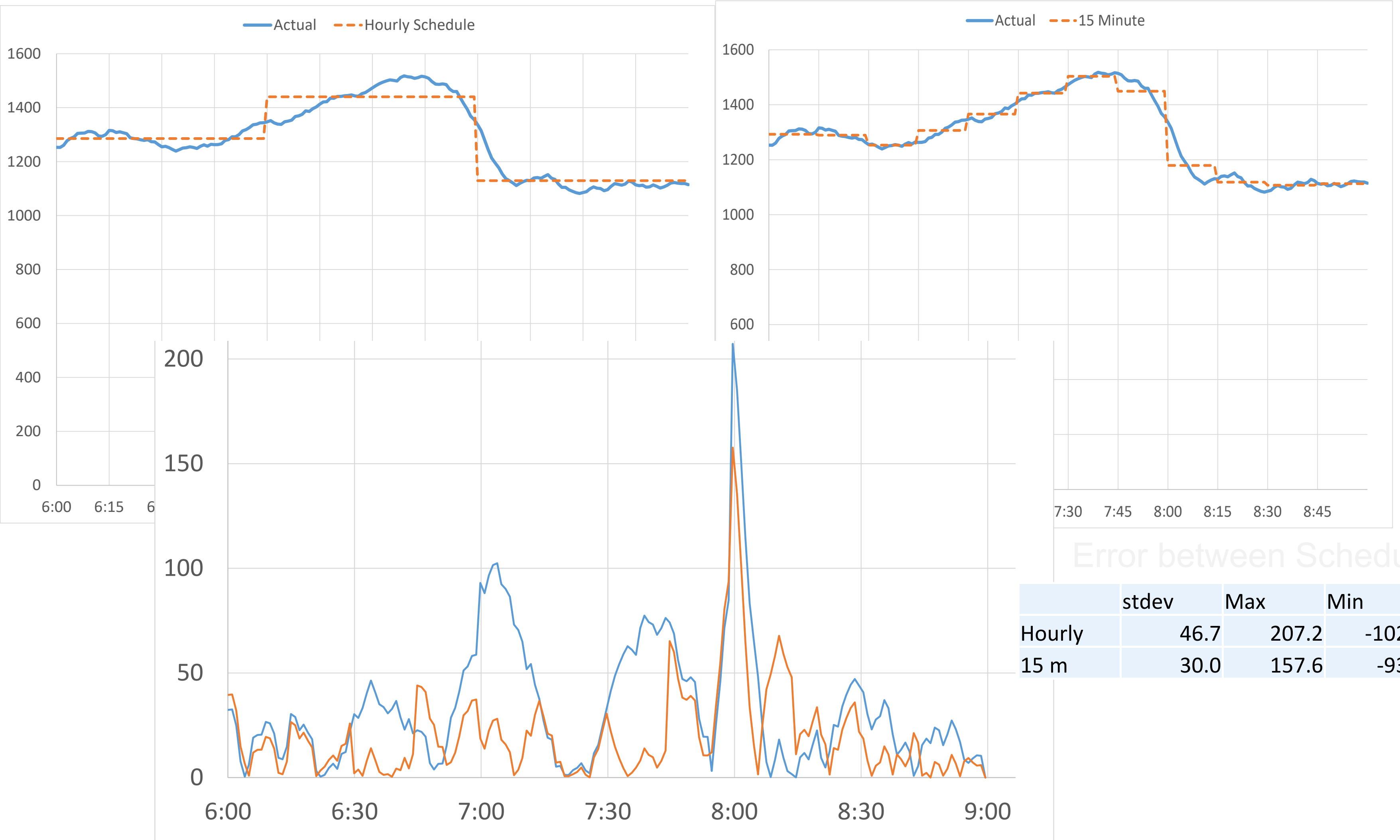
#7 Meeting Operating Reserve Needs Implicitly Through Advanced Scheduling Applications

Three Central Needs for Reserve	Explicit Reserve Requirement	Implicitly Scheduled Operating Reserve
1. Variability occurring within the interval	Reserve Requirements (e.g., regulation reserve)	Shorter scheduling intervals
2. Variability anticipated beyond the interval	Reserve Requirements (e.g., flexible ramping reserve)	Time-coupled multi-period dispatch w/ longer look-ahead horizons
3. Uncertainty of future conditions	Reserve Requirements (e.g., contingency reserve)	Stochastic or robust unit commitment and dispatch meeting multiple scenarios

https://www.epri.com/research/products/000000003002021781

EPRI Research has shown that the impacts of variability and uncertainty coming from different sources of the power system can be met either through explicit reserve requirements, or implicitly through advanced scheduling features. The operating reserve requirement that is used, and that we are more familiar with for different reserve types is then determined based on what the residual need is that is not met through implicit means. This means that certain features in the scheduling applications can reduce the need for explicit reserve, and that the requirements for those reserve should take into account the scheduling parameters themselves. Examples of three reserve needs and how they are met through explicit or implicit means is shown in the table. One can see that, if it were feasible, a scheduling application that had extremely short scheduling intervals, with horizons that extended very far into the future, and with a large list of possible scenarios that were modeled within the scheduling procedure, it is possible that no explicit reserve were necessary at all.

#7 Reserve needs depend on scheduling parameters



Error between Schedule and Actual

	stdev	Max	Min	MAE	95th %ile
Hourly	46.7	207.2	-102.4	33.9	77.2
15 m	30.0	157.6	-93.6	20.0	49.7

A system that uses a scheduling model with 15-minute intervals requires less explicit reserve than one that uses a scheduling model with hourly intervals

This is a simple illustrative example of how explicit reserve can be visualized as the residual need that is not captured through the scheduling application. THe two charts on the top show the actual net load and the schedules that are met with a 1-hour scheduling interval resolution and a 15-minute scheduling interval resolution, respectively. The table on the lower right shows statistics of the delta between the actual and scheduled values for both cases. One can observe the need being lower with 15-minute scheduling intervals, meaning that the reserve to meet this variability that is used as a requirement in the explicit reserve product would be lower with 15-minute scheduling intervals than it would be with hourly scheduling intervals, with all else being equal. This doesn't mean that the impacts are lesser, just that the amount needed to be met that are not met implicitly are lower with the higher (shorter) interval resolution. Similar examples could be used for other scheduling advancements such as long horizons and multi-scenario models.

#8: Implicit Scheduling achieves greater economic efficiency than reserve, but may lack incentives

	Cost	Ramp	Capacity
G1	20$/MWh	2 MW/min	100 MW
G2	30$/MWh	2 MW/min	100 MW
G3	80$/MWh	2 MW/min	100 MW

Load	I1	I2
Time Coupled Multi-Period (TCMP)	200 (19)	218
Single Period w/ Ramp Product	200 (19)	218

TCMP	I1	I2 (adv.)
G1	100	100
G2	91	100
G3	9	18
LMP ($/MWh)	30	(130)80
Flexi price ($/MWh)	N/A	N/A

https://ieeexplore.ieee.org/document/7192736/

Insight #7 described how reserve needs can be met explicitly with reserve requirements or implicitly through advanced applications and modified scheduling procedures. EPRI research has shown that meeting the needs implicitly through scheduling applications has benefits in terms of economic efficiency and reliability due to how the features are used and incorporated into modeling. However, implicitly meeting these needs may not lead to price formation that would lead to incentive compatible results, and prices that align with the schedules that are produced. This may mean while efficient results are achieved in theory, there may lack incentives for market participants to follow the schedule results, at least with no other modifications. We show this example comparing the need for “inter-interval variability”, variability occurring interval to interval into the future. The need can be met explicitly with a reserve product that ensures reserve is available to meet the future change in conditions, or it can be met implicitly through time-coupled multi-period dispatch and including the horizon within the optimization. The table here shows the results of the TCMP with the three generators and two-period economic dispatch.

#8: Implicit Scheduling achieves greater economic efficiency than reserve, but may lack incentives

	Cost	Ramp	Capacity
G1	20$/MWh	2 MW/min	100 MW
G2	30$/MWh	2 MW/min	100 MW
G3	80$/MWh	2 MW/min	100 MW

Load	I1	I2
Time Coupled Multi-Period (TCMP)	200 (19)	218
Single Period w/ Ramp Product	200 (19)	218

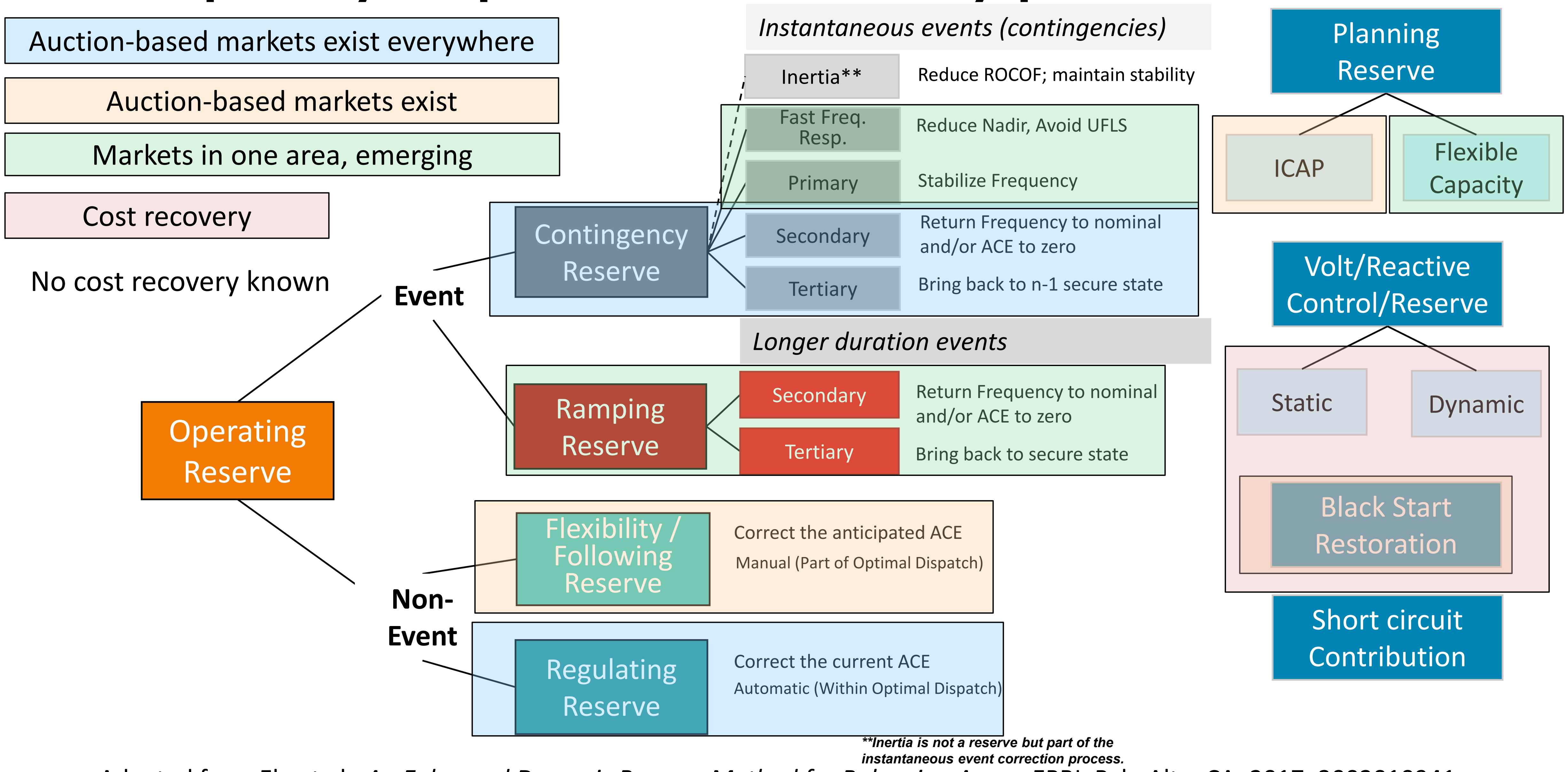
https://ieeexplore.ieee.org/document/7192736/

TCMP	I1	I2 (adv.)
G1	100	100
G2	91	100
G3	9	18
LMP ($/MWh)	30	(130)80
Flexi price ($/MWh)	N/A	N/A

Comparison (Same Costs, Same Schedules)	Time-coupled multi-period	Flex ramping capability product
G3 cost	$2,160	$2,160
G3 revenue	$1,710	$2,660
G3 profit (rev – cost)	$-450	$500

For this example, the TCMP and explicit reserve for variability cases have the same costs and same schedules for all resources, although this is not always the case. Typically, the TCMP can better anticipate the costs in the future and provide a more economic result. However, with the schedules being exactly the same, we observe that the settlement is not. The result with explicit reserve has the advantage of producing a reserve price that is paid to the reserve providers. The TCMP case ends up depressing the price in the second interval due to the market design choice that only the first interval of the real-time market is binding for settlement. The price in the second interval is likely to be depressed even though the opportunity cost of G2 is already incurred. This leads to a negative earnings for G3 over the two intervals in the TCMP case, which would result in a disincentive, or non-transparent make-whole payments. A solution might be to price the decision in interval 1 similar to a reserve, which has been proposed as the cross-interval marginal price (CIMP).

#9: Frequency response markets may provide benefits



Adapted from Ela et al., *An Enhanced Dynamic Reserve Method for Balancing Areas*, EPRI, Palo Alto, CA: 2017. 3002010941.

This chart has some of the examples of how different ancillary services are compensated in electricity markets. From those that have competitive auctions in all U.S. market regions, to those that have no identified recovery of costs. Primary Frequency response is one of those services that is not often incentivized in organized electricity markets or that has competition for suppliers. Generators have historically had the capability as inherent capability, but the response is still based on having enough head room capacity to provide, and the unit of measure is not necessarily in MW, due to the dependency to frequency.

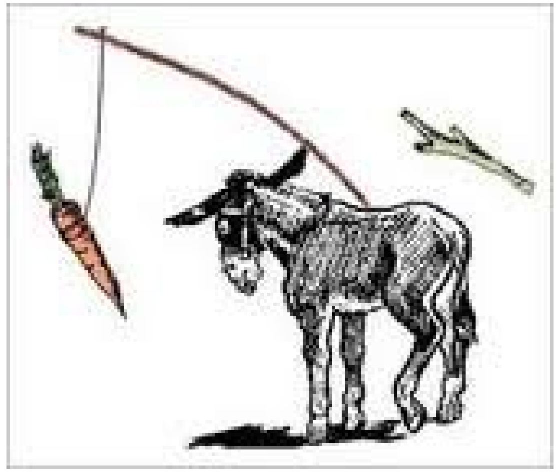
#9: Frequency response markets may provide benefits

- 3% Penalty Band over generation schedule
- 60 Hz system
- 5% Droop curve setting
- 0 Hz Dead band

$$\frac{1\ pu\ power}{0.05\ pu\ frequency} = \frac{0.03\ pu\ power}{X\ pu\ frequency}$$

$$X = 0.0015\ pu\ frequency = 90\ mHz\ for\ 60\ Hz\ system$$

- Any deviation greater than 90 mHz, a generator **<u>automatically</u>** will be penalized with a functioning turbine governor enabled!
- (1) Eliminate disincentive (2) Need for incentives?



E. Ela, A. Tuohy, M. Milligan, B. Kirby, and D. Brooks, "Alternative approaches for a frequency responsive reserve ancillary service market," *The Electricity Journal*, vol. 25, no. 4, pp. 88-102, May 2012.

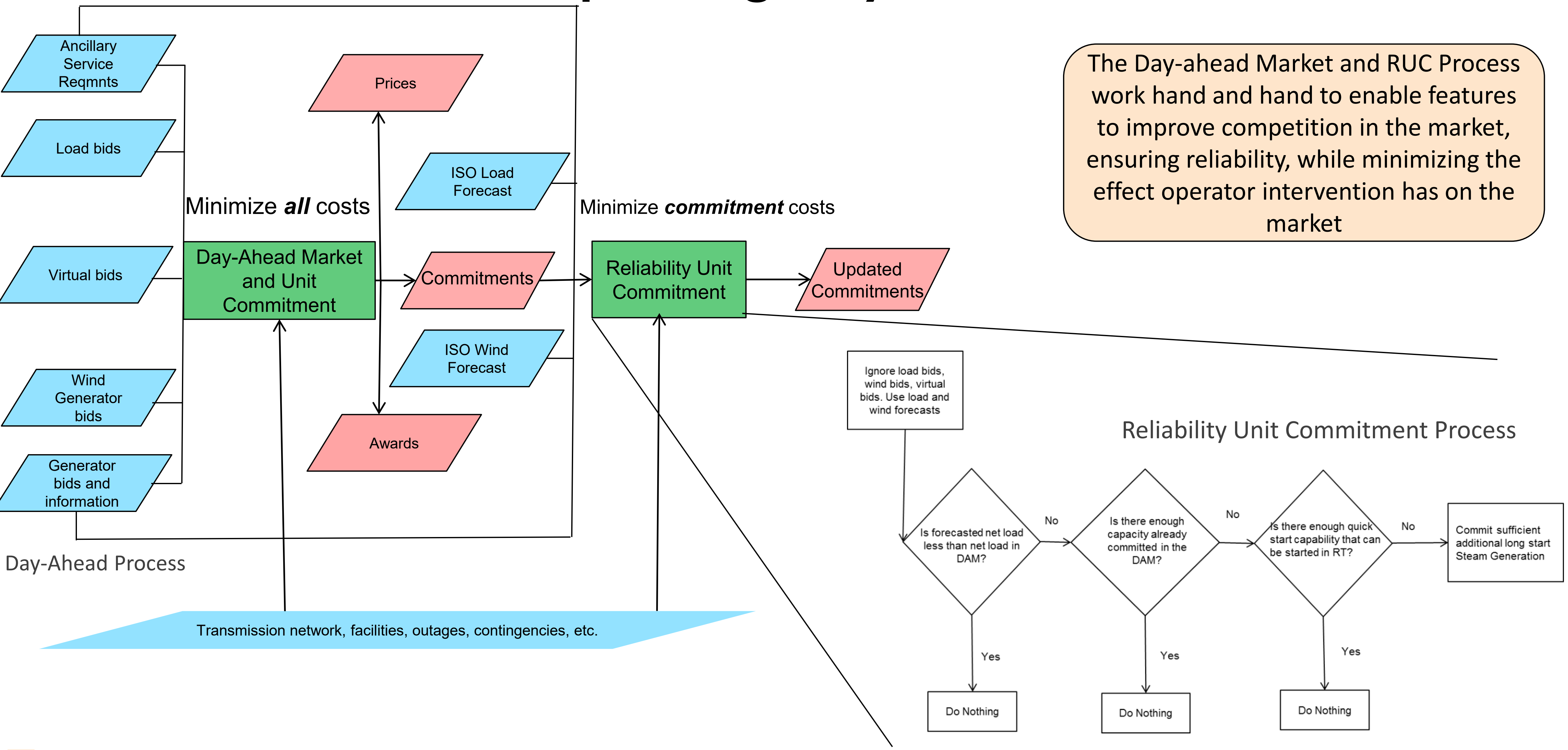
It has been even observed that there could be a disincentive to provide the service. Given typical droop settings of 5%, and knowledge of at least one market area which has a penalty if a resource deviates from its market signal by more than 3% of its capacity, it can be observed that a 90 mHz deviation in frequency would automatically cause any frequency response capable resource of getting a financial penalty. Perhaps more importantly is knowing that any deviation can move a resource closer to that tolerance band and getting penalized for a smaller deviation otherwise. So should eliminating a disincentive like this be all that is necessary, or should an incentive be required as well.

#9: Frequency response markets may provide benefits

- Eliminate penalties during frequency events
- Eliminate penalties if frequency response provided
- Add PFR capability to spin or regulation product
- Separate PFR ancillary service market in MW/Hz

Ever since the NERC BAL-003 standard was modified to include a frequency response obligation (FRO) for balancing authority areas to meet, having sufficient frequency response has become more important. In regions that might not have sufficient frequency response during certain conditions to meet the FRO, a frequency response ancillary service market could provide benefits to ensure that the amount is met and that when the system is constrained, and resources are moved out of merit to meet the obligation, those resources are compensated. It also can provide a signal for improved frequency response capabilities in the future, such as for grid forming inverter capability. The incentive can be done in a number of different ways as shown in the slide here, including embedding the requirement in other ancillary service market products. However, a separate primary frequency response (PFR) product that explores the amount in MW/Hz from different resources, and correctly models the contribution, is likely the most efficient if it can be observed that the benefits outweigh the costs of designing and administering such a market product.

#10 Limitations to improving day-ahead forecasts



Day-ahead forecasts are critical to maintaining reliability and are a necessary data set for operating the electric grid. However, it is important to recognize the limitation that improving day-ahead forecasts beyond the current state-of-the-art accuracy may have for improving economic efficiency. The reasons for this are shown in this chart. In U.S. organized electricity markets, day-ahead processes include a day-ahead market and a subsequent reliability unit commitment (RUC) process. The RUC process takes the existing commitments from the market, and makes any additional unit commitments when it finds that there is no other alternative. The only place where ISO forecasts are used are in the RUC process so that the ISO does not force a position onto any market participant within the day-ahead market. As system operators want to avoid interference with the market as much as possible, they will modify the RUC process to make it so already-committed resources and those can defer commitment until real-time are used before any commitment decisions. Only the no-load and start-up costs are really considered so that if any commitment decisions must happen, the most economic are selected. Since the forecasts are only used in this process where limited economic data is used, improved accuracy may not have a great impact on the overall economic outcome of the day-ahead commitment process.

Bonus: Price of ancillary services under zero fuel cost systems is complex

	Cost	Min Gen	Capacity/Forecast
VER1	0$/MWh	0 MW	100 MW
VER2	0$/MWh	0 MW	50 MW
G1	$500 (NL) + 50$/MWh	40 MW	100 MW

Load	Reserve
120	30

VER can provide reserve

Scenario 1	P	R
VER1	85	15
VER2	35	15
G1	0	0
LMP ($/MWh)	0	
Reserve price ($/MW-h)	0	
Make whole payment ($)	0	

We already had our ten insights, but we would like to share this bonus insight. Here we provide some preliminary insights into the impact that very high levels of renewables that have zero fuel costs may have on the ancillary service markets and ancillary service prices. Sometimes, expectations are that with high levels of zero-fuel-cost resources integrating to the system to provide energy that the wholesale energy prices will decline, but that ancillary service prices may have to rise to make up for those declines. As ancillary service prices are heavily dependent on lost opportunity costs to provide energy with the existing ancillary service market designs of today, it isn't that clear what the outcome will be. In this example, there are two variable energy resources with zero operating cost and one generator with a $500 No-load and $50/MWh operating cost. The load is 120 MWh and reserve is 30 MW. When the VER can provide the reserve, the generator is not needed, and assuming the VERs offer energy and reserve at $0, prices for both energy and reserve are set to $0/ MWh.

Bonus: Price of ancillary services under zero fuel cost systems is complex

	Cost	Min Gen	Capacity/Forecast
VER1	0$/MWh	0 MW	100 MW
VER2	0$/MWh	0 MW	50 MW
G1	$500 (NL) + 50$/MWh	40 MW	100 MW

Load	Reserve
120	30

VER can provide reserve

Scenario 1	P	R
VER1	85	15
VER2	35	15
G1	0	0
LMP ($/MWh)	0	
Reserve price ($/MW-h)	0	
Make whole payment ($)	0	

VER cannot provide reserve. Traditional LMP

Scenario 1	P	R
VER1	65	--
VER2	15	--
G1	40	30
LMP ($/MWh)	0	
Reserve price ($/MW-h)	0	
Make whole payment ($)	2500	

Given the variable and uncertain nature of these resources, there may be some rules limiting how much reserve these resources are eligible to provide. Taking the other bookend, we assume that the VER are not eligible to provide any reserve. In this situation, using traditional LMPs, the energy price and reserve price are still both $0/MWh. As the VERs are curtailed, they would provide the next increment of demand on the system at an incremental cost of $0, and since there are no opportunity costs of losing out on profit with a $0/MWh energy price, the price of reserve in this situation would also be set to $0/MWh. There would in fact be a large make-whole payment in this situation to G1 for the costs that it does not recover.

Bonus: Price of ancillary services under zero fuel cost systems is complex

	Cost	Min Gen	Capacity/Forecast
VER1	0$/MWh	0 MW	100 MW
VER2	0$/MWh	0 MW	50 MW
G1	$500 (NL) + 50$/MWh	40 MW	100 MW

Load	Reserve
120	30

VER can provide reserve

Scenario 1	P	R
VER1	85	15
VER2	35	15
G1	0	0
LMP ($/MWh)	0	
Reserve price ($/MW-h)	0	
Make whole payment ($)	0	

VER cannot provide reserve. Traditional LMP

Scenario 1	P	R
VER1	65	--
VER2	15	--
G1	40	30
LMP ($/MWh)	0	
Reserve price ($/MW-h)	0	
Make whole payment ($)	2500	

VER cannot provide reserve. ELMP*

Scenario 1	P	R
VER1	65	--
VER2	15	--
G1	40	30
LMP ($/MWh)	0	
Reserve price ($/MW-h)	25	
Make whole payment ($)	1750	

*Formulation dependent

There is some concern over price formation when a majority of the funds exchanged go through side payments and not transparent price signals. For energy payments, proposals such as extended locational marginal price (ELMP) or fast-start pricing (FSP) have been put in place to incorporate some of the operational fixed costs into prices. A similar mechanism may be used in the future to improve the pricing of ancillary services. In the third example here, we apply a version of the ELMP pricing (which can vary depending on how it is formulated), where we relax the minimum generation and integer constraint to get a non-zero price for reserve. This non-zero price essentially incorporates the no-load cost and variable cost of G1, while the energy price is still based on the VERs being the marginal resource. If prices are in fact designed like this, we could end up seeing more intervals with $0/MWh energy prices while ancillary service prices are non-zero. Further exploration on how often this occurs is recommended.

Appendix (EPRI Response to FERC AD21-10)

Summary

- The flexibility response time need (e.g., 10-minute versus 30-minute versus hour ramp product) can be considered a continuum, but if needed, the most critical time horizon can be determined by 1) the largest need or inflection point 2) where lack of associated flexibility may currently exist. (Paragraph 2)
- Forecasting dynamic reserve is similar to forecasting load or Variable Energy Resource(s) (VER) output, and can be used to improve reliability and economic efficiency. (Paragraph 5)
- Technology agnostic eligibility requirements for ancillary services and flexibility can lead to efficient competition and innovation. (Paragraphs 7 and 35)
- Extended downward sloping Operating Reserve Demand Curves (ORDCs) beyond the minimum contingency reserve requirement can be beneficial if designed adequately, as can flexibility reserve. These two flexibility mechanisms may be identical under different names; however, the current implementation of both may lead to different outcomes. (Paragraphs 9-11, 24)
- Reserve products may be better defined by the decision processes compared to response times (e.g., day-ahead reserve released in hour-ahead). This can align the reserve holding to the decisions that can be made according to the scheduling process. (Paragraphs 14 and 17)
- Longer-term reserve may be beneficial because forecast errors are larger in these time periods. While not intuitive, reserve during this time frame may also improve economic efficiency by reducing costs, in addition to any reliability benefit it may have. (Paragraphs 13-14, 16-17)
- The price of ancillary services under zero-fuel cost futures is complex and unclear. The interval-based lost opportunity cost when energy price is zero may also be zero. Non-convex pricing, inter-temporal opportunity costs, and shortage pricing design can all impact ancillary service prices in the future. (Paragraph 15)
- Designing and introducing products and participation models may demonstrate benefits in studies; however, the design (including stakeholder debate), implementation, and administration can also be costly. (Paragraphs 19 and 49)
- Frequency response ancillary service markets may provide benefits in regions that foresee difficulty in meeting the RTO's/ISO's frequency response obligation. (Paragraph 18)
- Implicit ways of meeting operating reserve such as multi-period dispatch and shorter market intervals, have efficiency gains but require additional consideration in regards to incentives. (Paragraphs 21-22, 43-45)
- Because day-ahead forecasts by the RTO/ISO are not used directly in the day-ahead market process, there may be limited benefits from more advanced forecasting techniques in this time frame. (Paragraph 41)
- Multi-period real-time models can have economic efficiency benefits. Pricing needs to be improved due to complications of only using the binding interval for settlement and additional consideration is necessary for large storage penetrations. (Paragraphs 43-47)

.

In the appendix we include the entire EPRI response to questions asked by the Federal Energy Regulatory Commission. The questions asked by FERC are shown in the heading and the responses follow. Most of the responses provide additional context to the insights shown in the main body of this report.

Question 1.1.1. RTOs/ISOs and other industry experts generally agree that power systems will require greater flexibility from system resources in the future. What operational capabilities or services will be most valuable to RTO/ISO operators in the future as the resource mix and net load profile changes and why? Is there a desirable reaction time, sustained performance duration, etc. expected from a resource?

1. **Response**. The need for greater flexibility is apparent in most studies that evaluate future system reliability needs and is notably observed on systems that have experienced higher growths of VER in practice. Flexibility has many definitions, but at EPRI, when focusing on active power flexibility, it is described as the operational maneuverability of power across the set of resources available to the system. Individual resources have greater flexibility when they can adjust output more rapidly, have more adjustable range between minimum and maximum limits, can sustain output for longer, and have fewer constraints between operational modes (e.g., lower start-up times, lower minimum down times, etc.). Each of these attributes will be desirable on a future system with increased variability and uncertainty, and it is important that incentive signals and planning studies provide indication on those attributes that may be more desirable for each region and horizon.

2. Desirable reaction or response times are also system dependent and to the extent possible must be considered as a continuum rather than discrete needs. Designing a ten-minute product for ramp capability must ensure that 20 minutes of ramping capability does not appear as worthless. If designing products explicitly for a specific response time is the practical solution to incentivize the availability or investment in ramp capability, then there are two important factors to consider: first, at which response time the magnitude of change or ramp rate requirement may be the largest relative to others; and second, the current availability (or lack thereof) of resources that are able to meet that response time. In a recent study for the CAISO, EPRI found that 3-hour net load changes had almost double the magnitude (MW) of 1-hour net load changes on the CAISO system for the time period studied. Because the needs were double and not triple, the 1-hour horizon had a greater ramp need (MW/min), which on the surface would have led to the conclusion that the 1-hour horizon was the more critical response time for designing a flexible capacity product. When investigating the actual *feasible* flexibility to provide ramping for a 1-hour horizon and a 3-hour horizon, the amounts were nearly identical. Most of the online resources would already use up the majority of their available ranges in one hour, such that they had no additional flexibility to offer in three hours. Additionally, there were not many resources in the region that had start-up times greater than 1 hour and less than 3 hour that could contribute to 3-hour flexibility needs that could not contribute to 1-hour flexibility needs. In conclusion, while the 1-hour horizon had a larger ramp need, the 3-hour horizon was more important to incentivize flexibility builds because there was a greater need for more resources to meet the larger magnitude. It is important to look at both perspectives and also to be flexible in case these factors change through time.

Electric Power System Flexibility: Challenges and Opportunities. EPRI, Palo Alto, CA: 2016. 3002007374. Available: https://www.epri.com/research/products/000000003002007374.
Flexibility Assessment for the California ISO: Evaluating Flexibility Needs and Systemwide Feasible Installed Flexible Capacity. EPRI, Palo Alto, CA: 2019. 3002013725. Available: https://www.epri.com/research/products/000000003002013725.

Question 1.1.2. To what extent will the "traditional ancillary services" defined in Order No. 888 and existing energy market designs continue to ensure reliability as the resource mix changes in RTO/ISO markets in the future? Will traditional ancillary services provide the appropriate types and adequate quantities of operational flexibility RTOs/ISOs need to manage both expected (e.g., reasonably predictable) and unexpected (e.g., inherently uncertain and captured in forecast errors) variability in net load? Will existing RTO/ISO energy and ancillary services market designs that generally compensate certain traditional ancillary services resources based on the opportunity cost of foregone energy sales – for example, spinning and non-spinning reserves – give resources a sufficient economic incentive to offer their flexible capabilities to the RTO/ISO?

3. **Response**. The traditional ancillary services were primarily put in place for contingency events and for regulation service that utilized automatic generation control (AGC) to respond to normal variations in load within an economic dispatch interval. Recent changes including more VER and greater supply impact from extreme weather events have led to greater variability and uncertainty on both the demand and supply side. Changes to the supply mix have also led to certain attributes, previously inherent in traditional generating technologies, requiring additional choices for emerging technologies to have similar capabilities. The former situation can lead to changes to existing or new ancillary services that can accommodate the higher probability and frequency of larger power balance conditions. Examples include extended downward sloping operating reserve demand curves placed on existing ancillary services or new ramp capability services. The latter situation can lead to changes to existing or new ancillary services to promote incentives for resources that do not have inherent capabilities that are needed for system reliability to invest in the equipment needed for similar capabilities. Examples include primary frequency response, fast frequency response, and inertia.

4. Utilizing marginal cost pricing for ancillary services as is currently done will allow for those prices to be based on availability costs, lost opportunity costs, and administratively-set shortage pricing. With prices incorporating availability costs and lost opportunity costs, resources have the incentive to standby, forego the provision of energy, and provide the capacity for the service. When called upon to deploy and convert the capacity reserved into energy, energy prices are typically higher due to the shortage condition and the resource is incentivized to respond. The activation of shortage pricing, including prices set based on longer multi-step ancillary service demand curves, can also provide the incentive to respond and may also incentivize resources to build and invest with flexibility. ISOs and stakeholders may benefit from continuing to evaluate whether ancillary service markets incentivize the investment in flexible resources, combined with the incentives for flexibility built within energy markets, capacity markets, and other products or contracts. It is important to note that when a zero-fuel cost resource is setting the energy market price to zero, the lost opportunity cost that may be used to set the ancillary service capacity price is going to also be zero in most circumstances. Thus, shortage pricing and demand curve design will be important.

[4] We note that some ISOs/RTOs may reduce the reserve requirements used in market clearing during reserve deployments, which may cause the energy price not to increase substantially.

Question 1.1.3. How should RTOs/ISOs define the system's need for operational flexibility, now and in the future? To what extent is operational flexibility needed on a bi-directional basis (i.e., both up and down) versus a unidirectional basis (i.e., only up or down)? How do these needs compare to the services provided by traditional ancillary service products?

5. **Response**. Operational flexibility needs can generally be identified through observations and forecasting. Forecasting the load and VER output in the future can determine the anticipated variability need for the associated time horizon in question. Historical information can also provide information on the uncertainty in that horizon, from which system operators may use probability distribution functions from historical forecast errors and assign a confidence interval that they hope to have sufficient flexibility to achieve. EPRI research has focused on the concept that operating reserve needs can be forecasted, just like load or VER output can be forecasted. By using historical data and advanced forecasting techniques, system operators can forecast the need for reserve based on different needs (events compared to normal conditions) and time horizons based on their risk criteria (or those of reliability standards such as NERC BAL standards).

6. Flexibility and reserves are called upon and deployed in both the upward and downward direction to meet under-generation conditions (low frequency or negative Area Control Error (ACE)) and over-generation conditions (high frequency or high ACE), respectively. This is true for both traditional reserves like contingency reserve and newly introduced reserve products like ramp products. However, the quantity needed and the frequency of when this flexibility is needed differs between these two conditions. In addition, reserve products and ancillary service markets that are enforced as capacity and ramp constraints in market clearing models, are put in place to influence the market clearing solution to improve the likelihood that reserve can be met. Ample downward flexible capacity is typically available as an artifact of unit commitment and economic dispatch procedures that meet load and ensure transmission deliverability. If a reserve product, such as a downward capacity reserve product, is unlikely to change the commitment and dispatch decisions whether it was enforced or not, then there is little value in creating it as a product in the market. The design and operation of new ancillary service products, including the time invested in stakeholder deliberations, can be costly and thus RTOs/ISOs may only want to evaluate when there is a clear benefit. Finally, new reserve products can also slow down market clearing solutions, that can threaten the ability to clear the market within market clearing time windows.

Question 1.1.4. Could variable energy resources or new resource types (e.g., storage, hybrid, and co-located resources) be operated or dispatched differently from the status quo to provide greater operational flexibility to the RTO/ISO, if so, how? Given the evolving resource mix, are the current eligibility requirements for each resource type to provide ancillary services appropriate?

7. **Response**. Most of these emerging resource types have demonstrated the capability to provide operational flexibility. While it is unclear whether any operational changes are necessary for these resources, it is important to recognize that proper price formation which aligns prices to decisions for each individual resource will guide each resource to provide flexibility when it is needed and when it is economical for that resource to do so. Eligibility requirements that are technology agnostic, that are clearly defined based on the needed capability or attribute, or based on performance as observed on the system is a path that will allow all technology types that are interested to meet those requirements and provide a greater set of participants to compete to provide these flexibility services.

Question 1.2.1. Contingency reserves are provided by existing 10- and 30-minute reserve products and are designed to ensure the system can recover from a contingency (e.g., a generator or transmission outage). How will the procurement of additional contingency reserves help RTO/ISO operators manage routine operational flexibility needs (e.g. needs driven by net load variability and uncertainty)?

8. **Response**. Additional reserve, above those that may be required to meet a minimum requirement, to meet a large resource forced outage can support additional operational flexibility needs, as long as the deployment of the reserve are not restricted by the event type. If more 10-minute or 30-minute reserves are procured than may be needed to meet contingency requirements, and a net load forecast error requires that contingency reserve to be converted to energy to make up the loss, the system can still meet the minimum contingency reserve requirement in real-time after that deployment.

Question 1.2.2. What are the benefits of procuring contingency reserves beyond the minimum reserve requirement through a given ancillary service product? If employing such a method, how should RTOs/ISOs determine the market's demand for contingency reserves (both the quantity and willingness to pay) beyond the minimum reserve requirement of a given contingency reserve product? What principles should RTOs/ISOs follow if they consider revising the shape of the ORDC for a given contingency reserve product (e.g., introducing additional steps or graduation to the ORDC curve)? For example, should the willingness to pay for such additional reserves be based on the Value of Lost Load times the loss of load probability with a given quantity of the reserve product associated with the ORDC, the cost of actions operators would take to procure additional reserves, or some other valuation method? How should customer willingness to pay be incorporated?

9. **Response**. The benefits of procuring reserve above the minimum reserve requirement may include putting value to additional capacity that can be used for reasons other than contingencies and allow for sufficient capacity to be available in case of a contingency when other conditions such as forecast errors are present and consume the reserve capacity. Without this employed, capacity reserved above the minimum requirement is considered valueless. Procuring reserve above the minimum reserve requirement has a similar benefit to introducing a new product, such as a ramp product, that is used for the same reasons.

10. Just like setting the requirements for a separate product, such as flexibility products, the quantities for setting reserve above the minimum contingency reserve can be determined using reserve forecasting techniques. The steps of the curve can be partitioned using different confidence/exceedance levels. For example, covering the 60^{th} percentile of the uncertainty set of conditions can be one step on the curve and covering the 90^{th} percentile could be a separate step, with a lower price setting, further down the curve. There are other methods to employ. There will be different implementations of ORDCs and more research is required to share on what practices may lead to desired results for which systems and under which scenarios.

Question 1.2.4. To what extent do RTOs/ISOs use contingency reserves to manage noncontingency related operational uncertainties (e.g., expected and unexpected net load variability)? If such reserves are used for this purpose, should this alter an RTO/ISO's approach to establishing the maximum height and shape of the ORDC? Under such approaches, how do prices in the ORDC appropriately reflect the marginal reliability value contingency reserves provide?

11. **Response**. To EPRI's knowledge, this differs across different RTOs/ISOs. It is EPRI's understanding that some RTOs/ISOs, such as those in the WECC region, are only able to use contingency reserve for large disturbances (forced outages). Others are able to use it for non-contingency reasons, such as when there is significant ACE or frequency deviation resulting from non-contingency conditions. Although not explicitly a required feature, ISOs/RTOs may often attribute a higher penalty value to a reserve (or other) violation that would lead to a NERC standard violation and resulting financial penalty. This may be one reason why it might affect the ORDC shape and prices.

Question 1.2.5. Is there a particular point at which procuring reserves beyond the minimum reserve requirement can reduce or conflict with the objectives of shortage prices? What is an appropriate balance between raising shortage prices and procuring reserves beyond the minimum reserve requirement given that procuring additional reserves can reduce the probability of the RTO/ISO experiencing a shortage?

12. **Response**. Shortage pricing has two objectives. It can signal resources to provide as much energy or reserve as possible in the short-term to reduce the shortage by as much as possible. It can also provide needed investment signals for resources to recover capital costs, due to the nature of co-optimization (or price adders in the case of ERCOT), which typically carries the reserve shortage prices to the energy prices. Procuring reserve above the minimum reserve requirement with a lower shortage price to the shortage price set for insufficient reserve to meet the minimum requirement, but still relatively high compared to average ancillary service prices and energy prices on the part of the demand curve does not appear to conflict with these objectives, as long as these objectives are considered during the design of the curves. The extended demand curve can incentivize resources in the short-term to start-up and commit or provide energy/reserve commensurate with the value of those reserves above the minimum. The higher prices during when the reserve is above the minimum reserve requirement but still on the demand curve will coexist with lower prices and fewer periods where there exist shortages of the minimum reserve, because this greater amount of reserve procured can reduce that risk of being short on the minimum reserve requirement. This careful balance may allow for the investment signal objective to continue in a similar manner as it would without procuring reserve above the minimum requirement. Thus, these objectives of shortage pricing remain generally intact with the added potential benefit of improved reliability. However, the design of the prices, at what quantities they are set, and curve shape will be critical to achieving this balance. ISOs/RTOs can work closely with stakeholders, including consumer representatives and reliability experts, and may desire to perform analysis and simulation, and in some cases be able to adapt and make adjustments over time from experience if it appears the desired outcomes have not been reached.

Question 1.3.1. Ramp products, as distinguished from traditional ancillary service products, are relatively new ancillary services that are in place in CAISO and MISO, and approved for implementation in SPP. Ramp products are generally not designed to address contingencies but are instead a mechanism to position the system efficiently to meet forecasted ramping needs in future intervals at least cost on an expected basis. RTO/ISO ramp products procure ramp on a short-term basis (e.g., for intervals of 10 or 15 minutes), but longer-term ramp products are being considered. For example, SPP is considering a longer-term ramp product and the California Department of Market Monitoring has advised CAISO to consider a longer-term ramp product. What drives the need for, and what are the benefits of, a longer-term ramp product compared to the existing shorter-term ramp products or traditional reserve products?

13. **Response**. The short-term ramp products that are in place in CAISO and MISO and approved for implementation in SPP are put in place to meet short-term "almost real-time" issues, even if they are also enforced in the day-ahead or other forward horizons. They are set primarily to meet the 5- to 20-minute expected needs and 10- to 30-minute ahead forecast errors. Without these products, these ISOs would see a higher amount of price spikes in the energy market because there is not enough *committed* capacity to meet the short-term changes in net load. Even though these short-term products are primarily required to be online only, they do create commitments of resources so that they are online to hold the reserve. However, it is well known that net load forecasts have much larger magnitude errors in the longer horizons such as day-ahead. While these short-term products may improve operation in the context of day-ahead forecast errors, that is not their objective.

14. The longer-term products of an hour or more may be designed to support longer-term variability and longer-horizon forecast errors. The magnitude of the need for this type of service may be greater than the existing short-term products, while the ramp rate requirement is less. By having both products in place, either or both constraints may bind, and the ISO/RTO may be better able to promote sufficient capacity to meet the longer-horizon ramp needs while also having sufficient ramp to meet the shorter-horizon needs.

Question 1.3.1. Ramp products, as distinguished from traditional ancillary service products, are relatively new ancillary services that are in place in CAISO and MISO, and approved for implementation in SPP. Ramp products are generally not designed to address contingencies but are instead a mechanism to position the system efficiently to meet forecasted ramping needs in future intervals at least cost on an expected basis. RTO/ISO ramp products procure ramp on a short-term basis (e.g., for intervals of 10 or 15 minutes), but longer-term ramp products are being considered. For example, SPP is considering a longer-term ramp product and the California Department of Market Monitoring has advised CAISO to consider a longer-term ramp product. What drives the need for, and what are the benefits of, a longer-term ramp product compared to the existing shorter-term ramp products or traditional reserve products?

14. **Response (continued).** Another important consideration for the longer-horizon products and short-horizon reserve is the difference in decisions that can be made by the ISO/RTO or balancing area authority to provide capacity available to meet each need. EPRI research has focused on defining reserve products based on the scheduling process in which capacity and ramp is held along with the scheduling process in which it released, and possibly deployed if needed. For example, regulation is held in day-ahead and deployed in AGC. Contingency reserve is held in day-ahead and deployed through operator action or contingency dispatch programs. Rather than define the reserve by the arbitrary time horizon, defining it through these scheduling process may better achieve the actions that the ISO/RTO needs to take to secure sufficient flexibility on the system. For example, holding reserve during the day-ahead commitment process and then releasing in the real-time dispatch process (or hour-ahead scheduling process), will allow for sufficient flexible resources to be available and used in case day-ahead forecast errors result, but not overprocure when the reserve is no longer needed. It is important to note that not all reserve has to be online. In fact, eligible reserve characteristics should be based on what decisions are made during the "release and deployment" stage of the reserve. For example, if reserve is held in the day-ahead and released for potential deployment in the hour-ahead commitment process, then any resource that can be committed in the hour-ahead commitment process should be eligible to provide that reserve in the offline state in the day-ahead process.

Question 1.3.2. Will establishing reserve and ramp prices based on foregone energy revenues provide such signals in a system with a high penetration of variable energy resources, many of which have low or zero marginal costs? If not, what other options exist to ensure sufficient compensation for resources providing reserve and ramp capability? Historically, the prices for the ramp products in CAISO and MISO have often been zero. Are ramp prices expected to increase over time as system needs evolve? If so, what specific conditions might cause ramp prices to increase? Will any expected ramp price increases be sufficient to incent and appropriately compensate the ramp capability RTOs/ISOs and others expect will be needed due to the changing resource mix?

15. **Response**. Marginal cost pricing in electricity markets which include thousands of different unique constraints is complex. It can be even more complex when there exist other rules, for example those related with fast-start pricing, that set the price at different values than what would be determined through the dual solution of the security constrained economic dispatch software. During instances where VER set wholesale energy prices across an RTO/ISO region, there is not foregone energy revenue for any resource that is remaining on the system, and thus no lost opportunity cost to set any upward ancillary service capacity price to a non-zero value. If the energy price is zero, then it will be very likely that the upward ancillary service capacity price is also zero in the current marginal cost pricing paradigm. Presumably this means that VER are curtailed as the marginal resource, and that some VER across the system can end their curtailment to provide upward flexibility if VER in other parts of the system ramp down (or load increases or a contingency occurs) to cause an upward flexibility need. However, on a system with a high penetration of VER, there may be more storage and demand response that are providing flexibility and contributing to ramping needs. The marginal cost of these resources is complex because lost opportunity costs can be incurred due to foregone revenues in future market intervals. There may more intervals with energy and reserve prices set at penalty prices dispersed with the intervals that are set at zero, leaving the average prices at values that may not be that different from today or even higher. Lastly, it may be possible that fuel-based thermal resources are remaining on the system and dispatched at their minimum generation limits but are needed and providing reserve and the possibility of upward flexibility while VER are technically the marginal resource and setting the energy price at zero. The marginal cost of the reserve in these cases based on current practice may still be set at zero; however, there is a substantial cost incurred by keeping these resources at the minimum generation limit and if directed by the ISO/RTO to remain online, they must be compensated to recover at least those costs. This, along with principles that follow some of the pricing design relating to fast-start pricing in the energy market, may lead to setting the ancillary service prices at a non-zero value that can provide the needed incentives for these types of resources to remain in the market and provide these needed services. In conclusion, while theory may suggest lower prices for ancillary services along with those for energy under a zero-fuel cost dominated supply fleet, it is more nuanced and may require more study from ISOs/RTOs and researchers to better prepare for outcomes.

- E. Ela et al., “Electricity Market of the Future: Potential North American Designs Without Fuel Costs”, *IEEE Power and Energy Magazine*, vol. 19, no. 1, January/February 2021.

Question 1.3.3. CAISO is considering a Day-Ahead Energy Market Enhancement proposal that seeks to ensure that the day-ahead market clears sufficient resources to address expected net load variability and uncertainty that arises between day-ahead and real-time. What are the expected advantages and disadvantages of revising the day-ahead market construct in this way to procure additional operational flexibility?

16. **Response**. As discussed in the response to Question 1.3.1 of This Response, EPRI research recommends defining reserve based on the scheduling process that the capacity and ramp capability is held, as well as the scheduling process when it is released and potentially deployed. The CAISO Imbalance Reserve Product is one of the first products to focus on day-ahead uncertainty of net load. Forecast errors for load and VER output are generally much greater in the day-ahead timeframe than they are in the 10-minute to hour-ahead time frame, which is the horizon of most other flexibility and ramp products. One important distinction of products covering day-ahead forecast errors compared to shorter-term forecast errors, is that there is no need to require all resources providing that service to be online. Offline resources that can be committed in the real-time scheduling processes can provide this service to reduce costs and deploy only when they are needed. It is also important to note some challenges associated with multi-settlement markets of ancillary services. The reserve that is held to accommodate day-ahead uncertainty is no longer needed once entering the real-time scheduling processes (e.g., an hour ahead of real-time). Thus, there is no real-time imbalance settlement, as all that reserve held in the day-ahead for day-ahead purposes is no longer needed (unless it is also needed for shorter-term uncertainty needs). Market design may need to account for this.

17. Another consideration is that reserve can be used for and provide a multitude of benefits. While the industry is most familiar with holding reserve for reliability reasons, such as providing the capacity to meet an imbalance when it occurs to reduce ACE, frequency error, and potential involuntary load shedding, there may be other reasons as well. One such reason may be to reduce costs. Reserve held in the day-ahead time frame may accomplish this goal. The day-ahead scheduling process includes more decision opportunities than the real-time commitment scheduling process, which includes more than the real-time dispatch, which in turn includes more opportunities than the AGC. When conditions change in real-time, there may be options available to meet the imbalance through quick-start commitments, dispatch of resources on inefficient portions of their heat rate curves, or demand response. Each of these choices tend to have high costs. Holding reserve in the day-ahead time frame may reduce the costs as more inexpensive resources can be committed that can be used to meet the imbalance more cost-effectively than relying on just-in-time options. An EPRI study for the Hawaiian Electric Company demonstrated that improved reserve requirements may improve reliability while at the same time reduce operating costs. This is an important consideration and ISOs/RTOs and their stakeholders may benefit from looking at both cost savings and reliability metrics when evaluating the benefits of new products or strategies.

- See summary PowerPoint of case study at https://www.ferc.gov/sites/default/files/2020-08/W3B-3_Ela.pdf.

Question 1.3.4. The Electric Reliability Council of Texas, Inc. (ERCOT) has proposed to procure fast-responding, limited duration products to address primary frequency control issues associated with declining system inertia. CAISO also intends to initiate a stakeholder process to discuss, among other options, compensating internal resources for the provision of primary frequency response. What are the merits of such reforms and should they be considered in other regions?

18. **Response**. Markets that competitively procure primary frequency response can have merits when it is discovered that the system does not have the ability to meet its needs inherently. This has been demonstrated qualitatively and quantitatively. With NERC BAL-003-2, each balancing area authority, including each ISO/RTO in the U.S., must meet a minimum amount of frequency response, the frequency response obligation (FRO), based on interconnection criteria that is allocated to each balancing area. Primary frequency response (PFR) is measured in MW/0.1Hz rather than in MW like other active power ancillary services. Having sufficiency contingency reserve or regulation in a region may result in the region also having sufficient primary frequency response, but it is not guaranteed. While FERC Order 842 required newly interconnected resources to have the *capability* to provide PFR, it did not require resources to provide it. Different resources will provide it at different times based on system conditions and economics, just like other ancillary services, and commitment status and capacity head room both will impact the amount that an individual resource can provide. Droop settings (the amount of response by frequency deviation in per unit of the maximum capacity of a resource) and governor dead bands (the level of frequency deviation below which would lead to no response) may also influence the quantities. Not all resources that have capacity headroom will provide PFR, particularly those that were interconnected prior to FERC Order 842. Markets that incentivize response to meet the FRO can ensure response is sufficient while potentially incentivizing new capabilities to provide the response efficiently. It is important to note that in regions where PFR is sufficiently above the FRO due to the resource mix, existing ancillary service markets, and/or reliability standards, a market may not be necessary for the service.

E. Ela et al., "Alternative Approaches for Incentivizing the Frequency Responsive Reserve Ancillary Service," *The Electricity Journal,* 25(4), May 2012. Available: https://www.nrel.gov/docs/fy12osti/54393.pdf.
E. Ela et al., "Market Designs for the primary frequency response ancillary service—Part I: Motivation and design," *IEEE Trans. Power Syst*. Vol. 29, no. 1, January 2014.
E. Ela et al., "Market Designs for the primary frequency response ancillary service—Part II: Case Studies," *IEEE Trans. Power Syst*. Vol. 29, no. 1, January 2014.

Question 1.3.5. What other new products not yet discussed at this conference, do you think could increase operational flexibility in RTOs/ISOs? Can capacity markets or other, potentially new, "intermediate" forward market constructs that clear between existing capacity market auctions and the day-ahead timeframe help ensure that RTO/ISO operators have sufficient operational flexibility in real time? For example, can a new shorter-term forward market to procure expected operational flexibility needs held closer to the delivery period (e.g., three months ahead as opposed to three years ahead) and with a more granular delivery period than the annual capacity market (e.g., monthly or seasonal delivery period, or a delivery period based on the hours of an RTO/ISO's morning or evening ramp as opposed to the annual delivery period of most RTO/ISO capacity markets) help ensure that RTO/ISO operators have sufficient operational flexibility in real time?

19. **Response**. Designing and implementing new products can be costly and time consuming. What works in each region may be different. Product horizons (e.g., month-ahead, day-ahead, seasonal, etc.) should align generally with when the decisions have to be made if applicable. On a future decarbonized supply system, it is possible that more long-duration storage resources may be beneficial to meet both reliability and environmental goals. Design of existing products can be evaluated to ensure that these resources will have the incentives to store energy when it is ample and low cost, and supply when it is needed to meet reliability and lower costs during these critical time periods. More analysis may be beneficial to understand what new products, if any, may provide economic efficient and reliable solutions.

Question 1.4.1. To date, most RTOs/ISOs have pursued new ramping products or ORDC reforms, but not both. What are the tradeoffs to consider when deciding between these two approaches and how do they interact? Should these two types of reforms be considered substitutes or complements? Does the opportunity-cost-based method of establishing reserve and ramping product prices send appropriate long-term signals to resources to invest in or maintain flexible capabilities?

20. **Response**. EPRI has been conducting research and development on ORDC and flexibility products for several years and currently have ongoing projects evaluating these exact questions. In EPRI's current understanding, extended downward sloping ORDCs for existing contingency reserve products and separate flexibility products, in general with similar parameters, may accomplish the same objectives. They can be designed almost identically to one another. However, the parameters and design of each make them quite different in practice. For example, (1) an ORDC for 10-minute synchronous contingency reserve that extends to 500 MW beyond the minimum contingency reserve with 10 steps at 50 $/MW-h increments from $500/MW-h to $0/MW-h and (2) a 10-minute online flexibility product that has a requirement of 500 MW but stepped penalty prices from $500/MW-h to $0/MW-h at 50 MW increments that, importantly, cascades prices with the separate 10-minute contingency reserve, would largely result in a similar set of decisions and prices. In practice, the parameters for each design differ and therefore create different implementations of each solution with different results. The table below shows a few examples of different parameters for each mechanism. EPRI is working on a large effort that includes the ISOs/RTOs and national laboratories, funded by the U.S. Department of Energy, to explore the similarities and differences of these two strategies in more detail, including detailed market clearing simulation. EPRI aims to provide more of those insights to the ISOs/RTOs, their stakeholders, and FERC when a more complete analysis is ready to be shared.

Question 1.4.3. What other market design issues and tradeoffs should RTOs/ISOs, stakeholders, and regulators consider when designing and implementing reforms to energy and ancillary services markets to increase operational flexibility?

21. **Response**. One other important factor that has been explored in the EPRI research is that there are multiple ways of ensuring flexibility including both implicit and explicit means. Explicit means include reserve requirements, where the amount of reserve or flexibility that is needed is determined exogenously, and a reserve constraint is added to the scheduling or market clearing software. Implicit means include various features to the scheduling or market clearing tools that will inherently schedule the flexibility directly within the software. For example, see the table below. For the three primary needs of operating reserve: (1) variability occurring within the interval (intra-interval variability) can be met implicitly through shorter scheduling intervals (e.g., a 15-minute day-ahead commitment rather than an hourly day-ahead commitment); (2) variability occurring beyond the interval (inter-interval variability) can be met implicitly through time-coupled multi-period dispatch with extended look-ahead horizons (e.g., real-time dispatch with hourly look-ahead); and (3) uncertainty can be met implicitly through allowing the scheduling or market clearing software to meet more than just the expected scenarios (e.g., multi-scenario commitment, stochastic unit commitment).

22. The implicit methods have some advantages and disadvantages. Generally, due to their nature and how they are used with more direct inclusion of costs in the objective function, they can allow for more efficient solutions than the solutions that may come from exogenous reserve requirements. They may also incorporate deliverability constraints more easily, though this may be dependent on the solution technique. However, the implicit scheduling options may require more data and more solve time for the market clearing solutions. Finally, and perhaps most importantly, they may not have the same incentives that reserve constraints and reserve markets may have as part of their established features. A concrete example is provided in the response to Question 2.3.2. When flexibility needs are met through implicit scheduling mechanisms, there typically is not a straightforward "clearing price" for that flexibility that is a result of the market clearing solution. It is possible that an efficient incentive signal can be formed through the solution, but this requires thought and analysis. Given that these implicit means are, from a cost-minimization perspective, theoretically superior to reserve requirements, it can be beneficial to explore these methods so that they may work in a practical fashion, where data needs, computational limits, and incentive signal barriers can be addressed.

E. Ela, "Comparing performance of explicitly and implicitly scheduled operating reserve." Presented at the FERC Software Conference. June 28, 2016. Available: https://cms.ferc.gov/sites/default/files/2020-05/20160629114654-1%2520-%2520FERC%2520Reserve%2520explicit%2520implicit%25202016.pdf.

Question 2.1.4. Do current RTO/ISO offer rules, market power mitigation practices, and reference levels prevent or discourage resources from including in their offers the additional costs, if any, that resources incur from being more flexible (e.g., longer-term wear and tear on natural gas resources due to increased cycling, battery warranty considerations, etc.)? Are such costs difficult to quantify? If so, please explain why. How should RTOs/ISOs review such costs to ensure that resources' energy and ancillary services supply offers are competitive?

23. **Response**. For thermal powerplants, especially older plants, these additional costs are difficult to quantify and often require significant expense to quantify adequately. It has been demonstrated that thermal plants that operate flexibly have lower reliability. Plants often use typical costs extracted from a similar population of energy resources, and attempt to offer flexibility based on iterative testing and risk assessment of the units.

24. The most significant costs from flexible operation come in discrete events. For example, a last stage blade liberation from a steam turbine may occur after prolonged operation at low load. Should this occur, there is a very large cost to the flexible operation. If the plant were able to identify and prevent this event through a maintenance intervention, the cost is less in comparison. Rules that limit opportunities for maintenance interventions may limit the ability of a plant to cost effectively offer flexibility.

25. Flexibility costs are difficult to quantify. As an example, in thermal powerplants, the high temperature pressure parts all have finite lifetimes generally designed based on time at temperature (creep). When operating flexibly, the life of these components may become limited by fatigue. With miles of tubing and piping, and thousands of welds in each thermal powerplant boiler or heat recovery steam generator and associated piping, it is not simple to quantify the life consumption associated with flexible operation. Power plants do not have extensive enough metal temperature instrumentation to quantify this consumption. In addition, the instrumentation required to monitor high temperature systems has limited life and thus is a significant cost to maintain. When high temperature pressure components begin to fail there are reliability issues and repairs, and depending on the location, costs and outage time can vary widely. When larger components, e.g., a superheat outlet header, reaches end of life, there are long duration outages required to replace. Due to these and other complexities with flexible operation, it is exceedingly difficult to quantify costs on a specific unit. Rather, average data across a fleet of similar units is typically used to approximate the cost. Reliability of flexibly operated thermal plants is consistently less than those that operate at constant outputs with limited starts/stops, with a significant reduction in reliability as the plant ages while operating flexibly. The cost of reduced reliability due to flexible operation is also difficult to quantify.

26. There is also not a clear input to include some of these costs in market offers. These can possibly be included in start-up costs where applicable, but in other cases, these costs must be assigned to incremental energy, which is not necessarily the unit of measurement to which the cost applies. Often, these costs will need to be strategized into energy or ancillary service cost offers and adjusted based on increased flexible operation over time.

- N. Kumar et al., "Power Plant Cycling Costs." NREL/SR-5500-55422. July 2012. Available: https://www.nrel.gov/docs/fy12osti/55433.pdf.

Question 2.2.2. To what extent do existing RTO/ISO energy and ancillary services market rules require standalone variable energy resources to respond to dispatch instructions (e.g., curtailment)? To what extent are standalone variable energy resources technically capable of being "dispatchable?" Is there a distinction between being dispatched down and being curtailed? Under what circumstances can a standalone variable energy resource be dispatched up versus down?

27. Response. Most RTO/ISO energy and ancillary services market rules require that standalone VER be able to respond to dispatch instructions to ensure economic and reliable system operations. The ISOs currently have utility-scale wind follow the real-time dispatch when they need to dispatch down, and many of the ISOs also do this for utility-scale solar. This is typically not yet done in a similar manner for wind or solar on the distribution system. What makes standalone VER unique is that they have a *dynamic* (i.e., changes with time) upper operating (MW) limit that cannot be predicted in advance with perfect accuracy. Standalone VER are technically capable of being "dispatchable" and providing several ancillary services. In fact, the majority of the modern-day VER are flexible and can be operated at any dispatch point between zero MW and its dynamic upper operating limit with fast response rates and at near-zero operating costs. With these parameters in mind, standalone VER connected to the transmission system are indeed dispatchable and can be part of an operator's set of decisions to balance load, manage transmission congestion and prepare for system ramps and flexibility needs.

28. Technically, there is no distinction between being dispatched down and being curtailed. The two are considered synonyms by most, but the latter has typically carried with it a negative connation. The overarching goal should always be to ensure that a resource is operated in a manner that is most cost-effective from the system and individual perspective and to devise ways to incentivize it accordingly. There may be instances in which a standalone VER may need to be curtailed to promote reliable operations or dispatched down to accommodate changing system conditions.

29. VER may be curtailed or dispatched down due to transmission congestion, low load, and minimum generation constraints, or to promote sufficient available rampable capacity. For instance, due to limited options to manage the transmission congestion, sometimes VER may need to be curtailed to avoid transmission system overloads. During such instances, the VER must be dispatched down or else be penalized. Once a VER is dispatched down, it can then be dispatched up to its maximum upper limit by the ISO/RTO once the congestion has been resolved or whenever its zero-cost energy can again be accommodated. Importantly, if a VER is already at its dynamic upper operating limit (i.e., it is providing maximum available energy from the wind or sun), then it cannot be dispatched up in a meaningful and sustained manner.

E. Ela et al., Active power controls from wind power: Bridging the gaps, Tech. Rep. NREL/TP-5D00-60574, National Renewable Energy Laboratory, 2014. Available: https://www.energy.gov/sites/prod/files/2014/01/f6/Active%20Power%20Controls%20from%20Wind%20Power.pdf.
C. Loutan et al., *Demonstration of Essential Reliability Services by a 300-MW Solar Photovoltaic Power Plant,* NREL/TP-5D00-67799, March 2017. Available: https://www.nrel.gov/docs/fy17osti/67799.pdf.

Question 2.2.2. To what extent do existing RTO/ISO energy and ancillary services market rules require standalone variable energy resources to respond to dispatch instructions (e.g., curtailment)? To what extent are standalone variable energy resources technically capable of being "dispatchable?" Is there a distinction between being dispatched down and being curtailed? Under what circumstances can a standalone variable energy resource be dispatched up versus down?

30. **Response** (continued). The technical capability for VER to provide flexibility has been demonstrated through several research studies and pilot projects. These studies have shown the capability to provide dispatch, regulation service, primary frequency response, and voltage control among other services. Additionally, some ISOs also allow wind to provide regulation down service in regions where the regulation service is separated in upward and downward direction. The technical capability of wind and solar to provide the majority of ancillary services has been demonstrated by a number of studies. However, it is still rare in the U.S. for wind and solar to be participating in ancillary services markets. This is primarily due to (1) limited system operator confidence because of the uncertainty of the VER's availability to provide the ancillary service capacity and (2) because of economics and limited revenue potential due to foregone energy sales and foregone production credits.

31. It is possible that VER may be dispatched down from their upper operating limit to have the capacity to then provide an upward reserve service. Given that these facilities often earn additional revenues from production-based mechanisms for their environmental benefits (e.g., production tax credits and renewable energy credits) in parallel to wholesale electricity market revenues, typically, there are limited economic incentives for these resources to provide upward reserve service since that requires a resource to be dispatched down and forego both energy sales and these additional revenues. Thus, it is the economics that impacts the ability of a VER to provide upward response due to the limited instances where it is economic for both the system and VER. However, it is possible and may be more common as VERs grow their share of supply. For example, dispatching down VERs to provide ancillary services may reduce the need to commit an additional resource to do so. This results in accompanying system benefits in the form of reduced commitment costs that are potentially not always reflected in prices.

E. Ela et al., Active power controls from wind power: Bridging the gaps, Tech. Rep. NREL/TP-5D00-60574, National Renewable Energy Laboratory, 2014. Available: https://www.energy.gov/sites/prod/files/2014/01/f6/Active%20Power%20Controls%20from%20Wind%20Power.pdf.
C. Loutan et al., *Demonstration of Essential Reliability Services by a 300-MW Solar Photovoltaic Power Plant,* NREL/TP-5D00-67799, March 2017. Available: https://www.nrel.gov/docs/fy17osti/67799.pdf.

Question 2.2.3. To what extent do resource capabilities vary amongst different classes and vintages of variable energy resources (e.g., newer vs. older wind turbine models, onshore vs. offshore wind, fixed-tilt vs. tracking solar, etc.) and do offer rules currently reflect such differences, if any?

32. **Response**. Type 3 and type 4 wind turbines and modern photovoltaic facilities have controls that effectively allow them to be dispatched to any point between zero and their available power limit. Typically, almost all VER are allowed to bid any offer in the day-ahead market (DAM) with any upper operating limit and the ISO does not validate the offer with a forecast in the DAM. Instead, in the day-ahead reliability unit commitment (RUC) process that is often conducted following the DAM, the ISO seeks to make a commitment plan to satisfy the anticipated real-time conditions. It is during the day-ahead RUC stage that the ISO typically replaces the VER offers with its forecasts analogous to what it does with load to better understand the commitment needs of the following day. In real-time markets, VER submit offer costs that reflect their willingness to operate. ISOs also have real-time forecasts for VER that may be used directly in the market clearing software. Most ISOs directly use the real-time forecasts to establish the upper operating limits for such resources, but some ISOs may allow VER to provide their own forecast. When there is congestion or for any other economic and/or reliability reasons the system needs to curtail VER, a dispatch and curtailment will be sent and VER will need to follow this set point. Older VER technology, such as older wind turbines, may be exempt from these real-time curtailment rules.

Question 2.2.4. To what extent are emerging resource types, such as hybrids, storage resources, and distributed energy resource aggregations technically capable of providing existing ancillary service products or other reliability services? Acknowledging that some market rules are evolving due to Order Nos. 841 and 2222, do current RTO/ISO market rules for ancillary services and other reliability services, such as eligibility requirements, align with these emerging resource types' capabilities?

33. **Response**. Most emerging technologies that interface with the grid through an inverter show great performance with respect to fast response capability due to power electronics. These technologies may have substantial flexibility and fast response times while adjusting active power output. However, there exists a few technical limitations such as energy (i.e., state of charge) limits for storage, call limits for demand response, and telemetry and distribution system limits for distributed energy resources that warrant additional investigation by system operators. Market simulation studies may need to be conducted to better assess how to address these challenges and to evaluate the potential benefit of ancillary service provision from all these emerging technologies.

34. Blackstart service may be one of the few ancillary service products that have eligibility criteria such as a continuous energy duration requirement (or sustainment period) of about 16 hours at rated output that are challenging for some of these emerging technologies to satisfy. Emerging technologies may not be technically capable of providing blackstart service because it requires a sustained response for a significantly long duration. This may change in the future with improvements in technologies or evolution of ancillary service characteristics and requirements with changing resource mix. Alternatively, long duration energy storage may potentially play a role to provide blackstart service in the future as the conventional generation fleet retires.

Question 2.2.4. To what extent are emerging resource types, such as hybrids, storage resources, and distributed energy resource aggregations technically capable of providing existing ancillary service products or other reliability services? Acknowledging that some market rules are evolving due to Order Nos. 841 and 2222, do current RTO/ISO market rules for ancillary services and other reliability services, such as eligibility requirements, align with these emerging resource types' capabilities?

35. **Response (Continued).** Contemporary RTO/ISO market rules for ancillary services and other reliability services, such as eligibility requirements, generally may align well with the capabilities of these emerging technologies. For instance, due to FERC Order No. 841, most ISOs now allow standalone storage resources to provide ancillary services, provided storage resources are able to satisfy ancillary services duration requirements. In order to align with these emerging resource types' capabilities, ISOs are allowing for capacity derates to meet the duration requirements. The overall goal would be to look at performance in a technology agnostic manner and enable a way for a technology to prove its performance and provide the service and attributes that the system operator desires. If there are any restrictions on participating in a service, it is important to ensure that the limitations are based on technology attributes and performance, and to regularly allow such new technologies to prove their performance or ability to contribute in new ways.

36. Some ancillary service types have specific response rate requirements and duration requirements that may benefit from additional reevaluation of the need. Some of the existing RTO/ISO market rules for ancillary services and other reliability services were developed while considering the capabilities of traditional technologies. Further analysis of the present-day capabilities of emerging technologies that may eventually provide a majority of these services in the future, including advanced simulation to quantify contributions, may be beneficial to understand both reliability implications and economic impacts. Evaluating different future scenarios and different market designs may provide insights into these questions. The ancillary services are not all necessarily distinct, in that having a fast response with small duration may be very useful to have from the system perspective if it were to be combined with slow response with long duration. It might be beneficial to assess such modifications and the consequent impact from a system and asset owner perspective through case studies.

Question 2.2.5. What RTO/ISO energy and ancillary services market reforms could be adopted, if any, to ensure that new and emerging resource types are able to offer their full operational capabilities into RTO/ISO energy and ancillary services markets to help operators manage changing system needs? Would shortening the day-ahead market interval length increase the operational flexibility available from resources? What considerations (e.g., computing time) are important to consider when establishing the length of energy and ancillary services market intervals? RTOs/ISOs often require resources that provide ancillary services to be capable of doing so for a duration of 60 minutes. Does this eligibility requirement limit the pool of resources available to offer ancillary services into RTO/ISO markets? Would reexamining the need for this particular eligibility requirement present reliability concerns or raise other issues for operators? If so, please explain.

37. **Response**. Market designs that utilize participation models in both the energy and ancillary service markets that enable the full operational flexibility for emerging technologies and incentivize them to offer that range may be beneficial. These should also give the system operator an indication of how much flexibility is feasible for emerging resources to provide, so that the operator has the confidence to rely on these resources going forward. In other words, participation models that both allow full flexibility to be provided but also alert system operators and scheduling tools when flexibility is constrained due to physical limitations may be beneficial to pursue.

38. Shortening the day-ahead market interval length may provide potential benefits by more efficiently allocating resources and their costs to providing flexibility needed for expected conditions (i.e., variability but not uncertainty). This would have a similar, yet potentially more efficient, result compared to including a reserve requirement to accommodate the 15-minute intra-interval variability within each day-ahead hourly interval (see response to Question 1.4.3 and Paragraphs 24-25 in This Response). However, as is noted, this can have substantial computational time increases that may render the option infeasible. Its value is also highly dependent on how available and accurate the data is on a fifteen-minute interval basis for a day-ahead horizon. If the data is merely a linear interpolation of what is occurring on the hourly basis, and if day-ahead forecasts do not allow for additional information at 15-minute granularity, then it may have a diminishing return.

Question 2.2.5. What RTO/ISO energy and ancillary services market reforms could be adopted, if any, to ensure that new and emerging resource types are able to offer their full operational capabilities into RTO/ISO energy and ancillary services markets to help operators manage changing system needs? Would shortening the day-ahead market interval length increase the operational flexibility available from resources? What considerations (e.g., computing time) are important to consider when establishing the length of energy and ancillary services market intervals? RTOs/ISOs often require resources that provide ancillary services to be capable of doing so for a duration of 60 minutes. Does this eligibility requirement limit the pool of resources available to offer ancillary services into RTO/ISO markets? Would reexamining the need for this particular eligibility requirement present reliability concerns or raise other issues for operators? If so, please explain.

39. **Response**. (Continued). Presently, most of the RTOs/ISOs impose a continuous energy duration of about 30 to 60 minutes for contingency reserves post-activation. Based on the region or jurisdiction under consideration, reexamining the need for this particular eligibility requirement may present reliability concerns since the 60-minute duration requirement is based on relevant reliability standards. These values are typically based on the reliability standards of NERC or regional reliability entity, which explicitly or implicitly determine the duration requirements for ancillary service products. The NERC BAL-002 reliability standard for contingency disturbance conditions (also referred to as the Disturbance Control Standard (DCS)), requires the ACE to be returned to the lower of its pre-disturbance value or zero within 15 minutes. The DCS then imposes a requirement to restore its contingency reserve 105 minutes following the contingency event. The latter DCS requirement implies that once all resources have responded to the contingency and corrected the ACE, they must sustain output for 90 minutes longer before it is required to have other resources response and replace that reserve for a subsequent event. While this standard would demonstrate the need for at least a 60-minute duration for contingency reserve, it is unclear from where those values originated. In addition, if there are sufficient resources that can replace those reserve services more quickly, it may support having contingency reserve that need only sustain for shorter durations. While it is helpful to allow for changes to these duration requirements to allow for greater eligibility, any changes need to be very carefully considered, studied, and facilitated through the appropriate channels (including through NERC standards drafting teams).

40. The duration requirement for regulation used to be one hour initially based on the resolution of the day-ahead market, but is now modified to reflect the fact that ACE crosses zero in much shorter time periods and the hourly duration was not as necessary. This implies that the same direction of response should not be needed for more than that initial duration, resulting in majority of the ISOs modifying the duration requirement for regulation to now be 15 minutes. The modified duration requirement has further enabled emerging technologies such as storage resources with limited energy storage ability to provide regulation reserve more effectively.

Question 2.3.1. What are the challenges to incorporating uncertainty within the current RTO/ISO market software? For example, how can improvements in forecasting, the use of intra-day commitment processes that include a range of forecasts, or longer lookahead commitment and dispatch horizons result in more efficient unit commitment and dispatch in real time?

41. **Response**. RTO/ISO day-ahead forecasts for load, wind and solar are currently only used in the reliability unit commitment (RUC) process, also called the Reliability Assessment Commitment (RAC), forecast pass, and reserve adequacy analysis across the different RTOs/ISOs. These processes primarily are run subsequent to the day-ahead market, with a primary focus of starting and committing sufficient resources that require day-ahead notification time while minimizing residual unit commitment costs, such as start-up costs and no-load or minimum generation costs. Resources committed in the day-ahead market are not decommitted if, for example, the renewable forecasts used in the RUC are higher value than the renewable offer quantities used in the day-ahead market. In addition, it is often the case that the energy costs are ignored or largely discounted in the RUC (i.e., only the commitment costs are of concern). While these processes may be in place for valid reasons, this implies that improved day-ahead forecasts of load and renewables have limited impact on economic efficiency of the overall day-ahead commitment solution. This is a key challenge, in that the use of RTO/ISO forecasts are currently not explicitly used in market software, and so forecast improvements or how uncertainty is incorporated in market software is severely inhibited in the day-ahead time frame and there are limited economic efficiency benefits that can be realized. However, it is important to consider that improved forecasts may improve efficiency when used directly by loads, renewable resource assets, and financial participants such as virtual traders, when used in their offer strategy. Forecasts can also be used by ISOs in day-ahead reserve requirements that may improve efficiency as well.

42. In the real-time dispatch and most other intra-day processes such as intra-day reliability unit commitment and real-time unit commitment, these forecasts that are available to the RTO/ISO are used more directly in the RTO/ISO processes. This means that improved forecasts, and other enhancements of forecast application such as longer look-ahead horizons or multi-scenario (e.g., stochastic) scheduling processes, forecast utilization can have a more direct impact on both reliability and economic efficiency.

Question 2.3.2. Can changes to RTO/ISO unit commitment and dispatch software address the need to posture system resources optimally to meet expected and unexpected ramp and operational flexibility needs? How are these enhancements tailored to the expected magnitude of forecast errors in different time periods? How would multi-period dispatch modeling in the real-time market help address operational flexibility needs? What are the advantages and disadvantages of a binding as opposed to an advisory multi-period dispatch model? What are the computational burdens associated with such modeling enhancements?

43. **Response**. EPRI research has explored this topic and has adopted the use of the term "time-coupled multi-period economic dispatch" or "time-coupled multi-period unit commitment" to refer to these applications. The intervals of these tools are coupled through time-coupling constraints, the most common of which today are ramp constraints of ramp-constrained generation. This means that not only is the ISO's model looking into the future, but it is also making decisions now (or for the first interval) that are influenced by future system conditions. These models allow the RTO/ISO to determine real-time energy dispatch and ancillary service schedules for the upcoming market interval that may also better pre-position the system for an upcoming interval in the future, if that interval is contained within the horizon of the look-ahead. For example, a unit may be backed down in a way that appears out of merit in the first interval, so that the system has sufficient ramping capability in the next interval. This provides a benefit in promoting flexibility to meet the expected ramping needs of the future, so long as the need is within the horizon of the look-ahead horizon for the multi-period model. To EPRI's knowledge, while the ISOs almost all use some form of multi-period unit commitment in the real-time stages, only the NYISO and CAISO use a multi-period economic dispatch in the clearing model that determines real-time market dispatch schedules and real-time locational marginal prices (LMPs).

44. There are two important characteristics of the use of these models in real-time clearing. First, in general, these models in their current form are better at preparing the system for expected variability in future intervals than the current form of ramp or flexibility products that may be in place for similar reasons. This is because the objective function in the multi-period market model is minimizing the cost of holding capacity to meet the ramp needs *in addition to* the cost of deploying and converting energy to do so (see Question 1.4.3 and Paragraphs 23-24 of This Response). The ramp products in their current form on the other hand only include in their objective function the cost of holding the capacity *and not* the cost of deploying the capacity to meet the ramp need. Second, the multi-period models in their current form do not prepare the system to meet the potential ramp needs that are not forecasted. In other words, they only help to meet expected variability and do not currently help to meet uncertainty needs. Alternatively, ramp products can be used for both variability and uncertainty.

Question 2.3.2. Can changes to RTO/ISO unit commitment and dispatch software address the need to posture system resources optimally to meet expected and unexpected ramp and operational flexibility needs? How are these enhancements tailored to the expected magnitude of forecast errors in different time periods? How would multi-period dispatch modeling in the real-time market help address operational flexibility needs? What are the advantages and disadvantages of a binding as opposed to an advisory multi-period dispatch model? What are the computational burdens associated with such modeling enhancements?

45. **Response** (Continued). EPRI has conducted research on this topic. Time-coupled multi-period market clearing models used to clear the real-time market generally perform well in providing schedules that can meet expected ramp needs to enhance both reliability and economic efficiency. However, in the current form, there may be a price formation issue. In the two ISOs that use multi-period market clearing models to clear their real-time markets, it is EPRI's understanding that both may only use the prices and schedules of the first interval (also referred to as the binding interval) for settlement. This may present an issue because during a ramping period where a resource is backed down in the binding interval in order to meet the conditions of a future (advisory) interval, the prices that result are aligned with the decisions of both intervals. The price of the first interval is depressed which gives the resource the incentive to back down to provide more ramp capacity. The price of the second interval is elevated which incentivizes that resource to provide the energy later during that second interval when it is needed. On average, these prices will cover the cost of the ramping resource. However, when that second interval now becomes the first interval, the information of the preceding interval is effectively lost, and it is possible that the high price of that interval will no longer result. This is especially true if the ramp does not materialize. While some may think this is appropriate (i.e., why should the price stay high if the ramp does not materialize), it is important to realize that the decision to back down, and effectively provide reserve to meet that ramp, has already been made. Now, the resource will lose profit in that interval and may not have the incentive to follow the directions the next time. There are a few solutions that may improve upon these circumstances, including settling on all intervals of the look-ahead horizon in a multi-settlement arrangement, or pricing the marginal ramp constraint as an effective reserve price that is paid to the resources that are preparing for that anticipated ramp. These options are complex but important to evaluate since, when combined with multi-period economic dispatch, these solutions may provide an efficient way of promoting ramp capability and may be superior to introducing separate ancillary service products.

46. Regarding computational efficiency, multi-period models are more challenging and take longer to solve than single-period models, especially given the small time window of when the real-time market must solve and post. The more intervals, the more variables and time-coupling constraints that couple the intervals, and the longer it can take to solve. However, the advisory intervals do not have to be the same interval length as the five-minute real-time market, and ISOs and vendors can potentially evaluate the most efficient set of intervals to meet their needs and potentially avoid or reduce computational complications.

Question 2.3.2. Can changes to RTO/ISO unit commitment and dispatch software address the need to posture system resources optimally to meet expected and unexpected ramp and operational flexibility needs? How are these enhancements tailored to the expected magnitude of forecast errors in different time periods? How would multi-period dispatch modeling in the real-time market help address operational flexibility needs? What are the advantages and disadvantages of a binding as opposed to an advisory multi-period dispatch model? What are the computational burdens associated with such modeling enhancements?

47. **Response** (Continued). Finally, as mentioned earlier, ramp constraints are the primary time-coupling constraint of today. However, with more electric storage resources participating in the market and the use of state-of-charge constraints in the real-time market, this will add a time-coupling constraint that is important to include for improving economic efficiency and reliability but can further impact computational time. These constraints also need to be evaluated for pricing implications discussed earlier in Paragraph 47. RTOs/ISOs may benefit from testing and evaluation to see how solutions and computation are affected generally and with added storage and state-of-charge constraints for their particular system and existing set of market design and operational features.

Time-coupled multi-period unit commitment models have many more time-coupling constraints having to do with unit commitment, but the question posed here seems more focused on dispatch models and is therefore the focus of EPRI's response.

E. Ela and M. O'Malley, "Scheduling and pricing for expected ramp capability in real-time power markets," *IEEE Trans. Power Syst.*, vol. 31, no. 3 2016. Available: https://ieeexplore.ieee.org/document/7192736.

See also: https://cms.ferc.gov/sites/default/files/2020-05/20160629114654-1%2520-%2520FERC%2520ramp%2520pricing%2520Software%2520Conference%25202016.pdf.

D. A. Schiro, "Flexibility procurement and reimbursement: A multi-period pricing approach", Proc. FERC Tech. Conf., 2017, [online] Available: https://www.ferc.gov/CalendarFiles/20170623123635-Schiro_FERC2017_Final.pdf.

E. Ela and M. O'Malley, "Scheduling and pricing for expected ramp capability in real-time power markets," *IEEE Trans. Power Syst.*, vol. 31, no. 3, 2016. Available: https://ieeexplore.ieee.org/document/7192736.

Question 2.3.3. To what extent can software enhancements for modeling specific technology types (e.g., multi-configuration modeling of combined cycle units, storage, etc.) help address the system's changing operational needs?

48. **Response**. In theory, adding granularity to the characteristics of different technologies and their participation models has economic and reliability benefits. This was discovered initially by the U.S. industry when U.S. ISOs/RTOs first introduced unit commitment and three-part bidding. The optimal operation of a facility may depend on the costs across multiple intervals and reliability needs. If the ISO has the parameters including costs of the operation of the facility, it has the tools to determine the optimal operation. However, it is important to recognize that there are some characteristics of the operation of a facility that simply cannot be represented adequately in the ISO market clearing tools. These include non-linear characteristics and other internal characteristics that may simply impact the computation time of the ISO market tools too significantly to take advantage of, or is information that the asset owner considers private. The market participants may be able to internally factor these into their offer strategy efficiently to ensure optimal and feasible schedules for their resource. It is also important to recognize that ISO models minimize production costs and are not set to maximize profits for individual resources. These objectives usually converge but may not, especially due to non-convex operation, e.g., unit commitment.

49. A challenge associated with granular participation models is that ISOs may not have the time and money to develop them for each technology, and the software may not be able to handle complex granular models for each technology with the prevailing timelines to solve and post the solutions from such market clearing models. How to prioritize which models to develop and implement can be complex. Cost benefit analysis may be helpful to determine the value of these enhanced participation models. Some ISOs/RTOs have started to conduct cost benefit analysis studies of new enhanced participation models to determine practical paths forward. Advanced participation models also may counter the technology agnostic objective that RTOs/ISOs maintain. Having advanced granular models for one technology and not for another may provide a competitive advantage. Enabling a technology to use another technology's model may have unintended consequence and opportunity for gaming and complex market power screening methods. These possibilities can be factored into these cost benefit assessments when new granular participation models are being developed.

Question 2.3.4. Can multi-day-ahead markets or hour-ahead markets help address operational flexibility needs in RTOs/ISOs? What is the objective of such approaches, and are there potential drawbacks?

50. **Response**. While quite a different strategy with different reasons for its proposed use, multi-day-ahead markets can have a similar benefit to 15-minute interval day-ahead markets. They may add information for the market clearing tool to make more informed decisions. All of these potential benefits need to be weighed against each other, as it can be unwieldy to consider all of these changes to these tools at once. A useful conceptual exercise can be experimented with where one can think of what a market clearing tool might look like in the case of infinite computing power and infinite data availability. One could have commitment and dispatch tools that are updated every second, looking at second-by-second intervals into the next several days or weeks as a look-ahead horizon, and may have millions of scenarios that are evaluated in the model based on probabilistic forecasts. One might then walk back from that intense model and take features away one by one that provide little value, require unattainable data, or have too large an impact on computation time, until reaching the best tool for the particular RTO/ISO or region, given existing software and hardware capabilities. Multi-day-ahead markets are being investigated due to multi-day commitment needs (long-start steam turbine units or long-duration pumped storage hydropower resources) and for fuel inventory reasons. Hour-ahead advisory solutions already exist for most regions. If decisions do not require a day-ahead notification, which importantly may include generator staff constraints, there may be benefit to making those decisions later and closer to real-time, i.e., when forecasts are more accurate and the decisions are likely to be more efficient on average. RTOs/ISOs could consider whether these characteristics will be needed in the future on their particular system, given the amount of time to develop the rules and software, and compare to other features (such as shorter interval day-ahead, or multi-interval real-time) to determine the best solution.

TOGETHER…SHAPING THE FUTURE OF ENERGY®

www.epri.com

Export Control Restrictions

Access to and use of this EPRI product is granted with the specific understanding and requirement that responsibility for ensuring full compliance with all applicable U.S. and foreign export laws and regulations is being undertaken by you and your company. This includes an obligation to ensure that any individual receiving access hereunder who is not a U.S. citizen or U.S. permanent resident is permitted access under applicable U.S. and foreign export laws and regulations.

In the event you are uncertain whether you or your company may lawfully obtain access to this EPRI product, you acknowledge that it is your obligation to consult with your company's legal counsel to determine whether this access is lawful. Although EPRI may make available on a case by case basis an informal assessment of the applicable U.S. export classification for specific EPRI products, you and your company acknowledge that this assessment is solely for informational purposes and not for reliance purposes.

Your obligations regarding U.S. export control requirements apply during and after you and your company's engagement with EPRI. To be clear, the obligations continue after your retirement or other departure from your company, and include any knowledge retained after gaining access to EPRI products.

You and your company understand and acknowledge your obligations to make a prompt report to EPRI and the appropriate authorities regarding any access to or use of this EPRI product hereunder that may be in violation of applicable U.S. or foreign export laws or regulations.

3002031421

About EPRI

Founded in 1972, EPRI is the world's preeminent independent, non-profit energy research and development organization, with offices around the world. EPRI's trusted experts collaborate with more than 450 companies in 45 countries, driving innovation to ensure the public has clean, safe, reliable, affordable, and equitable access to electricity across the globe. Together, we are shaping the future of energy.

EPRI
3420 Hillview Avenue, Palo Alto, California 94304-1338 ▪ USA
800.313.3774 ▪ 650.855.2121 ▪ askepri@epri.com ▪ www.epri.com