Applied Energy 229 (2018) 69–79

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Data centers as a source of dynamic flexibility in smart girds

Ilari Alaperä[a,*], Samuli Honkapuro[b], Janne Paananen[c]

[a] Fortum Power and Heat Oy, Keilalahdentie 2-4, 021 Espoo, Finland
[b] Lappeenranta University of Technology (LUT), Skinnarilankatu 34, 53850 Lappeenranta, Finland
[c] Eaton EMEA, Koskelontie 15, 02920 Espoo, Finland

HIGHLIGHTS

- Data centers have a lot of excess capacity due to their redundant design.
- This capacity can be used to provide ancillary services, such as primary regulation.
- UPS systems are technically and economically viable in primary regulation.
- Additional stress to battery systems is within battery specifications.
- Participation to primary regulation can create significant revenue for data centers.

ARTICLE INFO

Keywords:
Demand response
Primary frequency regulation
Data center
UPS
Battery
Smart grid

ABSTRACT

Data centers have a significant potential to become a major source of flexibility in smart girds. They consume currently roughly 3% of all the electricity produced globally and are expected to only increase their consumption as the world becomes more connected and digitalized.

As data centers are required to operate without any interruptions, they use power protection systems and energy storages. This paper investigates the technical and economic feasibility of dual-purposing these power protection systems, the uninterruptible power supplies, and their batteries in data centers to perform primary frequency regulation services. While the topic of data centers and demand response has been extensively covered in the current scientific literature, the focus has been on the demand response enabled by server workload shifting or hardware-enabled peak shaving. Based on an extensive literature review, there is a knowledge gap in the literature concerning primary frequency regulation and dynamic response enabled by modern power electronics systems in data centers. In this paper, this knowledge gap is bridged by suggesting a novel approach of taking advantage of the bidirectional operations capabilities of the uninterruptible power supply systems, thereby enabling them to provide dynamic power response from their battery systems.

The feasibility of this approach is examined with the proposed method, which includes (1) an analysis of the required energy for primary regulation and the availability of this energy in a typical data center, (2) a simulation of activation events and their impact on the service life of the battery systems, (3) reaction speed and reliability considerations of the operations, and (4) an economic feasibility and balancing market analysis.

The results show that as primary frequency regulation is an energy nonintensive service and data center battery systems are by design oversized for redundancy reasons, typical data centers have more than ample amounts of energy to participate in the primary regulation without jeopardizing their own processes. The results also show that by maintaining reasonable levels of usage, the battery systems can be operated within their specifications, and the demand response operations will not cause premature aging of the battery systems. The reaction speed of the power electronics is found to be very high and easily meet the current market requirements.

While the achievable revenue from the primary regulation service is small compared for example with the electricity costs of the data centers, it is still significant as there is little to no impact on the daily business of the data centers.

* Corresponding author.
E-mail address: ilari.alapera@fortum.com (I. Alaperä).

https://doi.org/10.1016/j.apenergy.2018.07.056
Received 27 April 2018; Received in revised form 28 June 2018; Accepted 13 July 2018
Available online 31 July 2018

1. Introduction

As the amount of renewable and intermittent energy increases in electrical power systems globally, also the need for flexibility increases [1]. This flexibility has been traditionally provided by large power plants, but now, as an increasing proportion of production is becoming difficult or even impossible to adjust, new sources of flexibility are needed. Attention is turning to flexible consumption, and the topic of demand response has been addressed in multiple scientific publications. For example, Muhssin et al. and Xu et al. studied the usage of aggregated household consumption assets, such as refrigerators and heat pumps, as sources of ancillary services in [2,3], and Jia et al. investigated the usage of electric vehicles (EVs) in frequency regulation in [4]. Batteries and energy storages and their feasibility in grid support applications, such as primary regulation, have been extensively studied both in academic literature and commercial demonstrations. For example, Zheng et al. analyzed the economic and environmental benefits of different dispatch strategies of a large number of residential energy storage systems in [5], Shi et al. analyzed multipurpose usage of battery systems for peak shaving and frequency regulation in [6], and Cheng et al. and Brivio et al. covered the combined use cases of primary frequency regulation and energy arbitrage in [7,8]. The demand response potential of power-intensive industries has also been studied lately; Otashu et al. proposed a metric of analyzing the available load reduction in the industry in [9], and Ramin et al. presented a case example of a metal casting process in [10].

Data centers are the power-intensive industry of the modern age. They are highly redundant digital factories and among the largest energy consumers in the world, and their power consumption is expected to increase significantly over the following decades. This development is driven by increasing digitalization and a growing amount of data being transferred and processed [11,12]. To achieve the high uptime requirements, data centers are designed with significant amounts of inbuilt flexibility in the form of electrochemical storage (i.e. batteries) and redundant power electronic systems. These design choices and significant power consumption make data centers attractive candidates for demand response participation.

Currently, there are few data centers participating in grid support activities, and while some companies have already begun to commercialize these activities [13,14], the majority of the data center flexibility potential still remains an untapped resource. The participation of data centers in demand response has also been covered in the scientific literature. For example, Mamum et al. addressed the topic in [15,16], where demand response was studied from the perspective of performing peak-shaving with Li-ion batteries, while Li et al. focused on demand response enabled by IT load shifting in [17]. However, as the extensive literature review (presented below) shows, there is a knowledge gap in the current literature concerning data center participation in primary frequency regulation. In particular, the key questions still unanswered are: how data center power protection systems could enable dynamic regulation, and what is the technical and economic feasibility of such an approach.

This paper presents a novel way for the data centers to participate in grid support (specifically in primary frequency regulation) by actively using their uninterruptible power supply (UPS) systems and batteries to balance the grid, instead of shedding their loads by off-gridding their systems, or performing workload shifting. The technical and economic feasibility of this approach is analyzed methodologically from several viewpoints.

The main research questions of this paper are: could a data center with a battery system perform primary frequency response in an economically feasible way without significant risks to their primary business, and how much additional stress would be exerted on the existing batteries during these grid support operations.

The structure of the paper is the following: The results of a literature review are presented in the second section. The third section introduces primary frequency regulation. The fourth section explains typical UPS topologies and their inherent excess capacities deployed in data centers through selected example configurations. The fifth section explains the key differences between the commonly applied method of providing grid services and the proposed approach. The sixth section investigates the technical feasibility of performing dynamic upwards regulation with an UPS system. The seventh section presents results from a frequency analysis and a simulation model to estimate the additional stress exerted on the UPS systems and their batteries while participating in primary regulation, and references it to typical battery cycle life characteristics. The eighth section provides discussion, and conclusions are drawn in the final section.

2. Literature review

The subject of demand response (DR) in the data center space is extensively covered in the current scientific literature. The research can be divided into two main categories; (1) DR enabled by "IT knobs" (basically workload management of the servers) and (2) DR by data center hardware (e.g., UPS systems, air conditioning, generators, and additional on-site generation). By far, the majority of the research focuses on the first case, that is, server management, and specifically, how to enable implicit demand response. The term 'implicit demand response' refers to (1) optimizing the electricity consumption of a data center for example by peak shaving to reduce grid connection costs, (2) limiting server power consumption during periods of high electricity prices, or (3) spatial workload shifting to gain savings from regional price differences. Li et al. [17] modeled the effects of spatial and temporal spreading of IT workloads to take advantage of electricity price differences between different price regions and times of the day. Similar research was presented by Ruddy et al. in [18], where they introduce a methodology for shifting global demand and calculate the resulting cost and CO_2 emissions savings and potential DR revenue from capacity payments in the Irish electricity markets. Further, in [19], Liu et al. proposed several algorithms for optimizing electricity cost and the usage of renewable energy for data center operations.

A more extensively studied subtopic has been the role and potential of multitenant or colocation (colo) data centers, especially in emergency demand response (a type of ancillary service, where independent system operators (ISOs) contract resources to respond to their DR signals, dispatched when a grid power balance is in jeopardy). Tran et al. have investigated the subject and related topics in multiple publications; in [20], they presented a simulation model and related results for cost optimizing EDR activations in a mixed-use building with data center (server) loads (workload management), HVAC systems, and backup generators. In their other publications they have studied how to incentivize the colo tenants to participate in the EDR by proposing different schemes and analyzing their convergence rates [21,22]. Similar research was presented by Kishwar et al. in [23] and Sun et al. in [24]. Guo et al. also examined how to incentivize colo tenants' participation in the EDR and applied the Nash bargaining theory to coordinate the tenants' participation and revenue gain in [25]. Again, Zhan et al. [26] proposed (and mathematically proved) a pricing model for colo data center operators that would include a reward component for tenants with flexible processes. The colo operator would benefit from reduced grid tariffs (resulting from peak load reduction). There are also several conference publications on the topic of using server management for DR, each with a slightly different focus. Wang et al. presented a DR framework model (server workload management) in [27], and further, they introduced an electricity cost optimization algorithm in [28], which is much similar to the work presented by Baharm et al. in [29,30].

In addition to the above-mentioned work related to server load management, there are several extensive papers that address the topic of using hardware systems in data centers to perform demand response operations. The most relevant research with respect to the scope of this paper has been conducted by S. Gonvindan, A. Sivasubramaniam, B. Ungaokar, A. Mamun, I. Narayanan, and H. Fathy. These authors have published several high-value papers, such as [15], where they provide an in-depth study on the option of using learning algorithms to determine the aging process of lithium-ion batteries in data centers performing demand response operations. In [16] they studied the optimization of cost savings generated by performing peak shaving and Li-ion battery degradation as a result of constant usage and cycling. They also presented related research results in [31], where they extended the above-mentioned optimization to include physics-based models for battery performance and degradation in combination with stochastic models of data center demand. In [32] they posed a question: "Should we dual-purpose energy storage in data centers for power backup and demand response?" As the title suggests, they investigated whether it would make economic (in terms of total cost of ownership) and technical sense to dual-purpose UPS battery systems to perform demand response (peak shaving) alongside their primary function of providing backup power. Their conclusion is that provisioning lead-acid batteries for peak power load needed to handle power outages (backup power usage) already comes with a sufficient energy capacity to also handle DR operations. Furthermore, they have published several conference papers that study the subject of dual-purposing data center batteries for peak shaving, such as [33,34]. What is common for all the papers mentioned above is that they focus solely on implicit demand response (price signal driven).

Cupelli et al. investigated in [35] how a combination of a battery energy storage system (BESS; here a UPS battery system), HVAC, and IT workload management could be used in price- and incentive-based (implicit/explicit) DR. However, the incentive-based demand response uses a three-stage control signal, which is, by nature, significantly different from primary frequency regulation. Li et al. examined integrated power management of data centers and electric vehicles (EVs) in [36]. Their work features explicit DR operations, specifically frequency regulation. However, the addressed frequency regulation differs from the primary frequency regulation we have studied in this paper, as the one covered by Li et al. is an ISO-based control signal rather than a requirement to react to the grid frequency. Moreover, their paper focuses on presenting a control framework to incorporate different assets, rather than on the specifics of data center UPSs in frequency regulation.

Apart from the last few publications, the focus of academic research has so far been on implicit demand response, specifically peak shaving. The lack of research on explicit demand response, and particularly, primary frequency regulation, can mainly be explained by the fact that explicit demand response markets are still in the development phase, and active, commercially available markets can be found only in a couple of European countries and a few other locations globally [37–39].

Based on the above literature review, it can be concluded that to the authors' knowledge, there is a significant knowledge gap in the scientific literature concerning primary frequency regulation enabled by data center hardware, such as UPS systems and their batteries. In this paper, we aim to bridge that knowledge gap by providing a technically and economically feasible method for data center participation in primary frequency regulation services with the batteries of UPS systems.

3. Primary frequency regulation

The global market size for uninterruptible power system (UPS) batteries is estimated to be worth $5.5 billion annually [40]. Data centers represent a significant proportion of this market as they are one of the major UPS end users. For most of the time, these batteries are sitting at full charge, standing by for a grid disturbance or failure, during which they will deliver energy for the critical loads protected by the UPSs. The nature of the primary application (i.e., rare events with a short duration) of UPSs could allow them to be dual purposed for other applications, one of them being primary frequency containment reserves.

Primary frequency response (or frequency containment reserves, FCR) is a type of an ancillary service intended to mitigate the imbalances between electricity production and consumption on short timescales from a few seconds up to several minutes. In case of a longer disturbance, these primary reserves will be gradually replaced by slower reserve types (secondary and tertiary reserves). Typically, the grid operators (transmission system operations, TSOs, in the European power system) are responsible for upholding or acquiring these reserves [41–43].

3.1. Normal operations reserve and disturbance reserve

Primary regulation reserves can be divided into two main types; a normal operations reserve and a disturbance reserve. Normal operations reserves are intended to handle small deviations in frequency and are constantly active to prevent frequency from drifting away from the nominal window of operations. Disturbance reserves are intended to handle sudden changes in the balance between power generation and consumption, such as unexpected losses of electricity production of a power plant or disconnection of a major transmission line. These reserves are specified to be activated when the frequency has already deviated significantly from the nominal, and for that reason, the activation must be faster than that of the normal operations reserves. When normal operations reserves have to be activated within minutes, the disturbance reserves have to be activated within seconds [43,44].

The currently dominant battery technology (lead-acid based batteries) in data centers is not technically capable of providing normal reserve operations because of the cycle-life limitations and intolerance to operate at a partial state of charge (PSoC) [45]. However, Li-ion batteries have been a constant topic of discussion in data center forums for several years, and are forecasted to gain a significant market share in the future [46–49] mostly owning to their several technical advantages, such as a smaller footprint, a longer service lifetime, and a lower operations and maintenance cost. The discussion section presents the idea of using lithium-ion-based batteries with intent to provide normal reserve operations. To this end, the market requirements and market prices for normal reserves are briefly addressed in this article.

3.2. Primary frequency regulation in the Nordic Countries

For example, in the Nordic power system, the primary frequency regulation reserves are divided into two products; frequency containment reserves for normal operations (FCR-N) and frequency containment reserves for disturbances (FCR-D). FCR-N is a bi-directional normal operations reserve, whereas FCR-D is an only upwards regulating disturbance reserve.

The current regulation states that FCR-D starts to activate at 49.90 Hz and should be fully activated when frequency reaches 49.50 Hz. Activation between 49.90 Hz and 49.50 Hz is expected to be linear. The reaction speed is defined as 50% reaction within 5 s from a frequency deviation and 100% activation within 30 s [44,50].

FCR-N is defined to be active when the grid frequency is between 49.90 Hz and 50.10 Hz. Thus, the reserve is regulating with a full upwards power at 49.90 Hz and with full downwards power at 50.10 Hz. Current regulation defines the reaction speed requirement to be full activation within 3 min of a frequency change [44,50].

FCR-N is basically constantly activated, whereas FCR-D is activated only if the frequency goes below 49.90 Hz. Historically, frequency has been quite good in the Nordic power system; for example, the frequency was outside the nominal limit (49.90–50.10 Hz) for 14000 min in year 2016. 6500 min of these frequency anomalies were underfrequency situations (i.e., times when FCR-D would have been activated). It should be noted that while the frequency does go outside the nominal limits, large deviations are rare (i.e., FCR-D is activated, but with a limited power required). For example, the grid frequency in the Nordic

Table 1
Average market prices for primary reserves in the Nordic markets in 2017 [52,53].

Market	Normal Reserve		Disturbance Reserve	
	Local Market	Price [€/MW/h]	Local Market	Price [€/MW/h]
FIN	FCR-N	20.87	FCR-D	3.39
SWE	FCR-N	23.52	FCR-D	7.43

system went below 49.70 Hz (requiring an activation of 50% or more from FCR-D) eight times during 2016, with an average event duration of 6.43 s and a maximum event duration of 11.90 s [51].

3.3. Market prices

As transmission system operators acquire these reserves from open electricity markets, the market prices are subject to variation depending on supply and demand. Table 1 presents the average hourly market prices in the Finnish and Swedish markets in 2017. The Finnish market also has a year market, with a slightly lower price (2.8€/MW h in year 2017) [52]. These prices are later used to illustrate the level of revenue that the primary frequency regulation can provide.

4. Data center UPS topologies and their excess capacity

Data centers are designed to have a maximal uptime. This is achieved (for example) by building redundant power protection systems within data centers. As a result, data center power protection systems are significantly overdimensioned during normal operations.

4.1. Different topologies

To consider the amount of excess capacity (energy and power), it is necessary to understand different power protection topologies in data centers. Typical topologies deployed in data centers are N, N + 1, 2N, and 2(N + 1) (Fig. 1). In order for a data center to receive a tier III (or higher) classification from the Uptime institute, a redundant power delivery path is required, meaning that the at least a 2(N + 1) UPS topology should be implemented [54,55].

4.1.1. N – Topology

In the N topology there is practically no redundancy or excess capacity. A critical fault in the UPS will jeopardize the power supply of the critical loads.

4.1.2. N + 1 Topology

A N + 1 system is designed so that a failure of a single UPS device will not endanger the ability of the system to supply electricity to the critical loads. An example configuration in Fig. 1 shows a N + 1 system consisting of four 1 MW UPSs. This system can supply 3 MW of power at the maximum to IT -loads, even if one of the UPS devices encounters a critical failure.

4.1.3. 2N Topology

A 2N system has two independent power delivery paths that are supplying the critical loads. The loads will not lose power even if one path is compromised. The example configuration in Fig. 1 has two systems, each with three 1 MW UPSs supplying 3 MW of critical loads.

4.1.4. 2(N + 1) Topology

In the 2(N + 1) topology, two independent and redundant power delivery paths are supplying critical loads. The example configuration in Fig. 1 is a 2(N + 1) system made of two independent N + 1 systems (each with four 1 MW UPSs). This system is able to supply a maximum of 3 MW of power to the critical loads even in an event where another power path is out of commission and one device from another power path simultaneously encounters a critical failure.

4.2. Excess energy in different configurations

This subsection uses the above-mentioned system examples and illustrates the available excess energy in these systems. This excess energy is a result of redundancy. During normal operations, all the UPSs are online and the combined energy stored in their battery systems is larger than the load requirement.

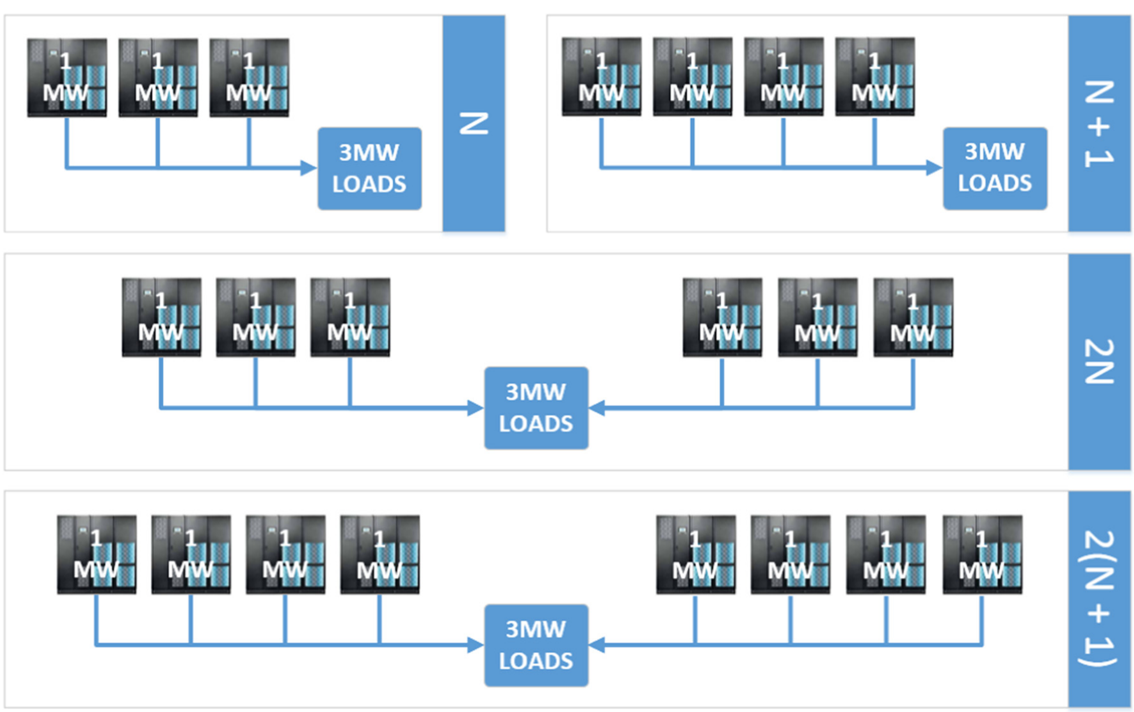

Fig. 1. Typical UPS topologies deployed in data centers [56].

Table 2
excess energy amounts inherent to different UPS system topologies (examples).

Design Topology	Number of 1 MW UPSs	Total amount of energy in the battery systems [kWh]	Excess energy in the battery systems [kWh]
N	3	500	0
N + 1	4	667	167
2N	6	1000	500
2N + 1	8	1333	833

To illustrate this, each UPS system is assumed to have a battery system with 10 min of autonomy time (a typical design value in data centers). For the purpose of this paper this is assumed to correspond to 167 kWh of energy (1 MW ∗ 10 min). For the critical loads expected to require full power autonomy for this 10 min, this is assumed to correspond to 500 kWh (3 MW ∗ 10 min). The third column in Table 2 presents the sum of energy in all the battery systems in each topology and the excess energy (4th column) is calculated by subtracting the energy requirement by the load from the sum.

It should be noted that the actual amount of energy available in lead-acid batteries depends on the discharge current (the higher discharge current, the less energy available), this relation is in accordance to well-known Peukert's Law [57]. For frequency response operations, the available energy amount is expected to be greater as typically the required discharge current that the batteries are subject to is less than in the main UPS application.

5. Different ways of using data center UPSs for primary regulation

5.1. Data center islanding

The most straightforward way for a data center to participate in primary frequency regulation is to install a frequency-controlled breaker after the grid connection point. If the frequency goes below a set threshold, the breaker will open and UPSs will feed energy to the critical loads until the on-site generators (genset) have had sufficient time to start.

The grid will see an upwards regulating effect, of the size of the power consumption of the critical loads, as they no longer get their power from the grid, but from on-site systems. In this case, the functionality of the power protection systems is the same as during a grid outage event. Fig. 2 illustrates the concept and the related energy flows. Doing this would allow the data center to participate in upwards regulation as a step-activated reserve.

Most, if not all, data centers currently participating in grid support activities apply this method. Some use a physical breaker in the critical power path, others take an activation signal from a frequency relay and use that information as a trigger for automation systems to perform the similar operation.

5.2. Dynamic upwards regulation

In dynamic upwards regulation, the UPS system would be used to modulate the power consumption from the grid with the help of a battery system. In case of a frequency disturbance, the UPS system will discharge energy from the battery system. Depending on the current power requirement of the critical loads and the regulation need, this discharged energy will be either consumed fully by the on-site loads or, if the regulation need exceeds the power consumption of the loads, power will be fed back to the grid. The UPS system has to be compatible with the functionality; mainly the rectifier has to be able to perform two-way operations to enable feeding back power to the grid. In case of a longer disturbance, the on-site back-up generators could be used to provide additional energy if the battery state of charge level is approaching the state of charge (SoC) level allocated for ancillary services. Fig. 3 illustrates the concept and the related energy flows.

The main difference from data center islanding is that in this approach the regulation power is not limited to the load power, but full UPS capacity can be utilized.

5.3. Potential market income of different data center topologies and participation methods

Table 3 illustrates the level of revenue that primary regulation could provide for data centers and how the revenue depends on the UPS topology (i.e., the level of redundancy). The revenues in the table are calculated for a 3 MW data center according to the example used in the previous section. Full availability and bid acceptance are assumed, and thus, the annual revenue is calculated simply by multiplying the available UPS power (P) with number of hours in a year (8760) and the average market price (p) (Eq. (1)).

$$Annual\ revenue = 8760 \times p_{avg.market\ price} \times P_{UPS} \quad (1)$$

6. Technical feasibility and reaction speed considerations

For a data center to participate in dynamic regulation, the UPS system has to be able to adjust its power consumption as described previously (i.e., the rectifier of the UPS needs to be able to perform bidirectional operations). Most of the modern UPS systems use IGBT (insulated gate bipolar transistor) based power electronics, which, at least in theory, are capable of performing these operations. It is more a question if these functions have been implemented in the control software of the UPS.

As an example, Fig. 4 presents the measurement results from a FCR-D prequalification measurement performed according to the Swedish TSO's (Svenska kraftnät) specification. The device under test (DUT) is a 100 kW 93PM UPS from Eaton. The UPS was subjected to a test signal (the orange line in Fig. 4). The test signal was a series of frequency values first going down from 49.90 Hz with 0.05 Hz steps, and once the frequency reached 49.50 Hz, upwards frequency steps of 0.05 Hz were

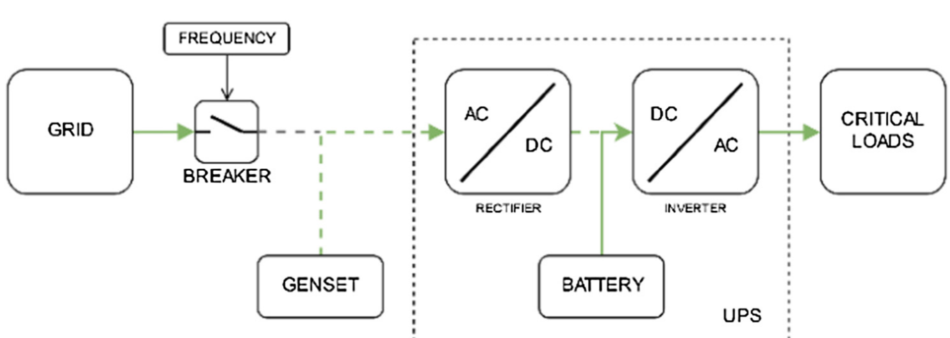

Fig. 2. Off-grid or islanding demand response power system configuration and the related power flows.

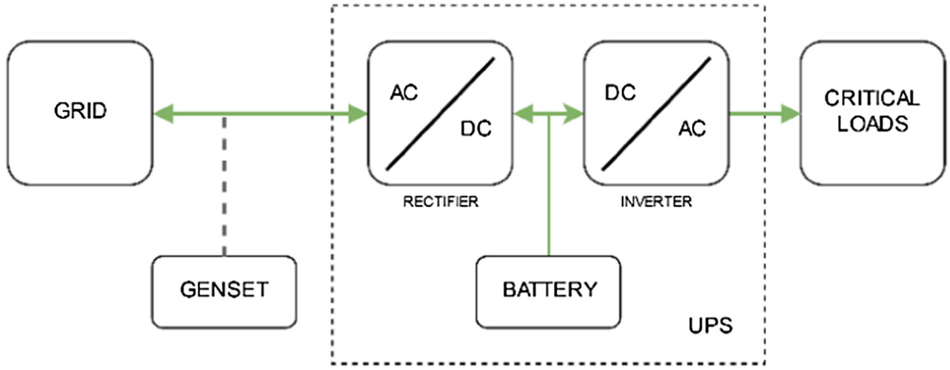

Fig. 3. Dynamic upwards regulation configuration and the related power flow.

Table 3
example of potential market revenue for a 3 MW data center with different UPS topologies in the Finnish and Swedish markets.

	Data center islanding		Dynamic regulation	
	Fin [k€/a]	Swe [k€/a]	Fin [k€/a]	Swe [k€/a]
N	90	195	90	200
N + 1	90	195	120	260
2N	90	195	180	390
2(N + 1)	90	195	240	520

issued until 49.90 Hz was reached. The blue line in the figure shows the UPS response, which is in compliance with the FCR-D activation requirements. It should be noted that the UPS was not connected to loads during the test, and thus, all the power was fed back to the grid.

6.1. Reaction speed

While the results of the prequalification test show that reaction to frequency changes is very fast, additional tests were performed under laboratory conditions in Eaton's facilities to thoroughly investigate the speed of reaction (among several other things). Fig. 5 is an oscilloscope capture from one of these tests. Voltage and current waveforms for the UPS input and output were recorded (one phase) along with the battery current. The communications channel (CAN) to issue the activation command was used as a measurement trigger. The DUT was a 200 kW 93PM UPS by Eaton. The UPS was loaded with a 100 kW constant load, and the activation signal that was given requested a full 200 kW activation.

The results show that the output waveforms (current and voltage) of the UPS remains unchanged during the event, indicating that operations have no impact on the critical loads. Before the CAN burst (top of the figure), the input and output currents are synchronized and uniform and the battery current is close to zero, as would be expected (i.e., UPS is feeding energy through the conversions to the loads and no energy is being drawn from the batteries). Quickly (in approx. 5 ms) after the CAN burst has been issued, the battery current changes followed by a change in the input current. The UPS starts to draw energy from the battery systems (shown by an increase in the battery current) and instead of drawing 100 kW from the grid and feeding that to the loads, the UPS is actually drawing 200 kW from the batteries and feeding 100 kW of it to the loads and 100 kW back to the grid. This can be observed from the 180-degree phase shift in the input current (and unchanged output current). The measurements show that it takes two to three

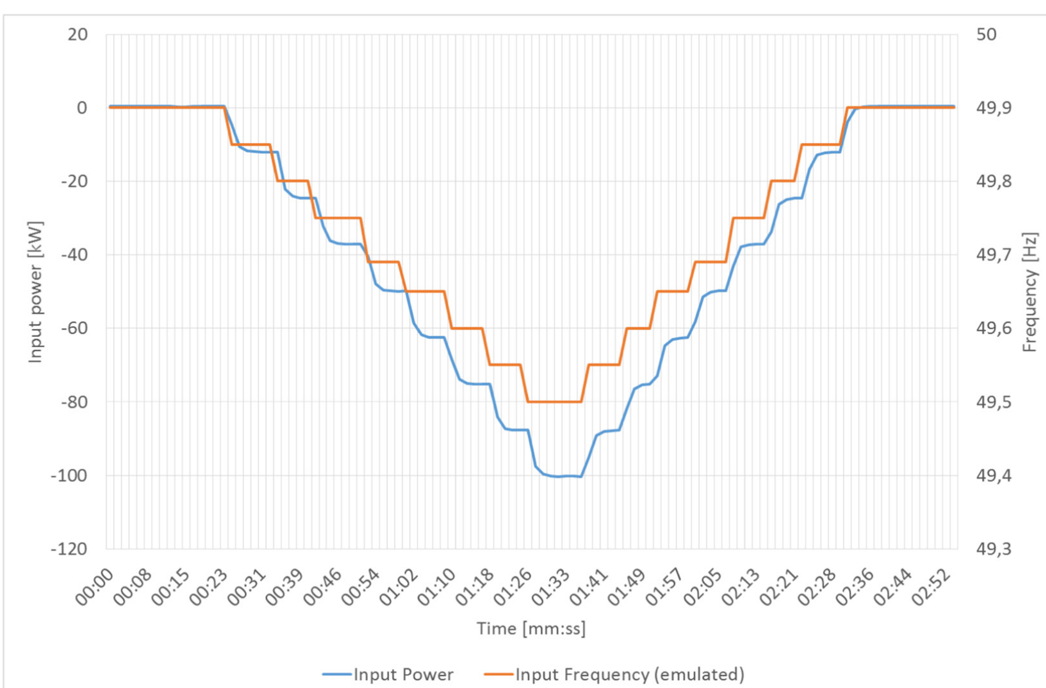

Fig. 4. Results of the linearity of the reaction tests.

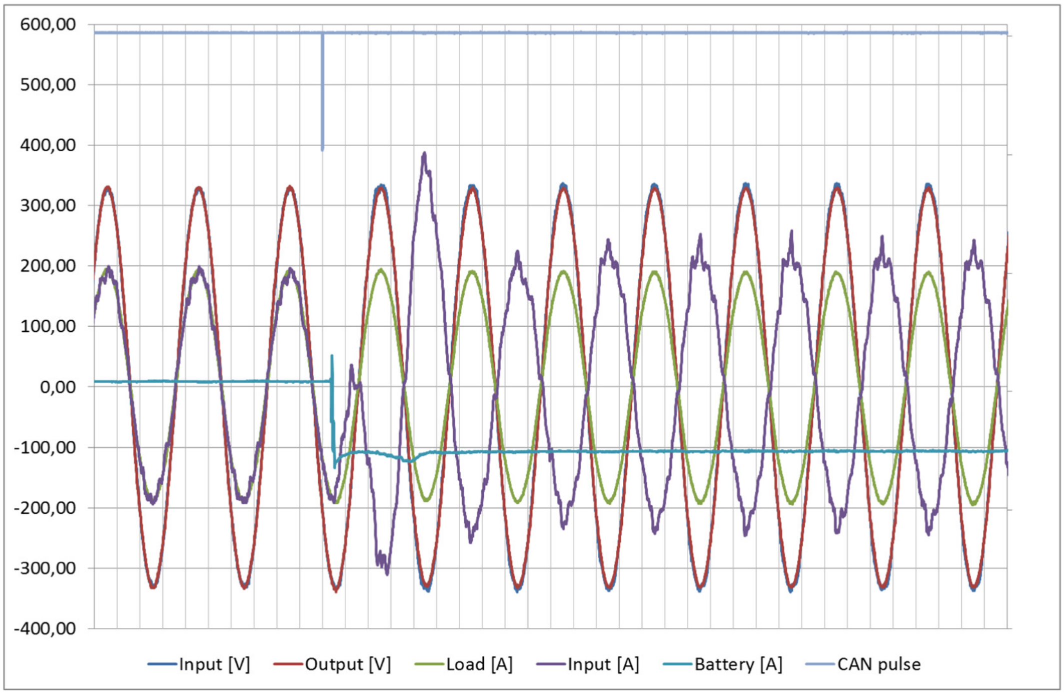

Fig. 5. Results of the speed of reaction tests.

cycles (50 Hz) for the UPS input to fully stabilize, but the UPS is outputting energy well within one cycle from the issuance of the command.

7. Frequency analysis, simulation model and results

7.1. Frequency analysis

A frequency analysis was conducted to investigate how sufficient the previously identified excess capacity would be in relation to the energy demand for primary regulation. Frequency data were analyzed and the energy demand (according to the FCR-D requirements) for each activation (an event where the frequency went below 49.90 Hz) was calculated. Frequency measurement data from the Nordic power systems (year 2015) were used as the input data. For reference, a similar analysis was also conducted for frequency data from the UK power system (year 2015). The regulating power was assumed to be 4 MW according to the N + 1 configuration presented previously.

The results of this preliminary analysis are presented in Fig. 6. The results show that the identified excess capacity (167 kWh limit in the figure) would have been more than enough in relation to the discharged energy and that the generators would have not been needed to start. The upper chart in Fig. 6 shows that for a Nordic data center the most energy requiring continuous activation in the year 2015 would have discharged 122 kWh from the battery systems. The lower chart shows that the corresponding discharged energy for a UK-based data center would have been 53 kWh. The discharged energy amounts are less than the value presented for the excess battery capacity in the previous sections (167 kWh). Thus, there would have not been a technical need to start the generators during these discharges.

7.2. Simulation model

A model was developed to simulate the state of charge (SoC) behavior of the batteries. The simulation input and outputs are described in Fig. 7. This model is modified based on the model presented in [58] so that it is more suitable for the case of the data center UPS. The SoC behavior information was used to study the additional stress exerted on the batteries while performing primary regulation operations.

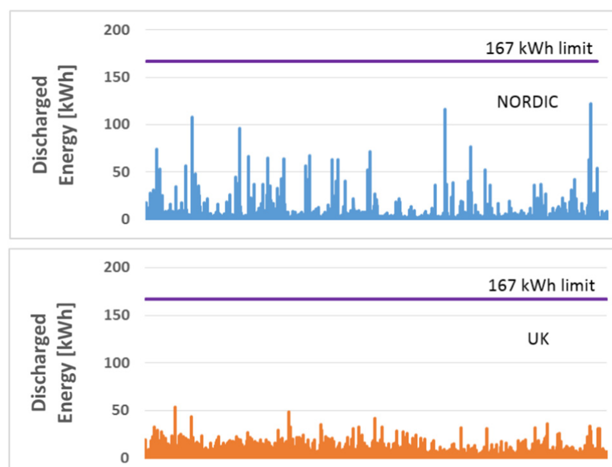

Fig. 6. Calculated discharged energies per activation for a N + 1 (4 × 1 MW) UPS system with the Nordic (above) and UK (below) frequency data from 2015.

7.3. Study of additional stress exerted to battery systems

The upper chart of Fig. 8 shows the SoC behavior of a single UPS battery system for a N + 1 (4 × 1 MW UPS) configuration, simulated with frequency data from year 2015 from the Nordic power system. In the simulations, the depth of discharge (DoD) was limited to 42 kWh (the excess capacity for a N + 1 UPS system according to Table 2 divided by the number of UPSs in the system). The simulation result shows that the UPS would have encountered roughly 200 charge/discharge events during that year. The corresponding results for a data center in the UK are presented in the lower chart of Fig. 8 showing approximately 270 charge/discharge cycles annually.

While the discharges are shallow (limited 42 kWh out of 167 kWh ~ roughly 25% SoC), the number of cycles throughout the expected service life of the battery systems would add up to significant figures in relation to the cyclic life expectancy: approx. 750 cycles with the DoD of 25% (Fig. 9) [59]. Limiting these cycles would be mandatory

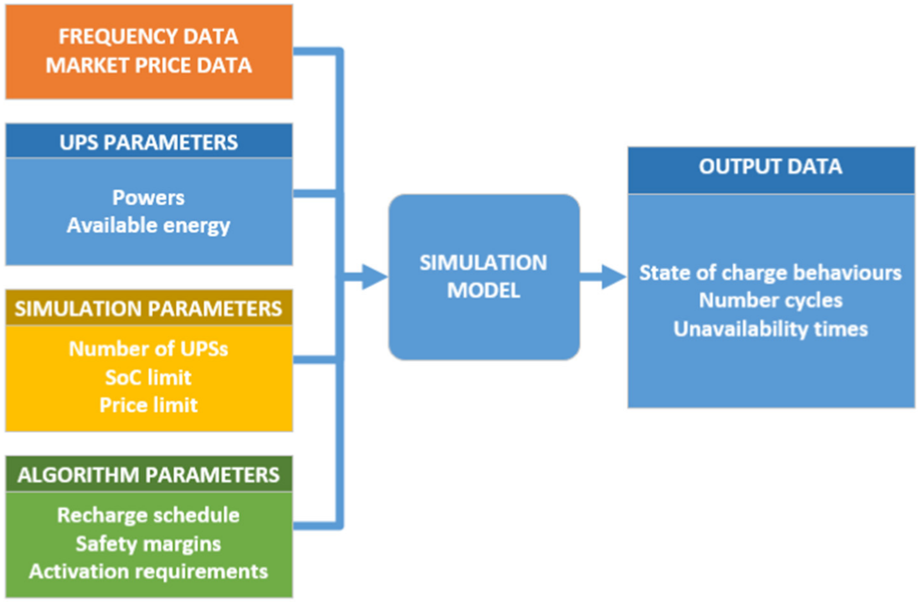

Fig. 7. SoC simulation model.

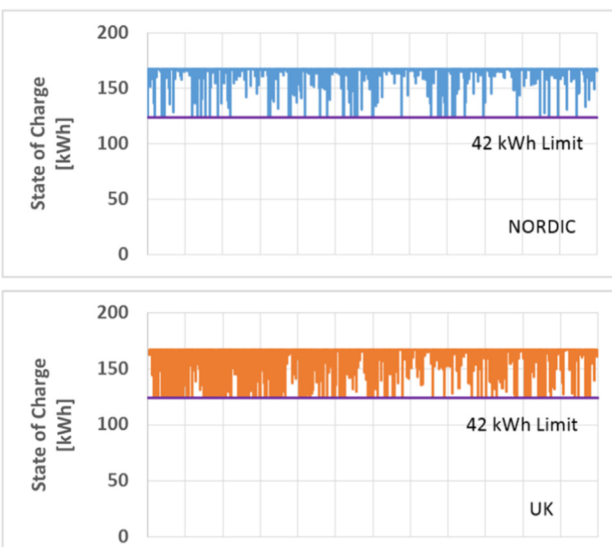

Fig. 8. State of charge behaviors of the most stressed UPS batteries, simulated with the 2015 frequency data from the Nordic (upper) and UK (lower) power systems.

to ensure that demand response participation would not significantly shorten the service life of the batteries. Depending on the target market structure, the cycle count can be effectively limited with aggregating several UPS and even data centers under one bid, thus reducing the stress for a single battery system. Another way to limit the cycle count is to bid for the hourly markets and to issue a limit price for participation.

Table 4 gathers the simulation results for different aggregate sizes and minimum bid prices. The effects of implementing different minimum bid prices for the relative market income are also studied. Price information used in the simulation is the hourly FCR-D prices for the year 2015 [52].

The results show that implementing minimum bid price levels will have a significant effect on the number of charge/discharge cycles that the batteries are subjected to by demand response operations. Implementing these minimum price levels will also have an effect on the relative market income, but as shown below, the decrease is moderate. Combining several UPSs into an aggregate will have a significant effect on the additional stress applied to individual battery systems. Additionally, the experienced cycle count can be significantly reduced if the market allows a selection of lower participation frequency (e.g., 49.80 Hz instead of 49.90 Hz).

A typical service life expectancy for a UPS battery system (VRLA, AGM, 10-year design life) is roughly between seven and eight years. In data center usage, these batteries are expected to encounter a very limited number of charge/discharge cycles a year; a design rule of thumb is that batteries will encounter an equivalent of one to two full discharge cycles a year, and thus, the batteries could be cycled significantly more without affecting the service life. As an example, adding 45 annual cycles with a DoD of 25% (per simulated result with 20 UPSs and a price limit of 10€/MW/h) would mean approximately 360 additional cycles during the service life of the battery system, still well within the specifications for cyclic performance, even considering the cycles the batteries will endure as a result of their primary operations.

8. Discussion

While data centers have multiple alternative ways to participate in demand response for example by managing IT loads, this paper focuses on primary regulation enabled by the inherent redundancy and energy storages in data centers. The purposed approach allows data center participation to demand response without impacting the power consumption profiles or the servers.

As shown in the paper, data centers can have a lot of underused assets that could be used for primary regulation. The exact amount of excess capacity depends heavily on the applied topology and the redundancy level of the data center. The figures presented in the paper serve as an example, but are highly related to real-life data center topologies and their power and back-up energy capacities. Further, in general, the more redundancy a data center has, the more excess capacity it will have to offer to the demand response markets.

The performed frequency analysis showed that the required amount of energy to perform primary regulation is limited, and the example configurations would have had more than ample amounts of excess energy capacity to participate in primary regulation and to have sufficient energy to support the critical loads at all times. Another important finding was the level of additional stress exerted on the batteries by the DR operations. It was found that by limiting the participation times (issuing reasonable price limits for the bids), the additional stress could

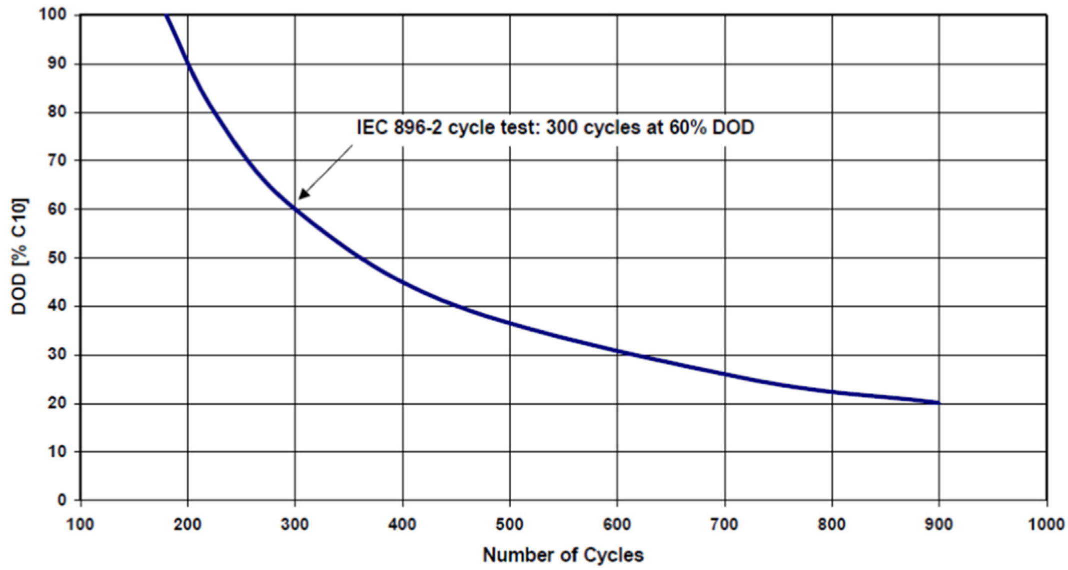

Fig. 9. Typical battery cycle life expectancy chart for a valve-regulated lead-acid (VRLA), absorbent glass mat (AGM) battery (Sprinter XP form Exide Technologies) [59].

Table 4
simulated number of charge/discharge cycles with different aggregate sizes, minimum bid prices and the effect of minimum bid prices on relative income.

Price limit	Aggregate size: 4 UPSs	Aggregate size: 20 UPSs	Aggregate size: 50 UPSs	Relative income [%]
No limit	205 cycles	130 cycles	122 cycles	100
5€/MW	140 cycles	86 cycles	81 cycles	94
10€/MW	72 cycles	45 cycles	42 cycles	83
15€/MW	58 cycles	37 cycles	34 cycles	79

be maintained within nominal specifications of the generally applied battery technology. The reliability aspect of participation was also investigated by several experiments in the laboratory, simulating various kinds of fault situations that could occur during DR operations. The simulation results show that, from the perspective of data centers, the suggested approach is technically feasible, and thus, answer the primary research question. The performance characteristics of the UPS systems are more than sufficient to provide response fast enough (~ 100 ms) to meet the requirements for primary frequency regulation in the observed markets (< 5 s). Based on this and the prequalification results from the tests performed for Svenska kraftnät, the suggested approach meets the technical requirements of the markets.

The economic feasibility of the approach was investigated by examining the potential market revenue in the Finnish and Swedish markets and comparing it with revenue from data center islanding. As the regulation power is independent of the data center power consumption, it will provide better economical results than data center islanding. The amount of revenue naturally depends on market prices, but for the example configurations, potential revenues (with 2017 prices) ranged from 90 k€ to 240 k€ in the Finnish markets, and from 200 k€ to 520 k€ in the Swedish markets. As previously shown in this paper, there are no significant additional investments and no additional wear and tear of the equipment, and thus, economic feasibility can be assumed.

This paper focuses on data centers with existing battery systems that apply the currently dominant lead-acid battery technology. While these batteries have been identified (in this paper) to be feasible in the upward-regulating primary disturbance reserve (FCR-D), they are not suitable for the more common and economically more valuable bidirectional normal operations primary reserve (FCR-N). This is mostly explained by the limited cycle life and inability to operate in a partial state of charge. UPSs equipped with a different battery technology, such as lithium ion (Li-ion), could be used in the normal operations reserves to provide better market access and higher revenue. For some time now, Li-ion batteries have been suggested as a replacement for traditional lead-acid batteries, but so far, only a few data centers have implemented them. Recent developments in the prices of lithium-ion batteries [60–62] alongside potential additional revenue from primary regulation could actually make a feasible business case for the data centers to change over from lead-acid batteries to lithium-ion ones. This is a subject of future work by the authors and will be covered for example in the forthcoming conference publication by the authors [63].

Recently, there has been a lot of discussion about the need to generate virtual inertia in order to compensate for the diminishing natural inertia on both the production and consumption sides of electricity generated by grid-connected rotating machinery [64]. This has led to debate on the role of primary frequency regulation, the adequacy of the current regulation, the possible need for tightening the reaction speed requirements, and demand for development of a new, faster market. New Zealand, for example, has a balancing product requiring a reaction time of one second or less [64]. Furthermore, the Norwegian TSO (Statnett) has recently announced a fast frequency reserve (FFR) pilot, which aims at investigating the feasibility of a fast-responding balancing product to support the grid during times of low inertia. The market requirement is full activation of the resource within 2 s [65].

It is noteworthy that at the moment, demand-side participation in primary regulation with aggregated assets is possible only in a few market areas in Europe (and the rest of the world), but recently, more and more markets have been opening up; more information on the development in Europe can be found for example in [37]. Further, there is a market unification process going on to uniform the FCR in Europe and a major market restructuring in the UK [66,67]. In addition to national regulation, there may be distribution-grid-related limits and/or regulation that may prohibit a data center from participating in the operations, as in some cases the data center might be inputting energy to the grid.

9. Conclusion

This paper investigates the technical and economic feasibility of data center participation in primary frequency regulation by adopting an approach where the UPS systems of the data centers are used to dynamically adjust the grid loading with energy from batteries.

Our research shows methodologically that the approach is technically feasible from the perspective of the data center, and it meets the requirements of the balancing market. Data centers have the required capacity, and further, the excess stress exerted on their systems and batteries can be maintained within the nominal operational limits. Reliability and security considerations have been taken into account, and the approach has been vigorously tested. The applied UPS technology has also been used for several years as part of autonomous battery test functions in the UPSs. The response speed of the UPSs is significantly faster than the current and foreseeable future market requirements.

As the suggested approach requires no significant additional investments and has practically no impact on the service and maintenance costs, while providing reasonable revenue stream, its economic feasibility can be demonstrated.

The knowledge gap identified in the literature on data centers and their participation in primary frequency regulation is addressed by presenting a technically and economically viable solution of using UPSs with lead-acid batteries. As discussed above, the dominant battery technology, however, limits the operations to a specific type of primary frequency regulation. Nevertheless, the price and technology development of Li-ion batteries will create new usages for the proposed approach and the applied technology. Therefore, future work on the topic is considered to include a study of the feasibility of Li-ion battery systems in data centers with intent of dual-purposing them for primary frequency regulation, as well as an analysis of opportunities in the area of inertia compensation (i.e., synthetic or virtual inertia) enabled by the extremely fast response of the power electronics.

References

[1] Yin R, Kara EC, Li Y, DeForrest N, Wang K, Yong T, et al. Quantifying flexibility of commercial and residential loads for demand response using setpoint changes. Appl Energy 2016;177:149–64.

[2] Mussin MT, Obaid ZA, Cipcigan LM, Sami SS. Potential of demand side response aggregation for the stabilization of the grids frequency. Appl Energy 2018;220:643–56.

[3] Xu Z, Østergaard J, Togeby M. Demand as frequency controlled reserve. IEEE Trans Power Syst August 2011;26(3):1062–71.

[4] Jia H, Li X, Mu Y, Xu C, Jiang Y, Yu X, et al. Coordinated control for EV aggregators and power plants in frequency regulation considering time-varying delays. Appl Energy 2018;210:1363–76.

[5] Zheng M, Wang X, Meinrenken CJ, Ding Y. Economic and environmental benefits of coordinating dispatch among distributed electricity storage. Appl Energy 2018;210:842–55.

[6] Shi Y, Xu B, Wang D, Zhang B. Using battery storage for peak shaving and frequency regulation: joint optimization for superlinear gains. IEEE Trans Power Syst 2017.

[7] Cheng B, Powell W. Co-optimizing battery storage for the frequency regulation and energy arbitrage using multi-scale dynamic programming. IEEE Trans Smart Grid 2017.

[8] Brivio C, Mandelli S, Merlo M. Battery energy storage system for primary control reserve and energy arbitrage. Sustain Energy Grids Netw 2016;6:152–65.

[9] Otashu JI, Baldea M. Grid-level "battery" operation of chemical processes and demand-side participation in short-term electricity markets. Appl Energy 2018;220:562–75.

[10] Ramin D, Spinelli S, Brusaferri A. Demand-side management via optimal production scheduling in powerintensive industries: the case of metal casting process. Appl Energy 2018;225:622–36.

[11] Grand View Research. Data center power market analysis by product (PDU, UPS, Busway), by end-use (IT & Telecommunications, BFSI, Government, Energy, Healthcare, Retail), by region, and segment forecasts, 2014–2025. Grand View Research; 2017.

[12] National Resources Defense Council. Data center efficiency assessment; 2014.

[13] Ciaran F. A data center perspective on demand response. 20 February 2013. [Online]. Available: http://www.datacenterdynamics.com/content-tracks/power-cooling/a-data-center-perspective-on-demand-response/73892.fullarticle [accessed 27 March 2018].

[14] Mission Critical. Webair and EnerNOC turn data centers into 'Virtual Power Plants' through demand response. 9 March 2011. [Online]. Available: https://www.missioncriticalmagazine.com/articles/83860-webair-and-enernoc-turn-data-centers-into-virtual-power-plants-through-demand-response [accessed 27 March 2018].

[15] Mamun A, Sivasubramaniam A, Fathy H. Collective learning of lithium-ion aging model parameters for battery health-conscious demand response in datacenters. Energy 2018.

[16] Mamun A, Narayananb I, Wang D, Sivasubramaniam A, Fathy HK. Multi-objective optimization of demand response in a datacenter with lithium-ion battery storage. J Storage Mater 2016;7:258–69.

[17] Li J, Bao Z, Li Z. Modeling demand response capability by internet data centers processing batch computing jobs. IEEE Trans Smart Grid 2015;6(2):737–47.

[18] Ruddy J, O'Malley M. Global shifting of data centers demand. 5th IEEE PES innovative smart grid technologies Europe (ISGT Europe), Istanbul. 2014.

[19] Liu Z, Lin M, Wierman A, Low S, Andrew LLH. Greening geographical load balancing. IEEE/ACM Trans Netw 2015;23(2):657–71.

[20] Tran NH, Pham C, Ren S, Hong CS. Coordinated energy management for emergency demand response in mixed-use buildings. IEEE international conference on ubiquitous wireless broadband (ICUWB), Montreal, QC, Canada. 2015.

[21] Tran NH, Pham C, Ren S, Han Z, Hong CS. Coordinated power reduction in multi-tenant colocation datacenter an emergency demand response study. IEEE international conference on communications (ICC), Kuala Lumpur, Malaysia. 2016.

[22] Nguyen MNH, Kim D, Tran NH, Hong CS. Multi-stage stackelberg game approach for colocation datacenter demand response. 19th Asia-Pacific network operations and management symposium (APNOMS), Seoul, South Korea. 2017.

[23] Ahmed K, Islam MA, Ren S. IEEE/ACM international conference on computer-aided design (ICCAD), Austin, TX, US. 2015.

[24] Sun Q, Wu C, Ren S, Li Z. Fair rewarding in colocation data centers Truthful mechanism for emergency demand response. IEEE 23rd international symposium on quality of service (IWQoS), Portland, OR, USA. 2015.

[25] Niu L, Guo Y, Li H, Pan M. A nash bargaining approach to emergency demand response in colocation data centers. IEEE global communications conference (GLOBECOM), Washington, DC, USA,. 2016.

[26] Zhan Y, Ghamkhari M, Xu D, Ren S, Moshenian-Rad H. Extending demand response to tenants in cloud data centers via non-intrusive workload flexibility pricing. IEEE Trans Smart Grid (Early Access) 2016.

[27] Wang C, Urgaonkar B, Wang Q, Kesidis G. A hierarchical demand response framework for data center power cost optimization under real-world electricity pricing. IEEE 22nd International Symposium on Modelling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), Paris, France. 2014.

[28] Wang R, Kandasamy N, Nwankpa C, Kaeli DR. Datacenters as controllable load resources in the electricity market. IEEE 33rd international conference on distributed computing systems (ICDCS), Philadelphia, PA, USA. 2013.

[29] Bahram S, Wong VWS, Huang J. Demand response for data centers in deregulated markets a matching game approach. IEEE international conference on smart grid communications, Dresden, Germany. 2017.

[30] Bahrami S, Wong VWS, Huang J. Data center demand response in deregulated electricity markets. IEEE Trans Smart Grid (Early Access) 2018.

[31] Mamun A, Narayanan I, Wang D, Sivasubramaniam A, Fathy HK. A stochastic optimal control approach for exploring tradeoffs between cost savings and battery aging in datacenter demand response. IEEE Trans Control Syst Technol 2018;26(1):360–7.

[32] Narayanan I, Wang D, Al Mamun A, Sivasubramaniam A, Fathy K. Should we dual-purpose energy storage in datacenters for power backup and demand response? 6th Workshop on power-aware computing and systems, hot power, Broomfield, CO, USA. 2014.

[33] Govindan S, Sivasubramaniam A, Urgaonkar B. Benefits and limitations of tapping into stored energy for datacenters. 38th Annu int symp computer architecture (ISCA), San Jose, CA, USA. 2011.

[34] Mamun A, Narayanan I, Wang D, Sivasubramaniam A, Fathy K. Battery health-conscious online power management for stochastic datacenter demand response. American control conference (ACC), Boston, MA, USA. 2016.

[35] Cupelli LJ, Schultz T, Jahangiri P, Fuchs M, Monti A, Muller D. Data center control strategy for participation in demand response programs. IEEE Trans Ind Inf (Early Access) 2018.

[36] Li S, Brocanelli M, Zhang W, Wang X. Integrated power management of data centers and electric vehicles for energy and regulation market participation. IEEE Trans Smart Grid 2014;5(5):2283–94.

[37] SEDC. Explicit demand response in Europe, mapping the markets 2017. Belgium: Smart Energy Demand Coalition (SEDC); 2017.

[38] PJM. PJM learning center – ancillary services market. [Online]. Available: http://learn.pjm.com/three-priorities/buying-and-selling-energy/ancillary-services-market/reserves.aspx [accessed 24 April 2018].

[39] AEMO. Guite to ancillary services in the national electricity market. AEMO; 2016.

[40] Technavio,. Global UPS battery market from 2017–2021. London: Technavio; 2017.

[41] Entso-e. Network code on load-frequency control and reserver. 28 June 2013. [Online]. Available: https://www.entsoe.eu/fileadmin/user_upload/_library/resources/LCFR/130628-NC_LFCR-Issue1.pdf [accessed 13 October 2017].

[42] Entso-e. Supporting document for the network code on load-frequency control and reserves. 28 June 2013. [Online]. Available: http://www.acer.europa.eu/Official_documents/Acts_of_the_Agency/Annexes/The%20Network%20Code%20on%20Load-Frequency%20Control%20and%20Reserves%20submitted%20on%2028%20June%202013.pdf [accessed 13 October 2017].

[43] Fingrid. Frequency control process. [Online]. Available: http://www.fingrid.fi/EN/ELECTRICITY-MARKET/RESERVES/RESERVETYPES/Pages/default.aspx [accessed 22 September 2017].

[44] Fingrid,. Application manual for frequency controlled reserves. Helsinki: Fingrid; 2017.

[45] Battery University. BU-804b: sulfation and how to prevent it. 22 Sebtember 2016. [Online]. Available: http://batteryuniversity.com/learn/article/sulfation_and_how_to_prevent_it [accessed 28 November 2017].

[46] Lansburg S, Jehoulet C. Innovative modular Li-ion battery system delivers high power support for data center UPS installations. Intelec 2013; 35th International

telecommunications energy conference, SMART POWER AND EFFICIENCY, Hamburg, Germany. 2013.
[47] Jung S-M, Ricci B, Chung G. Lithium ion battery system in data centers. 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), Osaka, Japan. 2015.
[48] Lansburg S, Siret C. Green and compact backup battery solution for a Green Datacenter. Telecommunications energy conference (INTELEC), Osaka, Japan. 2015.
[49] Schneider Electric. Bloomberg forecasts Li-ion batteries are poised for big gains in data center UPS systems. 20 July 2017. [Online]. Available: http://blog.schneider-electric.com/datacenter/2017/07/20/bloomberg-forecasts-li-ion-batteries-data-center/ [accessed 13 October 2017].
[50] Kraftnät Svenska. Regler för upphandling och rapportering av FCR-N och FCR-D. Stockholm: Svenska Kraftnät; 2017.
[51] Fingrid,. Frequency quality analysis, 2016. Helsinki: Fingrid; 2017.
[52] Fingrid. Fingrid, electricity market, frequency controlled reserves. [Online]. Available: http://www.fingrid.fi/fi/sahkomarkkinat/taajuusohjatutreservit/toteutuneettuntikaupat/vuosihinnat/Sivut/default.aspx [accessed 28 06 2017].
[53] Svenska kraftnät. Mimer, primärreglering. [Online]. Available: https://mimer.svk.se/PrimaryRegulation/ [accessed 22 March 2018].
[54] Uptime Institute LLC. Data center site infrastructure tier standard: topology. LLC: Uptime Institute; 2018.
[55] Uninterruptible Power Suppliers Ltd. The UPS role in the uptime institute's tier classification system. 16 November 2015. [Online]. Available: http://www.upspower.co.uk/the-ups-role-in-the-uptime-institutes-tier-classification-system/ [accessed 5 March 2018].
[56] Trash B. Right-sizing data center UPS redundancy and reliability. 5 August 2015. [Online]. Available: http://www.datacenterjournal.com/rightsizing-data-center-ups-redundancy-reliability/ [accessed 18 October 2017].
[57] Smart Gauge Electronics. An in depth analysis of the maths behind Peukert's Equation (Peukert's Law). Smart Gauge Electronics. 08 April 2013. [Online]. Available: http://www.smartgauge.co.uk/peukert_depth.html [accessed 04 December 2017].
[58] Alaperä I, Manner P, Salmelin J, Antila H. Usage of telecommunication base station batteries in demand response for frequency containment disturbance reserve Motivation, background and pilot results. Telecommunications energy conference (INTELEC), Gold Coast, Australia. 2017.
[59] GNB. AGM-handbook, Part 2. GNB, Network Power, Application Engineering; 2016.
[60] Bloomberg New Energy Finance,. BNEF new energy outlook 2017: Europe, Middle-East and Africa – June 2017. Bloomberg New Energy Finance; 2017.
[61] McKinsey & Company. Electrifying insights: how. Advanced Industries; 2017.
[62] The Economist. Daily chart. 14 August 2017. [Online]. Available: https://www.economist.com/blogs/graphicdetail/2017/08/daily-chart-8 [accessed 26 January 2018].
[63] Alaperä I, Honkapuro S, Tikka V, Paananen J. Dual-purposing UPS batteries for energy storage functions: a business case analysis. ICAE 10th international conference on applied energy, Hong Kong, China. 2018. (accepted for publication).
[64] Energy UK. Ancillary services report 2017. Energy UK; 2017.
[65] Statnett. Pilotprosjekt for raske reserver. 13 April 2018. [Online]. Available: http://www.statnett.no/Kundeportal/Kundeinformasjon/Pilotprosjekt-for-raske-reserver/ [accessed 20 June 2018].
[66] Entso-e. Public consultation on “FCR cooperation” potential market design evolutions. Entso-e; 2017.
[67] National Grid. Future of balancing service. Warwick: National Grid plc; 2017.