DTU Library

Decentralized and discretized control for storage systems offering primary frequency control

Ziras, Charalampos; Prostejovsky, Alexander Maria; Bindner, Henrik W.; Marinelli, Mattia

Link back to DTU Orbit

Decentralized and Discretized Control for Storage Systems Offering Primary Frequency Control

Charalampos Ziras[a,*], Alexander Maria Prostejovsky[a], Henrik W. Bindner[a], Mattia Marinelli[a]

[a]*Center for Electric Power and Energy, DTU - Technical University of Denmark, Risø Campus, Roskilde, Denmark*

Abstract

The provision of ancillary services is an additional revenue stream for the owners of inverter-equipped storage systems, such as batteries and electric vehicles. As real demonstrations have shown, Primary Frequency Control (PFC) is a suitable and economically viable service for small-scale Energy Storage (ES) systems. This paper proposes a decentralized stochastic control policy, which significantly reduces ES units' losses when providing PFC. The proposed controller can be tuned to obtain the desired service reserve provision errors, while achieving a balance between tracking accuracy and efficiency. An extension of the algorithm significantly reduces the switching rate of the devices by up to 95 %. Analytical expressions for the reserve errors and the switching rates, dependent on the aggregation size and the controllers' settings, are derived and verified by simulations. Simulation results show that the proposed controller can significantly reduce ES units' losses when they are providing PFC by $8 - 15.5$ %, while achieving the expected tracking performance.

Keywords: batteries; decentralized control; electric vehicles; energy storage systems; primary frequency control; stochastic control

1. Introduction

The number of Energy Storage (ES) units in modern power systems, such as Electric Vehicles (EVs) or battery systems, has significantly increased due to their decreasing costs. Their aggregated power capabilities and energy capacity is con-

---

*Corresponding author

Email address: chazi@elektro.dtu.dk (Charalampos Ziras)

siderable, rendering them an attractive resource for ancillary services. Offering such services benefits not only power system operation but also the owners of such units, as additional revenues can be generated. One such service is Primary Frequency Control (PFC). There are different implementations of PFC depending on the specific power system. However, they all intend to contain system frequency deviations occurring from sudden mismatches between production and consumption. In the Regional Group Nordic area, the implementation of this service is called Frequency-Normal Operation Reserve (FNR). Various studies have shown the profitability of such services for battery systems and EVs, as well as the power system [1, 2, 3]. In the case of small-sized ES units (which are the focus of this work), commercial entities called aggregators contract and manage sufficiently large numbers of units to participate in the electricity and ancillary services markets.

A number of challenges arise from the fact that the primary purpose of ES systems is not to offer system-wide services, and they are not designed accordingly. Moreover, each ancillary service has different requirements, and thus, the desired characteristics for the service provider are not the same. For example, some devices may have minimum power requirements or high standby losses. Additionally, maximum efficiency may be achieved at a narrow operating regime and units may exhibit a very low efficiency at low loadings. Finally, they might not be designed for very frequent large changes of their power output. In the case of FNR provision from a large number of ES systems, an effective control strategy must consider and tackle these aspects.

If the aggregator has perfect bi-directional communication with the units, many of these challenges can be addressed. However, communications in electric power systems are subject to various sources of uncertainty and delays stemming from the employed transmission technologies, relaying and the nature of packet-switching networks [4, 5]. In closed-loop feedback control these factors compromise stability margins [5] and require thorough analysis and tuning, as demonstrated for primary and secondary frequency control in [6, 7, 8]. Several different types of networks exist in power system communications, where the round-trip time of single data packages is typically in the order of seconds [4].

Extremely reliable communication links with latencies in the order of tens of milliseconds may be uneconomical to be established and maintained for a large number of small units. Even in that case, the communication architecture is not robust and may be vulnerable to failures, such as an IT-related breakdown. It is thus beneficial to employ a decentralized control structure, where units monitor system frequency and respond almost instantaneously by following a predefined control policy. A cheap, infrequent communication link with lower requirements can be used to exchange information between the aggregator and the units, without compromising service delivery. Another benefit of decentralized control is that the aggregator is not required to gather and process a large amount of data in real-time, and decision making for real-time control is transferred to the units. This relieves the aggregator from this task, which introduces additional response delays and uncertainties.

A number of decentralized EV strategies to offer frequency control have been proposed, where in most cases the EV charging power is controlled continuously in a droop-curve fashion. In [9] the authors employ a standard droop-curve to offer PFC with EVs, but superimpose a scheduled charging setpoint on the frequency response characteristic. The authors of [10] propose a hybrid centralized-decentralized control framework to offer secondary frequency control. The aggregator sends a switching probability with which EVs should change their power setpoints, and EVs decide to switch based on the results of a Bernoulli trial. The idea of employing stochasticity in the response of EVs is also used by the authors of [11], where a new decentralized frequency control scheme is proposed. Under this scheme each EV monitors the system frequency and switches between three states: charging, discharging or idle. A randomization in the responses is introduced to avoid undesired frequency oscillations.

To overcome low charging efficiencies at low loadings, the authors of [12] propose an autonomous stochastic control policy, albeit not for frequency control. Under this policy, at each predefined time interval (typically of 5 or 15 minutes) every EV stochastically decides to either charge at a power level with the highest efficiency, or does not charge at all. The authors of [13, 14] also recognized the drawback

of inefficient operation of ES systems providing PFC under a droop characteristic. They propose a clustering algorithm that distributes the units into the frequency spectrum, so that a predefined set of units respond according to the measured frequency and their individual State of Charge (SOC). However, the method relies on frequently updated information from the aggregator to allocate the clusters, increasing the reliance on the aggregator. Moreover, the proposed control method does not guarantee or estimate reserve provision errors, nor does it account for frequent switching actions with large changes of the setpoints. Similar to these works, in [15] a fuzzy controller is designed, where the SOC and the frequency deviation are used as inputs. However, no consideration for efficiency is made. The authors of [16] also propose strategies to maximize the efficiency of an aggregation of ES systems, but these strategies require coordination (thus they are not decentralized), and are not designed for frequency reserve provision.

This paper proposes a decentralized control policy that can be tuned to increase the ES systems' efficiency, while guaranteeing that the service is delivered with a predefined accuracy, depending on the number and size of the units. The contribution of this paper is twofold. First, a new, fully decentralized control policy for aggregated ES systems offering FNR is proposed. This policy allows ES systems to operate only on a set of predefined states, overcoming any minimum-power requirements and avoiding inefficient operating points. Furthermore, an extension of the main control algorithm significantly reduces the switching rate of the units, which is in general desirable, to reduce the wear of equipment and avoid deterioration of the network's power quality [17]. Second, analytical expressions for reserve provision errors are derived, both for homogeneous and heterogeneous, i.e., with different control settings, populations. Adequately small reserve errors are an important FNR requirement, and these expressions allow the aggregator to estimate these errors for any control settings, aggregation size and composition. Moreover, analytical expressions for the average switching rate over the reserve provision period are also provided. A preliminary version of our work was presented in [18]; here we formalize and extend the proposed control methods. More specifically, analytical expressions for the reserve provision errors and the average switching rates are

4

provided. The effect of standby losses during service provision is also examined, and the heterogeneity of the ES systems is accounted for through clustering.

The remainder of the paper is organized as follows. In Section 2 we provide some necessary background for FNR, and we describe the proposed control structure. In Section 3 we present the proposed decentralized control policy and an extension which reduces the average switching rate. In Section 4 we derive analytical expressions for the reserve errors and the average switching rates. The case study presented in Section 5 validates our theoretical results using real frequency data, and provides further insight, whereas Section 6 concludes the paper.

2. Decentralized FNR Scheme

2.1. FNR Preliminaries

Three quantities fully describe the expected FNR response of a unit i to a frequency deviation value Δf_k at time step $t_k = k \, \Delta T_c$, with $k \in \mathbb{N}$ and T_c being the control time interval. Upwards reserve capacity $P_{i,k}^{\mathrm{r,up}}$, downwards reserve capacity $P_{i,k}^{\mathrm{r,dn}}$ and power reference $P_{i,k}^{\mathrm{ref}}$. Power reference refers to the power consumed from or injected to the grid by the unit when $\Delta f_k = 0$, i.e., when no frequency response is required. This power reference is important in the cases of ES units because it can be used to compensate the losses of the unit due to the non-unitary charging and discharging efficiencies. In the case of EVs, a positive (by convention) power reference would be required to cover the driving energy requirements and bring an EV's SOC to an acceptable level at the end of service provision. Upwards and downwards reserve capacities refer to the maximum reserve provision in the positive (consumption) and negative (injection to the grid) direction. The requested power $P_{i,k}^{\mathrm{req}}$ is calculated as

$$
P_{i,k}^{\mathrm{req}} = \begin{cases}
P_{i,k}^{\mathrm{ref}} - P_{i,k}^{\mathrm{r,dn}} & \text{if } \Delta f_k < -0.1 \text{ Hz,} \\
P_{i,k}^{\mathrm{ref}} + P_{i,k}^{\mathrm{r,dn}} \Delta f_k / 0.1 & \text{if } \Delta f_k \in [-0.1, 0) \text{ Hz,} \\
P_{i,k}^{\mathrm{ref}} + P_{i,k}^{\mathrm{r,up}} \Delta f_k / 0.1 & \text{if } \Delta f_k \in [0, 0.1] \text{ Hz,} \\
P_{i,k}^{\mathrm{ref}} + P_{i,k}^{\mathrm{r,up}} & \text{if } \Delta f_k > 0.1 \text{ Hz.}
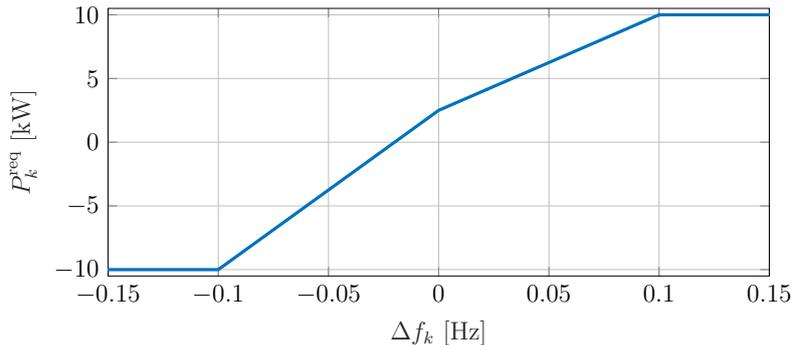\end{cases}
\tag{1}
$$

5

Figure 1: A unit's frequency response characteristic with $P_i^{\mathrm{r,dn}} = 12.5$ kW, $P_i^{\mathrm{r,up}} = 7.5$ kW and $P_i^{\mathrm{ref}} = 2.5$ kW.

There is no deadband in the provision of FNR, as is evident from (1). The most straightforward decentralized control method to provide FNR with inverter-equipped loads is applying a response characteristic similar to Fig. 1; in the rest of the paper we will refer to this control method with the commonly used term *droop control*. Further, the terms loads, units and ES systems will be used interchangeably. Each unit measures Δf_k locally and modulates its power output to the desired value. In Fig. 1 a unit's response characteristic is shown for a generalized case with a non-zero power reference and asymmetric reserve capacities. This characteristic refers to the response of one unit; to offer FNR, an aggregator needs to pool a sufficient number of units and the aggregate power reference and reserve capacities would be obtained by the summation of the individual characteristics. In Denmark, the minimum FNR capacity that can be offered amounts to 300 kW. Considering the maximum power capacity P_i^{max} of unit i, it is convenient to normalize the requested power as

$$p_{i,k} = \frac{P_{i,k}^{\mathrm{req}}}{P_i^{\mathrm{max}}}. \tag{2}$$

A droop controller will result in frequent operation of the ES units in low loadings for extended periods of time. This happens because the system's frequency deviations are approximately normally distributed. However, the efficiency of ES systems is lower on low loadings, with a maximum efficiency usually achieved close to 50 % of the unit's rated power. In Fig. 2 the efficiency curves of a converter are shown, along with a histogram of the system's frequency. 35 % of the frequency

deviation samples are within ± 0.02 Hz, and 50 % within ± 0.03 Hz. A unit providing FNR with a droop controller will operate 50 % of the time with a loading lower than 30 %, resulting in a low average efficiency. The main goal of this paper is to design a decentralized control policy which minimizes the units' operation time on low efficiency regimes, while ensuring a desired reserve delivery performance.
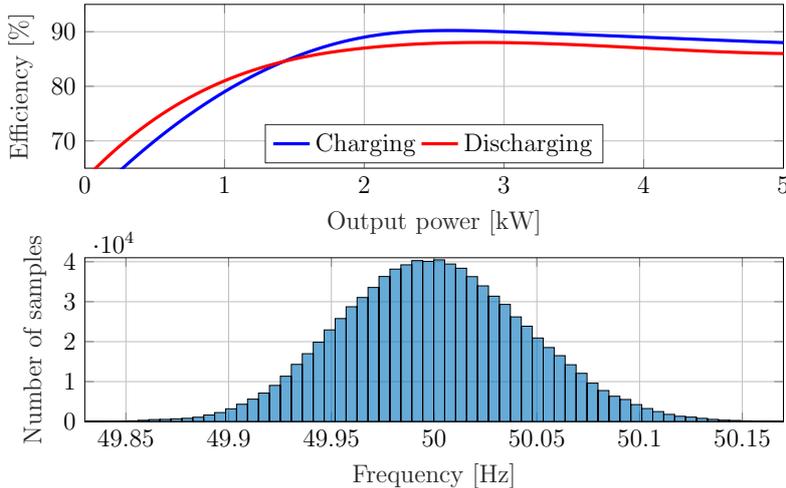

Figure 2: Upper subplot: converter efficiency curves taken from [19]. Lower subplot: histogram of frequency values in the Regional Group Nordic area.

2.2. Control Scheme

We consider a control structure which combines the advantages of both centralized and decentralized control. Prior to frequency reserve provision, the aggregator bids in the day ahead (DA) spot and reserve markets to obtain the DA energy schedule and reserve capacity for each hour of the following day. These values are aggregated and correspond to the whole portfolio. There is a substantial amount of work in the literature dealing with optimal bidding in the energy and reserve markets; bidding in those markets is outside the scope of this work. At a second stage, the aggregator ensures the reliable and accurate reserve provision by controlling its available portfolio. To improve economic performance and increase its operational flexibility while providing frequency reserves, the aggregator can utilize the intra-day market, as suggested in [20].

During service provision, all loads respond to the local frequency measurement in an entirely decentralized way, following a control policy described in the next section. Thus, loads can respond very fast without high local computational requirements, and without relying on fast and reliable communication links. This allows the use of more economic communication technologies, such as the available Wireless Local Area Network (WLAN). The aggregator gathers information regarding the availability and the SOC of the units, and applies a supervisory control by dispatching updated $P_i^{\mathrm{r,up}}$, $P_i^{\mathrm{r,dn}}$ and P_i^{ref} values to the units at minute-scale intervals. Therefore, the aggregator can communicate with the loads at a rate significantly slower than the FNR requirements, which are in the order of hundreds of ms, and without the need for very high reliability. These adjustment decisions are made centrally by the aggregator, with information available from all units and without the need for very fast decisions. It is thus possible to embed more advanced optimization and/or data-driven strategies in the supervisory controller, with a high level of information availability. The described control scheme is depicted in Fig. 3.
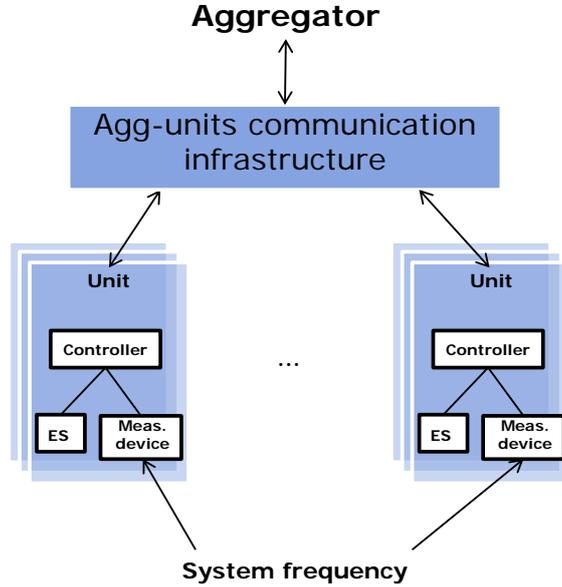

Figure 3: Schematic overview of the decentralized control approach, with infrequent communication between the aggregator and the units.

3. Proposed stochastic controller

Under the proposed control scheme, each load responds by stochastically changing its power output based on the measured frequency deviation Δf_k and a local control policy. Instead of using three operational modes (idle, fully charging or fully discharging), the proposed controller employs a customizable discretization of the responses. In Section 2 we explained that using a droop controller results in low average loadings and a lower overall efficiency under FNR provision. The proposed discretization offers several advantages compared to a droop controller or a 3-mode controller. First, it can achieve lower tracking errors for small ES aggregations compared to a 3-mode controller. Second, it can optimize the average efficiency of the units, by avoiding inefficient operating points. Third, it can overcome minimum power requirements of the units.

3.1. Basic stochastic switching algorithm

The actual power of each load at step k is denoted by P_k^{act}, and the normalized value against its capacity by y_k, where the unit index i is dropped for notation simplicity. We introduce discretization vector $\mathbf{u} = [u(1), u(2), \ldots, u(N)]^T$ to denote the N admissible normalized power values (indexed by j thereafter and in ascending order). For simplicity, but without loss of generality, we set $\Delta T_{\text{c}} = 1$ s. We define mapping $\mathbf{g}(p_k) : \mathbb{R} \to \mathbb{R}$ to map a value p_k to state $b_k \in \{2, \ldots, N\}$. To ensure that mapping $\mathbf{g}$ is surjective, meaning that a unique b_k value corresponds to each p_k, b_k is defined as

$$
b_k = \begin{cases} \exists! b_k = g(p_k) | u(b_k - 1) < p_k \le u(b_k), & p_k \ne -1 \\ 2, & p_k = -1. \end{cases} \tag{3}
$$

According to the definition of b_k in (3), y_k can take one of the two following values: $u(b_k)$ or $u(b_k - 1)$. The distance between the two states is denoted by d_k, which is equal to

$$
d_k = u(b_k) - u(b_k - 1). \tag{4}
$$

At each time step, all units generate a random number $r \in [0, 1]$ and switch to

either state b_k or $b_k - 1$. Switching probability ρ_k is calculated as

$$\rho_k = \begin{cases} \dfrac{p_k - u(b_k - 1)}{d_k}, & \text{if } p_k \geq 0 \\[2mm] \dfrac{u(b_k) - p_k}{d_k}, & \text{if } p_k < 0. \end{cases} \tag{5}$$

The algorithm (referred to thereafter as Algorithm 1) to determine how units stochastically change their output y_k according to the measured f_k is outlined in Algorithm 1. An example of this algorithm is given in Fig. 4 to better illustrate how b_k and $u(b_k)$ are derived at each time step. The response is discretized in five equidistant states, and at $k = 0$ the load consumes $y_0 = 0$. At $k = 1$ and for a $p_1 = 0.3$, $b_1 = 4$ according to (3), and the admissible states are $y_1 = 0$ (with $\rho_1 = 0.6$) and $y_1 = 0.5$ (with $1 - \rho_1 = 0.4$). In our example $r = 0.51$, which is smaller than ρ_1, and the load switches to $y_1 = 0.5$. At $k = 2$, with $p_2 = 0.6$ we have $b_2 = 5$. The load will end up either in $y_2 = 1$ (with $\rho_2 = 0.2$) or remain in $y_2 = 0.5$ (with $1 - \rho_2 = 0.8$). With $r = 0.32$ in our example, the load remains in $y_2 = 0.5$.
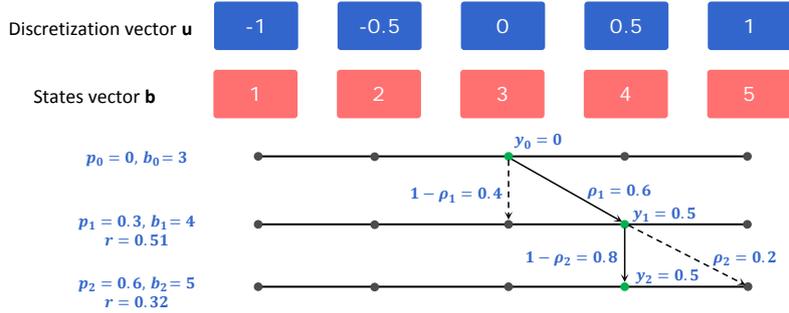

Figure 4: Illustrative example for the application of Algorithm 1. Lines indicate the admissible states: a solid line represents the transition to a new state, and a dashed line the other possible state. Green dots represent the realized state y_k based on the random roll r.

3.2. Switching minimization algorithm

A shortcoming of Algorithm 1 is that it results in a high switching rate. In general, a fast change of setpoints should not be an issue for inverters, which usually operate by frequently changing their output. However, if the changes of the setpoints are very large in magnitude and frequent, then this could potentially result in wear of the equipment and power quality issues in a distribution network. Below we

---
Algorithm 1 Basic stochastic switching algorithm

1: $p_k \leftarrow (1), (2)$
2: $b_k \leftarrow (3)$
3: $d_k \leftarrow (4)$
4: $\rho_k \leftarrow (5)$
5: $r \leftarrow$ random uniform number between $[0, 1]$
6: **if** $(p_k \geq 0) = (r \leq \rho_k)$ **then**
7:     $y_k \leftarrow u(b_k)$
8: **else**
9:     $y_k \leftarrow u(b_k - 1)$
10: **end if**
---

present an extension of the main algorithm, referred to as Algorithm 2, to reduce the excessive switching rate.

This requires keeping track of the ideal state densities at the previous time step. Let matrix $\mathbf{x} \in \mathbb{R}^{N_{\text{tot}} \times N}$ contain the ideal densities $x(k, b_k)$ for each step $k \in N_{\text{tot}}$ and state $b_k \in \{2, \dots, N\}$. Ideal state density refers to the percentage of loads in each state, if the stochastic processes were "perfect", i.e., the number of units was infinite. In the case study we will show that the errors introduced by the imperfect random numbers generation decrease with the square root of the aggregation size, and drop quickly to acceptable levels. The modified switching probability ρ_k^* is calculated as

$$\rho_k^* = \begin{cases} \dfrac{\Delta p_k}{d_k x(k-1, b_k - 1)}, & \text{if } \Delta p_k \geq 0 \\ \dfrac{-\Delta p_k}{d_k x(k-1, b_k)}, & \text{if } \Delta p_k < 0, \end{cases} \tag{6}$$

where $\Delta p_k = p_k - p_{k-1}$.

As already explained, at each step only two consecutive states are admissible. Ideal states $x(k, j)$ are calculated as

$$x(k, j) = \begin{cases} 0, & \text{if } j \neq b_k \text{ or } j \neq b_k - 1, \\ \rho_k, & \text{if } j = b_k \text{ and } p_k \geq 0, \\ 1 - \rho_k, & \text{if } j = b_k - 1 \text{ and } p_k \geq 0, \\ 1 - \rho_k, & \text{if } j = b_k \text{ and } p_k < 0, \\ \rho_k, & \text{if } j = b_k - 1 \text{ and } p_k < 0. \end{cases} \tag{7}$$

The modified algorithm is described in Algorithm 2. For illustrative purposes let $\mathbf{u} = [-1, -0.5, 0, 0.5, 1]$, as in Fig. 4. If $p_1 = 0.2$, approximately 40 % of the loads will have an output equal to 0.5 and 60 % equal to 0. Thus, $b_1 = 4$, $x(1, 4) = 0.4$ and

---
Algorithm 2 Switching minimization algorithm

1: $p_k \leftarrow (1), (2)$
2: $b_k \leftarrow (3)$
3: $d_k \leftarrow (4)$
4: $\rho_k^* \leftarrow (6)$
5: $r \leftarrow$ random uniform number between $[0, 1]$
6: **if** $|b_k - b_{k-1}| \geq 1$ **then**
7:     $\rho_k \leftarrow (5)$
8:     apply Algorithm 1, lines 6–10
9: **else**
10:     **if** $(\Delta p_k \geq 0) = ((y_{k-1} = u(b_k - 1)) \ \& \ (r \leq \rho_k^*))$ **then**
11:         $y_k \leftarrow u(b_k)$
12:     **else**
13:         $y_k \leftarrow u(b_k - 1)$
14:     **end if**
15: **end if**
16: $\forall j: \ x(k, j) \leftarrow (7)$

$x(1, 3) = 0.6$. A request $p_2 = 0.3$ requires that 60 % of loads end up in state 0.5 and 40 % in idle mode. If the algorithm is memory-less, then all loads will undergo the stochastic process outlined in Algorithm 1, with $\rho_2 = 0.6$, and 52 % of all loads will change state. Due to the increase of p_k, only loads in idle mode need to undergo the Bernoulli trial. A probability $\rho_2^* = 0.333$ (as calculated via (6)) should be applied only to the idle loads, resulting in a switching rate of 20 % and reducing switchings by 60 %. The difference of the two algorithms is illustrated in Fig. 5.
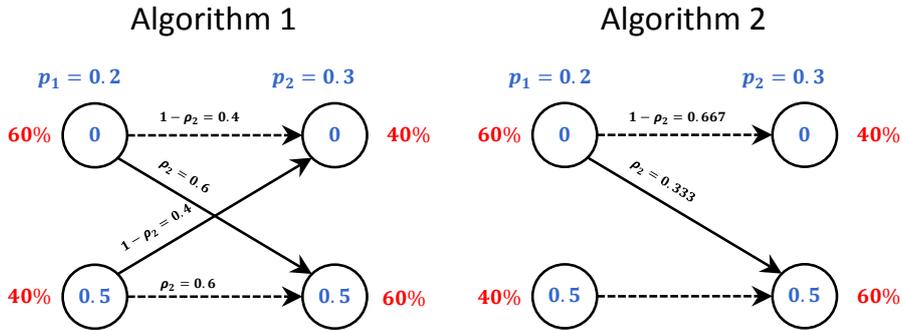

Figure 5: Switching actions under the two algorithms. Solid lines represent a change of state and dashed lines represent no switching. The percentages in red color show the ideal density of each state.

4. Theoretical error and switching rate calculations

In subsection 4.1 we first derive expressions for a homogeneous population, i.e., when all loads have the same controller settings. Next, we consider the case of a heterogeneous population. We use the Root Mean Square Error (RMSE) to assess the performance of the different controllers. For a reserve provision period equal to N_{tot}, an aggregation size equal to N_{agg}, and a reserve capacity $P^{\text{res}} = \sum_{i=1}^{N_{\text{agg}}} P_i^{\text{max}}$, the RMSE is given by

$$e_{\text{rmse}}^{\text{real}} = \sqrt{\frac{1}{N_{\text{tot}}(P^{\text{res}})^2} \sum_{k=1}^{N_{\text{tot}}} \left(\sum_{i=1}^{N_{\text{agg}}} (P_{i,k}^{\text{req}} - P_{i,k}^{\text{act}}) \right)^2}. \tag{8}$$

In subsection 4.2 we derive analytical expressions to calculate the theoretical values of the switching rates for the two proposed algorithms.

4.1. Reserve Error

We first consider the case where all units offer the same reserve capacity and have the same power reference. If this is not the case, then we can cluster the units so that they share the same characteristics, as we will show next. In the homogeneous case p_k and the corresponding switching probability ρ_k are common for all units. Given an arbitrary response discretization vector $\mathbf{u}$, it is straightforward to calculate ρ_k and d_k for each k according to Algorithm 1.

Proposition 1. The estimated RMSE of a homogeneous cluster can be calculated as

$$e_{\text{rmse}}^{\text{cl}} = \sqrt{\frac{1}{N_{\text{tot}}} \sum_{k=1}^{N_{\text{tot}}} d_k^2 \frac{\rho_k(1 - \rho_k)}{N_{\text{agg}}}}. \tag{9}$$

The derivation of (9) is provided in the Appendix. By using (9), it is possible to calculate the reserve RMSE values for any aggregation size and bin distances. This error decreases with the square root of the aggregation size N_{agg}. There is also a linear relationship between the error and the bin distances. As already mentioned, (9) can be used for a homogeneous cluster. Assume that the population has different controller settings, i.e., different response discretizations or superimposed power references, leading to different switching probabilities. We cluster the units so that

Table 1: Switching rate analytical expressions for both algorithms

Algorithm 1		Algorithm 2	
Expression for $\hat{s}_k$	Condition	Expression for $\hat{s}_k$	Condition
1	$|b_k - b_{k-1}| \geq 2$	1	$|b_k - b_{k-1}| \geq 2$
$(1-\rho_k)\rho_{k-1} + (1-\rho_{k-1})\rho_k$	$b_k = b_{k-1}$	$\rho_k^* x_\mathrm{l}$	$b_k = b_{k-1}$ & $\Delta f_k \geq \Delta f_{k-1}$
$\rho_k\rho_{k-1} + (1-\rho_{k-1})$	$b_k > b_{k-1}$ & $\Delta f_k \geq 0$ & $\Delta F_k \geq 0$	$\rho_k^* x_\mathrm{r}$	$b_k = b_{k-1}$ & $\Delta f_k < \Delta f_{k-1}$
$(1-\rho_k)(1-\rho_{k-1}) + \rho_{k-1}$	$b_k < b_{k-1}$ & $\Delta f_k \geq 0$ & $\Delta F_k \geq 0$	$\rho_k^* x_\mathrm{r} + x_\mathrm{l}$	$b_k > b_{k-1}$ & $\Delta f_k \geq 0$ & $\Delta F_k \geq 0$
$(1-\rho_k)(1-\rho_{k-1}) + \rho_{k-1}$	$b_k > b_{k-1}$ & $\Delta f_k < 0$ & $\Delta F_k \geq 0$	$(1-\rho_k^*)x_\mathrm{l} + x_\mathrm{r}$	$b_k < b_{k-1}$ & $\Delta f_k \geq 0$ & $\Delta F_k \geq 0$
$\rho_k\rho_{k-1} + (1-\rho_{k-1})$	$b_k < b_{k-1}$ & $\Delta f_k < 0$ & $\Delta F_k \geq 0$	$(1-\rho_k^*)x_\mathrm{r} + x_\mathrm{l}$	$b_k > b_{k-1}$ & $\Delta f_k < 0$ & $\Delta F_k \geq 0$
$(1-\rho_k)\rho_{k-1} + \rho_{k-1}$	$b_k > b_{k-1}$ & $\Delta F_k < 0$	$\rho_k^* x_\mathrm{l} + x_\mathrm{r}$	$b_k < b_{k-1}$ & $\Delta f_k < 0$ & $\Delta F_k \geq 0$
$(1-\rho_{k-1})\rho_k + \rho_{k-1}$	$b_k < b_{k-1}$ & $\Delta F_k < 0$	$\rho_k^* x_\mathrm{r} + x_\mathrm{l}$	$b_k > b_{k-1}$ & $\Delta F_k < 0$
-	-	$\rho_k^* x_\mathrm{l} + x_\mathrm{r}$	$b_k < b_{k-1}$ & $\Delta F_k < 0$

each cluster m (out of N_cl clusters) has the same switching probability $\rho_{k,m}$, nominal power P_m^max and bin width $d_{k,m}$ for each step k. If each cluster m contains N_m units and each unit's nominal power is P_m^max, then the cluster's nominal power is equal to $P_m^\mathrm{nom,cl} = N_m P_m^\mathrm{max}$. It is straightforward to show that the aggregation's RMSE $e_\mathrm{rmse}^\mathrm{agg}$ can be estimated as a combination of the individual errors according to

$$e_\mathrm{rmse}^\mathrm{agg} = \sqrt{\sum_{m=1}^{N_\mathrm{cl}} \left[e_\mathrm{rmse}^{\mathrm{cl},m} \frac{P_m^\mathrm{nom,cl}}{\sum_{m=1}^{N_\mathrm{cl}} P_m^\mathrm{nom,cl}} \right]^2}. \tag{10}$$

4.2. Switching Rates

We first present the case of Algorithm 1, i.e., the control policy which does not minimize the number of switching actions, or in other words the transition of a unit between two states. Switching $s_{i,k}$ is equal to one when $y_{i,k} \neq y_{i,k-1}$. The actual average switching rate is expressed as a percentage and is calculated as

$$s^\mathrm{av} \ [\%] = \frac{100}{N_\mathrm{agg} N_\mathrm{tot}} \sum_{k=1}^{N_\mathrm{tot}} \sum_{i=1}^{N_\mathrm{agg}} s_{i,k}. \tag{11}$$

To provide an analytical estimation of s^av, we first estimate the average rate $\frac{1}{N_\mathrm{agg}} \sum_{i=1}^{N_\mathrm{agg}} s_{i,k}$ at each step k. Based on these estimations, each denoted by $\hat{s}_k$, the estimated average switching rate over N_tot steps is given by

$$s^\mathrm{av,est} \ [\%] = \frac{100}{N_\mathrm{tot}} \sum_{k=1}^{N_\mathrm{tot}} \hat{s}_k. \tag{12}$$

In the following, we show how estimates $\hat{s}_k$ are calculated. Consider the example of Fig. 5. Since $\rho_1 = 0.4$, at $k = 1$ a share of loads equal to 40% are at state

0.5, and 60 % of the loads at state 0 (because $1 - \rho_1 = 0.6$). Under algorithm 1, a share of 52 % of the loads (because $\rho_1(1 - \rho_2) + (1 - \rho_1)\rho_2 = 0.52$) will switch. This can be generalized as $\hat{s}_k = (1 - \rho_k)\rho_{k-1} + (1 - \rho_{k-1})\rho_k$ for this particular case. Following a similar logic, expressions for $\hat{s}_k$ can be found for the other cases. Note that $\hat{s}_k$ is an estimate of the actual switching actions, based on the ideal outcome of the Bernoulli processes. For this reason, $\hat{s}_k$ converges to the actual value as N_{agg} increases to infinity. However, these inaccuracies average over time, and as we will show in the results, the estimated average switching rate converges to the actual value very quickly.

If Algorithm 2 is applied, then $x_{1,3} = 0.6$ and $x_{1,4} = 0.4$, $\rho_2^* = 0.333$. This time, the Bernoulli process is applied only on loads at state 0, and thus $\hat{s}_2 = \rho_2^* x_{1,3} = 0.5 \cdot 0.4 = 0.2$. In a generalized form $\hat{s}_k = \rho_k^* x(k-1, b_{k-1}-1)$. Again, expressions for $\hat{s}_k$ can be found for the other cases. All analytical expressions for $\hat{s}_k$ are summarized in Table 1 for both algorithms, where for notational simplicity $\Delta F_k = \Delta f_k \Delta f_{k-1}$, $x_{\mathrm{r}} = x(k-1, b_{k-1})$ and $x_{\mathrm{l}} = x(k-1, b_{k-1} - 1)$. For the ease of exposition, we have not included two special cases for Algorithm 1 in Table 1: for $b_k = b_{k-1}$, $\rho_k = 1$ and $\rho_{k-1} \neq 1$ then $\hat{s}_k = \rho_k$, whereas for $b_k = b_{k-1}$, $\rho_{k-1} = 1$ and $\rho_k \neq 1$ then $\hat{s}_k = \rho_{k-1}$.

5. Case Study

A frequency signal from the Nordic area from year 2016 was used to evaluate the proposed controllers. The resolution of the signal is one second, and its duration equal to one day: $N_{\mathrm{c}} = 1$ s and $N_{\mathrm{tot}} = 86400$. We consider an aggregation of battery systems offering FNR. Each battery system has a nominal power of 5 kW and is offering ± 5 kW of reserve, with a reference setpoint equal to zero. We use equidistant states for better illustration of our results, and in this context a 50 % granularity refers to a discretization of -5 kW, -2.5 kW, 0 kW (idle state), 2.5 kW and 5 kW.

5.1. Reserve provision errors

We first validate the two theoretical findings regarding the reserve RMSE from Section 4. Extensive simulations showed that both proposed algorithms result in

the same reserve errors, and the reported results were obtained by using Algorithm 1. The simulation results can be seen in Fig. 6. First, the reserve RMSE values decrease linearly with the distance of the states (response granularity), as expected. Second, errors decrease according to the square root of the aggregation size. For a response granularity of 100 %, the number of battery systems needs to increase from 10 to 1000 to reduce the RMSE from 12.72 % to 1.278 %. For the same number of battery systems (10), a finer granularity of 50 % reduces the RMSE by half, to a value of 6.395 %.
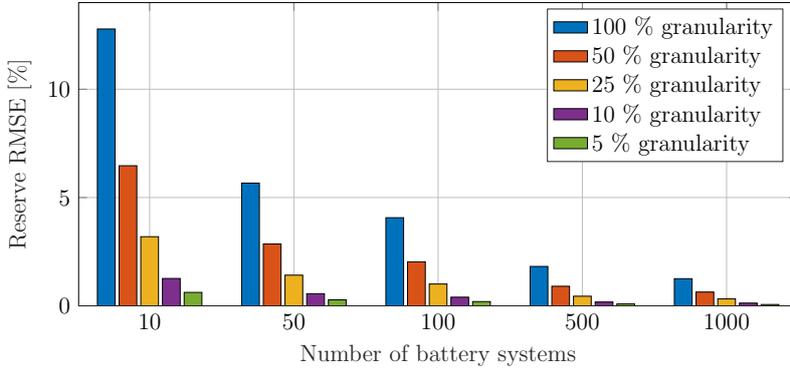

Figure 6: Simulation-based RMSE values for different aggregation sizes and response granularities.

Next, the theoretical RMSE values calculated by using (9) are compared with the actual RMSE values obtained via (8). The RMSE estimation error is defined as

$$e_{\mathrm{rmse}}^{\mathrm{est}} \ [\%] = 100 \frac{e_{\mathrm{rmse}}^{\mathrm{real}} - e_{\mathrm{rmse}}^{\mathrm{el}}}{e_{\mathrm{rmse}}^{\mathrm{el}}}. \tag{13}$$

In Fig. 7 the estimation errors are shown, where it can be seen that using analytical expression (9) results in very small estimation errors.

5.2. Switching rates

The results for the average switching rates over the reserve period are presented in Fig. 8. Algorithm 2 is able to considerably reduce the average switching rate, especially when the response granularity is large. This happens because in that case units are confined within the same two admissible states between two time steps more often. When this is the case, the stochastic switching process is applied to significantly fewer units under Algorithm 2, and the average switching rates decrease
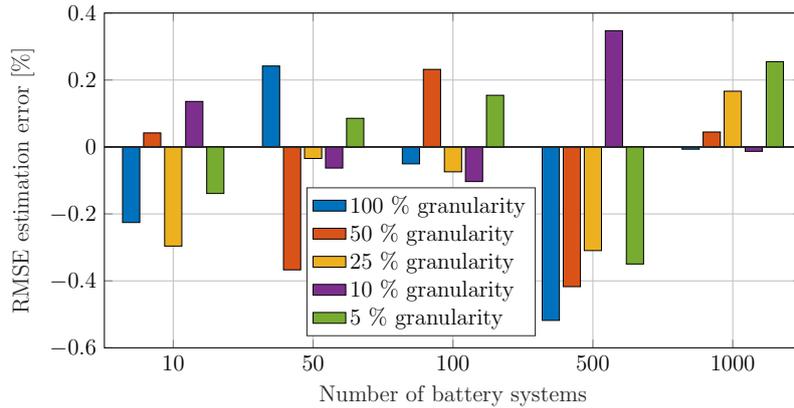
Figure 7: Estimation errors of theoretical and simulation-based RMSE values for different aggregation sizes and response granularities.

drastically. Additionally, the theoretical values (diamond-shaped markers in Fig. 8), which were calculated via the analytical expressions derived in Section 4, are very close to the actual values obtained by the simulations, resulting in very small estimation errors.
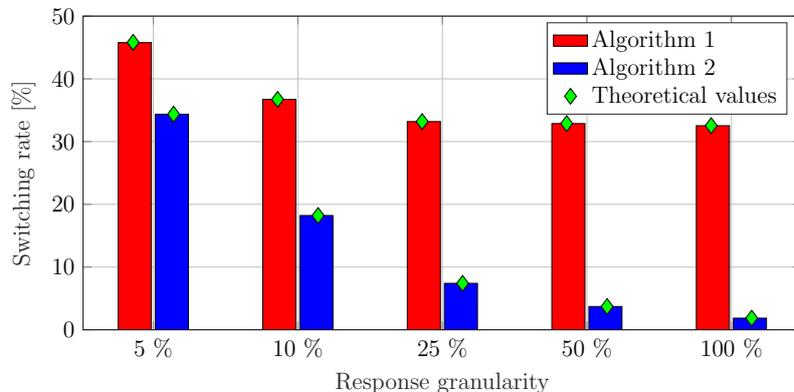

Figure 8: Average switching rates for the two proposed algorithms. The theoretical values are depicted with diamond-shaped markers.

In Fig. 9 the actual and the estimated average switching rates are plotted against time to get a better insight on the performance of the switching rate estimation expressions. The switching minimization algorithm was used, with an aggregation size of 1000 loads and a granularity of 100 %. The theoretical and actual rates are practically the same, with a negligible difference. The initial transient of the rates
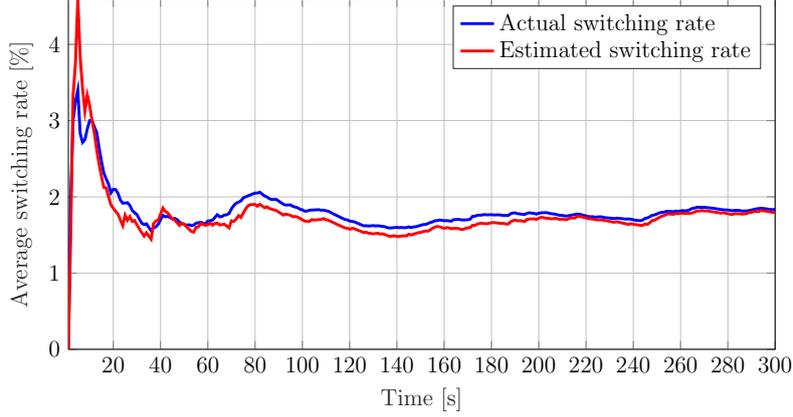
17

Figure 9: Estimated and actual average switching rate for 1000 units and a granularity of 100 % under Algorithm 2.

is attributed to the small number of samples used for the calculations. Both the actual and the estimated switching rates converge to a steady state value of 1.95 % in only 2.5 hours. It is therefore sufficient to use a 3 h frequency signal to obtain accurate results for the average switching rates, instead of 24 hours, as in the case of Fig. 8.

5.3. Losses under FNR provision

To analyze the impact of the proposed controller on the units' losses under FNR provision we use the converter efficiency curve from [19], shown in Fig. 2. Note that these efficiencies do not include battery losses, which are neglected because they are relatively small compared to the converter losses. However, it is straightforward to include them in the controllers' evaluation. Using the curves of Fig. 2, loss $l_{i,k}$ in kWh can be calculated for each power state $P_{i,k}$. Average losses per battery system L are calculated as

$$L = \frac{1}{N_{\text{agg}}} \sum_{k=1}^{N_{\text{tot}}} \sum_{i=1}^{N_{\text{agg}}} l_{i,k}. \tag{14}$$

Most ES systems have standby losses when they are idle. These losses depend on the design of the converters and have a non-negligible impact on a system's losses, even if they are small. We examine two cases: one where standby losses are equal to 60 W and one where they are zero. The results are shown in Fig. 10.

It is possible to substantially reduce the units' losses by defining a response
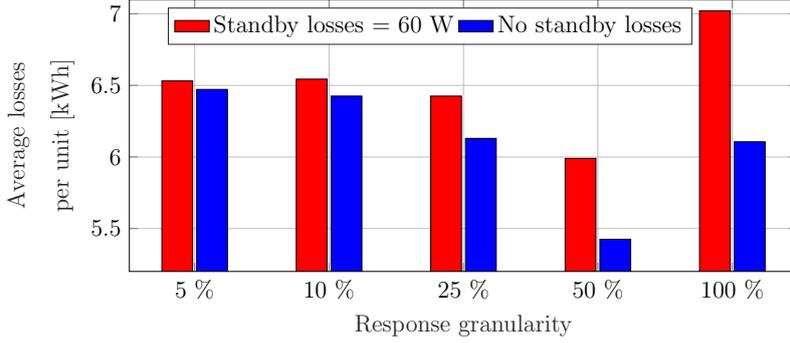
18

Figure 10: Losses per unit with and without standby losses, for a 24 hour FNR provision.

characteristic with five admissible states. Compared to the standard droop curve with 5 % steps (or equivalently 250 W), a loss reduction of 8.3 % is possible, when standby losses amount to 60 W. If standby losses are equal to zero, losses can be reduced by 15.5 %. If battery losses are added, the reduction is further increased. It is interesting to note that if standby losses exist, operating the systems with only three states (i.e., idle, fully charging or discharging) leads to higher losses compared to other response granularities. This happens because the most efficient operating regime is not utilized and the relatively high standby losses are more prevalent.

5.4. Analysis on a device level

So far only aggregated results have been presented, without examining the effect on a device level. The stochastic nature of the controllers leads to different individual responses and potentially to differing utilization levels. We define the load factor LF_i of unit i over the whole provision period as

$$LF_i \ [\%] = \frac{100}{P_i^{\max} \ N_{\text{tot}}} \sum_{k=1}^{N_{\text{tot}}} |P_{i,k}|. \tag{15}$$

We further introduce the Probability Density Function (PDF) $f_{LF}(LF)$, which describes the distribution of the different load factors among the population. Fig. 11 shows the PDFs for different discretizations and both algorithms, where it can be seen that values are normally distributed in all cases. Fewer discretization steps result in larger standard deviations due to the smaller average switching rates. This is more pronounced in the case of Algorithm 2, where the considerably smaller
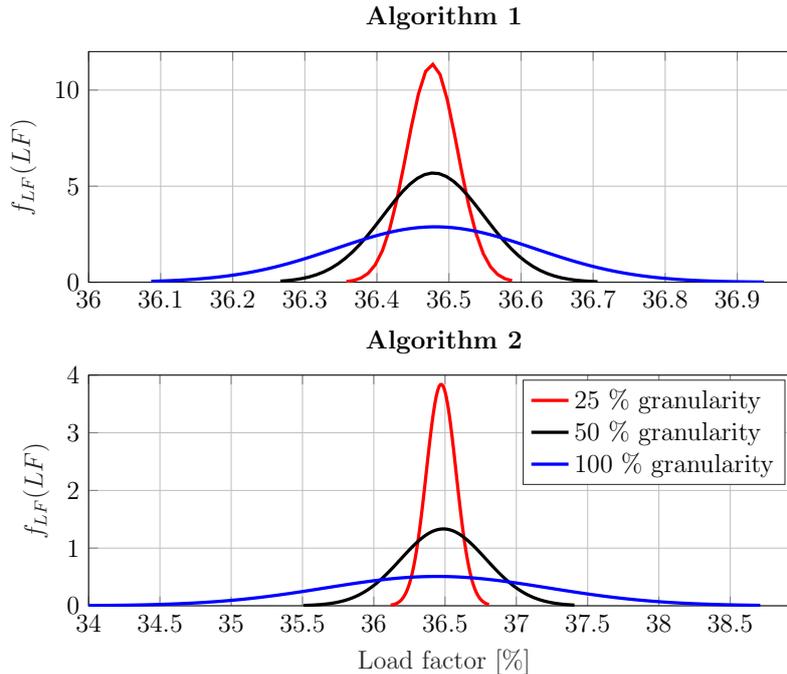
19

Figure 11: Load factor PDFs for three different discretizations and for both algorithms.

switching rates lead to a larger standard deviation. As units switch more often, the probability that some are utilized more or less than the average decreases, resulting in narrower distributions. Lowest losses are achieved with a discretization of 50 %, resulting in a standard deviation of 0.07 % and 0.3 % for Algorithm 1 and 2, respectively. These values indicate that all units are utilized fairly, with very small differences in their average absolute power loading.

Another important aspect is the distribution of the units' SOC at the end of reserve provision. As shown previously, response discretization affects losses and thus the individual $SOC_{i,k}$ states. To ease the comparison, the SOC is expressed in kWh and not as a percentage. To obtain comparable results we derive the PDF $f_{SOC}(SOC)$ from the SOC values of all units at the end of the reserve provision period, after the population's average SOC is subtracted. This is done to remove the offsets introduced by the different losses and obtain functions centered around zero for all cases. The simulation results are shown in Fig. 12.

The effect of the larger average switching rate on the distribution of the SOC values can be seen in the different results for the two algorithms. Algorithm 1,
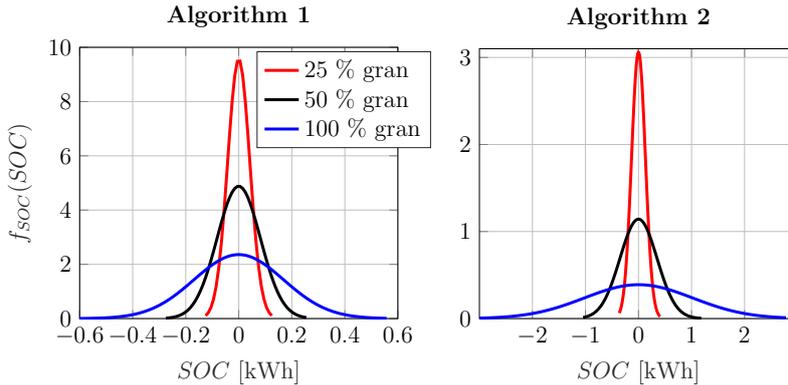
Figure 12: SOC PDFs for three different discretizations and for both algorithms.

with its larger switching rate, results in considerably smaller standard deviations, compared to Algorithm 2. Moreover, discretizations with fewer steps seem to result in wider SOC distributions. As a result, the accumulated impact on the SOC is larger in units which are utilized more or less than the average, especially when the admissible power states are more apart from each other (as in the case of a granularity of 100 %). However, for a granularity of 50 % and Algorithm 2, the standard deviation of the SOC values is only 0.35 kWh after 24 hours of service provision. These relatively small differences may cause some minor fairness issues, but the hierarchical control structure allows the aggregator to correct such differences during service provision, by sending different power setpoints to the units.

5.5. Heterogeneous populations

Table 2: RMSEs for a heterogeneous population

	Estimated RMSE [%]	Real RMSE [%]
Population 1 200 loads, ± 5 kW, 25 % granularity	0.7107	0.7099
Population 2 100 loads, ± 10 kW, 25 % granularity	1.0051	1.0074
Population 3 100 loads, ± 20 kW, 50 % granularity	2.0202	2.0143
Aggregation	1.0559	1.0522

In all presented cases a homogeneous population was considered. Let us consider an aggregation of 400 loads with different reserve capacities and granularity settings, as presented in the first column of Table 2. This case represents an aggregation of loads which collectively offers 4 MW of reserve capacity, comprised of units of different sizes and settings. The estimated RMSEs of the individual populations (calculated by considering their respective capacities) are calculated via equation (9), and are almost identical to the real values. The real RMSE of the aggregation is equal to 1.05 %, as is the estimated value calculated from (10). Despite the much larger error of the third population, the aggregation's RMSE is kept low. This happens because the individual errors are weighted quadratically based on their contributions to the total reserve capacity (see equation (10)).

It is important to note that the RMSE values of these three populations, or any other given population, can be estimated by simply using (9) once, and then adjusting the result according to the aggregation size and response granularity. For example, the error of population 3 is two times larger than population 1, due to the granularity of the response being two times larger. Moreover, the RMSE of any arbitrary composition of loads can be calculated by simply using (10), with very high accuracy as shown from the simulation results.

6. Conclusion

We presented a decentralized control policy for ES systems offering FNR, which addresses a number of real-life challenges. The settings of each unit within an aggregation can be tuned to achieve a balance between good tracking performance, low average switching rates and high efficiency. Tracking performance can range from 1.25 % to 12.5 % (in RMSE values) for 10 units, and from 0.062 % to 0.62 % for 1000 units. RMSEs were found to decrease linearly with the reduction of the distance of the admissible states and by the square root of the aggregation size. However, finer response granularities come at a cost in efficiency, resulting in considerably higher losses. For a typical efficiency curve with a maximum efficiency at 50 % loading, losses were minimized for a granularity with steps of 50 %. Losses can be reduced by 15.5 % if standby losses are neglected, and by 8.3 % if they are considered.

A variation of the proposed controller can drastically reduce the units' switching rate by up to 9 and 18 times for a granularity of 50 % and 100 %, respectively. This is a desirable characteristic because a high switching rate with large power differences (as is the case when the admissible states are few) can cause wear of the components and degradation of the network's power quality. This large reduction in the average switching rate results in a slight increase in the variance of the units' utilization, with the standard deviation of the units' load factor increasing from 0.07 % to 0.3 %, and that of the SOC from 0.08 kWh to 0.35 kWh, for a granularity of 50 %. The various trade-offs in the aggregation's overall performance can be balanced by appropriate tuning of the individual controllers and clustering, as shown in the case study with the evaluation of a heterogeneous population, as well as through the aggregator's corrective actions during service provision. Future work can be guided towards experimentally validating the reduction of losses in real ES systems.

Appendix

Proof of Proposition 1

The switching action of load i at step k is a binary random variable, denoted by $\chi_{i,k}$, and follows a Bernoulli distribution; the success probability of $\chi_{i,k}$ is equal to ρ_k. The normalized output can also be expressed as a random variable as

$$\tilde{y}_{i,k} = \chi_{i,k} u(b_k) + (1 - \chi_{i,k}) u(b_k - 1). \tag{16}$$

Due to (4), (16) can be simplified as

$$\tilde{y}_{i,k} = \chi_{i,k} d_k + u(b_k - 1). \tag{17}$$

Given the homogeneity of the cluster, i can be dropped from $p_{i,k}$. For the same reason, difference $(P_{i,k}^{\text{req}} - P_{i,k}^{\text{act}})$ in (8) can be replaced by $y_{i,k} - p_{i,k}$. Using (17), and replacing $\chi_{i,k}$ with its success probability ρ_k, p_k can be rewritten as

$$p_k = \rho_k d_k + u(b_k - 1). \tag{18}$$

The sum of differences $\sum_{i=1}^{N_{\mathrm{agg}}} \left[y_{i,k} - p_{i,k} \right]$ can now be expressed as

$$\sum_{i=1}^{N_{\mathrm{agg}}} \left[y_{i,k} - p_{i,k} \right] = \sum_{i=1}^{N_{\mathrm{agg}}} \left[\chi_{i,k} d_k + u(b_k - 1) \right] - N_{\mathrm{agg}} \rho_k d_k - N_{\mathrm{agg}} u(b_k - 1) \qquad (19)$$

$$= \sum_{i=1}^{N_{\mathrm{agg}}} \chi_{i,k} d_k - N_{\mathrm{agg}} \rho_k d_k.$$

By replacing (19) in (8), the estimated RMSE of the m-th homogeneous cluster $e_{\mathrm{rmse}}^{\mathrm{cl},m}$ is given by

$$e_{\mathrm{rmse}}^{\mathrm{cl},m} = \sqrt{ \frac{1}{N_{\mathrm{tot}}} \sum_{k=1}^{N_{\mathrm{tot}}} \left(\sum_{i=1}^{N_{\mathrm{agg}}} \frac{\chi_{i,k} d_k}{N_{\mathrm{agg}}} - \rho_k d_k \right)^2 }. \qquad (20)$$

By observing (20), one can notice that each term $\sum_{i=1}^{N_{\mathrm{agg}}} \left(\frac{\chi_{i,k}}{N_{\mathrm{agg}}} - \rho_k \right)^2$ corresponds to the variance of random variable $\chi_{i,k}/N_{\mathrm{agg}}$, for which the following equation holds

$$\sigma_{\chi_{i,k}}^2 / N_{\mathrm{agg}} = \left[\sum_{i=1}^{N_{\mathrm{agg}}} \frac{\chi_{i,k}}{N_{\mathrm{agg}}} - \rho_k \right]^2. \qquad (21)$$

Since each switching action is independent, all random variables $\chi_{i,k}$ are also independent. As a result, the estimated RMSE value can be expressed as the sum of the individual variances, multiplied by d_k, according to

$$e_{\mathrm{rmse}}^{\mathrm{cl},m} = \sqrt{ \frac{1}{N_{\mathrm{tot}}} \sum_{k=1}^{N_{\mathrm{tot}}} d_k^2 \frac{\sigma_{\chi_{i,k}}^2}{N_{\mathrm{agg}}} }. \qquad (22)$$

By replacing $\sigma_{\chi_{i,k}}^2$ with the Bernoulli distribution variance value $\rho_k(1 - \rho_k)$, the estimated RMSE can be calculated as

$$e_{\mathrm{rmse}}^{\mathrm{cl},m} = \sqrt{ \frac{1}{N_{\mathrm{tot}}} \sum_{k=1}^{N_{\mathrm{tot}}} d_k^2 \frac{\rho_k(1 - \rho_k)}{N_{\mathrm{agg}}} }. \qquad (23)$$

References

[1] J. Engels, B. Claessens, G. Deconinck, Combined stochastic optimization of frequency control and self-consumption with a battery, IEEE Transactions on Smart Grid 10 (2) (2019) 1971–1981 (2019).

[2] A. Thingvad, C. Ziras, M. Marinelli, Economic value of electric vehicle reserve provision in the nordic countries under driving requirements and charger losses, Journal of Energy Storage 21 (2019) 826 – 834 (2019).

[3] F. Teng, Y. Mu, H. Jia, J. Wu, P. Zeng, G. Strbac, Challenges on primary frequency control and potential solution from EVs in the future GB electricity system, Applied Energy 194 (2017) 353–362 (2017).

[4] P. Dambrauskas, M. H. Syed, S. M. Blair, J. M. Irvine, I. F. Abdulhadi, G. M. Burt, D. E. M. Bondy, Impact of realistic communications for fast-acting demand side management, CIRED - Open Access Proceedings Journal (1) (2017) 1813–1817 (2017).

[5] A. Khalil, A. Swee Peng, An accurate method for delay margin computation for power system stability, Energies 11 (12) (2018).

[6] C. Zhang, L. Jiang, Q. H. Wu, Y. He, M. Wu, Delay-dependent robust load frequency control for time delay power systems, IEEE Transactions on Power Systems 28 (3) (2013) 2192–2201 (2013).

[7] C. Zhao, U. Topcu, N. Li, S. Low, Design and stability of load-side primary frequency control in power systems, IEEE Transactions on Automatic Control 59 (5) (2014) 1177–1189 (2014).

[8] A. M. Prostejovsky, M. Marinelli, M. Rezkalla, M. H. Syed, E. Guillo-Sansano, Tuningless load frequency control through active engagement of distributed resources, IEEE Transactions on Power Systems 33 (3) (2018) 2929–2939 (2018).

[9] H. Liu, Y. Yang, J. Qi, J. Li, H. Wei, P. Li, Frequency droop control with scheduled charging of electric vehicles, IET Generation, Transmission & Distribution 11 (3) (2017) 649–656 (2017).

[10] M. G. Vayá, G. Andersson, Combined smart-charging and frequency regulation for fleets of plug-in electric vehicles, in: 2013 IEEE Power Energy Society General Meeting, pp. 1–5.

[11] M. R. Vedady Moghadam, R. Zhang, R. T. Ma, Distributed frequency control via randomized response of electric vehicles in power grid, IEEE Transactions on Sustainable Energy 7 (1) (2016) 312–324 (2016).

[12] W. Zhang, D. Zhang, B. Mu, L. Y. Wang, Y. Bao, J. Jiang, H. Morais, Decentralized electric vehicle charging strategies for reduced load variation and guaranteed charge completion in regional distribution grids, Energies 10 (2) (2017).

[13] J. Schlund, R. German, A control algorithm for a heterogeneous virtual battery storage providing FCR power, in: 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), 2017, pp. 61–66 (2017).

[14] J. Schlund, D. Steber, P. Bazan, R. German, Increasing the efficiency of a virtual battery storage providing frequency containment reserve power by applying a clustering algorithm, in: IEEE ISGT-Asia, 2017, pp. 1–8 (2017).

[15] K. Janfeshan, M. A. S. Masoum, Hierarchical supervisory control system for pevs participating in frequency regulation of smart grids, IEEE Power and Energy Technology Systems Journal 4 (4) (2017) 84–93 (2017).

[16] J. Choi, I. Choi, G. Ahn, D. Won, Advanced power sharing method to improve the energy efficiency of multiple battery energy storages system, IEEE Transactions on Smart Grid 9 (2) (2018) 1292–1300 (2018).

[17] S. Martinenas, K. Knezović, M. Marinelli, Management of power quality issues in low voltage networks using electric vehicles: Experimental validation, IEEE Transactions on Power Delivery 32 (2) (2017) 971–979 (2017).

[18] C. Ziras, A. Zecchino, M. Marinelli, Response accuracy and tracking errors with decentralized control of commercial v2g chargers, in: 2018 Power Systems Computation Conference (PSCC), 2018, pp. 1–7 (2018).

[19] T. Morstyn, A. V. Savkin, B. Hredzak, H. D. Tuan, Scalable energy management for low voltage microgrids using multi-agent storage system aggregation, IEEE Transactions on Power Systems 33 (2) (2018) 1614–1623 (2018).

[20] F. A. Qureshi, I. Lymperopoulos, A. A. Khatir, C. N. Jones, Economic advantages of office buildings providing ancillary services with intraday participation, IEEE Transactions on Smart Grid 9 (4) (2018) 3443–3452 (2018).