




Alternative Approaches for Incentivizing the Frequency Responsive Reserve Ancillary Service

E. Ela, M. Milligan, and B. Kirby
National Renewable Energy Laboratory

A. Tuohy and D. Brooks
Electric Power Research Institute

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

Technical Report
NREL/TP-5500-54393
March 2012

Contract No. DE-AC36-08GO28308

NATIONAL RENEWABLE ENERGY LABORATORY

Alternative Approaches for Incentivizing the Frequency Responsive Reserve Ancillary Service

E. Ela, M. Milligan, and B. Kirby
National Renewable Energy Laboratory

A. Tuohy and D. Brooks
Electric Power Research Institute

Prepared under Task No. WE11.0905

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

National Renewable Energy Laboratory
1617 Cole Boulevard
Golden, Colorado 80401
303-275-3000 • www.nrel.gov

Technical Report
NREL/TP-5500-54393
March 2012

Contract No. DE-AC36-08GO28308

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

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

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

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

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

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Alternative Approaches for Incentivizing the Frequency Responsive Reserve Ancillary Service

Erik Ela[1], Aidan Tuohy[2], Michael Milligan[1], Brendan Kirby[1], and Daniel Brooks[2]

[1]National Renewable Energy Laboratory

[2]Electric Power Research Institute

Abstract

Frequency responsive reserve is the autonomous response of generators and demand response to deviations of system frequency, usually as a result of the instantaneous outage of a large supplier. Frequency responsive reserve arrests the frequency decline resulting in the stabilization of system frequency, and avoids the triggering of under-frequency load-shedding or the reaching of unstable frequencies that could ultimately lead to machinery damage or system blackouts. It is a crucial service required to maintain a reliable and secure power system. Regions with restructured electricity markets have historically had a lack of incentives for frequency responsive reserve because generators inherently provided the response and on large interconnected systems, more than sufficient response has been available. This may not be the case in future systems due to new technologies and declining response. This paper discusses the issues that can occur without proper incentives and even disincentives, and proposes alternatives to introduce incentives for resources to provide frequency responsive reserve to ensure an efficient and reliable power system.

This work is part of the NREL and EPRI Active Power Control from Wind Power project
http://www.nrel.gov/wind/systemsintegration/projects_modeling_interconnection.html

I. Introduction

The control of electrical frequency of a synchronous interconnection is one of the primary measures to maintaining a reliable and secure power system. Power system operators must hold frequency as close as possible to its nominal level (60 Hz in the United States, 50 Hz in most other areas of the world) by adjusting the generation and load balance. A constant frequency at its nominal level indicates if the aggregate generation and load are in balance. Most large interconnections will generally try to keep frequency to within 50 mHz of the scheduled frequency. The frequency of the interconnection will start to deviate from its nominal level when there is an imbalance between the supply and demand. Any supply or demand imbalance can cause this deviation. Demand and generation can vary over short time scales causing slight frequency deviations from nominal. However, the most common cause of significant deviations is from the sudden loss of a large supplier (e.g., generator). This loss must be made up by the other resources on the grid to ensure the demand is still met. Frequency responsive reserve is the automatic response to frequency excursions, usually through turbine speed governors and frequency responsive demand response, which adjusts output or consumption to counter frequency deviations. This immediate response is needed to arrest frequency deviations before triggering under-frequency load shedding relays, triggering under- or over-frequency generation protection relays, or reaching unstable frequencies that could ultimately lead to a blackout.

Since the restructuring of the electricity sector, numerous markets have been designed to incentivize competition for providing different services. These complex markets integrate efficient economic principles with the engineering and physics of the power system [1]. In the United States, restructured markets have evolved toward a common set of market principles [2]. In this design, there usually exists a two-settlement system for forward and real-time markets, with co-optimized energy and ancillary services markets, locational marginal pricing for energy, and financial transmission rights markets in place for hedging. Current ancillary services markets include spinning reserve, nonspinning reserve, and regulation [3]-[4]. In the U.S., spinning reserve is reserve that must be online and available within ten minutes. Nonspinning reserve must also be available within ten minutes, but can be offline. Although the spinning reserve must come from resources that are synchronized to the grid, there are usually no enforceable requirements for the market participant to provide energy that is directly responsive to frequency, or for speed governors to be in an operational mode. Regulation is a type of ancillary service that is used to correct short-term imbalances in generation and load that happen during normal conditions. Regulation units usually must be equipped with automatic generation control. Since regulation is used to reduce the area control error at the direction of the system operator, it is not required to respond directly to frequency. The time scale for regulation is also longer than that relating to frequency responsive reserve. Few regions in the world have an explicit ancillary service market to procure frequency responsive reserve. One reason is likely because the conventional generating technologies historically had frequency responsive reserve as an inherent feature of the technology since installation. Another reason is likely because most systems, and especially those that are part of large synchronous, interconnected grids, had more response than was needed and, therefore, did not need to incentivize for more response. Both of these reasons may not hold true in future systems, based on current trends.

Recent studies have shown that the frequency response in the United States, and especially in the Eastern Interconnection, has been declining [5]. Reasons for this include high governor dead bands, generators operating in modes that do not offer frequency responsive reserve (e.g., sliding pressure mode), governors that are not enabled, a reduced percentage of direct drive motor load, and many other reasons (see [6]-[7]). Although at low levels today, significant penetrations of electronically-coupled renewable resources like wind and photovoltaic solar power can further reduce interconnection frequency response, if they are installed without additional enhancements that can provide frequency response [8]. However, the decline in frequency response may also be due to the electricity market design in some areas that may not incentivize frequency response, or in some cases offer disincentives.

The need for incentivizing frequency responsive reserve was one of the principal recommendations of an IEEE Task Force report on generation governing concerns [9]. Little attention has been given to frequency responsive reserve incentives since this initial report, and no U.S. region currently has a market for this service.[1] As regions begin to understand the need for a reliable frequency responsive reserve and create the standards to guide this need, it will become more important to ensure incentives are in place to assist the individual resources. In this paper, we discuss the issues and possible solutions of ensuring incentives are available for resources to provide frequency responsive reserve. In our analysis, we focus on Independent System Operator (ISO) and Regional Transmission Organization (RTO) regions, which already have restructured energy and ancillary services markets. However, in non-restructured areas (i.e., regulated utility areas), it is equally important for the balancing area operator to offer incentives for this response. Section II gives an overview of all of the mechanisms needed for reliable frequency control. In Section III, we give an overview of market design principles and discuss a possible flaw in the current market design where frequency response may be disincentivized. Next, Section IV describes potential changes to the existing market designs to eliminate disincentives and provide incentives for frequency responsive reserve. Section V provides a conclusion.

II. Frequency Control Overview

Frequency control is utilized with a variety of mechanisms. When a disturbance occurs, there is a set of standards and procedures that guide how balancing areas should respond. In North America, part of the response guidance is based on the North American Electric Reliability Corporation (NERC) Disturbance Control Standard (DCS) [10]. In Europe, the way in which transmission system operators respond is based on the Union for Coordination of Transmission of Electricity (UCTE)[2] Policy 1 [11]. Before discussing either of these policies, it is important to discuss the textbook example of the various mechanisms of response to a large disturbance event.

[1] According to [9], the only restructured area in the world with a market for this service is in New Zealand.

[2] The UCTE has been somewhat dissolved, with the European Network of Transmission System Operators for Electricity (ENTSO-E) now taking over as the entity with the responsibility of transmission level policies and standards. Policy 1 however, still refers to the UCTE Interconnection, which consists of the largest Interconnection in continental Europe.

Generation must always meet demand. When a large loss of supply occurs, due to a generator tripping offline, or due to a loss of a large line carrying imports into an interconnection, the amount of generation from that supply will instantaneously be lost. At the instant of the loss, the deficiency is met by the release of kinetic energy from the synchronous generators and motors on the system. This response is often termed inertial response. Since these machines are synchronous with the system's electrical frequency, their release of kinetic energy will slow down the rotational speed, thereby causing a decrease in the electrical frequency. The rate at which the frequency declines will be determined by the amount of inertia on the system, as well as the amount of the imbalance. Larger machines, and more machines, provide greater inertial response and will result in a slower decline. A higher amount of generation lost will require more kinetic energy and will, thus, increase the speed of decline. The slower the decline, the better, as this gives more time for the reserve to respond before reaching undesirably low frequencies.

The frequency responsive reserve (which is often also referred to as primary response) will then be triggered and is used to arrest the frequency decline and bring the frequency to a new steady-state level, below the nominal level. This response is autonomously triggered by turbine governors, inherent load response, and, in some cases, frequency responsive demand response [12], which sense the frequency deviation and, in response, increase generation (or decrease demand). After this quick response, a slower response can be supplied by resources on spinning and nonspinning reserve (often referred to as the secondary response), returning the frequency back to its nominal level. The NERC DCS and UCTE Policy 1 require that the frequency be returned to its nominal level in less than fifteen minutes.[3] Figure 1 shows the frequency following a loss of a large supplier, with the associated responses that provide a service to ensure frequency is held as close as possible to its nominal level to avoid reliability issues and under-frequency load shedding.[4] Note that, though not as common, a loss of a large block of load or a pumped storage hydro plant in pumping mode, can cause a similar effect, but as an over-frequency event. The responses are somewhat similar, but generation must be decreased to avoid over-frequency issues.

[3] The actual requirements are for the ACE, discussed later, to be returned to zero or its pre-contingency level, but since the secondary response is to be provided by the contingent balancing area, this can essentially result in a similar effect.

[4] The highest setting for under-frequency load-shedding in the continental United States is in the range of 59.3 Hz (ERCOT) to 59.7 Hz (Eastern Interconnection). On island systems like in Hawaii, it can be much lower.

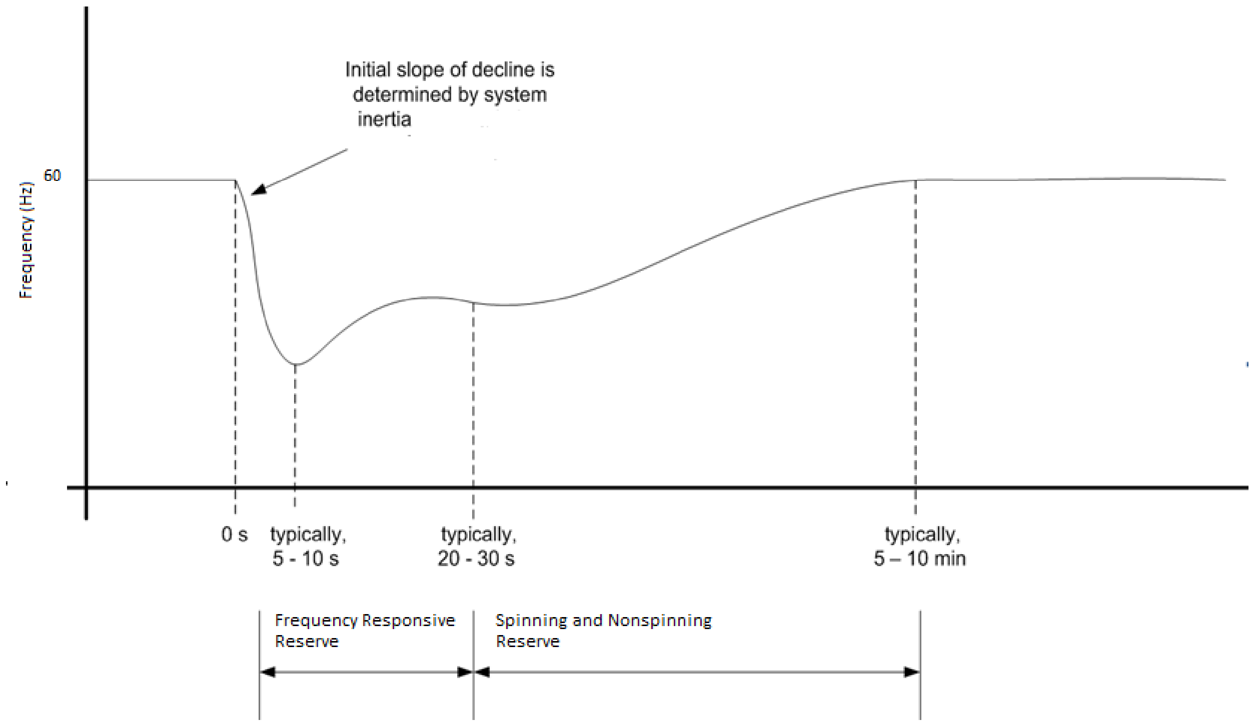

Figure 1: Frequency following a loss of a large supplier.

Lastly, these reserves should be restored so that they are prepared for a subsequent disturbance. This is achieved using reserve referred to as replacement reserve, supplemental reserve, or tertiary reserve, and can be attained through market responses to the event. Figure 2 shows an example of the energy that each service deploys during the event, and when the reserves should be restored and prepared for a subsequent disturbance. The timing on the x-axis displays requirements to deploy the reserves (when they are increased) and when the reserves “should” and “must” be restored (reserves are restored when they are decreased back to zero) and are based on NERC criteria. Additional references on how these ancillary services differ throughout the world can be found in [13]-[14].

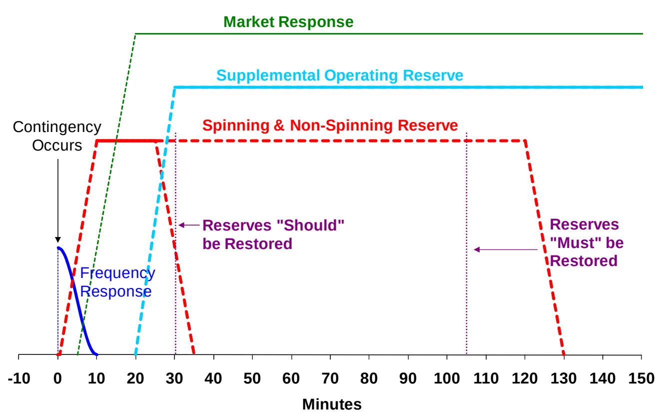

Figure 2: Reserve services deployment.

During normal conditions, when there has not been a significant disturbance or large loss on the grid, regulation is used to correct the imbalance of generation and load for each balancing area[5] in an interconnection. Regulation is generation (also sometimes demand response [15]) that is on automatic generation control (AGC). Units on AGC receive a control signal that is determined by a central operator, i.e., the ISO or balancing area authority. The signal, which the regulation providers receive, is based on correcting the area control error (ACE). ACE determines how much the balancing area is in imbalance with its generation and load. Positive ACE means the balancing area is over-generating and a negative ACE means it is under-generating. Most balancing areas that are part of large interconnected systems in North America, and elsewhere, will perform tie-line bias control [16]. With tie-line bias control, the ACE signal is determined based on Equation (1).

$$ACE = NI_A - NI_S - 10B(F_A - F_S) \qquad (1)$$

Where NI_A is the actual net interchange with neighboring balancing areas, NI_S is the scheduled net interchange, B is the frequency bias, F_A is the actual frequency, and F_S is the scheduled frequency. The net interchange is the sum of all interchanges of a balancing area out of (+) and into (-) the area. If the actual net interchange is greater than the scheduled net interchange, the balancing area is over-generating. The second part of the equation [$- 10B(F_A - F_S)$] is introduced to ensure that systems are providing frequency response, and that the AGC control does not counter its frequency response obligation. The frequency bias, B, is a constant, with the units of MW/0.1Hz, that represents the amount of frequency response that the balancing area has, or should have.[6] The value as it is shown in Equation (1) is a negative value, showing that generation should increase when frequency is low and decrease when frequency is high. We now give some numerical examples.

Scenario 1

NI_S = 500 MW	NI_A = 600 MW
F_S = 60 Hz	F_A = 60 Hz
B = -200 MW/ 0.1 Hz	ACE = 100 MW

In Scenario 1, the balancing area is transmitting 100 MW more power out of its balancing area than it has scheduled to transmit. Therefore, the balancing area is over-generating by 100 MW, as seen in its 100 MW ACE. The AGC would send signals to the resources on regulation to reduce output and, thereby, reduce the 100 MW ACE back to zero.

Scenario 2

NI_S = 500 MW	NI_A = 600 MW
F_S = 60 Hz	F_A = 59.95 Hz
B = -200 MW/ 0.1 Hz	ACE = 0 MW

[5] A balancing area is an area within an interconnection that is responsible to balance its generation and load and keep its inter-tie line schedules with other areas to schedule. Interconnections in North America range from one balancing area (ERCOT and Hydro Quebec) to over 60 (Eastern Interconnection).

[6] Some areas may have a variable frequency bias that changes based on conditions to better reflect the actual system frequency response.

In Scenario 2, the balancing area is again transmitting 100 MW more power out of its balancing area than it has scheduled to transmit. However, this time the actual frequency is below the scheduled frequency. According to the balancing area's frequency bias, B, it is providing the correct amount of frequency response to assist the interconnection. By incorporating the 10 coefficient, we have: (-2000MW/Hz)*(-0.05 Hz) = 100 MW of frequency response that is needed. Therefore, the 100 MW of over-generation from the balancing area is necessary, and so the AGC should not attempt to correct the imbalance between actual and scheduled interchange, hence the 0 MW of ACE.

It is important to understand the contribution of individual resources providing frequency responsive reserve, to distinguish it from spinning reserve and regulation. For generating units, a turbine speed governor is usually in place for both thermal and hydro generating units. The governor will autonomously sense frequency changes, as it senses the speed of the turbine, and will change the position of steam valves, combustion turbine fuel flow, or water gates, to adjust the power output in response. For a stable response from the various resources contributing frequency response, a speed droop is used. The droop curve determines the pickup from each resource with respect to the frequency change. For example, Figure 3 shows a 5% droop curve. For a 5% change in frequency, which on a 60 Hz system reflects a 3 Hz change, the governor will adjust by 100% of its capacity. This determines the slope of the curve and how much the governor will adjust based on varying deviations of frequency. Typical droop curves range from 3-5%.

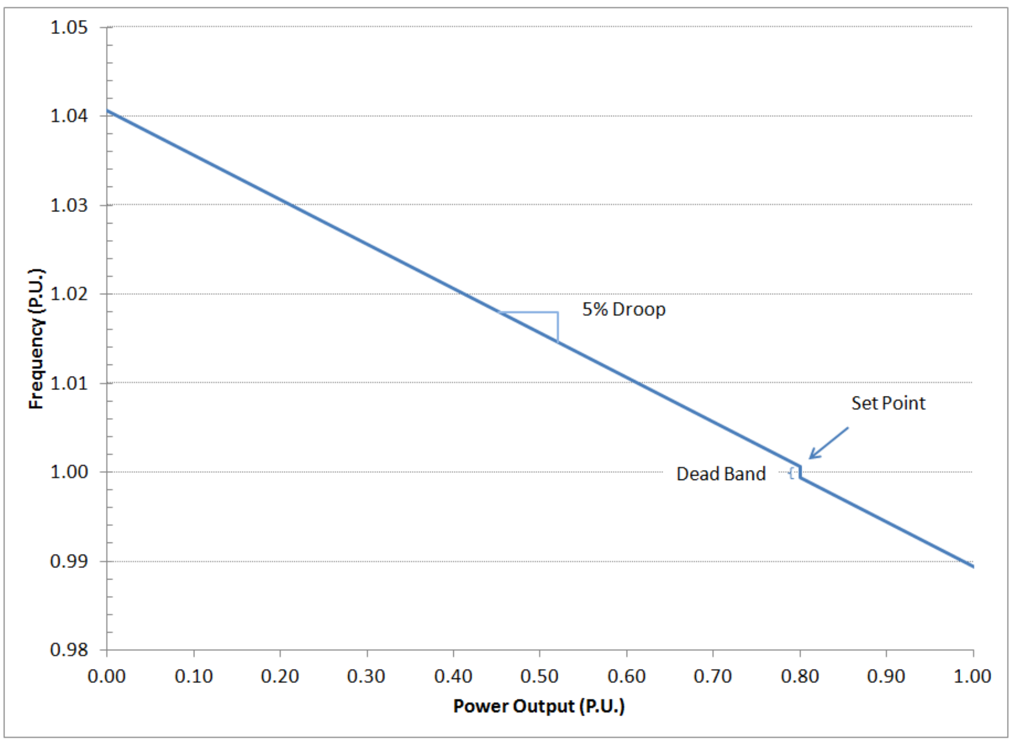

Figure 3: Governor Droop Curve in per unit values. 1 p.u. frequency would equate to 60 Hz in the United States, and 1.01 p.u. would equate to 60.6 Hz. A 100 p.u. power output would equate to the capacity of the plant.

Usually, generators will also carry a governor dead band (depicted in Figure 3). The governor dead band ensures that the governor is not operating during relatively minor frequency changes, as its autonomous operation is only desired for large disturbances. Typical governor dead bands are in the range of 0.02 – 0.06% of nominal frequency (e.g., 12-36 mHz on a 60 Hz system). When frequency deviations are below these levels, the generator does not provide frequency response through its governor.

The amount of frequency responsive reserve is also dependent on the set point of the generating unit and its minimum and maximum capacity levels, as well as if it is operating in a mode that allows for frequency response or an enabled governor. If a generator is generating at its maximum capacity, it cannot provide any frequency response during under-frequency events, regardless of its droop curve, dead band, and its governor being enabled. Therefore, the "head room" is very important when determining how much frequency responsive reserve is available. Units also may be operating in modes that do not provide frequency response or block the governor response. These include turbine follow, integrated boiler turbine controls, and sliding pressure modes [6]. Extremely high dead bands and limiters can also block governor response.

The last important component of frequency responsive reserve is the timing when the frequency response is achieved. Different technologies have different inherent time constants for the frequency response to be fully activated. For example, when steam valves are opened on a steam turbine generator, the change in electrical output that results is not instantaneous. This delay is important as the faster the frequency response is delivered, the sooner the frequency change can be arrested, and a higher minimum frequency, or nadir, can be achieved, resulting in less potential for under-frequency load-shedding. In addition, the response should be sustained long enough for the secondary response to return the frequency to its nominal setting. If there is a withdrawal of the frequency responsive reserve before the secondary reserve response, the frequency will start to decline from the steady-state level, placing it again at risk of under-frequency load-shedding. Resources operating in outer-loop megawatt control will have this affect; shortly after they provide frequency response, they are driven back to their original MW set point.

Although not discussed in detail, frequency responsive demand response can assist in frequency response, both inherently and intentionally. Some loads are frequency dependent, like for example fans and pumps that are directly driven by synchronous or induction motors, and will, therefore, inherently reduce demand during frequency declines (and increase when frequency increases). The amount that the load inherently changes with frequency is referred to as the load damping constant, given in the percent change in the load per the percent of frequency change. For most systems, this value ranges from a 1-2% load change per 1% frequency change. In ERCOT, Loads Acting as a Resource (LAaR) is the demand response that provides an intentional load response, either manually as directed by the ERCOT operator, or through intentional under-frequency relays, set at frequencies significantly higher than under-frequency load-shedding relays for fixed load, but still at frequencies that are low enough so that they are only reached during very large disturbances. These demand resources are paid for their service and can help avoid significant frequency declines.

The payment for ERCOT LAaR might be one of the few instances when a resource is paid for providing frequency responsive reserve, since there are instances in which the resources explicitly respond to frequency deviations. However, the resources are paid under the "Responsive Reserve Service" market (similar to spinning reserve), and at slower timescales. Otherwise, technologies that provide any of these services are not incentivized to provide the frequency response service. There is often no requirement for generators to be operated such that their governors are enabled or provide effective response. If the requirement does exist in areas, it is difficult to monitor and enforce the rule as many technologies may be operating in modes that do not provide adequate response. This is likely one of the major reasons for the decline of frequency response in the United States, especially in the Eastern Interconnection. If resources have no incentive (or requirement) to provide the service, and it costs them to provide it, either directly or indirectly through increased wear and tear, there would be little motivation for them to provide the service.

III. Markets and Incentives

A well-designed electricity market can be difficult to design and must carefully consider the alignment of the market incentives with the needs of the power system [1], [17]. Unintended consequences can occur, reducing the effectiveness of the market or causing undesired impacts in separate related markets. The objective of a market is to elicit an incentive for providers to supply the desired product in an economically efficient manner. The market will create competition, which will initiate market participants to minimize their costs of providing the particular service. This means that a well-designed market will induce the least-cost solution, subject to the various physical constraints and objectives of the market.

When a new market is introduced, there will typically be short-run and long-run adjustments. Initially, market suppliers may have limited ability to respond because operational changes or equipment needs or modifications may be required. After the suppliers make these changes, the market is likely to function smoothly (in the absence of market power).

Rules can be used in cases where markets are too difficult to design, or if the market would be fundamentally flawed as a result of market power or other concerns. For example, the ancillary service of voltage support is typically a cost-based service [18]. Voltage support does not have a market with dynamic pricing and competitive suppliers. Voltage support is serviced with injecting and absorbing reactive power, and reactive power does not travel far on the transmission system. Since reactive power does not travel far, there are few suppliers of voltage support when it is needed in a particular location, limiting the competition for that service. This type of limitation is the main reason behind creating rules and standards for supplying voltage and reactive power support rather than introducing a dynamic market. Complexity, cost of administration, and low diversity in costs from competitors (i.e., if all suppliers would inherently cost the same to supply the service and innovations to reduce costs are not possible) are all reasons why rules may be more practical than introducing markets for certain services.

Disincentives in the current market design to provide frequency response

Sometimes, in the absence of markets for particular services, there may be flaws when that service is coupled with other incentives. We will highlight an example of how current energy market designs may be giving disincentives to potential suppliers of frequency responsive reserve. In most ISO regions, energy is settled at prices and schedules that are produced from a security-constrained economic dispatch model. The energy markets are based on two settlement systems, so that a forward market will create initial schedules, and a real-time market will accommodate the differences in demand from what was expected in the forward market. The day-ahead market is the common forward market. Day-ahead schedules and day-ahead locational marginal prices, or LMP, are produced. The LMP is the price settled for each location on the transmission grid, and is based on the marginal cost of supplying energy at the location. In the real-time market, differences from the day-ahead market are settled at the real-time LMP, which can be quite different than the day-ahead LMP at the same location, based on the new characteristics of the demand, and generation and transmission available. Therefore, the full settlement of a generator is based on Equation (2) where DA is day-ahead, RT is real-time, Pg is the generation schedule, i is the index for the generator, and h is the index for the hour.

$$Payment_{i,h} = Pg_{i,h}^{DA} * LMP_{i,h}^{DA} + (Pg_{i,h}^{RT} - Pg_{i,h}^{DA}) * LMP_{i,h}^{RT} \qquad (2)$$

The only time that payment may differ from this situation is when generators do not produce the same amount of energy as they are scheduled in the real-time market. Even though it is a real-time market, the schedules are produced for some time in the future (usually 5-10 minutes ahead). Therefore, it cannot predict the energy that the generators will produce nor can it predict the exact demand. Since generators have particular control over the output that they produce, most ISOs provide disincentives for the generators to produce energy differently from what they were told to produce. In some cases, this may be a change of LMP to reflect the actual conditions (i.e., the prices are given ex post, where prices are based on actual outcomes, rather than ex ante, where prices are based on expected outcomes). In other cases, financial penalties may be assessed to those market participants who stray from the schedule they were given. Both methods should work in incentivizing generators to follow their schedule as closely as possible.

As discussed in the last section, a properly functioning governor will respond autonomously to frequency deviations, without the intervention of a control room operator. This is what they are designed to do and this response is vital for power system reliability. In conversations with ISO employees, and in various manuals, tariffs, and user guides of the U.S. ISOs, the authors have found that few, if any, ISOs incorporate frequency as input into the energy settlements system. A frequency deviation will cause a generator with a turbine governor to adjust its output to assist the interconnection in arresting the frequency decline (or frequency increase in the case of a loss of load or a loss of a pumped storage hydro facility), and its energy output will deviate from its real-time energy schedule. This can cause either a change in the LMP that reflects the ex post pricing, which may harm the generator, or it can be financially penalized for deviating from its energy schedule.

As an example, one U.S. ISO has a 3% tolerance band around its financial penalties. This means that if a generator deviates from its schedule by more than 3% of its operating capacity above or below its energy schedule, it will receive a financial penalty for not following schedule. Assuming, that a generator with a properly functioning governor has a 5% droop curve, we have the following:

$$\frac{1\, p.u.power}{0.05\, p.u.frequency} = \frac{0.03\, p.u.power}{X\, p.u.frequency}$$

$$X\ =\ 0.0015\, p.u.frequency\ =\ 90\, mHz\ for\ a\ 60Hz\ system$$

This means that any time the frequency deviates more than 90 mHz (i.e., if frequency dips below 59.91 Hz or above 60.09 Hz), any generator with a properly functioning governor and a 5% droop curve will automatically be penalized under this rule, if frequency is not incorporated within the energy settlements system. This is a significant disincentive. If the generator has no other incentive to provide frequency response, which occurs when there is no frequency responsive reserve ancillary service market, it will have no motivation to enable its governor or operate in modes to provide frequency response. Since frequency response is a crucial service to the power system, this can have drastic affects if this turns into a trend, which some researchers believe it has [9].

There are some important discussion points that are worth mentioning. First, a 90 mHz deviation, especially in the Eastern Interconnection, is extremely rare. A very large loss of supply or loss of two or more large generating units would have to occur for a frequency deviation of this magnitude to occur in large interconnections. Second, frequency usually begins to return to its nominal level a few minutes after the disturbance event. Penalties may be based on an hourly average production, meaning that a deviation based on the frequency response would likely have little impact on deviations from the hourly schedule. The point is, regardless of these caveats that may lessen the impact of these disincentives from providing frequency response, there still is a disincentive, and without any incentive for providing frequency response or any standard or grid code enforcing the response, generators will have every reason to disable their governors or operate in a way that provides little or no frequency responsive reserve. When the governors are disabled, the frequency response would then have to come from the secondary (AGC) response, a much slower response, which may cause degradation in the overall system frequency response, and put the system at higher risk for under-frequency load-shedding and other issues.

IV. Market Designs Modifications to Incentivize the Frequency Responsive Reserve Ancillary Service

We now discuss some potential resolutions and new or modified market designs to provide alternative levels of incentives, with varying levels of market design effort. There are two, usually conflicting goals, in market design. The more complexities that are included to characterize the responses of the market participants and system will better reflect what is needed as the desired response. It should also limit market participant gaming. However, the more complex the market design, the more difficult and expensive it will be to implement and obtain regulatory approval, and the benefits may be diminished in smaller markets. With this in mind, we have proposed four market design proposals for consideration. If any of the ISOs were to choose any of these proposals, or a combination of them, they would likely work out specific details through the regular ISO stakeholder process to determine which design best fits the particular region.

Proposal 1: Elimination of penalties during frequency events.

The simplest proposal would eliminate penalties during frequency events, as discussed in the previous section. For example, in the New York Independent System Operator (NYISO) Market Services Tariff [19], there is a current rule that over-generation penalties do not apply during reserve pick up periods. Reserve pick up periods occur when reserves are needed for a disturbance that occurred in the NYISO balancing area. This rule could be extended so that over-generation penalties do not apply when frequency is below some threshold, and under-generation penalties do not apply when frequency is above some threshold during any instant within the settlements interval. This would assure market participants that enabling the governor and providing frequency responsive reserve cannot harm them financially.

The benefit of this change is that it is relatively simple to implement. No new markets or changes to scheduling software are required, and there are no changes to the resulting prices or schedules. However, it would mean that frequency would have to be recorded in the settlement system, and that some basic logic would have to be added to the settlements system. The logic would simply check every penalty assessed to see if there were any frequency excursions that would have caused governors to cause the penalty, and void the penalty if so. A simple threshold could be based on a typical governor dead band. If the frequency deviation is within the governor dead band, the governors do not react and, therefore, there is no reason to void the penalty. If the frequency deviation is outside the dead band, then the governors react and, therefore, there is a reason to ensure that no resources with functioning governors are penalized during these times. While this proposal may eliminate disincentives, there is still no incentive for resources to enable governors and provide frequency response.

Proposal 2: Specific accountability of frequency response to avoid penalties.

The second option applies additional parameters to the settlements system, so that only those resources that offer the frequency responsive reserve avoid penalties. In this proposal, in addition to the frequency being used in the settlements system, governor droop and governor dead band are added. The settlements system now includes the logic to determine what the resource's output should be during frequency excursions and during normal conditions. Instead of avoiding penalties during frequency excursions, the settlements system knows what the frequency responsive units should have provided based on the frequency and its response characteristics. Resources without a governor or operating in a mode that does not provide frequency response should not avoid these penalties.

The benefit of this change is that it ensures proper frequency response, and still retains the original goals of penalties and settlements for those resources that are not following schedules, whereas Proposal 1 may eliminate the penalties for resources regardless of whether they are actually helping the system during the frequency excursions. This can better incentivize resources to provide the desired frequency response. However, there is added complexity in that additional parameters and logic must be added to the settlements software. The complexity of the software logic will depend on the time resolution of the retained frequency measurements. Also, if this proposal is implemented without any rule that certain resources are required to enable governors, there is still no incentive for those resources to enable governors and provide the frequency response.

Proposal 3: Incorporate Frequency Responsive Reserve Requirement within Spinning Reserve Requirement.

The third proposal can likely be implemented in conjunction with either Proposal 1 or Proposal 2. This proposal would require that any resource providing spinning reserve and participating in the existing spinning reserve market enable its governors and provide frequency responsive reserve with enforcement. There is a connection with spinning reserve and frequency responsive reserve since both are responsive to disturbances on the system. The spinning reserve would then be required to respond both to system operator commands and autonomously to frequency deviations. This offers an additional incentive for resources to provide frequency responsive reserve in contrast with Proposal 1 or Proposal 2 by themselves. The resources would have to enable the governors to earn revenue in the new spinning reserve market.

This is also a relatively easy market design change to implement. However, it is not clear that this change by itself will obtain the correct amount of frequency response. Resources differ in the amount of frequency response they are capable of providing. Scheduling sufficient spinning reserve might not result in sufficient frequency responsive reserve being available. Further, the amount of available frequency responsive reserve could change as the mix of resources providing the spinning reserve changed. Imposing specific frequency response capability requirements on spinning reserve resources would likely limit the spinning reserve supply and increase the price [20]. Energy prices could also be impacted since spinning reserve resources also supply energy. Locational constraints for spinning reserve and frequency response often differ, further complicating a simple joint supply requirement. Spinning reserves in many areas have locational requirements to avoid overloading the transmission system during contingency events. Frequency responsive reserves may have completely different location-based requirements, or may have none at all, which would add a further complexity, if the two were paired together in one market. Other than these issues, the proposal would be relatively easy to implement, making an incentive available for resources to provide frequency responsive reserve.

Proposal 4: Separate Frequency Responsive Reserve Ancillary Service Market.

The fourth proposal is the implementation of a new frequency responsive reserve ancillary service market within the ISO. This proposal is also recommended in [20] and, less explicitly, in [9]. The market would incorporate the reliability requirement of a minimum amount of frequency responsive reserve, similar to spinning reserve. This requirement, in both MW and MW/0.1 Hz, would ensure enough head room and enough sensitivity to avoid under-frequency load-shedding following some large, credible interconnection-wide event. The resources would offer their droop curves, response range, time delays, and governor dead bands, and the market would select the least cost optimization that meets the specified reliability requirements. Since frequency responsive reserve is tightly coupled with energy and other reserves, it would be beneficial to co-optimize this service with energy, as is done with the other ancillary services.

This proposal provides the incentives in Proposal 3 to provide frequency responsive reserve. However, it avoids the issues of pairing two services at the same price when they are in fact different services, as well as the other issues (e.g., measurement of compliance, and locational requirements). The major issue is the complexity of implementing this new market, with regard to both software and regulatory complexity. The market software would have to be enhanced to incorporate these new parameters and requirements, as is done for example in [21]-[23]. It would also have to go through large regulatory hurdles to introduce a new market to consumers, which would require the approval of stakeholders and federal regulators.

Comparison

Table I gives an overview of the four market design proposals discussed in the previous sections. Each has different ways of eliminating disincentives, providing incentives, and increasing market design complexity. Each of these proposals has different advantages, and can also be paired with another to achieve the best efficiency. The authors believe that the most efficient results can be achieved with Proposal 4 paired with some version of Proposals 1 or 2. However, the benefits will likely be system-specific. The costs and regulatory hurdles of Proposal 4 can make it less attractive and perhaps not appropriate in certain markets. However, this option will establish clear incentives for resources to enable governors and provide frequency responsive reserve, while eliminating any disincentives that penalize resources in the energy market for providing the response. Importantly, if there is no cost for providing the service during certain instances, which should occur during times when enough resources are available to provide the service and compete in the market, then it should result in a price of zero. This often occurs with the other ancillary services like spinning reserve, especially at night when the supply-side competition is high, and costs are low. If this market was created, it would be important to monitor the activities and outcomes, through normal market monitoring procedures, to ensure proper competition exists and resources are not gaming the market through loopholes or other means.

Table I: Overview of Frequency Responsive Reserve Market Design Proposals

Market Design Proposal	**Eliminate Disincentives?**	**Provide Incentives?**	**Complexity L:Low M:Medium H:High**	**Limitations**
Proposal 1: *Elimination of penalties during frequency events*	Yes, penalties would no longer be assessed during frequency disturbances	No, but could be paired with Proposal 3 or 4	L: Very minor changes to settlements rules	Penalties will be avoided regardless of whether the resource is providing frequency response
Proposal 2: *Specific accountability of frequency response to avoid penalties*	Yes, for resources that can properly provide frequency responsive reserve	No, but could be paired with Proposal 3 or 4	L-M: Somewhat complicated logic, but only to the settlements system	No requirement to enable governors
Proposal 3: *Incorporate Frequency Responsive Reserve Requirement within Spinning Reserve Requirement*	No, but could be paired with Proposal 1 or 2	Yes, resources that provide frequency response can bid in spinning reserve market	M: Settlements system, monitoring requirements, and regulatory complexities would result	Price signals may not be clear due to joining of two services in one market
Proposal 4: *Separate Frequency Responsive Reserve Ancillary Service Market*	No, but could be paired with Proposal 1 or 2	Yes, correct implementation would give clear incentives for resources to provide frequency responsive reserves	H: Many different software programs would be impacted significantly; Large regulatory process	Complexity of market will be expensive and will take time for market participants to learn

Standards and Requirements vs. Markets

Any new market designs should be compatible with the physics and reliability needs of the power system. Well-functioning markets do not have significant market power issues and provide the product in an economically efficient manner. A frequency responsive reserve ancillary service market is nonsensical, without a justifiable standard or requirement for frequency response. Currently, outside ERCOT, no interconnection in North America has a frequency response standard, though it is currently being developed [24]. The standard will likely be consistent with the current European frequency response policy [11]. The European policy sets an interconnection-wide amount of frequency responsive reserve[7] of 3000 MW. This amount is to be fully deployed when a frequency deviation reaches 200 mHz, which also gives a requirement in MW/Hz (or MW/0.1Hz). The contribution of each balancing area is then based on total energy generated in a year by the balancing area divided by the total energy generated in a year by the entire interconnection. Other requirements discuss the governor dead bands and speed of deployment. Each of these requirements could be utilized as minimum requirements, or achieved through market designs that incentivize the desired response characteristics. For

[7] In Europe this is called primary control reserve.

example, to achieve the frequency response characteristic requirement in MW/Hz, the market may increase the pay for a generator of more MW/Hz than another, e.g., a generator with 4% droop would get paid more than a generator with a 5% droop, even if the same MW capacity is offered, because it is offering more frequency response. The design must be careful to incentivize the right responses. Finally, any of these requirements should be coordinated in regions throughout the interconnection. Unlike spinning reserve and regulation, frequency response is an interconnection-wide issue, and coordination of the new market rules should be consistent throughout the interconnection. This becomes difficult on an interconnection like the U.S. Eastern Interconnection, where there are many balancing areas of different sizes and not all areas have restructured market designs. Also, other ancillary service markets and requirements are constantly evolving and these changing requirements should be considered in any new market designs. For example, the proposed reliability-based control requirement may change regulation requirements [25], and the new pay-for-performance requirements for the regulation requirement in a recent FERC order [26] are two ways the regulation market may change and impact the rules and market design of a frequency responsive reserve market.

While a system standard is absolutely required for a market to exist, unit-specific standards can exist as an alternative to a market. For example, if every generating unit on the system was required to have its governor enabled with specific characteristics, the need for a frequency responsive reserve market would be eliminated. The original movement toward deregulating the electricity sector initiated from the fact that new supply-side technologies could offer energy at various ranges of costs, and deregulation helps to reduce costs borne by consumers. This trend is seen in the technologies that provide frequency responsive reserve. Many technologies have different ways of providing frequency response. For example, nuclear generators rarely have governors as they are generally operated with load limiting. Combined cycle gas turbine technology has a very different frequency response than conventional steam turbine technologies. Wind generators, which historically did not provide any frequency response, can be equipped with power electronics that can control the blade pitching to provide frequency response [27]. However, adding this control will likely come at a cost. By introducing frequency responsive reserve incentives, it would be up to the wind plant market participant to decide whether or not to include this control, based on how much revenue it believes it can earn. Photovoltaic solar or electronically interfaced storage would face the same question. This new trend of different market participants being able to supply different forms of frequency responsive reserve at different costs, may reveal a frequency responsive reserve market that is a more efficient alternative to unit-specific requirements.

V. Conclusions

This paper discusses the needs for introducing incentives and eliminating disincentives for resources on the electricity grid to provide frequency responsive reserves. In some areas, especially in the Eastern Interconnection of the United States, the frequency response has been in a constant decline for the past 20 years. Given that load and generation have been increasing during this time, this trend should not occur. If the trend continues, the system may be at risk of under-frequency load shedding, especially during instances of system islanding. The authors believe one of the most important ways to end this trend is to eliminate any disincentives that may convince market participants to provide little or zero frequency response, and potentially to introduce new incentives that may entice current resources to provide the response, or new resources to ensure that the capability is installed. Four potential designs are detailed. While we believe any of the four designs should eventually lead to a better frequency response on a system, the design for any particular system will depend on the characteristics of that system.

VI. References

[1] R. Baldick, U. Helman, B. Hobbs, and R. O'Neil, "Design of efficient generation markets," *Proceedings of the IEEE,* vol. 93, no. 11, pp. 1998-2012, Nov. 2005.

[2] W. Hogan, *Competitive Electricity Market Design: A Wholesale Primer*, Cambridge, MA: John F. Kennedy School of Government, Harvard Univ., [Online]. Available: http://www.hks.harvard.edu/fs/whogan/empr1298.pdf. Dec. 1998.

[3] E. Hirst, and B. Kirby, *Creating Competitive Markets for Ancillary Services*, ORNL/CON-448, Oct. 1997.

[4] H. Singh, and A. Papalexopoulos, "Competitive Procurement of Ancillary Services by an Independent System Operator," *IEEE Transactions on Power Systems*, vol. 14, no. 2, pp 498-504, May 1999.

[5] J. Ingleson, and E. Allen, "Tracking the Eastern Interconnection frequency governing characteristic," *Proceedings of the IEEE Power and Energy Society General Meeting*, Minneapolis, MN, July 2010.

[6] R. Schulz, "Modeling of governing response in the Eastern Interconnection," *Proceedings of the IEEE Power Engineering Society Winter Meeting,* New York, NY, February 1999.

[7] S. Virmani, "Security impacts of changes in governor response," *Proceedings of the IEEE Power Engineering Society Winter Meeting,* New York, NY, February 1999.

[8] J. Eto et al., *Use of frequency response metrics to assess the planning and operating requirements for reliable integration of variable renewable generation*, Prepared for the Federal Energy Regulatory Commission, LBNL-4142E, December 2010.

[9] IEEE Task Force on Large Interconnected Power Systems Response to Generation Governing, *Interconnected Power System Response to Generation Governing: Present Practice and Outstanding Concerns*, IEEE Special Publication 07TP180, May 2007.

[10] North American Electric Reliabillity Corporation, Reliability Standards for the Bulk Electric Systems of North America, June 2010.

[11] Report: ENTSO-E, UCTE Operational Handbook Policy 1, Load-Frequency Control and Performance, March 2009.

[12] Shun-Hsien Huang, John Dumas, Carlos González-Pérez, and Wei-Jen Lee, "Grid security through load reduction in the ERCOT Market," *IEEE Transactions on Industry Applications,* vol. 45, no. 2, pp. 555-559, March/April 2009.

[13] Y. Rebours, D. Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary services-part I: technical features," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp 350-357, Feb. 2007.

[14] E. Ela, M. Milligan, and B. Kirby, *Operating Reserves and Variable Generation*, NREL/TP-5500-51928, August 2011.

[15] B. Kirby, M. Starke, and S. Adhikari, *NYISO Industrial Load Response Opportunities: Resource and Market Assessment*, New York State Energy Research and Development Authority, Oak Ridge National Laboratory, June 2009.
[16] N. Jaleeli, L.S. VanSlyck, D.N. Ewart, L.H. Fink, and A.G. Hoffmann, "Understanding automatic generation control," *IEEE Transactions on Power Systems*, Vol. 7, no. 3, pp. 1106-1122. Aug 1992.
[17] S. Stoft, *Power System Economics*, New York: Wiley, 2002.
[18] B. Kirby and E. Hirst, *Ancillary-Service Details: Voltage Control*, ORNL/CON-453, Oak Ridge National Laboratory, Oak Ridge TN, December 1997.
[19] NYISO Market Services Tariff, accessed December 2011.
[20] H. Ilian, "Relating primary governing frequency response to future operating reliability," *Proceedings of IEEE Power Engineering Society General Meeting*, July 2007.
[21] K. Papadogiannis and N. Hatziargyriou, "Optimal allocation of primary reserve services in energy markets," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 652-659, February 2004.
[22] J. Restrepo and F. Galiana, "Unit commitment with primary frequency regulation constraints," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1836-1842, Nov. 2005.
[23] M. Mashhadi, M. Javidi, and M. Ghazizadeh, "The impacts of capabilities and constraints of generating units on simultaneous scheduling of energy and primary reserve," *Electrical Engineering*, vol. 93, no. 3, pp. 117-126, February 2011.
[24] NERC, BAL-003-1 – Frequency Response and Frequency Bias Setting, Draft 2, October 24, 2011, accessed December 2011.
[25] North American Electric Reliability Corporation, Standard BAL-007-1, BAL-008-1, BAL-009-1, BAL-010-1, BAL-011-1, BAL-012-1, Standards authorization request, July 2004. Available: http://www.nerc.com/docs/standards/sar/BAL_007-012_03-11-05.pdf.
[26] *Frequency Regulation Compensation in the Organized Wholesale Power Markets*, Order No. 755, FERC Stats. & Regs. October 20, 2011.
[27] I. Erlich, and M. Wilch, "Primary frequency control by wind turbines," *Proceedings of IEEE Power and Energy Society General Meeting*, July 2010.