IEEE Access
Multidisciplinary ⁞ Rapid Review ⁞ Open Access Journal

Received August 19, 2020, accepted September 12, 2020, date of publication September 29, 2020, date of current version October 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027722

New Ancillary Service Market for ERCOT

PENGWEI DU[1], (Senior Member, IEEE), NITIKA V. MAGO[1], WEIFENG LI[1], (Member, IEEE), SANDIP SHARMA[1], QINRAN HU[2], (Member, IEEE), AND TAO DING[3], (Senior Member, IEEE)

[1]Electric Reliability Council of Texas, Taylor, TX 76574, USA
[2]Southeast University, Nanjing 210096, China
[3]Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Pengwei Du (pengwei.du@ercot.com)

ABSTRACT Ancillary services (AS) are the services necessary to support the transmission of electric power from generators to consumers given the obligations of control areas and transmitting utilities within those control areas to maintain reliable operations of the interconnected transmission system. As a result of the increasing penetration of renewable resources, the new reliability needs are emerging so that the changes made to the AS market become necessary. This paper presents a new AS framework being implemented at the Electric Reliability Council of Texas (ERCOT) in order to address the primary frequency control issues associated with the declining system inertia. This new design can balance the need for both reliability and economics while opening up the AS market to the traditional and non-traditional resources, including load resources with under-frequency relays and energy-limited resources like batteries. This paper also introduces a new way to quantify the benefits of the new AS design using ERCOT model and data.

INDEX TERMS Ancillary service, primary frequency control, system inertia, load resources with under-frequency relays, energy-limited resources, battery.

ACRONYM

AS:	ancillary service
HASL:	high ancillary service limit
HDL:	high dispatch limit
HSL:	high sustainable limit
FFR:	fast frequency response
LR:	load resource
NSRS:	non-spin reserve service
PFR:	primary frequency response
RRS:	responsive reserve service
RRSC:	RRS schedule
SCED:	security constrained economic dispatch
UDBP:	updated desired base point
UFR:	under-frequency relay
UFLS:	under frequency load shed
VDI:	verbal dispatch instruction
VER:	variable energy resource

I. INTRODUCTION

Following Federal Energy Regulatory Commission (FERC) order 888, the U.S. power industry began its restructuring process [1]. Essentially, the competition was introduced among the generation companies after the vertically integrated utility was unbundled [2]. To date, wholesale electricity markets in U.S., with a goal to increase the power system's economic efficiency without compromising its reliability, have been considered successful [3]. In U.S., the market operations are administratively managed by independent system operators (ISOs). While the market design differs cross the regions, all of those markets procure and manage an array of ancillary service (AS) products to ensure that they can balance the supply and demand for energy in real-time. In this regard, ASs are necessary to support the transmission of energy from resources to loads while maintaining reliable operation of the transmission service provider's transmission system [4]. It is also a prevailing practice to co-optimize the provision of energy and ASs to achieve the most efficient capacity allocation of resources.

AS market has evolved with changes in regulatory policy, technological innovations, and economic conditions since the start of electric restructuring [4]–[6]. The AS products are designed according to the needs of reliable operation for power systems. Because the different systems may have different reliability needs, AS products could be different among ISOs.

The associate editor coordinating the review of this manuscript and approving it for publication was Ning Kang.

Today, variable energy resources (VERs) have grown dramatically in their installed capacity. While VERs have tremendous benefits, a high penetration of VERs creates a new challenge to maintaining the reliability and security of a power grid [7], [8]. Therefore, moving toward a future power system with a lower carbon footprint, it is necessary to make changes to the AS market design as the new reliability need arises. Recently, some changes in the AS market design have been implemented. For example, California ISO introduced a 5-minute ramping product and Midcontinent ISO added a 10-minute ramping product to their AS market [7]. Both of them are mainly used to mitigate the variability and uncertainties of VERs.

VERs are also connected to the grid through an inverter so that they do not contribute to the system inertia. Power system inertia is defined as the ability of a power system to oppose changes in system frequency due to resistance provided by rotating masses [9]. Inertia is dependent on the amount of kinetic energy stored in rotating masses of synchronously interconnected machines. As VERs do not provide inertia, the large-scale integration of VERs could lead to a lower system inertia. A declining system inertia could result in a difficulty in regulating the system frequency [10]–[13], which is considered as one of the major barriers to the reliable integration of VERs at the large scale.

Various works have been performed to address the need of improving the secure and reliable operation of a future power grid under the low system inertia conditions. In [14], the frequency stability challenges at ultra-high wind penetrations were examined and a system non-synchronous penetration ratio was defined to help to identify system operational limits. The authors in [15] proposed a framework for assessing renewable integration limits concerning power system frequency performance using a time-series scenario based approach. The study in [16] established the instantaneous penetration level limits of VERs. In [17], [18], the frequency response change trend of each U.S. interconnection due to the increasing penetration level of VERs was examined. All of these studies pointed to a great need of the primary frequency response (PFR) capability as the system inertia decreases. PFR denotes the autonomous reaction of system resources to change in frequency [19]. In the past, the main contributor to PFR has been the governor response of synchronous generation units.

Given the significance of PFR, researchers also studied how to incentivize synchronous generators to provide such a PFR capability. The basic principle of scheduling and pricing of coupled energy and primary, secondary, and tertiary reserves is discussed in [20]. A simplified dynamic model is introduced to determine the minimum spinning reserve requirement that is used as part of the constraints in economic dispatch for a pool-based power market [21]–[23]. The dependency between the system inertia and PFR is approximately taken into account in the scheduling process [24]. More detailed models of the governor responses are adopted in [25], [26] to calculate the pricing of PFR influenced by different dynamic characteristics of the governors. The problem formulation accounting for PFR constraints in unit commitment is described in [27]. Those theoretic studies focused primarily on the provision of PFR from synchronous generators without considering other viable resources. In [28], [29], an energy, inertia and reserve co-optimization formulation was proposed in which the PFR requirement can be met by synchronous generators and load resources.

However, these theoretic works reported in [20]–[29] have been performed assuming that no changes are applied to existing AS products. The practical effort to reinvent the AS products is still lacking, but it is imperative to incentivize the resources like batteries to provide fast frequency response capabilities. One exception is the national electricity market (NEM) in Australia, where the 6-second frequency control ancillary service (FCAS) product is used to arrest the frequency decline and the 60-second product is used to stabilize the system frequency [30].

The Electric Reliability Council of Texas (ERCOT) is an ISO serving over 23 million customers in Texas. As a single Balancing Authority (BA) without synchronous connections to its neighboring systems, ERCOT relies purely on its internal resources to balance power shortages and variations. More than 35 GW of wind generation and 10 GW of solar generation will be connected to the ERCOT grid by 2021. This paper presents a new framework to re-design AS market, which is being implemented at ERCOT in the anticipation of the reliability need arising from a declining system inertia. The salient features of the new AS market include

1) The new framework considers the balancing need of the grid cross a variety of time scales so that the array of AS products will be coordinated holistically when deployed.
2) The new design of AS market is able to accommodate difference resources in providing PFR capability although the characteristics, opportunity costs, economics, and practicality of these resources vary.

This paper is organized as follows. Existing AS market at ERCOT is reviewed in Section II. Section III presents the inertia trend and primary frequency control at ERCOT. The new AS market is discussed in Section IV. The details for the new AS product, i.e., fast frequency response (FFR), are provided in Section V. Section VI and VII describe the methods to determine the maximum amount of FFR allowed and the benefits of FFR, respectively. The conclusion is given in Section VIII.

II. EXISTING ANCILLARY SERVICE MARKET AT ERCOT

ERCOT is running a security constrained economic dispatch (SCED) market every 5 minutes to operate the system at least cost while managing the reliability. In order to manage reliability, SCED must dispatch resources to balance generation with load demand, while operating the transmission system within established limits.

However, load and generation are constantly changing, due to daily load patterns, instantaneous load variations, changes

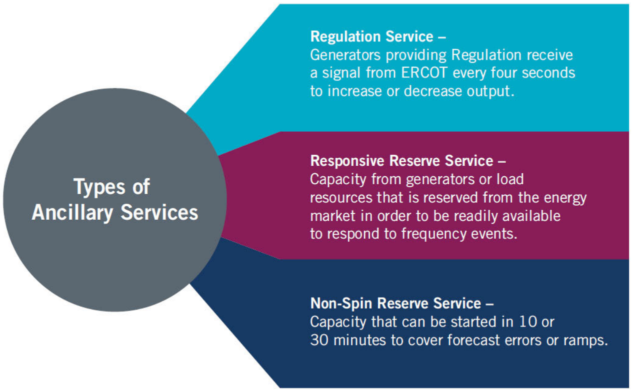

FIGURE 1. Overview of AS products at ERCOT.

in variable generation output and disconnection of generators. Thus, ASs are procured in the day-ahead market at ERCOT to ensure that the reserve capacity is available in real-time to address variability that cannot be covered by the five-minute energy market. Once the day-ahead award for ASs is cleared, it will become physically binding in real-time, i.e., those resources need to fulfill their AS obligations in real-time once awarded in the day-ahead market. There is an ongoing effort to change the market design such that the provision of ASs will be co-optimized with the energy in real-time after 2024.

An efficient and well-functioning AS market is critical to provide incentives to these qualified resources so that the sufficient resource capacity is reserved which can be deployed in a timely manner to restore the balance between the load and generation. The wholesale market design and rules have been evolving in the past. Despite this, the AS products procured in the AS market remain unchanged since the inception of the AS market. These AS products, which are primarily characterized by their response time and deployment mechanism, have been introduced to meet North American Electric Reliability Corporation (NERC) requirements. Similar to other AS markets, AS products at ERCOT consist of regulation service (up and down), responsive reserve service (RRS) and non-spin reserve service (NSRS) before the AS market was restructured – see Fig. 1. The major characteristics of these AS products are briefly described as follows.

A. REGULATION SERVICE

Regulation service is needed to correct actual frequency to scheduled frequency and to ensure NERC reliability requirement (BAL-001) is met. Regulation service is provided by the resources that can be deployed every four seconds to compensate for the load or generation variations within the SCED time intervals, i.e., ERCOT sends load frequency control (LFC) signal every four seconds to increase or decrease power output to the generators providing regulation services. The amount of regulation service needed is determined by the historical usage of regulation service and 5-minute net-load[1] changes [32].

[1] net-load is the load minus the aggregated wind and solar generation.

B. RESPONSIVE RESERVE SERVICE (RRS)

Per NERC standard BAL-003 "Interconnection Frequency Response Obligation", ERCOT must procure a sufficient amount of RRS to avoid activation of under frequency load shed (UFLS) at 59.3 Hz for loss of two largest generation units within the ERCOT interconnection.

RRS can be provided by the capacity from generators or load resources that are readily available to respond to frequency excursions during unit trips. Two types of frequency response characteristics exist currently. First, the governors of thermal generating units begin to respond "immediately" but take a few seconds to provide significant deployment (since they require more steam or more combustion). Second, load resources (LRs) providing RRS have under frequency relays that respond in approximately 0.5s when frequency drops below 59.7Hz.

C. NON-SPIN RESERVE SERVICE (NSRS)

NSRS is provided by those resources that can be started in 10 or 30 minutes to cover net-load forecast errors or ramps. These resources consist of generation resources capable of being ramped to a specified output level within 30 minutes and load resources that are capable of being interrupted within 30 minutes and that are capable of running (or being interrupted) at a specified output level for at least one hour. NSRS is required to meet NERC BAL-002 (to restore contingency reserve within 90 minutes).

NSRS may be deployed to replace loss of generating capacity, to compensate for net-load forecast uncertainty on days in which large amounts of spinning reserve are not available online, to address the risk of net-load ramp, or when there is a limited amount of capacity available for SCED. Historically, the need for NSRS has occurred during hot or cold days with unexpected changes in weather or following large unit trips to replenish reserves.

III. INERTIA TREND AND PRIMARY FREQUENCY CONTROL AT ERCOT

As more VERs are integrated to the grid, the system inertia could decline. This section presents an overview of inertia trend and how the primary frequency control is coordinated at ERCOT.

A. INERTIA TREND AT ERCOT

The power system's primary frequency response performance is highly dependent on the system inertia and how the primary frequency control is activated.

Rotating turbine generators and motors that are synchronously connected to the system store kinetic energy. In response to a sudden loss of generation, kinetic energy will automatically be extracted from the rotating synchronous machines causing the machines to slow down as the system frequency is decaying. Inertia response provides an important contribution to reliability in the initial moments following a

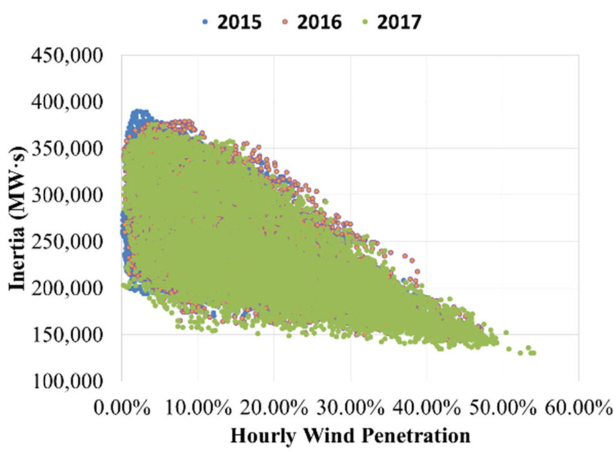

FIGURE 2. Inertia vs. hourly wind penetration for 2015-2017.

generation or load trip event, and determines the initial rate of change of frequency [9]–[11].

The amount of the system inertia depends on the number and size of generators and motors synchronized to the grid. It is difficult to account for the contribution of motor loads to the system inertia as the statuses of motors are unknown to system operators [19]. Therefore inertia response of motor load can be included into load damping constant.

The system inertia from all online synchronous generators is calculated as

$$M_{sys} = \sum_{i \in I} H_i \cdot MVA_i \tag{1}$$

where H_i and MVA_i is the inertia constant and installation MVA capacity of synchronous machine i, respectively, and I is the online synchronous unit set.

The system inertia at ERCOT decreases over the years as shown in Fig. 2, which depicts the hourly system inertia and the corresponding wind penetration, i.e. the portion of load supplied by wind generation, between 2015 and 2017. The lowest inertia in each year has dropped from 152 GW· s in 2015 to 130 GW·s in 2017. The decrease in the inertia was partially attributed to the reduction in the net-load. If there is an abundant wind or solar generation coinciding with a low load condition, the wholesale energy market prices can be low or even negative. Under these circumstances, synchronous generators may be offline for economic reasons. As a result, a low net-load case could lead to fewer generators committed online, then reducing the system inertia.

B. OVERVIEW OF FREQUENCY CONTROL COORDINATION AT ERCOT

Figure 3 depicts an overview of how the primary frequency control is coordinated for a severe under-frequency event at ERCOT. When the frequency drop below 59.91 Hz, the RRS capacity carried by the generation units is released to SCED. After a delay of one minute, SCED will be re-run with the updated resource limits so that the base points for these RRS generation units will be increased to assist in restoring the frequency. If the frequency continues to drop below 59.8 Hz, the hydro resources operated as synchronous condensers will respond autonomously. If the frequency stays below 59.7 Hz, LRs will be tripped offline to arrest the decline of the frequency. Once the frequency is restored back above 59.98 Hz, the deployment of RRS will be recalled. At ERCOT, the amount of the hydro resources operated as synchronous condensers is relatively small compared to the generation units and LRs in the provision of RRS

C. RRS AT ERCOT BEFORE RE-DESIGN OF AS MARKET

Having a sufficient amount of RRS is critical to arrest frequency excursions within a few seconds following generation unit trips. Both the generators providing governor response and LRs with under-frequency relays are eligible to participate in RRS market. The following describes their features in the provision of primary frequency control.

When RRS is provided by online synchronous generators through governor response or governor-like actions to arrest frequency deviations, it is termed as a PFR service. As a single BA, ERCOT must comply with the BAL-003 standard. The frequency response obligation for ERCOT is 413MW/0.1Hz. To meet this requirement, ERCOT requires every resource with a speed governor to put the governor in service whenever the resource is online. In addition, the droop setting should not exceed 5% and the frequency response dead band should not be no more than ±0.018 Hz.

LRs on under-frequency relay (UFR) can also provide RRS if they can be self-deployed to provide a full response within 30 cycles after the frequency meets or drops below certain threshold (59.7Hz). LRs on UFR are equipped with an under-frequency relay to arrest the quick frequency decline. As required by ERCOT, the response time for LRs should be less than 500 ms (including the frequency relay pickup delay and the breaker action time). This makes the response of LRs more effective to mitigate the decline of frequency compared to the generators because a generator needs a few seconds to react to the change in the frequency to provide the primary frequency response. Therefore, the deployment of LRs is able to improve the frequency nadir and is instrumental in preventing frequency from dropping below the involuntary UFLS threshold when losing large generation units.

Historically, ERCOT has relied more on PFR and LRs to protect the grid against the large disturbances. Nowadays, there is an increasing interest in deploying energy storage resources in a large scale in the future power grid. This opens up a new opportunity to utilize these fast-acting resources to enhance the primary frequency control performance. ERCOT is in the process of developing new market rules to allow batteries to respond at 59.85 Hz to provide a FFR reserve, which is described in details in the next Section..

IV. NEW ANCILLARY SERVICE MARKET

As the system inertia is declining, fast frequency response can significantly improve the frequency control performance.

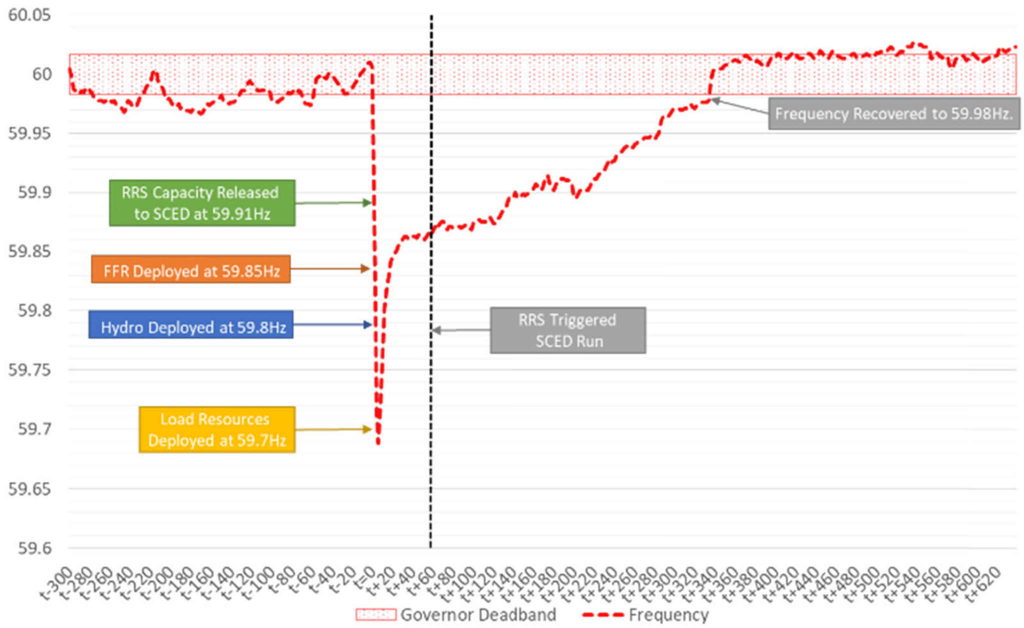

FIGURE 3. Coordination of primary frequency control at ERCOT.

On the other hand, the past market rules at ERCOT required that RRS should be deployed within 10 minutes upon the receipt of the deployment instructions. This requirement does not explicitly incentivize the resources which can quickly deliver the primary frequency response before the frequency nadir (point C). In addition, this could create unnecessary barriers for emerging technologies like batteries to participate in the RRS market. To this end, it is imperative to re-design the AS market, especially RRS, in order to improve both economic and reliable operations for a future power grid with a large amount of VERs.

The fundamental principles to guide the design of such new AS market are given as follows.

1) To meet NERC reliability requirement so as to maintain a satisfactory frequency control performance
2) To allow batteries to be awarded FFR to improve the efficiency of the AS market and reduce the operational cost
3) To accommodate the characteristics of a variety of new market entrants
4) To coordinate the deployment actions of AS products at different time scales.

Based on these principles, a new AS market framework has been proposed and is being implemented at ERCOT, which consists of four AS products, i.e., regulation service, RRS (new), ERCOT contingency reserve service (ECRS) and NSRS – see Fig. 4.

The key factor distinguishing these new AS products is still the response time required for each reserve. Regarding regulation service, there is no change proposed to its qualifications and deployment mechanism.

Within this new framework, RRS (old) will be unbundled into two products: RRS (new) and ECRS. RRS (new) includes three subsets: PFR, LRs and FFR. All of three RRS subsets are procured with an intention to arrest large frequency excursions following a generation trip. More detailed description of FFR will be presented in Section V.

ECRS is a new AS product introduced to restore RRS (new) responsibility once RRS resources are depleted or to mitigate a reliability concern if there is a deficiency in the ramping capacity. By design, ECRS can be dispatched by SCED and should respond within 10 minutes to the deployment instructions.

ECRS is required to meet NERC BAL-002 (to recover frequency within 15 minutes). The minimum amount of ECRS is determined in two steps. In the first step, dynamic simulations are performed to identify the frequency response (point B) when a single largest unit outage causes the frequency nadir (point C) to drop at 59.7 Hz (LRs will not be tripped in this case). Figure 5 shows different point B as a function of the system inertia.

In the second step, to fulfill the obligation of restoring point B frequency to above 59.98 Hz within the 10 minutes, the amount of ECRS required is calculated by

$$p_{ECRS} = \left(59.98 - f_{BPoint}^{M_{sys}}\right) \cdot p_{Load} \cdot \frac{10}{100} \tag{2}$$

where $f_{BPoint}^{M_{sys}}$ is the B point frequency for the system inertia (M_{sys}) and p_{Load} is the load demand.

The qualification criterion and the response time for NSRS will remain unchanged. However, the minimum amount of NSRS required will be modified. Before the new AS design

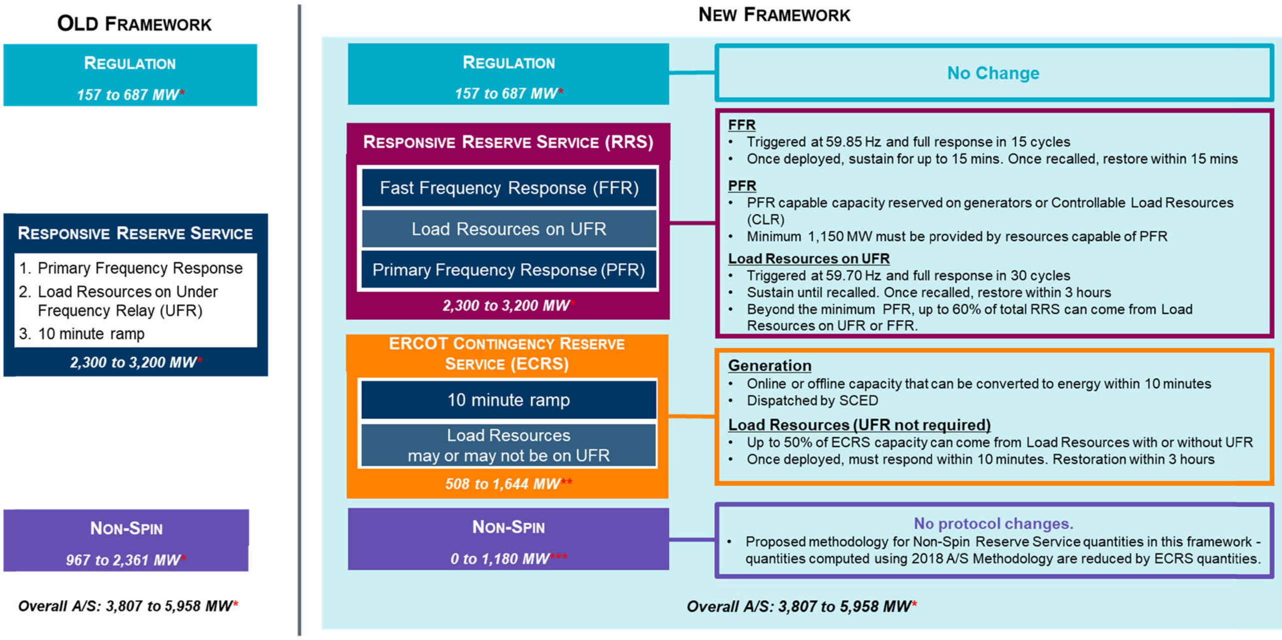

FIGURE 4. New AS products at ERCOT.

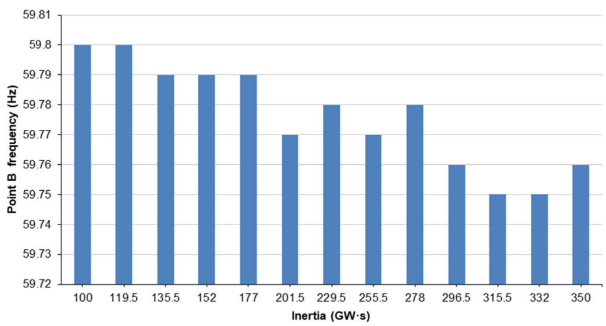

FIGURE 5. Point *B* vs. inertia as a result of a single largest unit outage.

is in effect, the amount of NSRS needed is given by

$$p_{NSRS} = \Omega\left(\varepsilon_{load,3h} - \varepsilon_{wind,3h} - \varepsilon_{solar,3h}\right) \tag{3}$$

where $\varepsilon_{load,3h}$, $\varepsilon_{wind,3h}$, $\varepsilon_{solar,3h}$ is the 3-hour ahead load, wind and solar generation forecast error, respectively, Ω is the percentile function applied to the net-load forecast error data set $\left(\varepsilon_{load,3h} - \varepsilon_{wind,3h} - \varepsilon_{solar,3h}\right)$ [33].

After the new AS design is implemented, the amount of NSRS needed will be

$$p_{NSRS} = max(0, \Omega\left(\varepsilon_{load,3h} - \varepsilon_{wind,3h} - \varepsilon_{solar,3h}\right) - p_{ECRS}) \tag{4}$$

Equation (3) represents the way how NSRS was determined before ECRS is introduced. Since ECRS is a 10-minute AS product, it can substitute NSRS (30-minute AS product) in order to meet NERC BAL-002 requirement. Therefore, the minimum amount of NSRS needed will be significantly reduced after the introduction of ECRS as indicated by (4).

V. FAST FREQUENCY RESPONSE (FFR)

To be qualified for the provision of FFR, a resource should be able to be automatically deployed and provide its full response within 15 cycles after the frequency meets or drops below a preset threshold (59.85 Hz) or be deployed via a verbal dispatch instruction (VDI) within 10 minutes. FFR resources must sustain a full response for at least 15 minutes once deployed. When a resource providing RRS as FFR is deployed, it shall not recall its capacity until the system frequency is greater than 59.98 Hz or they have been sustainably deployed over 15 minutes. Once recalled, the resources providing FFR must restore their full FFR responsibility within 15 minutes after the cessation of deployment or as otherwise directed by ERCOT.

Table 1 shows a comparison of key features between FFR and LRs. In comparison to LRs, FFR will be deployed earlier and faster when a loss of generation event happens. Earlier response from FFR will aid in preserving LRs providing RRS for more severe events. With a trigger set at 59.85 Hz, FFR will deploy more frequently than LRs in the response to the under-frequency events. As FFR resources provide a frequency response, most frequency events will not trigger the LR deployment, thus preserving the frequency response capabilities from LRs to be available for the next severe event.

Shorter restoration time for FFR resources also limits ERCOT's exposure (i.e. inability to respond) to next event with a similar magnitude. FFR resources upon the deployment completion will reset themselves and become available to respond to another event within 15 minutes. In contrast, LRs providing RRS are allowed to reset themselves with 3 hour after a RRS deployment and become available to respond to another event. If a frequency event triggers

TABLE 1. Comparison between FFR and LR.

	FFR	LR
Frequency Trigger	59.85 Hz	59.7 Hz
Response Time	15 cycles	30 cycles
Sustained Deployment Period	15 minutes	1 hour
Restoration Time	15 minutes	3 Hours
Deployment during EEAs	VDI	VDI

the LR providing RRS to be deployed, for the 3-hour duration following such an event, ERCOT may not have adequate frequency responsive resources to respond to another large disturbance.

During energy emergency alert (EEA) conditions, FFR resources may be deployed through VDI. As required, FFR resource should not withdraw energy from the grid until the EEA event has ended.

The response time and the frequency trigger are two key design parameters for the FFR reserve, which not only impact the overall frequency control performance but also influence the market liquidity and efficiency. The determination of these two parameter should consider the performance trend of the ERCOT grid to ensure that FFR resources are effective in protecting the system frequency. In addition, the capabilities of the new technologies should be taken into account in order to attract as many participants in the RRS market as possible. Figure 6 shows the historical trend of ERCOT system described by the magnitude of the power losses and the corresponding frequency nadir point. A selection of the frequency trigger at 59.85 Hz can protect against the large disturbances without a need to activate LRs while avoiding the unintended responses to numerous insignificant events. The response time required for FFR is 15 cycles, which is attainable by fast-acting energy storage resources or fast protection relays/breakers.

A. QUALIFICATIONS OF FFR AND PERFORMANCE EVALUATIONS

ERCOT will qualify resources that can provide FFR via a qualifications test. A resource must meet the following two requirements in order to pass the test:

1) A resource must respond within 10 minutes of receiving test instruction;
2) A resource's response must be within 95% and 110% of the minimum between the RRS obligation and the maximum deployment allowed to respond.[2]

ERCOT may revoke a resource's FFR qualification if it has two performance failures during actual deployments (manual or frequency triggered) within a rolling 365-day period. The

[2]The maximum deployment allowed to respond equals to telemetered high sustained limit (HSL)–low sustained limit (LSL) for modeled generation resource (generators and batteries) or maximum power consumption (MPC)-low power consumption (LPC) for modeled load resource.

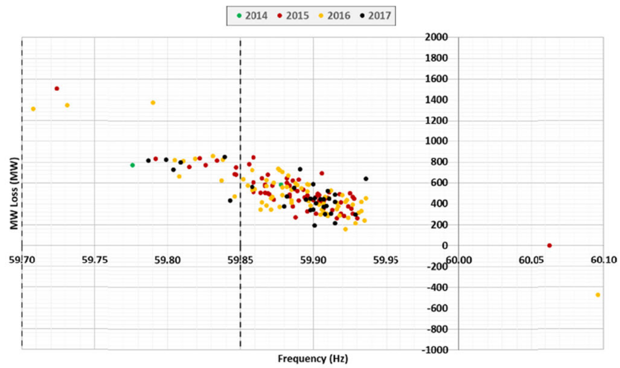

FIGURE 6. MW loss vs. frequency nadir at ERCOT.

actual performance of a FFR resource is evaluated using following metric:

- A FFR resource must be deployed in 15 cycles (or 10 minutes for verbal deployment) after the frequency reaches the trip threshold;
- A resource must sustain the response for at least 15 minutes or till ERCOT recalls deployment, whichever occurs first;
- A resource must be reset and made available for next event within 15 minutes after the deployment is ended.

For each FFR deployment event, the following data will be collected for evaluating FFR response performance:

- High speed event data from the FFR resources that are not deployed via breaker action;
- High speed event data from the recorders at the primary and back-up facilities;
- High speed event data from phasor measurement units available to ERCOT;
- Telemetry data for all resources providing FFR during the event;
- Recording of frequency and power output with a resolution of no less than 30 samples per second.

B. TELEMETRY DATA REQUIREMENT FOR DEPLOYMENT AND RECALL OF FFR

The FFR resources respond autonomously to the local frequency excursions so that no centralized mechanism is needed to activate the deployment of the FFR. However, certain coordination is necessary when the deployment of FFR resources is recalled in order not to harm the frequency recovery. One example of telemetry data for a FFR resource is given in Fig. 7. Once the frequency is below 59.85 Hz, ERCOT will send the FFR resource the deployment instruction so that the FFR resource will update its RRS schedule (RRSC) and inform ERCOT the new RRSC, which makes the reserved FFR capacity available to SCED. After the execution of next SCED is completed, ERCOT will provide the FFR resource a new updated desired base point (UDBP), ramping from pre-event level to the full deployment. However, the actual FFR deployment has occurred prior to

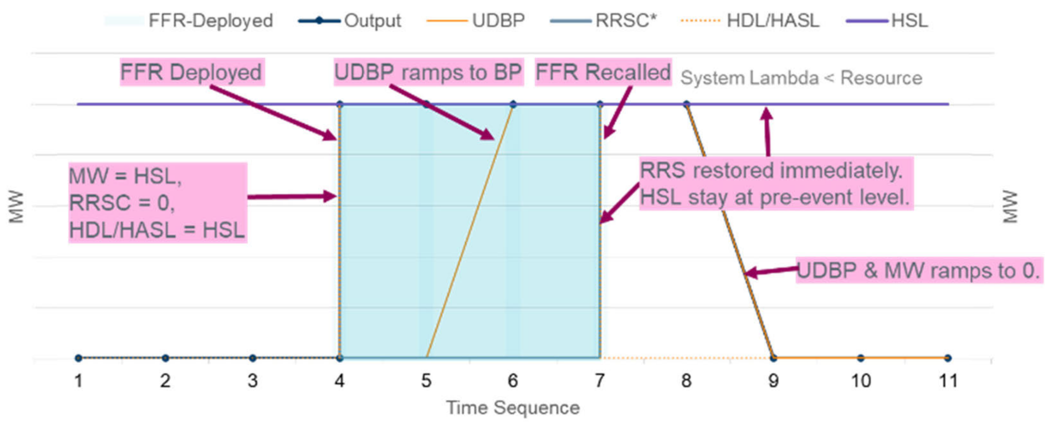

FIGURE 7. Telemetry example for deployment and recall of FFR (for illustration only) (UDBP: updated desired base point, RRSC: RRS schedule, HSL: high sustainable limit, HDL: high dispatch limit, HASL: high ancillary service limit).

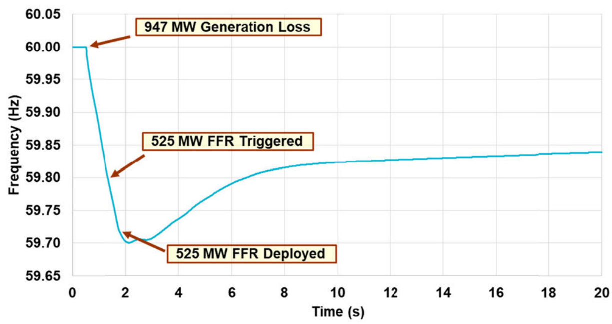

FIGURE 8. Simulated frequency response for a 947 MW of generation loss.

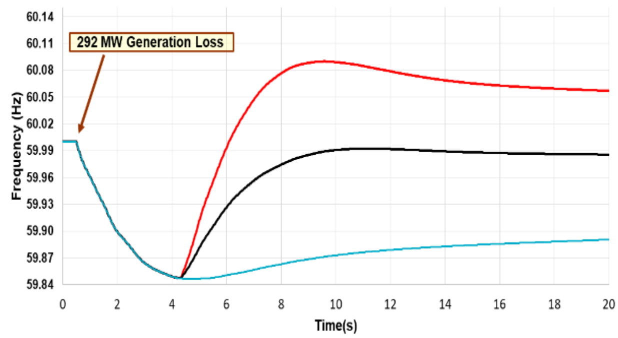

FIGURE 9. Frequency overshoot due to deployment of FFR (red: 420 MW FFR, black: 225 MW FFR, blue: 0 MW FFR).

this so that the FFR resource is not required to follow UDBP when deployed. After the cessation of 15 minutes, the FFR resource is recalled so that its HSL telemetry data returns back to pre-event level. The next SCED run will update the UDBP, ramping down to 0 MW, which the FFR resource must follow. In this way, a FFR resource gradually and smoothly reduces its output to 0 MW once recalled.

VI. MAXIMUM AMOUNT OF FFR ALLOWED

The deployment of FFR could quickly arrest the frequency decline. One example is given in Fig. 8, which simulates the frequency response of the ERCOT grid when a generation producing a 947 MW of power output is tripped offline. A full-detailed ERCOT dynamic model is used in the dynamic simulation, which is composed of over 500 generators and 10,000 buses. For this particular disturbance, 525 MW of FFR resources were deployed and as a result, the frequency nadir has been prevented from dropping below 59.7 Hz. This clearly demonstrates the effectiveness of FFR in avoiding the activation of the first-stage UFLS.

However, a reliability concern could arise if an excessive amount of FFR resources is deployed, leading to a potential frequency overshoot. Figure 9 depicts the overshoot of frequency due to activation of different amount of FFR when the system inertia is 100 GW·s (100 GW·s is the minimum system inertia requirement for ERCOT). In this case, 292 MW of generation was lost at 0.5 second, causing the frequency to drop below 59.85 Hz. If 420 MW of FFR has responded to this, the frequency will recover more quickly but the highest frequency can reach as high as 60.10 Hz. In the case when an excessive amount of FFR reacts to this even, this post-disturbance over-frequency response may cause another need to deploy regulation-down resources in order to bring the system frequency back to 60 Hz.

However, when the system inertia is extremely low, over-generation may likely happen, i.e., thermal units may have to be operated close to their minimum output limit and thus have a limited capability of reducing their power production in the response to an over-frequency event. To mitigate this concern, ERCOT currently limits the maximum system-wide FFR responsibility to 420 MW.

VII. BENEFITS OF FFR

The earlier and faster response of FFR can benefit the grid's reliability and operational efficiency. In particular, a new approach is presented in this Section, which can qualify the

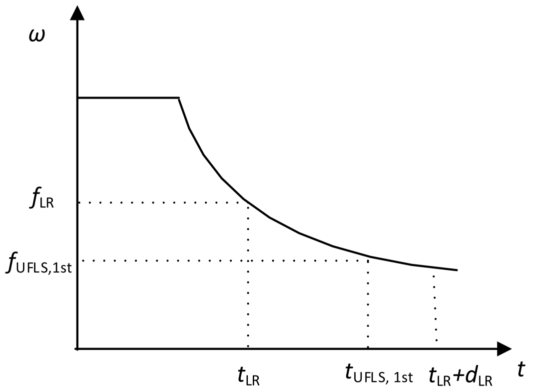

FIGURE 10. Frequency response when $t_{LR} + d_{LR} > t_{UFLS,1st}$ (t_{LR} and $t_{UFLS,1st}$ are when the frequency reaches 59.7 Hz and 59.3 Hz, respectively).

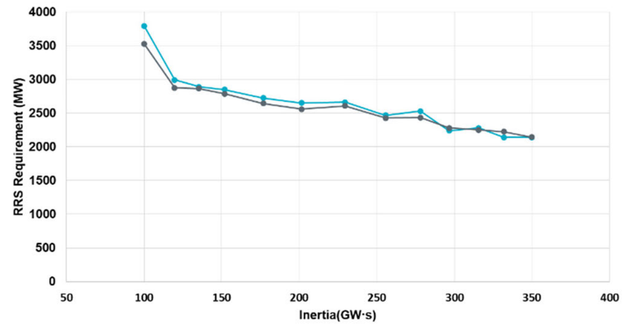

FIGURE 11. Reduction in RRS requirement due to inclusion of FFR (blue: PFR+LR, black: PFR+LR+FFR).

benefits of FFR to the grid operations in two terms: 1) the reduction in critical inertia and 2) RRS cost saving.

A. IMPACT OF FFR OVER CRITICAL INERTIA

In a low inertia grid like ERCOT, the grid frequency could decline very quickly following a resource trip. LRs begin to react once the frequency drops below their triggering frequency, f_{LR} (59.7 Hz). However, it takes a delay, d_{LR} (30 cycles), for LRs to completely open their breaker. During this period of d_{LR}, the system frequency will continue to decrease. If the rate of change of the frequency is very high, there could be a possibility that the frequency can drop below the trigger of first-stage UFLS (59.3 Hz) before the breaker of LRs can be fully opened. This scenario can be depicted in Fig. 10 and be mathematically described as

$$t_{LR} + d_{LR} > t_{UFLS,1st} \tag{5}$$

To prevent this from happening, LRs need to have sufficient amount of time to respond to the frequency excursions. One remedy to this is to maintain a minimum amount of inertia (critical inertia), M_{min}, on the grid, which ensures that the condition $t_{LR} + d_{LR} < t_{UFLS,1st}$ is satisfied. When the system inertia is above this minimum inertia level, M_{min}, LRs are still effectively tripped offline to arrest the frequency drop. Otherwise, the frequency will drop below the first-stage UFLS triggering frequency before LRs can provide sufficient frequency response, following the worst contingency, i.e., the simultaneous loss of two largest online generators.

Through the dynamic simulation, the critical inertia at ERCOT is estimated to be 100 GW·s when FFR resources are not considered.[3] With the appropriately selected trip settings stated in Section V, FFR can help to reduce this critical inertia. If 420 MW of FFR can be fully deployed with 15 cycles, it can reduce the critical inertia from 100 GW·s to 88 GW·s. This is because an earlier and faster response from FFR can decrease the rate of change of the frequency, and thus provide LRs more time to deliver their full response. A reduction in the critical inertia can facilitate more renewable resources to be reliably integrated in the ERCOT grid. Otherwise, thermal units have to be uneconomically committed to maintain the system inertia above the critical inertia even though the abundant energy could be produced from wind or solar renewable resources.

B. RRS COST SAVING WITH FFR RESOURCES

RRS is procured to ensure that the sufficient capacity is available to respond to frequency excursions once a unit trips. Prior to the next operational year, ERCOT sets the minimum RRS requirement for the expected grid operations, which varies by hour of the day and by month [31]. One key factor influencing how much RRS is needed is the system inertia. This dependence is considered here when deciding the need for RRS. The basic approach to determine RRS requirement for each hour of the next year consists of two steps: 1) to project the system inertia conditions for the next year by creating a time-series data for 8760 hours, and 2) to map from the projected inertia condition to the RRS need. This two-step approach is briefly described as follows and more details can be found in [31].

The first step is to calculate future inertia conditions for each four-hour interval of the next year. This is based on expected diurnal load and wind/solar patterns for the same hour block of the same month in the past 2 years. A percentile is then applied to this data set to produce a future inertia condition for each 4-hour block of the same month for the next year.

The second step is to determine the minimum amount of RRS requirement based on the future 8760-hour inertia conditions, through a look-up table. Without including FFR, the RRS need corresponding to an inertia level for twelve different cases is derived from dynamics studies in which the loss of two largest online units is simulated– see the blue line in Fig. 11. It can be seen that the RRS need decreases dramatically as the system inertia increases. This is because when there are more generators online under heavy loading conditions, i.e., the overall system inertia is higher, the rate of change of frequency following the disturbance is much

[3]Additionally, simulated conditions show the wide-area voltage oscillations at inertia below 100 GW·s.

smaller than in a low-inertia condition. As a result, the aggregated slow-acting governor-like PFR response has enough time to react to the loss of generation.

After 420MW of FFR resources are used, the total RRS requirement will be reduced, as shown in the black line in Fig. 11. This decrease in the RRS requirement is larger at the low system inertia levels and becomes diminished as the system inertia increases.

In order to evaluate the potential benefits of FFR, it is estimated that if the new AS market were in effect in 2019, and FFR offers of 420 MW were procured in each hour, there may be a lower quantity of RRS procured in 7,272 hours out of 8,016 hours for January through November 2019. The reduction in RRS quantities would vary between 0 MW and 81 MW during these 8,016 hours resulting in an overall reduction during this period of 244,712 MWh in required RRS quantities.

The historical RRS market clearing prices for capacity (MCPCs) from January 1, 2018 through November 30, 2018 is used as a proxy to estimate the magnitude of the financial impact of the decrease in the minimum RRS requirements. Assuming 244,712 MWh in required RRS quantities for January through November 2019, the reduction in total RRS quantity by hour is multiplied by 2018 hourly MCPCs for RRS results in an estimated savings of $3,426,088. The estimated cost savings is expected to increase in the future as ERCOT experiences higher frequency of lower inertia periods, and procures higher quantities of RRS.[4]

VIII. CONCLUSION

AS market awards those services critical to maintain the reliability and security of the power systems through a market mechanism. Currently, the power industry is undergoing a major transformation. The increasing penetration of VERs will dramatically change the reliability need of a future power grid and thus calls for re-design of AS market. On the other hand, it is desirable to achieve a balance between reliability and economics while opening up the AS market to both traditional and non-traditional market participants.

This paper presents a new AS framework being implemented at ERCOT in order to address the primary frequency control issues associated with the declining system inertia. The essential innovation is to divide RRS into three sub-categories (PFR, LRs on UFR and FFR) so that it explicitly awards those resources which can be quickly deployed to arrest the frequency drop following the loss of the large online generation units. To meet NERC BAL-002, ECRS is also introduced to recover frequency within 15 minutes. This new design makes it suitable for energy-limited resources like batteries to offer FFR in the AS market.

In addition, the detailed description of designing key performance metrics and qualifying the benefits of FFR was provided in the paper. It was found that the earlier and faster response of FFR can bring significant benefits to the grid's reliability. Introduction of FFR can reduce the critical inertia from 100 GW·s to 88 GW·s and the estimated cost saving amounts to $3,426,088 if the new design were implemented for January through November 2019. On the other hand, if an excessive amount of FFR resources is deployed, it could lead to a potential frequency overshoot. Therefore, the maximum system-wide FFR responsibility at ERCOT is limited to 420 MW.

The experiences gained at ERCOT can be beneficial to other regions which also face the similar challenges when dealing with a high penetration of VERs. Future work will focus on the study and analysis of the performance of FFR and its impact over the market operations and grid reliability.

[4]In addition, with an increase in the quantity of resources qualified to provide RRS and able to submit offers for RRS, it is likely the clearing price of RRS will decrease.

REFERENCES

[1] *Promoting Wholesale Competition Through Open Access Nondiscriminatory Transmission Services by Public Utilities*, FERC, document 888, 1996.

[2] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. Hoboken, NJ, USA: Wiley, 2003.

[3] E. Litvinov, F. Zhao, and T. Zheng, "Electricity markets in the united states: Power industry restructuring processes for the present and future," *IEEE Power Energy Mag.*, vol. 17, no. 1, pp. 32–42, Jan. 2019.

[4] G. Yann Rebours, S. Daniel Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary services—Part I: Technical features," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 350–357, Feb. 2007.

[5] E. Ela and A. Tuohy, "The evolution of ancillary services to facilitate integration of variable renewable energy resources: A survey of some changes to the ancillary services and ancillary service markets," Electric Power Res. Inst., Washington, DC, USA, Tech. Rep. 3002008987, 2016.

[6] Z. Zhou, T. Levin, and G. Conzelman, *Survey of US Ancillary Services Markets*, Argonne National Lab (ANL), Lemont, IL, USA, 2016.

[7] P. Du, R. Baldick, and A. Tuohy, *Integration of Large-Scale Renewable Energy into Bulk Power Systems*. Springer, 2017.

[8] J. Matevosyan and P. Du, "Wind generation scheduling and coordination in ERCOT nodal market," in *Integration of Large Scale Renewable Energy into Bulk Power System: From Planning to Operation*. 2017.

[9] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.

[10] J. H. Eto, "Use of frequency response metrics to assess the planning and operating requirements for reliable integration of variable renewable generation," Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep., 2011.

[11] *NERC Frequency Response Initiative*, Apr. 2010.

[12] M. Dreidy, H. Mokhlis, and S. Mekhilef, "Inertia response and frequency control techniques for renewable energy sources: A review," *Renew. Sustain. Energy Rev.*, vol. 69, pp. 144–155, Mar. 2017.

[13] *Essential Reliability Services and the Evolving Bulk-Power System-Primary Frequency Response*, FERC, document 842, 2018.

[14] J. O'Sullivan, A. Rogers, D. Flynn, P. Smith, A. Mullane, and M. O'Malley, "Studying the maximum instantaneous non-synchronous generation in an island system—Frequency stability challenges in Ireland," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2943–2951, Nov. 2014.

[15] A. S. Ahmadyar, S. Riaz, G. Verbic, A. Chapman, and D. J. Hill, "A framework for assessing renewable integration limits with respect to frequency performance," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4444–4453, Jul. 2018.

[16] M. Yu, A. J. Roscoe, A. Dyśko, C. D. Booth, R. Ierna, J. Zhu, and H. Urdal, "Instantaneous penetration level limits of non-synchronous devices in the British power system," *IET Renew. Power Gener.*, vol. 11, no. 8, pp. 1211–1217, Jun. 2017.

[17] Y. Liu, S. You, J. Tan, Y. Zhang, and Y. Liu, "Frequency response assessment and enhancement of the US power grids toward extra-high photovoltaic generation penetrations—An industry perspective," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3438–3449, May 2018.

[18] S. You, Y. Liu, J. Tan, M. T. Gonzalez, X. Zhang, Y. Zhang, and Y. Liu, "Comparative assessment of tactics to improve primary frequency response without curtailing solar output in high photovoltaic interconnection grids," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 718–728, Apr. 2019.
[19] P. Du and J. Matevosyan, "Forecast system inertia condition and its impact to integrate more renewables," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1531–1533, Mar. 2018.
[20] R. Doherty, A. Mullane, G. Nolan, D. J. Burke, A. Bryson, and M. O'Malley, "An assessment of the impact of wind generation on system frequency control," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 452–460, Feb. 2010.
[21] T. Xu, W. Jang, and T. Overbye, "Commitment of fast-responding storage devices to mimic inertia for the enhancement of primary frequency response," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1219–1230, Mar. 2018.
[22] G. Delille, B. Francois, and G. Malarange, "Dynamic frequency control support by energy storage to reduce the impact of wind and solar generation on isolated power System's inertia," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 931–939, Oct. 2012.
[23] A. Molina-García, F. Bouffard, and S. Daniel Kirschen, "Decentralized demand-side contribution to primary frequency control," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 411–419, Feb. 2011.
[24] R. Doherty, G. Lalor, and M. O'Malley, "Frequency control in competitive electricity market dispatch," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1588–1596, Aug. 2005.
[25] E. Ela, V. Gevorgian, A. Tuohy, B. Kirby, M. Milligan, and M. O'Malley, "Market designs for the primary frequency response ancillary service—Part I: Motivation and design," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 421–431, Jan. 2014.
[26] E. Ela, V. Gevorgian, A. Tuohy, B. Kirby, M. Milligan, and M. O'Malley, "Market designs for the primary frequency response ancillary service—Part II: Case studies," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 432–440, Jan. 2014.
[27] W. Li, P. Du, and N. Lu, "PFR ancillary service in low-inertia power system," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 5, pp. 920–930, Mar. 2020.
[28] W. Li, P. Du, and N. Lu, "Design of a new primary frequency control market for hosting frequency response reserve offers from both generators and loads," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4883–4892, Sep. 2018.
[29] C. Liu and P. Du, "Participation of load resources in day-ahead market to provide primary-frequency response reserve," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5041–5051, Sep. 2018.
[30] N. Miller, D. Lew, and R. Piwko, "Technology capabilities for fast frequency response," GE Energy Consulting, Tech. Rep., 2017.
[31] P. Du and W. Li, "Frequency response impact of integration of HVDC into a low-inertia AC power grid," *IEEE Trans. Power Syst.*, early access, Apr. 27, 2020, doi: 10.1109/TPWRS.2020.2990304.
[32] P. Du, H. Hui, and N. Lu, "Procurement of regulation services for a grid with high-penetration wind generation resources: A case study of ERCOT," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 16, pp. 4085–4093, Dec. 2016.
[33] N. Rajbhandari, W. Li, P. Du, S. Sharma, and B. Blevins, "Analysis of net-load forecast error and new methodology to determine non-spin reserve service requirement," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2016, pp. 1–5.

PENGWEI DU (Senior Member, IEEE) received the Ph.D. degree in electric power engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently with Electric Reliability Council of Texas, Austin, TX, USA. Prior to this, he was a Senior Research Engineer with the Pacific Northwest National Laboratory, Department of Energy. He was a recipient of the IEEE PES Power System Dynamic Performance Committee Best Paper Prize in 2016 and the IEEE PES Best Paper Prize. He is also a Subject Editor of the *IET Generation Transmission and Distribution* and an Associate Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS.

NITIKA V. MAGO received the B.E. degree from the Mumbai University's Veermata Jijabai Technological Institute (VJTI), Mumbai, India, in 2005, and the M.S.E.E. degree from The University of Texas at Austin, Austin, TX, USA, in 2007. She is the Manager of Balancing Operations Planning with Electric Reliability Council of Texas (ERCOT), Taylor, TX.

WEIFENG LI (Member, IEEE) received the B.S. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2012, and the M.S. and Ph.D. degrees in electrical engineering from North Carolina State University, Raleigh, NC, USA, in 2015 and 2018, respectively. He is currently with Electric Reliability Council of Texas, Austin, TX, USA, and was a recipient of the IEEE PES Best Paper Prize. His research interests include renewable integration, ancillary services, and advanced data analytics for power grid operation and planning.

SANDIP SHARMA received the M.S.E.E. degree in electrical engineering specializing in power systems from the University of Texas at Arlington in 2006. He is the Director of Market Operations with Electric Reliability Council of Texas (ERCOT).

QINRAN HU (Member, IEEE) received the B.S. degree in electrical engineering from Chien-Shiung Wu Honors College, Southeast University, Nanjing, China, in 2010, and the M.S. and Ph.D. degrees in electrical engineering from the University of Tennessee, Knoxville, TN, USA, in 2013 and 2015, respectively. He was a Postdoctoral Fellow of Harvard University, Cambridge, MA, USA, from 2015 to 2018. He joined the School of Electrical Engineering, Southeast University, in October 2018. His research interests include power system optimization, demand aggregation, and electricity market.

TAO DING (Senior Member, IEEE) received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 2009 and 2012, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2015. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN, USA. He is currently an Associate Professor with the State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China.