

Improving frequency response with synthetic damping available from fleets of distributed energy resources

Hani Mavalizadeh Luis A. Duffaut Espinosa Mads R. Almassalkhi

Abstract—With the increasing use of renewable generation in power systems, responsive resources will be necessary to support primary frequency control in future low-inertia/under-damped power systems. Flexible loads can provide fast-frequency response services if coordinated effectively. However, practical implementations of such synthetic damping services require both effective local sensing and control at the device level and an ability to accurately estimate online and predict the available synthetic damping from a fleet. In addition, the inherent trade-off between a fleet being available for fast frequency response while providing other ancillary services needs to be characterized. In this context, the manuscript presents a novel, fully decentralized, packet-based controller for diverse flexible loads that dynamically prioritizes and interrupts loads to engender synthetic damping suitable for primary frequency control. Moreover, the packet-based control methodology is shown to accurately characterize the available synthetic damping in real-time, which is useful to aggregators and system operators. Furthermore, spectral analysis of historical frequency regulation data is used to produce a probabilistic bound on the expected available synthetic damping for primary frequency control from a fleet and the trade-off from concurrently providing secondary frequency control services. Finally, numerical simulation on IEEE test networks demonstrates the effectiveness of the proposed methodology.

Index Terms—Decentralized control, distributed energy resources, fast frequency response, packet-based energy management, frequency responsive loads, synthetic damping.

I. INTRODUCTION

The penetration of renewable energy sources (RES) such as wind and solar is increasing rapidly as a part of the global effort to reduce greenhouse gas emissions. However, increased RES use leads to more variability in electricity generation due to RES' intrinsic uncertainty. In addition, replacing conventional synchronous generators with inverter-based renewable generation reduces the inertia, i.e., the power system's ability to oppose changes in frequency [1]. A significant deviation from nominal frequency can lead to an outage of generation units which subsequently causes further frequency deviation and, in severe cases, can result in a total system blackout. A recent example was the Texas blackout in 2021, which led to approximately 155 billion dollars in loss [2]. Therefore corrective actions are crucial, to control the power system frequency. Frequency control mechanisms include primary, secondary, and tertiary frequency control. Primary frequency control also called fast frequency response is largely automatic and instantaneous and occurs over the first few seconds following a grid disturbance event. The secondary frequency control brings the frequency back to the nominal value by adjusting

the output of generating units within a few minutes after a frequency event. Tertiary frequency control restores the power reserve of the generators used for the secondary frequency control [3]. In this paper, the focus is on primary control since secondary and tertiary control do not significantly influence the transient frequency dynamics.

Distributed energy resources (DER) are widely considered an effective and scalable way to provide primary frequency control [4], [5]. DER coordination can be used to provide synthetic damping, which is defined as the percentage change in the total DER consumption in response to frequency change [1], as well as inertia, improving the stability of the power system. Coordinating DERs to provide primary frequency control has been studied for many years [6], [7], [8], [9]. One of the first works on frequency responsive loads was presented in 1980 called frequency adaptive power and energy re-scheduler (FAPER) [6]. In this method, the dynamic state of thermostatically controlled loads (TCLs) is used to prioritize TCLs for frequency response. For example, devices with high temperatures will be prioritized to be turned off during an under-frequency event. In addition, the bound on temperature are frequency-dependent meaning that for higher frequency deviations, more devices participate in the frequency response. Probabilistic FAPER was introduced in [7] by injecting random delays in devices switching on/off, which helped address synchronization concerns with FAPER, i.e., avoided large groups of DERs attaining nearly the same temperature and, thus, responding nearly identically and causing large power swings.

Different types of DERs can be coordinated for primary frequency control. TCLs (e.g., electrical water heaters (EWHs) and refrigerators) can be turned off for short periods without a considerable effect on the temperature, which provides some flexibility used for a rapid change in load. Also, TCLs form a large portion of the power system load [8], and therefore, coordinating them provides considerable capacity for the power system operator. Another significant advantage of using TCLs is that they are highly responsive, making them an appropriate option for primary frequency control where fast response is required. In [10], smart EWHs are used to compensate for the uncertainty in wind and solar energy, peak shifting, and frequency response. A dynamic model is presented in [9] for different types of TCLs that adapt and improve a direct load control (DLC) scheme for primary frequency regulation in hybrid isolated microgrids. Frequent on/off switching of TCLs increases wear-and-tear and should be minimized during TCL coordination schemes as discussed in [11].

Different control architectures have been proposed for DER coordination, which differs in the level of communication requirements, quality of service (QoS), number of cycling, and level of grid awareness. In [12], a novel method is proposed that relies on transient phase offset to achieve a fast response to primary frequency control. The method enables the power system operators to use resources closer to the frequency deviation source. Such a technique also

The authors are with the Department of Electrical and Biomedical Engineering, University of Vermont, Burlington, VT 05405 USA (Email: {hani.mavalizadeh, lduffaut, malmassa}@uvm.edu). This work was supported by the U.S. Department of Energy's ARPA-E award DE-AR0000694. M. R. Almassalkhi is co-founder of startup company Packetized Energy, which has commercialized aspects related to PEM.

ensures that the stability of the system is preserved. In [13], a fully decentralized leaky integral controller for frequency restoration is presented. The use of decentralized control led to the elimination of communication delays and failures (e.g., lost messages) associated with centralized control schemes. Instead, [13] uses communication between local loads, which decreases the communication structure costs significantly. Other methodologies employ adaptive controllers that adapt to online measurements. One example of such techniques was introduced in [14], where an adaptive control framework is built based on the time-space distribution characteristics of the frequency in the power system. Also, the frequency response control is transformed from decentralized feedback control to centralized feed-forward control. Moreover, adaptive controllers have been shown to reduce problems with actuation delay. Finally, hierarchical optimization-based DER coordination schemes were developed in [15] with the advantage that AC network constraints can be considered. This is the so-called *grid-aware coordination*.

The authors have presented a fully decentralized proportional controller in [16] to provide synthetic damping from a TCL fleet. In this manuscript, the results from [16] are extended to arbitrary reference signals. While the proposed control scheme is tested on a timer-based prioritization scheme [17], it can be applied to other *fitness-based* DER prioritization schemes (e.g., [18]) as well. Each DER measures frequency locally and based on the designed control law, a frequency-dependent threshold on the timer is calculated. Based on the calculated threshold, devices determine whether to participate in primary frequency control or not. To the best of the author's knowledge, this is the first work to provide an analytical estimate of the amount of synthetic damping that can be extracted from a DER fleet. The main contributions of this manuscript are listed below:

- The decentralized energy packet interruption controller initially proposed in [16] is now generalized to consider bi-directional DERs and rate-of-change of frequency (RoCoF) to improve the fleet's response during either under- or over-frequency events.
- Using limited information available to the coordinator, the synthetic damping available from a fleet of packetized resources can be precisely estimated in real-time.
- Spectral analysis of historical AGC data is used to develop and compute a probabilistic lower bound on the expected synthetic damping available from a fleet. This lower bound can be used to analyze the trade-off between a fleet's expected primary and secondary frequency control capabilities.
- Simulation-based analysis is provided on practical considerations for packet-based DER coordination and synthetic damping, such as local measurement resolution and sensing/controller delays.

The remainder of the paper is organized as follows: a brief description of the dynamic grid model and packet-based coordination is provided in section II. The proposed control law is described in section III. In section IV the trade-off between FFR services and other ancillary services is characterized. Practical considerations are discussed in section V. Finally, section VI provides the conclusion.

II. PRELIMINARIES

In this section, the network model is presented and the concepts of packet coordination and packet interruption are provided.

A. Network Model

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph representing the topology of a transmission network, where $\mathcal{V} := \{1, \dots, N_n\}$, is the set of N_n nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of branches, such that if i and j are connected, then $(i, j) \in \mathcal{E}$. The frequency dynamics of the network are governed by the swing equations [19]:

$$\Delta \dot{\theta}_j = \Delta \omega_j, \quad (1a)$$

$$M_j \Delta \dot{\omega}_j = \Delta P_j^G - \Delta P_j^L - \Delta P_j^{\text{DER}} - D_j \Delta \omega_j + \sum_{i:(i,j) \in \mathcal{E}} \Delta P_{ij}, \quad (1b)$$

where θ_j and ω_j are the voltage angle and angular velocity at bus j , respectively, and ΔP_j^G , ΔP_j^L , and ΔP_j^{DER} are deviations in the generation, uncontrollable load, and controlled DER from nominal at bus j , respectively. Inertia [sec] and damping [MW/Hz] are described by M_j and D_j , respectively, while P_{ij} denotes the power flow [MW] between areas i and j , respectively. The j^{th} generator's turbine dynamics is modeled in (2),

$$\tau_j \Delta \dot{P}_j^G = -\Delta P_j^G - \frac{\Delta \omega_j}{R_j}, \quad (2)$$

where R_j is the generator's governor droop coefficient [Hz/MW] and τ_j is the turbine time constant [sec]. Synchronous generator droop controllers usually have a deadband of 36 mHz [20]. To model the DERs response, the basics of packet-based coordination are presented next.

B. Packet-based DER coordination

Packet-based DER coordination is enabled by a cyber-physical system that coordinates incoming and asynchronous discrete grid-access requests for energy from individual DERs [17], [21], [22], [23]. The DER's asynchronous requests are central to packet-based coordination and are explained next.

Definition 1. (*Energy packet*) An energy packet is a fixed-duration and fixed-power epoch of energy consumed (or delivered) by a DER.

In packet-based schemes, each DER requests an energy packet based on its need for energy (NFE). DERs considered in this manuscript represent residential water heaters, EV chargers and/or residential batteries, whose energy (e.g., temperature and SoC) dynamics are much slower than the frequency response. This means that during the primary frequency control period, the specific model of the DERs is not significant, as long as their power consumption is adjustable. For example, if an air conditioner (A/C) measures a room temperature in the summer above some desired set-point, then the room temperature is too high and the device's NFE increases. This leads to more frequent requests for energy packets to cool down the room. Similarly, if the A/C measures a low room temperature, the NFE decreases and the device will not request an energy packet often, if at all. The energy packet requests then arrive from devices asynchronously and each request is either accepted or rejected by the coordinator based on aggregate demand and a market or grid reference signal. When a request for an energy packet is accepted, an internal timer for the switched device is triggered and the DER charges or discharges until the timer's absolute value equals the packet length (or epoch length). The local timer for DER n is described as

$$t_n[k+1] = \begin{cases} t_n[k] + \Delta t, & \text{if } C_n[k] = 1 \\ t_n[k] - \Delta t, & \text{if } C_n[k] = -1 \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where Δt is the sampling time, the number of bins is $n_p := \lceil \delta / \Delta t_B \rceil$ and Δt_B is the timer bin width. The packet duration (epoch) is denoted δ and typically is set between 60 to 600 seconds. Without loss of generality, one can choose $\Delta t = \Delta t_B$. When the n^{th} DER has its charging request accepted at time k , then $C_n[k+1] = C_n[k+2] = \dots C_n[k+n_p] = 1$. On the other hand, if a discharging request packet is accepted at time k , $C_n[k+1] = C_n[k+2] = \dots C_n[k+n_p] = -1$. Otherwise, $C_n[k+1] = 0$. In fact, C_n is 1 when the device is charging, -1 when the device discharges, and 0 when the device is OFF.

Even though the coordinator does not have access to the individual DERs' internal timers, it knows how many requests were accepted at each time step in addition to packet height P_n^{rate} for each request, which permits the coordinator to construct an accurate estimate of the DER fleet's aggregate timer. In general, the coordinator needs to consider four different timers using (3); *i*) Charge-only timer which includes devices that only can be charged, such as TCLs and ACs; *ii*) Discharge-only timer which includes devices that can only be discharged, such as solar panels and stand-alone gen-sets; *iii*) Charging bi-directional devices, such as ESS; *iv*) discharging bi-directional devices. In this paper, the focus is on DER fleets of TCLs and ESSs, therefore, its three corresponding timers are (*i, iii, iv*). Since all accepted DERs start their packet at the first bin, the linear timer dynamics is defined by

$$\begin{aligned} x_{\text{tcl}}^{\text{ch}}[k+1] &= Mx_{\text{tcl}}^{\text{ch}}[k] + Bq_{\text{tcl}}^{\text{ch}}[k], \\ x_{\text{ess}}^{\text{ch}}[k+1] &= Mx_{\text{ess}}^{\text{ch}}[k] + Bq_{\text{ess}}^{\text{ch}}[k], \\ x_{\text{ess}}^{\text{dis}}[k+1] &= Mx_{\text{ess}}^{\text{dis}}[k] + Bq_{\text{ess}}^{\text{dis}}[k], \end{aligned} \quad (4)$$

where $x_{\text{tcl}}^{\text{ch}}, x_{\text{ess}}^{\text{ch}}, x_{\text{ess}}^{\text{dis}} \in \mathbb{R}^{n_p}$ are binned distributions of power for charge-only TCLs, bidirectional charging ESS, and bidirectional discharging ESS, respectively, while $q_{\text{tcl}}^{\text{ch}}[k], q_{\text{ess}}^{\text{ch}}[k], q_{\text{ess}}^{\text{dis}}[k] \in \mathbb{R}$ are the total power of accepted charging TCL, charging ESS, and discharging ESS requests during time step k , respectively. That is, $q_{\text{tcl}}^{\text{ch}}[k] = \sum_{n \in I[k]} P_n^{\text{rate}}$, where $I[k]$ is the set of DERs with accepted requests at time k . The timer dynamics are defined by $M \in \mathbb{R}^{n_p \times n_p}$, which is a lower triangular matrix with 1's on the lower off-diagonal and zero elsewhere, while $B := [1, 0, \dots, 0]^T \in \mathbb{R}^{n_p}$. Thus, when a request is accepted, the accepted DER enters the first bin, and at each time step it propagates through the timer. The number of devices completing their packets at time-step $k+1$ is equal to the number of devices in the last bin of the timer distribution. During a frequency event, the distribution can be considered constant because the timer states evolve slower than the grid frequency. That is, if a frequency event occurs at k , states $x_{\text{tcl}}^{\text{ch}}[k], x_{\text{ess}}^{\text{ch}}[k], x_{\text{ess}}^{\text{dis}}[k]$ can be assumed constant since frequency response is a fast event (i.e., $< 10\text{s}$).

It is clear that the timer states are a function of past coordinator packet acceptance rates. During the frequency event, packets actively participate in the frequency response based on their internal timer states. Therefore, the concept of packet participation is presented next by extending the packet interruption defined in [16].

Definition 2. (*Packet participation*) *The packet's participation is the forced change of DER n 's local state $C_n[k]$ before the end of its epoch length (i.e., $t_n < \delta$) due to a frequency deviation event.*

To explain the role of packet participation, first, consider the power draw for a general DER n . Its power consumption at time step k is $P_n[k] \in [\underline{P}_n, \bar{P}_n]$, where $\underline{P}_n = 0$ for TCLs and $= -P_n^{\text{cap}} < 0$ for (discharging) ESS and $\bar{P}_n = P_n^{\text{cap}} > 0$ when ON (TCL) or charging

(ESS). P_n^{cap} is the power rating of DER n . DER n is then participating in FFR, if, for example, during an under-frequency event, DER n changes its consumption from \bar{P}_n to \underline{P}_n . It is important to note that the formulation generalizes to the case when the DER reduces its power to a value larger than \underline{P}_n , as long as the coordinator is aware of the DER's available power change. If there is no ESS with $\underline{P}_n < 0$, packet participation refers to packet interruption as defined in [16] for TCLs only. The coordinator continuously monitors three distinct timers in real time. These timers encompass the binned power values of charging TCLs, charging bi-directional DERs, and discharging DERs, as illustrated in (4). By adding the values across the timer bins, the total power within each timer is obtained. To determine the total consumption of the entire fleet at time k , the total discharging power is subtracted from the total charging power as shown below,

$$P^{\text{DER}}[k] := \mathbf{1}_{n_p}^T (x_{\text{tcl}}^{\text{ch}}[k] + x_{\text{ess}}^{\text{ch}}[k]) + \mathbf{1}_{n_p}^T x_{\text{ess}}^{\text{dis}}[k]. \quad (5)$$

Based on measured grid frequency, the flexible (net) demand, P^{DER} , can then be actively modified via packet participation by selectively interrupting and/or "toggling" packets (e.g., charging at $P_n[k] = P_n^{\text{cap}} > 0$ toggles to discharging at $P_n[k] = -P_n^{\text{cap}} < 0$). The selection of which packets participate during any given frequency event will be based on a fully decentralized DER control law.

Remark. *Note that in some packet-based coordination schemes, devices can be interrupted before the completion of their packet to maintain quality of service, i.e. turned off in case of excessive temperature or turned ON in case of low temperature (also called opt-out) as discussed in [17]. In (4) the number of opt-outs is assumed negligible, which is reasonable when the DER fleet operates near nominal demand. The coordinator can further ensure that this assumption is valid by constraining operations to only track power reference signals close to its fleet's nominal power. That is, tracking a signal with a relatively large amplitude can cause the devices to deviate from the set point, which in turn leads to more opt-outs and interruptions.*

The proposed decentralized frequency control scheme is provided in the next section, where the DERs leverage information about their local timer state to participate in primary frequency control.

III. PROPOSED DECENTRALIZED CONTROL LAW

This section presents a fully decentralized control law, which prioritizes DER participation based on a local timer and dynamic state. The designed controller creates additional damping from DER fleet which is added to the conventional system damping, i.e., D_j in equation (1b).

A. Local DER control law

The overall layout of the proposed controller is shown in Fig. 1. A deadband with a size of f_{db} is defined such that if $|\Delta f[k]| < f_{\text{db}}$ DER does not respond to a frequency deviation. f_{db} is assumed 36 mHz to match a typical synchronous generator droop controller's deadband. A fleet's maximum participation is achieved at f_{max} when all of the active devices have participated in primary frequency control. η_{min} determines the portion of the timer that does not participate in FFR. In other words, devices will be locked for $\eta_{\text{min}}\delta$ seconds after their request is accepted. The proposed local control law is:

$$\eta_n[k] = \begin{cases} 1, & |\Delta f[k]| < f_{\text{db}} \\ g_2(K_P f^{\text{eff}}[k] + K_D \mathcal{D}(f^{\text{eff}}[k])), & f_{\text{db}} \leq |\Delta f[k]| \leq f_{\text{max}}, \\ \eta_{\text{min}} & |\Delta f[k]| > f_{\text{max}} \end{cases} \quad (6)$$

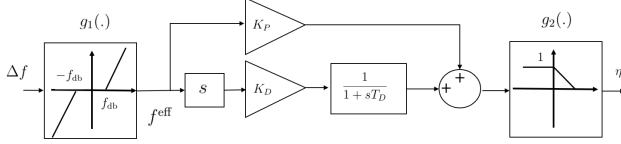


Fig. 1: Block diagram of proposed derivative-proportional control law. $g_1(\cdot)$ and $g_2(\cdot)$ are defined in (7)

where $K_P = 1/(f_{\max} - f_{\text{db}})$ and K_D are the design parameters representing proportional and derivative gains, respectively, and T_D is the derivative time constant. The derivative in the Laplace domain is denoted by s . $f^{\text{eff}}[k] := g_1(\Delta f[k])$ with $\Delta f[k]$ the deviation from nominal frequency (e.g., 60 Hz) and functions

$$g_1(x) := \begin{cases} 0, & |x| \leq f_{\text{db}} \\ |x| - f_{\text{db}}, & f_{\text{db}} \leq |x| \leq f_{\max} \end{cases}, \quad (7)$$

$$g_2(x) := \min\{\max\{-|x| + 1, \eta_{\min}\}, 1\}.$$

Note that $\mathcal{D}(f)$ is the backward discrete-time difference operator as expressed in (8). It estimates the rate of change of frequency (RoCoF) over the standard 500 ms window [24].

$$\mathcal{D}(f[k]) = \frac{f[k] - f[k - \alpha_w / \Delta t_s]}{\alpha_w} \quad (8)$$

where α_w is the window size in [sec]. Thus, each device calculates its η_n based on the locally measured frequency and participates in the frequency response, if $t_n[k]/\delta \geq \eta_n[k]$ for charging packets and $-t_n[k]/\delta \geq \eta_n[k]$ for discharging packets. Immediately after a typical frequency event, the frequency deviation is zero while the magnitude of RoCoF is largest (i.e., $\mathcal{D}(f^{\text{eff}})$ is a monotonic function and $\mathcal{D}(f^{\text{eff}})$ approaches 0 exponentially). Therefore, the aggregate power response is initially due to the differential term $K_D \mathcal{D}(f^{\text{eff}})$ in (6). However, since $\mathcal{D}(f^{\text{eff}}) \rightarrow 0$ exponentially (top plot in Fig. 2), the proportional term $K_P f^{\text{eff}}$ becomes dominant, resulting in a linear decrease in aggregate power with respect to the frequency deviation. Finally, after the frequency reaches its nadir point and starts to recover, no more DERs participate, as shown in the bottom plot of Fig. 2. As illustrated in Fig. 2, the local control law can effectively coordinate packet participation at scale to improve the frequency response with higher K_D values leading to more responsive (and aggressive) DER participation. Some remarks on controller tuning are presented in Subsection IV-D. The next subsection makes use of the timer definition and the proposed control law to determine the available synthetic damping in real-time.

B. Real-time estimation of damping

Since the coordinator determines how many devices are accepted during each time step and the packet height, i.e., P_n^{rate} is known for any packet request, $x_{\text{cl}}^{\text{ch}}[k]$, $x_{\text{ess}}^{\text{ch}}[k]$, and $x_{\text{ess}}^{\text{dis}}[k]$ can be accurately estimated by the coordinator in (effectively) real-time. Furthermore, to overcome any inaccuracies associated with the communication or actuation delays, the coordinator can use feedback in the form of a simple acknowledgment sent (asynchronously) from each device when its operating state transitions. In addition, the frequency of the system can be measured by the coordinator, and from (6), a single

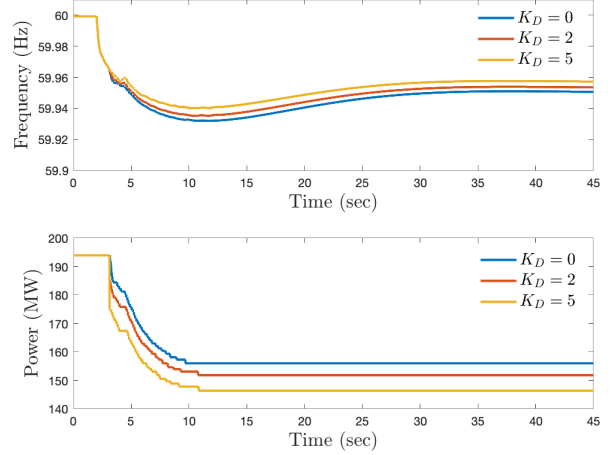


Fig. 2: Frequency response of 200,000 TCLs for different values of K_D .

$\eta[k]$ can be calculated for the entire fleet in real time. Moreover, given that the number of ON devices in each timer bin is known, the coordinator can then determine the available load reduction in response to frequency deviation without the need for additional communication with devices. For example, (9) determines the amount of available power for an under-frequency event.

$$\Delta P^{\text{DER}}[k] = \sum_{i=1}^{\tilde{K}} (x_{\text{icl}}^{\text{ch}}[k - \delta / \Delta t + i - 1] + 2x_{\text{ess}}^{\text{ch}}[k - \delta / \Delta t + i - 1]), \quad (9)$$

where $\tilde{K} := \lfloor \eta[k] \delta / \Delta t \rfloor$. Thus, from the aggregate fleet power and any potential system frequency event (i.e., a deviation with nadir Δf_{nadir}), the coordinator can simply and, in real-time, estimate the available synthetic damping from a DER fleet as

$$D^{\text{syn}}[k] = \frac{\Delta P^{\text{DER}}[k]}{\Delta f_{\text{nadir}} - f_{\text{db}}}. \quad (10)$$

As seen in Fig. 2, after reaching the nadir frequency, the frequency begins to recover, which results in an increase in $\eta_n[k]$ according to (6). However, it is important to highlight that even as $\eta_n[k]$ increases, DERs that have already been interrupted will not switch back on again. That is, the drop in load is sustained. Therefore, the amount of damping added from the fleet of DERs is determined by the nadir frequency, as described in (10). Fig. 3 illustrates the accuracy of the synthetic damping estimate compared with actual damping provided by the DER fleet for 10 different frequency events. The top figure shows the change in power versus the change in frequency for one of these realizations.

Calculation time to find synthetic damping for a fleet of 200,000 DERs from Eq. (6), (9) and (10) takes less than 500 μs which is acceptable for real-time applications. This allows the coordinator to accurately and continuously estimate the available synthetic damping across a number of what-if scenarios (e.g., different frequency nadir and ROCOF pairs). Clearly, 100s of these calculations could be executed every 60 seconds to provide the coordinator/ISO with an accurate and real-time estimate of available synthetic damping capability from a fleet of DERs. The

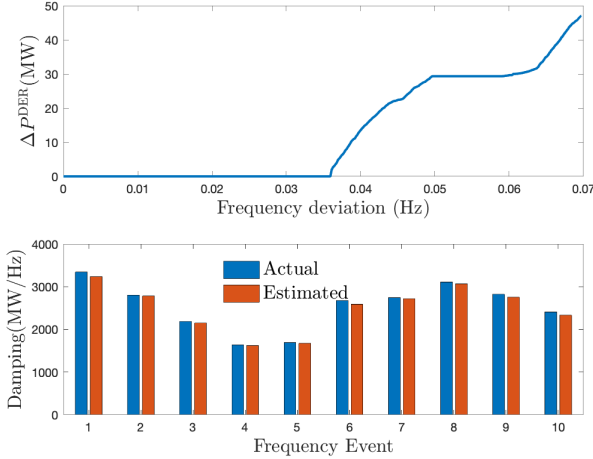


Fig. 3: (Top) A single frequency event with $\Delta f_{\text{nadir}} = 0.07\text{Hz}$ yields a large load reduction in the aggregate DER fleet as a function of the frequency deviation, packet timer distribution, and designed control law. (Bottom) Comparing the actual damping with the real-time estimate of synthetic damping for 10 different frequency events.

ISOs could then use these data/capabilities to evaluate stability margins/contingencies/ancillary services needs/etc.

In this section, a method is provided for the coordinator to accurately estimate the synthetic damping available from a fleet of DERs in any real-time operating condition (i.e., with an arbitrary, but known timer distribution). However, the coordinator may want to know a day- or hour ahead how much synthetic damping will be available from the fleet, in which case the timer distribution is unknown. Thus, in the next section, a probabilistic prediction of the available synthetic damping from a fleet of DERs participating in frequency regulation (e.g., PJM’s Reg-D) is developed that captures a range of operating conditions via the amplitude of the Reg-D regulation signal. The method in [16] is generalized to a bidirectional fleet by leveraging a specific packet-acceptance policy from [25] to ensure a unique mapping between the reference signal and the power-timer distribution. This enables an analytical characterization of synthetic damping statistics. It also can toggle the load between charging and discharging modes which double the synthetic damping available. Consequently, this permits us to analyze the trade-off between expected primary (damping) and secondary frequency control capabilities and (statistically) guarantee a DER fleet’s ability to deliver synthetic damping.

IV. CHARACTERIZING THE AVAILABLE SYNTHETIC DAMPING

Here, a probabilistic framework incorporating historical AGC data is used to characterize the available synthetic damping that can be guaranteed (i.e., a lower bound) from a fleet of DERs that are coordinated via ON/Charge/Discharge packets and providing a certain MW-level of frequency regulation (AGC) services. The process of mapping a representative AGC signal from the fleet’s timer distribution to the change in the fleet’s aggregate power due to a frequency event is outlined next. It is based on a spectral decomposition of historical AGC (PJM Reg-D) data and was inspired by [26], [27].

To map timer states to changes in power, ΔP^{DER} , for a given frequency event, the (conveniently) designed frequency-dependent timer threshold, η is leveraged.

Remark. The minimization policy, as presented in [25], aims to track the reference signal with the fewest number of DERs. This policy guarantees that the number of DERs in the timer is always lower compared to other control policies. Consequently, it establishes a lower limit on the flexibility (i.e., synthetic damping) achievable for a given fleet.

However, the shape of the timer distribution is unknown in advance as it depends on the fleet’s operating conditions (i.e., the reference signal and the number of available packet requests). Under the assumptions that (A1) a sufficient number of packet requests are available to the DER coordinator for effective aggregate power reference tracking (i.e., negligible tracking error); (A2) a fixed packet-request-acceptance policy (e.g. minimize the number of accepted packets) is adopted [25]; and (A3) the power reference signal is known ahead of time, then the exact timer distribution can be constructed over the duration of a packet epoch and the available synthetic damping can be estimated. However, if the coordinator wants to *predict* the available synthetic damping ahead of time (e.g., for possible FFR markets or planning studies), the exact AGC power reference will be unknown (i.e., assumption A3 will not hold), which implies that the timer distribution will be unknown. To overcome this challenge, historical AGC data is used to characterize the statistics of the timer distribution and thus to provide a probabilistic lower bound on the synthetic damping availability. Hence, the methodology for characterizing the synthetic damping availability from a fleet of DERs consists of the following steps:

- (A) Decompose historical AGC signals into its N most salient harmonics.
- (B) From the spectral decomposition and under assumptions A1 and A2, determine the statistics of the corresponding timer distribution.
- (C) Using the timer distribution statistics, determine a probabilistic lower bound on the number of packets in each timer.
- (D) Compute the probabilistic lower bound on the available synthetic damping from the fleet.

A. Spectral decomposition of AGC data

Following [27], where a clustering technique is able to categorize similar 2-hour samples of historical AGC data based on spectral analysis, a 2-hour AGC sample is considered. That is, the *methodology* is representative of a historical 2-hour AGC signal. However, the methodology can readily be applied to larger AGC data sets via methods presented in [27].

Consider the spectral (Fourier-based) decomposition of a 2-hour-long historical AGC sample signal of 2-second resolution into its N most salient harmonics as

$$AGC[k] \approx A \sum_{h=1}^N H_h[k], \quad (11)$$

where $H_h[k] := c_h \cos(2\pi h f_0 k - \phi_h)$. The coefficient c_h for the h -th harmonic is scaled between 1 and -1, while phase shift is denoted by ϕ_h , and f_0 is the signal’s fundamental frequency. A is the amplitude of the reference signal in MW. The reference signal is then defined as $P^{\text{ref}}[k] = P^{\text{nom}} + AGC[k]$ where P^{nom} is the power set-point that maintains the average SoC of the fleet stationary and $AGC[k]$ is obtained from (11). The goal is to map P^{ref} to the

coordinator's timer distribution (under assumptions A1 and A2). The procedure is detailed next for under-frequency events. The derivation for over-frequency events follows similarly. An expression for the total power of accepted requests at time k for each harmonic h , $q_h^+[k]$, must be found first. $q_h^+[k]$ determines the power in the first bin of the timer at time k for harmonic h . This is then used to find $q^+[k]$ which is the total number of accepted requests at time k .

It is convenient to decompose H_h into two functions an increasing function ($Y_h[k]$) and a decreasing function ($Z_h[k]$). That is, $H_h[k] := Y_h[k] - Z_h[k]$, where

$$Y_h[k] = \begin{cases} H_h[k], & \text{if } H_h[k] - H_h[k-1] > 0 \\ 0, & \text{otherwise} \end{cases} \quad (12a)$$

$$Z_h[k] = \begin{cases} -H_h[k], & \text{if } H_h[k] - H_h[k-1] < 0 \\ 0, & \text{otherwise} \end{cases}. \quad (12b)$$

Under assumption A1, $Y_h[k] - Y_h[k-1]$ and $Z_h[k] - Z_h[k-1]$ define, respectively, the net increase and decrease in DER aggregate power reference signal at each time step k . Now, if enough packet requests are assumed available, the coordinator can accept enough of them to match this increase or decrease. Thus, at each k , each of the reference signal's harmonics can be matched with enough number of accepted packet requests entering the coordinator's timer. Note that when the fleet includes both charging and discharging requests, the mapping becomes non-unique (since charging and discharging packets can effectively *cancel* each other out). To ensure a unique mapping between (harmonic) reference signals and accepted requests, a minimizing packet accepting the policy at the coordinator is employed that essentially does not select both charging and discharging requests at the same time [25]. Thus, under assumptions A1 and A2, $q_h^+[k]$ is written as,

$$q_h^+[k] = (f(Y_h[k] - Y_h[k-1]) - f(Z_h[k - n_p] - Z_h[k-1 - n_p])) + (f(Y_h[k - n_p] - Y_h[k-1 - n_p])), \quad (13)$$

where $f(x) = x$ for $x \geq 0$ and $f(x) = 0$, otherwise. In (13), $f(Y_h[k] - Y_h[k-1])$ determines the increase in the reference signal while the second and third terms determine the number of expired discharging and charging packets, respectively.

The next theorem characterizes the statistics of the charging timer for the minimization policy.

Theorem 1. Let δ , Δf , P^{cap} , and n_p , η_{min} , f_{max} , f_{db} and K_D be fixed for a given fleet under decentralized control policy (6). The mean and standard deviation of q^+ are given by:

$$\mathbb{E}(q^+) = n_u + \frac{N_d A \sum_{h=1}^N h c_h \Delta t}{T}, \quad (14)$$

$$\sigma^2(q^+) = \sum_{h=1}^N \left(\frac{1 - e^{-\frac{T^2}{6h^2}}}{2} (2\pi \Delta t f_h A c_h)^2 + \frac{2N_d A^2 h c_h^2 \Delta t}{T} \right), \quad (15)$$

where $n_u = P^{nom}/(n_p)$, $N_d = 2$ for ESS fleet and $N_d = 0$ for TCL fleet.

Proof. The proof is done by construction. The properties of Y_h and Z_h are utilized to simplify (13) and find the mean and standard deviation. From (12a) and (12b) and the definition of f , it can be seen that

$$f(Y_h[k] - Y_h[k-1]) = \begin{cases} 0, & k \in \frac{iT}{2} \quad i = \pm 1, \pm 2, \dots \\ Y_h[k] - Y_h[k-1], & \text{otherwise} \end{cases} \quad (16)$$

In the case of the ESS fleet, the above equation implies that there are two discontinuities in $f(Y_h[k] - Y_h[k-1])$ during a period. That is $N_d = 2$. In the case of a TCL fleet, since there are no discharging requests, $Z_h[k] = 0$, and $H_h[k] = Y_h[k]$, which leads to $N_d = 0$. Note also that $Y_h[k] - Y_h[k-1] = 0$ when $Y_h[k] = 0$ and in any other case

$$Y_h[k] - Y_h[k-1] = A c_h \cos(2\pi f_h k \Delta t) - A c_h \cos(2\pi f_h (k-1) \Delta t) \approx -2\pi f_h A c_h \Delta t \sin(2\pi f_h k \Delta t), \quad (17)$$

where $\omega_h = 2\pi f_h$. Using (16) and (17), one can find the mean and variance of q_h^+ directly from the mean and variance of $\sin(2\pi f_h k \Delta t)$, where $k \sim \mathcal{U}(0, 1/f_h)$ represents the random time of failure. The interest here is to find the mean and variance of $f(Y_h[k] - Y_h[k-1])$, which can be obtained using the corresponding mean and variance of $Y_h[k] - Y_h[k-1]$ given that these expressions only differ by N_d over each period. Recalling that for a set X_1, X_2, \dots, X_N of mutually independent normal random variables with corresponding means $\mu_1, \mu_2, \dots, \mu_N$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2$ one has that

$$Y = \sum_{h=1}^N c_h X_h \sim \mathcal{N} \left(\sum_{h=1}^N c_h \mu_h, \sum_{h=1}^N c_h^2 \sigma_h^2 \right), \quad (18)$$

and assuming that the dependence between harmonics of the AGC decomposition is negligible, then the mean and variance of the timer variable q^+ can be obtained by adding the mean and variance of each harmonic. In addition, the expected value of the power of the accepted requests corresponding to tracking P^{nom} is $n_u = P^{nom}/(n_p)$ [16]. Therefore, using (18), q^+ has a Gaussian distribution with mean and standard deviation given by (14) and (15). \square

B. Finding probabilistic lower bound on available damping

Theorem 1 is now used to compute a lower bound on the total power in each bin of the timer, P_{min} , analytically. That is, P_{min} is estimated to be

$$P_{min} = \mathbb{E}(q^+) - F \sigma(q^+). \quad (19)$$

where F is a *safety factor* determined by the operator. P_{min} estimates, at each time step, the minimum power inside the timer. Higher F leads to a more robust estimate but at the same time results in a more conservative estimate of the fleet's available damping. F is usually defined by the information available on the underlying distribution of the uncertainty, e.g., distribution, statistics, etc. If more information is available about the distribution of uncertainty, less conservative estimation can be made. The probability of being within F standard deviations is defined as $\rho := P(q^+[k] > \mathbb{E}(q^+) - F \sigma(q^+))$. If no information about the distribution of n_{min} is available (only mean and standard deviation are known), then using the Borel-Cantelli inequality [28], F is

$$F = \left(\frac{1 - \rho}{\rho} \right)^{\frac{1}{2}}. \quad (20)$$

The bounds obtained by Borel-Cantelli inequality are the worst-case scenarios and are unlikely to be encountered in practice. By assuming that there exists evidence that the distribution is unimodal, the Chebyshev generating function (CGF) can be used as shown below

$$F = \left(\frac{1-\rho}{e\rho} \right)^{\frac{1}{1.95}}. \quad (21)$$

[29]. Furthermore, if the distribution is Gaussian, the safety factor is found as follows:

$$F = \sqrt{2} \operatorname{erf}^{-1}(1-2\rho). \quad (22)$$

where ρ is the probability of violation of the lower bound on P_{\min} calculated by (19). By choosing the desired level for ρ and based on the level of information available to the coordinator, F is selected from the above equations.

In [16] a method to calculate the synthetic damping for a uniform timer distribution was presented. By assuming that all of the timer bins are at P_{\min} , one can use the method described in [16], to calculate a constant value for lower-bound damping. The next theorem is used to find the probabilistic lower bound on damping from P_{\min} .

Theorem 2. *Given Theorem 1, for a given contingency with known $\Delta f_{\text{nadir}} \in (f_{\text{db}}, f_{\text{max}})$ and RoCoF, the minimum damping for under-frequency events is,*

$$D^{\text{min}} = n_p P_{\min}^{\text{eff}} \left(K_P + \frac{K_D R^{\text{max}}}{\Delta f_{\text{nadir}} - f_{\text{db}}} \right). \quad (23)$$

where $P_{\min}^{\text{eff}} = P_{\min}$ for TCLs and $P_{\min}^{\text{eff}} = 2P_{\min}$ for ESS.

Proof. The proof is by construction. From (6), the total change in the fleet's aggregate power for a uniform distribution is:

$$\Delta P^{\text{DER}} = (1 - \eta_{\text{nadir}}) P_{\text{ON}}, \quad (24)$$

where η_{nadir} is the calculated η at nadir frequency. P_{ON} is the total power of ON devices and is calculated as $n_p P_{\min}$. In under-frequency events, charging devices participate in frequency response, and discharging devices do not participate. From control law in (6), (24) and the damping definition [1], the estimated damping is $D^{\text{min}} = \Delta P^{\text{DER}} / (\Delta f_{\text{nadir}} - f_{\text{db}})$, which leads to $n_p P_{\min} \left(K_P + \frac{K_D R^{\text{max}}}{\Delta f_{\text{nadir}} - f_{\text{db}}} \right)$ for $f_{\text{db}} \leq |\Delta f_{\text{nadir}}| \leq f_{\text{max}}$. As mentioned in III-A, the magnitude of $\mathcal{D}(f^{\text{eff}}[k])$ is at its maximum at the beginning of the disturbance and it decreases exponentially. Therefore, it is possible to replace $\mathcal{D}(f^{\text{eff}}[k])$ with the known maximum RoCoF of the event, R^{max} . Since ESS charging devices can be toggled to discharging during the frequency event, they can provide twice their capacity. This is captured by $P_{\min}^{\text{eff}} = P_{\min}$ for TCLs and $P_{\min}^{\text{eff}} = 2P_{\min}$ for ESS. If the frequency deviation is less than f_{db} , damping is zero as indicated by (6). Finally, if frequency deviation exceeds f_{max} , all of the available power is shed, resulting in $D^{\text{min}} = (P^{\text{cap}} n_p n_{\min}) / (\Delta f_{\text{nadir}} - f_{\text{db}})$. \square

Observe that from Eqs. (20), (21) or (22) the probability of violation of P_{\min} (ρ) can be found based on the level of information available about the timer distribution. To relate ρ to the probability of violation of D^{min} , the following remark is used.

Remark. *It is straightforward to show that the probability of violation of the calculated minimum damping in Theorem 2, is always smaller or equal to the probability of violation of P_{\min} (ρ).*

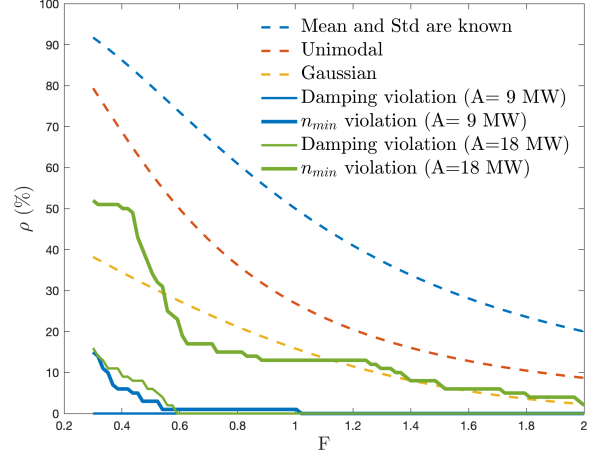


Fig. 4: Probability of violating the bounds

Fig. 4 shows the estimated probability of violating the bounds for different information available versus the actual percentage of violations. 9 and 18 MW are chosen for A which are equal to 5% and 10% of the nominal power. The green and blue curves show the percentage of violations obtained in simulations whereas the blue, red, and yellow dashed lines are obtained from Eqs. (20), (21) and (22), respectively.

Remark. *It should be noted that the lower bound on damping is derived under the assumption of negligible tracking error. Based on the previous work, an epoch length of 3 minutes or less satisfies this assumption [30]. For higher amplitudes of AGC, the tracking error increases since the reference signal has higher fluctuations around the nominal power. Therefore, for higher AGC amplitudes, because of the higher tracking error, the probability of violation of the calculated bounds is higher. This can be seen in Fig. 4 by comparing the violation probabilities for A being 5% and 10% of the nominal power.*

Fig. 5 shows the accuracy of the estimated damping versus the true damping for different amplitudes of AGC. A fleet of 200,000 EWHs with 4.5 kW capacity each is used in a two-area power system. The blue curve shows the mean of the true damping for 100 different realizations, while the green and purple curves show the mean minus 1 and 2 standard deviations, respectively. The dashed line indicates the estimated lower bound on damping calculated by (23). Similar results are given for an ESS fleet in Fig. 6 when $K_D = 5$. It should be noted that for each fleet, the reference signal is scaled around the fleet's nominal power. Therefore, the reference signal used to generate figures 5 and 6 is different which leads to a difference in the provided damping as seen in the figures.

To find the number of harmonics (N) required to reconstruct the AGC signal a compromise between computation burden and accuracy must be taken into account. The goal is to find a N that gives an acceptable reconstruction error for all of the 1-hour samples of the AGC data. To do so, the reconstruction error is calculated for all of the 1-hour samples in one year. The simulation results show that for $N = 23, 35, 53$ the RMSE of reconstruction error is always lower than 30%, 20%, and 10%, respectively. Therefore, by choosing $N = 53$, it can be guaranteed that the reconstruction error for any day of the year is lower than 10%. In this paper, $N = 100$ is

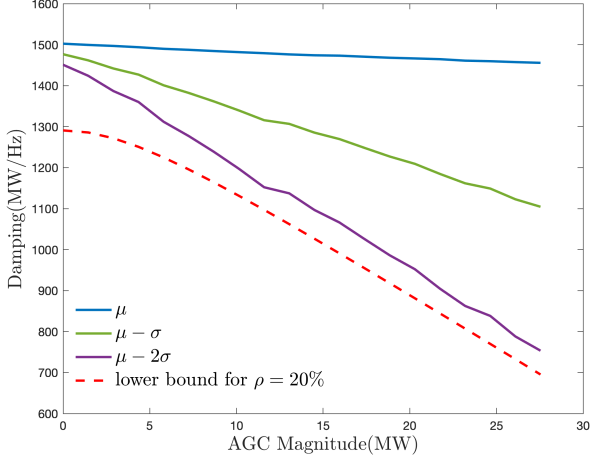


Fig. 5: The actual vs estimate damping lower bound for 200,000 TCLs

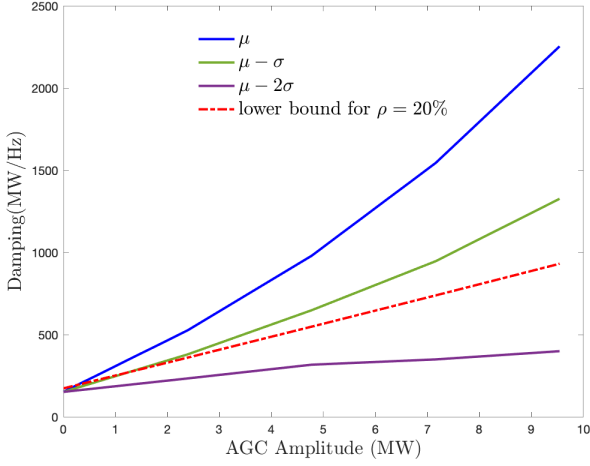


Fig. 6: The actual vs estimate damping lower bound for 200,000 batteries

chosen which limits the construction error for any given day to 6%.

C. FFR versus frequency regulation trade-off

From Theorem 1, (19) and (23), it can be seen that tracking a larger AGC signal (larger A) leads to higher variance. This always translates to lower n_{\min} and D^{\min} for TCL fleet. For the ESS fleet, since $\mathbb{E}(q^+)$ is a function of A , as mentioned in Theorem 1, higher A does not necessarily lead to lower synthetic damping. In this section, a procedure to determine the proper A , for a TCL fleet is presented to maximize the total profit. The same procedure can be applied to the ESS fleet, as well.

If the prices of frequency regulation is β^{Reg} [\$/MW] and FFR damping is β^{FFR} [\$/MW/Hz], then the total revenue can be written as $\beta^{\text{Reg}}A + \beta^{\text{FFR}}D^{\min}$. If one defines $O := A\beta + D^{\min}$, where $\beta := \beta^{\text{Reg}}/\beta^{\text{FFR}}$, then it is straightforward to show that maximizing O maximizes the total revenue. By replacing (14) and (15) in (19), one gets

$$P_{\min} = n_u - 2F\pi\Delta t f_0 A \sqrt{\sum_{h=1}^N h^2 c_h^2 \left(\frac{1 - e^{-\frac{\tau^2}{6h^2}}}{2} \right)} \quad (25)$$

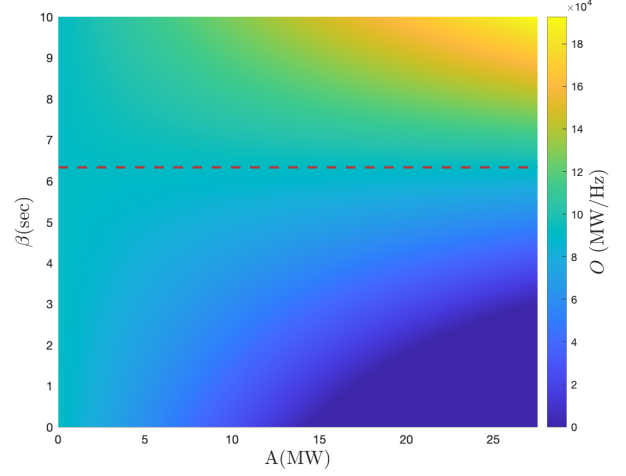


Fig. 7: The impact of ancillary service prices on the total normalized revenue, O [MW/Hz]. The red line indicates β^{thr} above which larger AGC magnitudes, increase revenue.

Now, substituting (25) in (23) gives

$$O = A\beta + n_p \left(n_u - 2F\pi\Delta t f_0 A \sqrt{\sum_{h=1}^N h^2 c_h^2 \frac{1 - e^{-\frac{\tau^2}{6h^2}}}{2}} \right) \cdot \left(K_P + \frac{K_D R^{\max}}{\Delta f_{\text{nadir}} - f_{\text{db}}} \right). \quad (26)$$

To maximize O in (26), the derivative with respect to A is calculated as $\frac{\partial O}{\partial A} = \beta - \beta^{\text{thr}}$, where

$$\beta^{\text{thr}} = n_p 2F\pi\Delta t f_0 \sqrt{\sum_{h=1}^N h^2 c_h^2 \frac{1 - e^{-\frac{\tau^2}{6h^2}}}{2}} \left(K_P + \frac{K_D R^{\max}}{\Delta f_{\text{nadir}} - f_{\text{db}}} \right). \quad (27)$$

From (27), it can be seen that β^{thr} is a function of frequency events. Then, for a set of credible system contingencies $\{\mathcal{C}_1, \dots, \mathcal{C}_{N_c}\} \in \mathcal{C}$, with known $\Delta f_{\text{nadir},c}$, R_c^{\max} and probability w_c , $\beta_1^{\text{thr}}, \dots, \beta_{N_c}^{\text{thr}}$ can be calculated from (27). Finally, a weighted average for β^{thr} is calculated as follows:

$$\overline{\beta^{\text{thr}}} = \sum_{c=1}^{N_c} w_c \beta_c^{\text{thr}}. \quad (28)$$

The normalized revenue, O [MW/Hz], is presented in Figure 7 for different β and A . Total revenue in dollars [\$] is obtained by $\beta^{\text{FFR}} O$. The following conclusions can be derived from (26), (27), and (28):

- 1) If $\beta \geq \overline{\beta^{\text{thr}}}$, $\frac{\partial O}{\partial A} > 0$, which means that to maximize the profit, the largest possible value for A must be chosen, subjects to the fleet being able to track without a considerable increase in the tracking error.
- 2) If $\beta < \overline{\beta^{\text{thr}}}$, $A^{\text{opt}} = 0$.

In the proposed method, DERs are used in two modes. When the frequency is close to nominal (i.e., $|\Delta f[k]| < f_{\text{db}}$) the fleet tracks the AGC signal, and when frequency deviation exceeds f_{db} fleet goes to primary frequency control mode. Larger AGC amplitudes mean that more resources are used for AGC tracking. It can be seen from Fig. 5 that for a TCL fleet, this leads to higher variation in

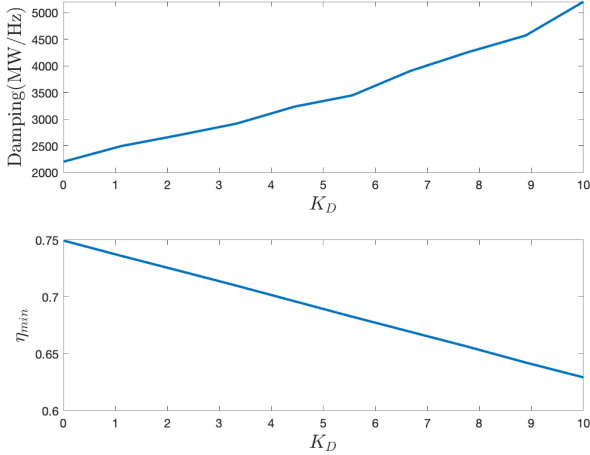


Fig. 8: Unused timer and minimum damping for different values of K_D .

damping which limits the ability to guarantee a minimum value for damping. Therefore, a fundamental trade-off between primary and secondary frequency control capability exists as seen in (25).

Another observation from (27) is that by using derivative control (i.e., increasing K_D), β^{thr} increases. This means that providing primary frequency control will be profitable for lower FFR prices (β^{FFR}).

D. Tuning of the controller parameters

The coordinator is assumed to only have access to its own DER information and not that of the grid operators or other coordinators. As such, the tuning of K_P, K_D is based on the coordinator's fleet information and published system-wide reliability metrics, such as frequency nadir and RoCoF. For credible contingencies in the system, the initial post-contingency rate of change of frequency (RoCoF) (R^{max}) and nadir frequency are considered by the coordinator to characterize the network's frequency response and the DER fleet's available capacity to respond. It might be necessary to lock the devices for a certain time after turning them ON for reliability issues or to avoid excessive switching. This can be done by setting $\eta_{\text{min}} > 0$. Using the designed controller, the constraint on η_{min} leads to the following inequality:

$$\begin{aligned} \eta &= 1 - \frac{\Delta f_{\text{nadir}} - f_{\text{db}}}{f_{\text{max}} - f_{\text{db}}} - K_D R^{\text{max}} \geq \eta_{\text{min}} \\ \Rightarrow K_D &\leq \frac{1}{R^{\text{max}}} \left(1 - \frac{\Delta f_{\text{nadir}} - f_{\text{db}}}{f_{\text{max}} - f_{\text{db}}} - \eta_{\text{min}} \right). \end{aligned} \quad (29)$$

In addition to (29), the coordinator needs to design K_P and K_D with the corresponding minimum damping from (23) in mind relative to a desired predefined damping value.

To illustrate these results, η_{min} and minimum damping are plotted versus K_D in Fig. 8 for the simulation setup of Fig. 2. For example, Fig. 8 depicts the relationship between η_{min} , K_D and expected available damping. When $\eta_{\text{min}} \leq 0.67$, then $K_D \leq 8.30$, which means that the expected available damping will be less than 4300 MW/Hz. The next section provides insight into practical considerations for packet-based DER coordination and primary frequency control via simulation-based analysis.

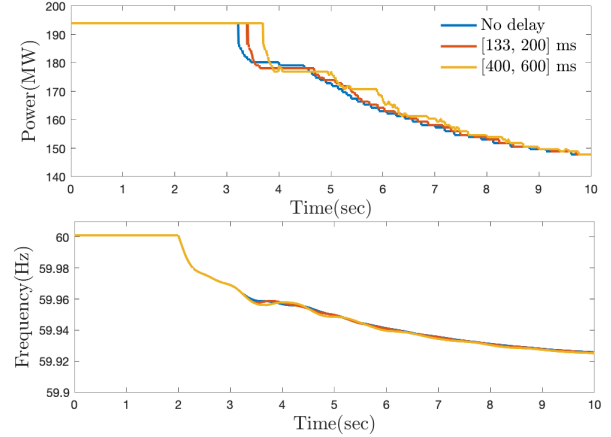


Fig. 9: The frequency response for different actuation delays for a population of 200,000 DERs. The disturbance occurs at $t=2$ s.

V. SIMULATION AND PRACTICAL CONSIDERATIONS

In this section, different practical considerations are tested to verify the performance of the proposed frequency-responsive controller for FFR. Different frequency measurement resolutions (between 1 mHz to 100 mHz) and random actuation delays (between 133 ms to 600 ms) are tested, and the effect on primary frequency control is analyzed. Moreover, simulation results are provided to determine how many DERs are needed to provide equal damping to an average droop-controlled generator in the network. The proposed decentralized controller is tested on the New England 39-bus system [31]. $K_D=2$ unless otherwise specified. A generation outage of 250 MW occurs at bus 30 at $t=2$ seconds. All of the DERs are connected to bus 20. The simulation setup is provided in table I.

TABLE I: Simulation Parameters

Parameter	Value
Fleet size	200,000
Δt	10 ms
$(\Delta f_{\text{db}}, \Delta f_{\text{max}})$	(36, 200) mHz
$T_n^{\text{min/set/max}}$	48.8/52.0/55.2 C°
(δ, MTTR)	(3,3) mins

A. Actuation delay

The effect of actuation delays is shown in Fig. 9. With no actuation delay, the fleet responds to frequency deviation immediately after frequency deviation exceeds f_{db} . While the delay slightly affects the transient behavior, the impact on final frequency (and damping) is negligible. The results illustrate the acceptable performance of the proposed approach even with a 400 to 600 ms delay which is the typical delay value in practice.

B. Frequency measurement resolution

Fig. 10 shows the impact of frequency measurement resolution on frequency response. The root mean square error (RMSE) of power interruption is 0.84 MW for 1 mHz resolution, 4.25 MW for 10 mHz resolution, and 34.78 MW for 100 mHz resolution. It

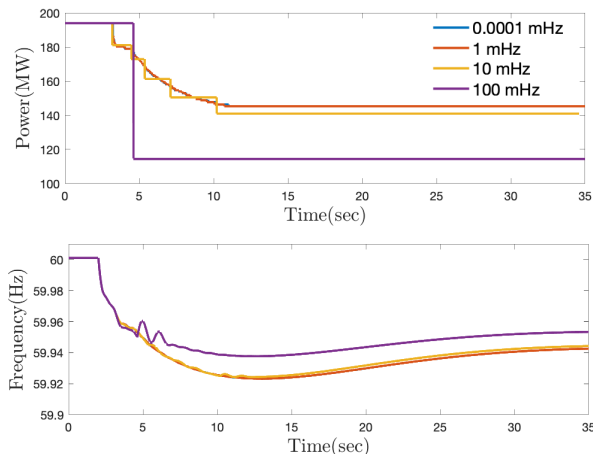


Fig. 10: The frequency response for different frequency measurement resolutions for a population of 200,000 TCLs.

can be seen that for 10 mHz frequency measurement resolution, the results are close to the actual values. Therefore, the effect of measurement inaccuracy can be neglected if the measuring devices' accuracy is at least 10 mHz.

C. Relating the scale of aggregate DER response

When coordinating aggregations of DERs, it is of value to understand how many DERs are needed to replicate the synthetic available from a realistic power plant. To answer this question, an experiment is performed in which a generator in the IEEE 39-bus test system is tripped at bus 30, disconnecting 250 MW power. Then, the frequency at bus 39 is measured in two cases: *i*) 1000 MW Generator at bus 39 with 5% droop coefficient and no responsive loads *ii*) deactivate the droop control at bus 39 and replacing it with 80,000 EWHs, each of which has a 4.5 kW power rating. The simulation results are presented in Fig. 11. The yellow curve shows the frequency response at bus 39 without droop control and without DER coordination. The red curve shows the frequency response with 5% droop control at bus 39, and the blue curve shows the frequency response when the droop controller is replaced with 80,000 EWHs with 4.5 kW power. The results show that 80,000 coordinated EWHs can provide a fast frequency response equivalent to a 1000 MW generation unit with a 5% droop coefficient.

VI. CONCLUSION

A fully decentralized frequency-based DER controller is designed and an analysis is presented that enables a DER aggregator to precisely estimate the synthetic damping available (online) from a fleet of aggregated DERs. To understand the impact of participating in ancillary services (i.e., frequency regulation) while also guaranteeing available synthetic damping from the fleet, a comprehensive analysis is provided to characterize a probabilistic lower bound on the available synthetic damping. This bound enables trade-off analysis between the fleet's ability to provide frequency regulation versus synthetic damping. Finally, practical considerations of the proposed decentralized control scheme are presented in a simulation-based study concerning the effects of actuation delays and frequency measurement resolutions. Future research directions include adapting the

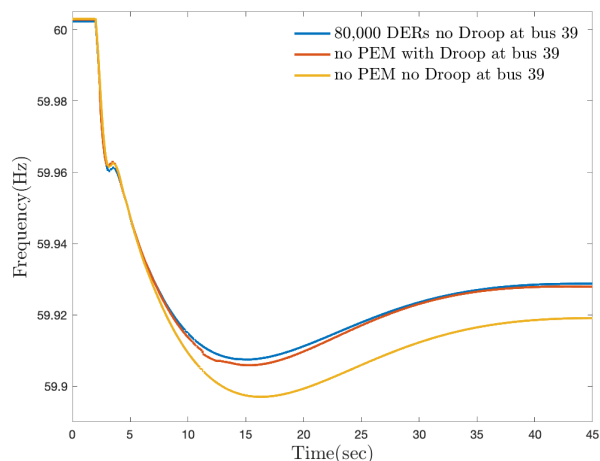


Fig. 11: Frequency response, with/without droop control at bus 39 and with/without DER coordination.

decentralized controller parameters based on spatial grid information to differentiate and prioritize certain locations/buses/feeders, as a way to incorporate the proposed synthetic damping with existing under-frequency load-shedding (UFLS) schemes. Another venue of interest represents the development of market mechanisms for incentivizing and valuing synthetic damping in low-inertia power systems.

REFERENCES

- [1] Electric Reliability Council of Texas, "Inertia: Basic concepts and impacts on the ERCOT grid," 2018.
- [2] J. W. Busby, K. Baker, M. D. Bazilian, A. Q. Gilbert, E. Grubert, V. Rai, J. D. Rhodes, S. Shidore, C. A. Smith, and M. E. Webber, "Cascading risks: Understanding the 2021 winter blackout in Texas," *Energy Research & Social Science*, vol. 77, p. 102106, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2214629621001997>
- [3] S. Zhang, H. Liu, F. Wang, T. Yan, and K. Wang, "Secondary frequency control strategy for BESS considering their degree of participation," *Energy Reports*, vol. 6, pp. 594–602, 2020, 2020 The 7th International Conference on Power and Energy Systems Engineering. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352484720316085>
- [4] L. Guo, C. Zhao, and S. H. Low, "Graph laplacian spectrum and primary frequency regulation," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 158–165.
- [5] B. K. Poolla, D. Groß, and F. Dörfler, "Placement and implementation of grid-forming and grid-following virtual inertia and fast frequency response," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3035–3046, 2019.
- [6] F. C. Schweppe, R. D. Tabors, J. L. Kirtley, H. R. Outhred, F. H. Pickel, and A. J. Cox, "Homeostatic utility control," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 3, pp. 1151–1163, 1980.
- [7] M. Brokish, "Adaptive load control of microgrids with non-dispatchable generation," Master's thesis, Massachusetts Institute of Technology, <https://dspace.mit.edu/handle/1721.1/53205>, 2009.
- [8] H. Zhao, Q. Wu, S. Huang, H. Zhang, Y. Liu, and Y. Xue, "Hierarchical control of thermostatically controlled loads for primary frequency support," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2986–2998, 2018.
- [9] W. Mendieta and C. A. Cañizares, "Primary frequency control in isolated microgrids using thermostatically controllable loads," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 93–105, 2021.
- [10] T. Clarke, T. Slay, C. Eustis, and R. B. Bass, "Aggregation of residential water heaters for peak shifting and frequency response services," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 22–30, 2020.
- [11] A. R. Coffman, N. Cammardella, P. Baroah, and S. Meyn, "Flexibility capacity of thermostatically controlled loads with cycling/lock-out constraints," in *2020 American Control Conference (ACC)*, 2020, pp. 527–532.
- [12] M. H. Syed, E. Guillo-Sansano, S. M. Blair, A. J. Roscoe, and G. M. Burt, "A novel decentralized responsabilizing primary frequency control," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3199–3201, 2018.

- [13] E. Weitenberg, Y. Jiang, C. Zhao, E. Mallada, C. De Persis, and F. Dörfler, "Robust decentralized secondary frequency control in power systems: Merits and tradeoffs," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 3967–3982, 2019.
- [14] C. Jin, W. Li, J. Shen, P. Li, L. Liu, and K. Wen, "Active frequency response based on model predictive control for bulk power system," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3002–3013, 2019.
- [15] M. Almassalkhi, S. Brahma, N. Nazir, H. Ossareh, P. Racherla, S. Kundu, S. P. Nandanoori, T. Ramachandran, A. Singhal, D. Gayme, C. Ji, E. Mallada, Y. Shen, P. You, and D. Anand, "Hierarchical, grid-aware, and economically optimal coordination of distributed energy resources in realistic distribution systems," *Energies*, vol. 13, no. 23, 2020. [Online]. Available: <https://www.mdpi.com/1996-1073/13/23/6399>
- [16] H. Mavalizadeh, L. A. Duffaut Espinosa, and M. R. Almassalkhi, "Decentralized frequency control using packet-based energy coordination," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2020, pp. 1–7.
- [17] M. Almassalkhi, L. A. Duffaut Espinosa, P. D. H. Hines, J. Frolik, S. Paudyal, and M. Amini, *Asynchronous Coordination of Distributed Energy Resources with Packetized Energy Management*. New York, NY: Springer New York, 2018, pp. 333–361.
- [18] S. P. Nandanoori, S. Kundu, D. Vrabie, K. Kalsi, and J. Lian, "Prioritized threshold allocation for distributed frequency response," in *2018 IEEE Conference on Control Technology and Applications (CCTA)*, 2018, pp. 237–244.
- [19] P. W. Sauer and M. A. Pai, *Power systems dynamics and stability*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [20] B. J. Kirby, "Frequency control concerns in the north american electric power system," U.S. Department of Energy, Office of Scientific and Technical Information, Tech. Rep., 2003. [Online]. Available: <https://www.osti.gov/biblio/885842>
- [21] L. A. Duffaut Espinosa and M. Almassalkhi, "A packetized energy management macromodel with quality of service guarantees for demand-side resources," *IEEE Transactions on Power Systems*, pp. 3660–3670, 2020.
- [22] S. Brahma, A. Khurram, H. Ossareh, and M. Almassalkhi, "Optimal frequency regulation using packetized energy management," *arXiv preprint arXiv:2107.12939*, 2021. [Online]. Available: <https://arxiv.org/abs/2107.12939>.
- [23] M. Almassalkhi, J. Frolik, and P. Hines, "Packetizing the power grid: The rules of the internet can also balance electricity supply and demand," *IEEE Spectrum*, vol. 59, no. 2, pp. 42–47, 2022.
- [24] North American Reliability Control Corporation, "Essential reliability services task force measures framework report," 2015.
- [25] L. Duffaut Espinosa, A. Khurram, and M. Almassalkhi, "Reference-tracking control policies for packetized coordination of heterogeneous DER populations," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 6, pp. 2427–2443, 2020.
- [26] S. Brahma, H. Ossareh, and M. Almassalkhi, "Statistical modeling and forecasting of automatic generation control signals," in *IREP Bulk Power System Dynamics and Control Symposium*, 2022.
- [27] R. Haffen, K. Subbaro, V. Viswanathan, and M. Kintner-Meyer, "Requirements for defining utility drive cycles: An exploratory analysis of grid frequency regulation data for establishing battery performance testing standards," University of Melbourne, Tech. Rep., 2011.
- [28] S. Bartolomeo, "Data-driven chance constrained optimization," Master's thesis, ETH-Zürich, <https://www.research-collection.ethz.ch/handle/20.500.11850/154768>, 2017.
- [29] B. Stellato, "Data-driven chance constrained optimization," Master's thesis, Autom. Control Lab., ETH-Zurich, 2014.
- [30] A. Khurram, L. A. Duffaut Espinosa, and M. R. Almassalkhi, "A methodology for quantifying flexibility in a fleet of diverse DERs," in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6.
- [31] A. Moeini, I. Kamwa, P. Brunelle, and G. Sybille, "Open data IEEE test systems implemented in simpowersystems for education and research in power grid dynamics and control," in *2015 50th International Universities Power Engineering Conference (UPEC)*, 2015, pp. 1–6.