By Janne Paananen

Grid-Interactive Data Centers Enabling Energy Transition

Data center's hidden potential to provide essential grid services of a future power system.

IMAGE LICENSED BY INGRAM PUBLISHING

THE INTERMITTENT NATURE OF POWER generation from variable renewable energy sources is a well-understood and discussed topic in public. When an increasing number of variable renewable energy sources become connected to the electricity grid, more assets for balancing and flexibility are required to level out the variations in production of electricity. These variations are typically predictable based on wind and solar radiation forecasts and typically last for hours or days. Therefore, this long-term balancing relies on energy-intensive assets that can provide flexibility or balancing for long durations, while the speed of response is not critical.

Digital Object Identifier 10.1109/MELE.2023.3291195
Date of current version: 29 August 2023

2325-5897/23©2023IEEE

A less discussed and often unknown aspect of nonsynchronous renewable energy sources is their impact on electrical grid inertia. These converter-based power generation technologies do not contribute to system inertia and are causing additional challenges for electrical grid reliability and frequency stability.

To maintain grid frequency at a nominal 50 or 60 Hz, the system demand and supply must meet. When there's more electricity consumption (demand) than production (supply), grid frequency will drop and vice versa. The balance between demand and supply must always be maintained, and the frequency deviations resulting from momentary imbalances shall be contained within specified limits to ensure system reliability.

The initial rate of change of frequency following a disturbance is impacted by system inertia and the relative size of a fault.

System Inertia

A system's ability to regulate frequency is also impacted by grid inertia, which is the kinetic energy of the electricity grid and its natural ability to resist changes in frequency. Traditional synchronous generators and motor loads contribute to system inertia by having their spinning masses (rotors) directly coupled to system voltage and frequency through a magnetic field. This inertia in the electrical system naturally helps correct momentary system imbalances. When grid frequency increases, the spinning mass accelerates and absorbs energy. Equally, when grid frequency decreases, spinning mass decelerates and releases stored kinetic energy. By absorbing and releasing energy during frequency transients, inertia is naturally reacting to momentary imbalances and attenuates frequency deviations. The more inertia a system has, the better it resists variations in system frequency. This has importance, especially during large system imbalances caused by large disturbances, i.e., contingency events.

As traditional power generation technologies are replaced with nonsynchronous converter-based sources such as photovoltaics and wind turbines, system inertia in the electricity grid reduces, making it more challenging to contain the frequency (see Figure 1). With higher system nonsynchronous penetration (SNSP) and lower system inertia, the resulting frequency deviations from large disturbances are faster and higher. The initial rate of change of frequency (RoCoF) following a disturbance is impacted by system inertia and the relative size of a fault. The larger the fault or lower the inertia, the faster the frequency will deviate from the prefault value. With high SNSP levels, the frequency variations following large disturbances can be too fast to be contained with turbine generators or other traditional assets used for frequency regulation.

The challenges caused by low inertia must be mitigated to allow higher SNSP and a reduced inertia floor in a synchronous area (SA). This then allows the avoidance of curtailment of renewable power generation and allows retirement of additional fossil fuel-based power generation capacity from the system, thus reducing electricity grid emissions.

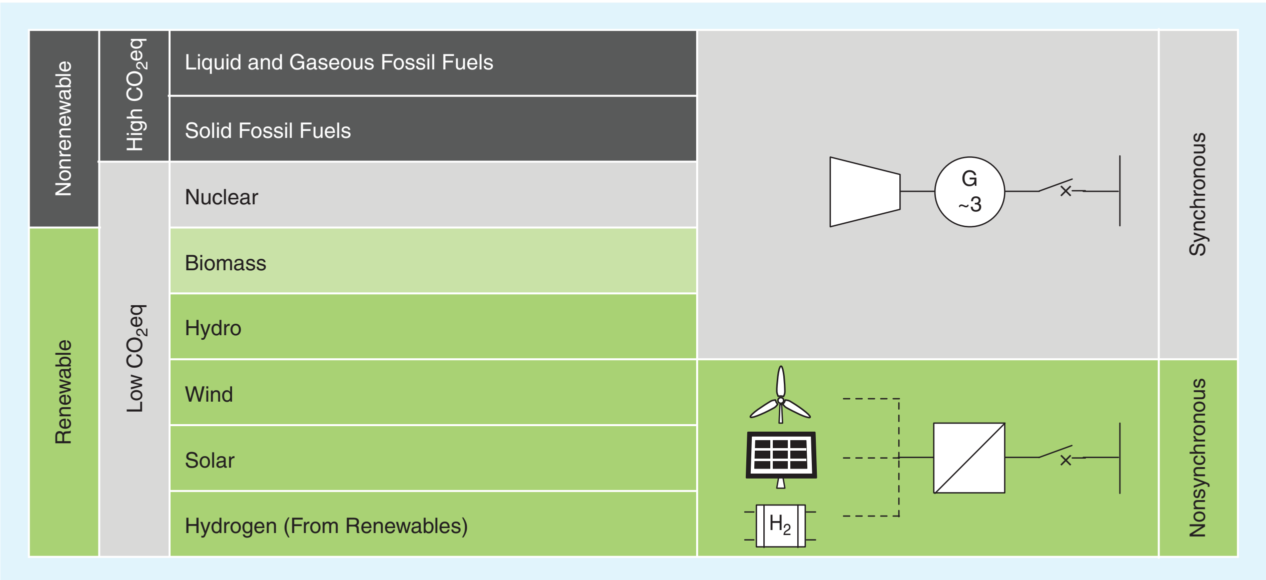

Figure 1. Various power generation technologies and their characterization from a sustainability and inertia perspective.

System Balancing and Role of Reserves

The SA is an electricity grid interconnected with ac lines that shares the same frequency. Examples of these in Europe are Nordic SAs (Finland, Sweden, Norway, and East Denmark), Ireland SAs, Great Britain SAs, and Continental Europe SAs. In the United States, the three SAs are known as the *Western*, *Eastern*, and *Electric Reliability Council of Texas interconnections*.

All the assets connected to the electricity grid that change their active power demand or supply impact the system's balance and frequency within the whole SA. As the SA shares the same frequency, frequency regulation is also a joint effort for the whole SA; even the system-balancing responsibilities can be split for specific load frequency control areas or balancing authorities.

Long-term balancing of the electrical system to ensure that supply meets demand at every hour or bidding slot is based on forecasts. Enough production capacity is offered and sourced from the energy market to meet the expected consumption in each bidding slot. Forecasts are based on time and day, temperature, wind, solar radiation, and so on. As these forecasts can have inaccuracies, and there's some variation in demand inside bidding slots, the remaining slight and short-term imbalances are corrected by reserve assets acquired the through ancillary services market to maintain system balance and frequency at a nominal value. Reserves are also used to manage the unexpected system imbalances caused by a large disturbance, i.e., a contingency event.

These reserves are generally categorized as primary, secondary, and tertiary reserves, also known as *frequency containment reserves* (*FCRs*), *frequency restoration reserves* (*FRRs*), and *replacement reserves* (see Figure 2). The reserves are not dimensioned at the system level to manage normal variations in renewable and other power generation but rather to manage a largest expected fault (a reference) within the SA. In the literature, this event is often referred as a *reference incident*, *dimensioning incident* (*DI*), or *largest single infeed*. The DI is typically defined as a *sudden disconnection of the largest generating unit(s)* or a *high-voltage dc transmission line*.

In a smaller SA, the relative size of a DI is bigger, and it is harder to contain frequency deviation following a contingency event.

The size of the DI varies among SAs as well its relative size compared to overall demand and supply in the electricity grid. In a smaller SA, the relative size of a DI is bigger, and it is harder to contain frequency deviation following a contingency event. The challenge to containing the frequency within specified limits during contingency events is also negatively impacted by the reduced inertia in a power system as this further increases the RoCoF after a fault, as discussed earlier.

Fast Frequency Response

Fast frequency response (FFR) is an efficient method used to mitigate low system inertia. The purpose of these fast-acting reserves is to provide very a fast active power response to the larger frequency deviations caused by large disturbances and allow time for other FCRs to react. As opposed to long-term balancing, this application requires power-intensive assets capable of providing a short but fast response to momentary frequency deviations.

When the response is fast enough (i.e., full activation within a second or so) and the active power response is sustained for long enough to allow other reserves to engage, the efficacy of containing the frequency deviation and nadir is far greater than the efficacy of the inertial

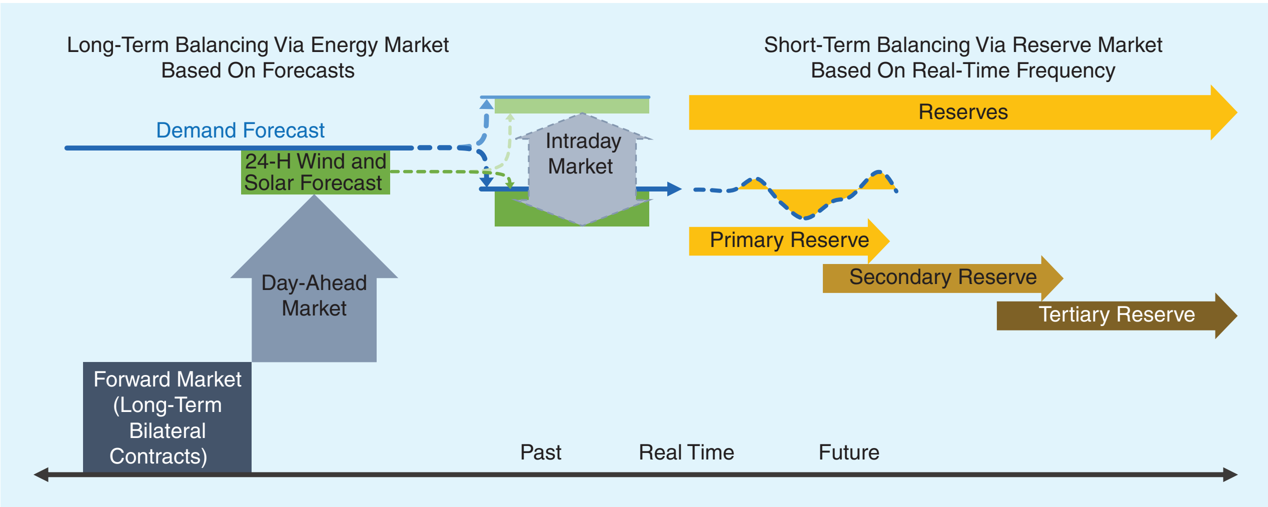

Figure 2. Balancing the principle of the electricity grid through energy and ancillary services markets.

response from the spinning reserves. Therefore, FFR is a cost- and resource-efficient method used to mitigate low inertia in the electricity grid, and much lower capacities of FRRs are needed compared to traditional spinning reserves.

FFR can be provided by various converter-based systems with coupled energy storage, or with interruptible loads that can be quickly disconnected. For example, grid-interactive uninterruptible power supplies (UPSs) in data centers have been studied and piloted to provide FFR with good results.

The type of response and exact application of a grid-interactive UPS greatly impacts the required design of the batteries used.

Grid-Interactive UPSs

A grid-interactive UPS is capable of adjusting and moving its load demand between the main supply and energy storage, typically batteries. This capability can be leveraged for various purposes, for example, for traditional demand-response applications such as time of use and peak shaving. It can also be leveraged to provide ancillary services for electricity grid operators, such as frequency regulation.

As power flows through the UPSs and between main supplies and energy storage is controlled with fast-switching power electronics, a grid-interactive UPS is well suited to provide time-critical services such as FFR. Also, when required and with correct control algorithms, the power flow between mains and energy storage can be seamlessly controlled to provide an accurate and dynamic response proportional to frequency deviation. Furthermore, when leveraging the bidirectional capabilities of a modern rectifier, the response can be independent of the load connected to a UPS output, maximizing the potential to provide system services (see Figure 3).

The type of response and exact application of a grid-interactive UPS greatly impacts the required design of the batteries used. Traditional demand-response applications such as time of use and peak shaving require a long duration response and a high number of energy-intensive cycles, while applications such as FFR, which is power intensive, require a short and fast response. The former probably requires an additional investment for larger batteries with cyclic capabilities, similar to those used with battery energy storage systems (BESSs). The latter is better suited for typical UPS (power) batteries, including lead acid technology, and requires only minimal additional investments if any.

Various academic research has been done on data centers and demand response. Al Kez et al. studied data center capability to provide FFR in the Irish electricity grid and concluded in their extensive study that data centers "can provide superior performance in mitigating the severity of frequency deviation" in power systems with low inertia. Alaperä et al. have studied the business case of grid-interactive UPSs in data center versus dedicated grid-scale BESSs, showing the financial benefits when leveraging existing assets. Also, Eaton has piloted the use of grid-interactive UPSs to provide FFR in Nordic SAs with Norwegian transmission system operator Statnett and Fortum (aggregator) in a live data center in 2018, proving the suitability of UPS technology for FFR.

The concept of using a grid-interactive UPS to provide system services and frequency containment is not only an idea. The first data centers are already participating in ancillary services markets, providing FFR with grid-interactive UPSs. For example, Microsoft in Ireland is leveraging its grid-interactive UPS in its data center to participate in the *Delivering a Secure Sustainable Electricity System (DS3)* programme and provide FFR. Similarly, few of the colocation data center operators in the Nordics have been using their grid-interactive UPS assets to offer FRRs to local ancillary services markets for some time already.

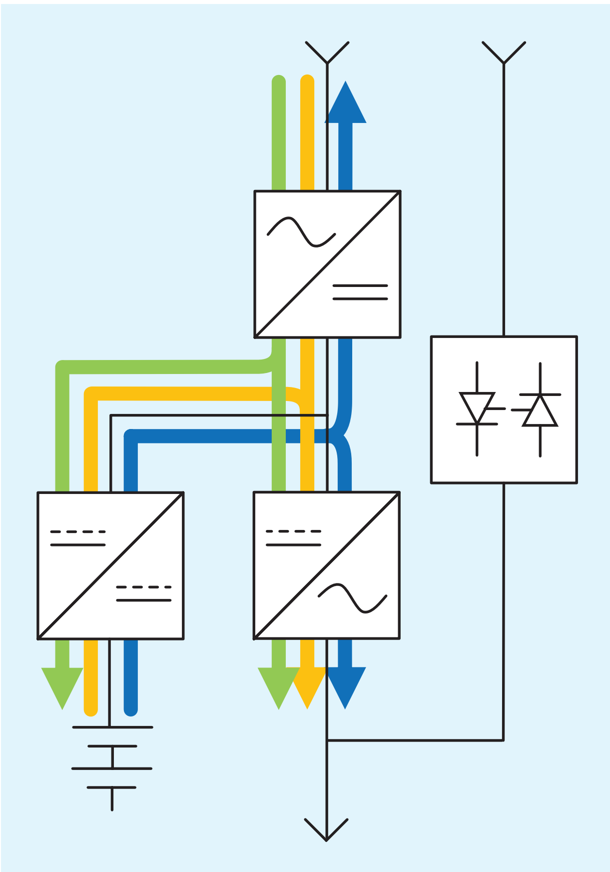

Figure 3. The power flow of a grid-interactive UPS supporting load and charging batteries to increase demand (green), supporting load and discharging batteries to reduce demand (orange), and supporting load and injecting power to mains (blue).

Data Center Potential to Provide FFR

There is often a misconception about the overall potential of data centers and grid-interactive UPSs to provide system services like FFR to help to manage large contingency events in electricity grids with low system inertia.

When talking about the potential of data centers to provide system (ancillary) services, one has to compare the data center's grid-interactive equipment's power capacities against the required volumes of specific reserve types from the ancillary services markets.

There are differences between SAs in respect to their power generation mix and amount of inertia, impacting the required type and volumes of reserve assets from ancillary services markets. Also, the size of the DI relative to minimum system demand varies further, impacting the required volumes of fast-acting reserves. Finally, the amount of data centers with grid-interactive technologies to provide services varies among systems. Therefore, the potential to provide system services shall be studied on a case-by-case basis for each SA.

To estimate the potential of grid-interactive data centers to provide system services, one has to know the following:

- the quantity of required reserves from the markets acquired by system operator(s)
- the quantity of data center capacities within the system
- the quantity of grid-interactive equipment installed within data centers, taking into account power distribution topologies
- the actual power demand (utilization rate) of data centers as this impacts the maximum active power response when not injecting power back to grid.

A grid-interactive UPS and data center can provide an active power response to regulate grid frequency by moving its demand between the electricity grid and the connected energy storage. The maximum response provided is limited by the amount of demand from the grid, i.e., connected loads. In this case, an evaluation of the overall potential to provide grid services is based on the actual reported demand of connected data centers or the estimated demand based on their load utilization rates and nominal power capacities.

The response from grid-interactive UPSs can also be load independent and exceed the power demand of connected loads when injecting power from UPSs and batteries to the grid. Leveraging this bidirectional capability allows utilization of more UPS capacity for grid services and is limited by UPS capacity itself. In this case, the estimate is based on the installed UPS capacity within data centers impacted by used power distribution topologies.

Power Distribution Topologies

Commonly used data center power distribution topologies are tier 3, tier 4, distributed-redundant, and block-redundant topologies. All of these are redundant, allowing a failure or maintenance in one of the power paths without interruption of critical loads. Tier 3 (or 4) has two (active) power distribution paths, both capable of supporting maximum load. The associated UPS capacity is ≥ 200% of the design value of critical loads. In distributed- or block-redundant topology, the loads are split between a higher number or paths, having a lower level of overhead in UPS capacity.

Based on 125 reviewed data center projects and designs in Europe, the typical UPS capacity versus design capacity of critical IT loads (UPS-to-IT-load ratio) is 137%. This is explained by a high number of hyperscale data centers with tens of megawatts as nominal power capacities that are using distributed- or block-redundant topologies, or variations of these. Many smaller and regional data centers use dual-path-redundant topologies with higher power infrastructure overhead, but they present a smaller share in overall data center power capacity (see Figure 4).

Utilization Rate and Power Usage Effectiveness

The load utilization rate tells the actual load demand versus the design IT load demand, while power usage effectiveness (PUE) tells how much additional energy is consumed for cooling and auxiliary loads on top of the energy consumed by IT loads. The data center's server and load utilization rates and the PUE vary and can be rather difficult to estimate without access to data center internal energy and power metering data. However, for example, within The European Code of Conduct for Energy Efficiency in Data Centers scheme, the participating data centers are reporting an overall IT design load in kilowatts, overall IT load energy consumption, and overall data center energy consumption. The average load utilization rate among participant data centers was 46%, and the average PUE was 1.64 in 2016. Similarly, in a recent study, BloombergNEF estimated average data center demand to be 43.5% of design power capacity, while the Uptime Institute has reported a plateauing of the average PUE of surveyed data centers, with 1.55 as an average value in 2022.

Although the reported PUE tells how much additional energy is consumed outside IT loads in real life, the design PUE indicates the share of cooling and auxiliary loads on top of IT loads in the design capacity. The common design PUE of modern data centers is 1.30, meaning 30% of the additional demand by other equipment on top of every kilowatt of IT loads in data center design. When IT load capacities are estimated based on data center design connection capacity, use of design PUE is more meaningful. Average reported PUE can be used when IT load is estimated based on reported data center demands at typical utilization rates (partial loads).

Data Center Power Demand and Capacities

Numerous studies have been done about global and regional data center power demand and energy

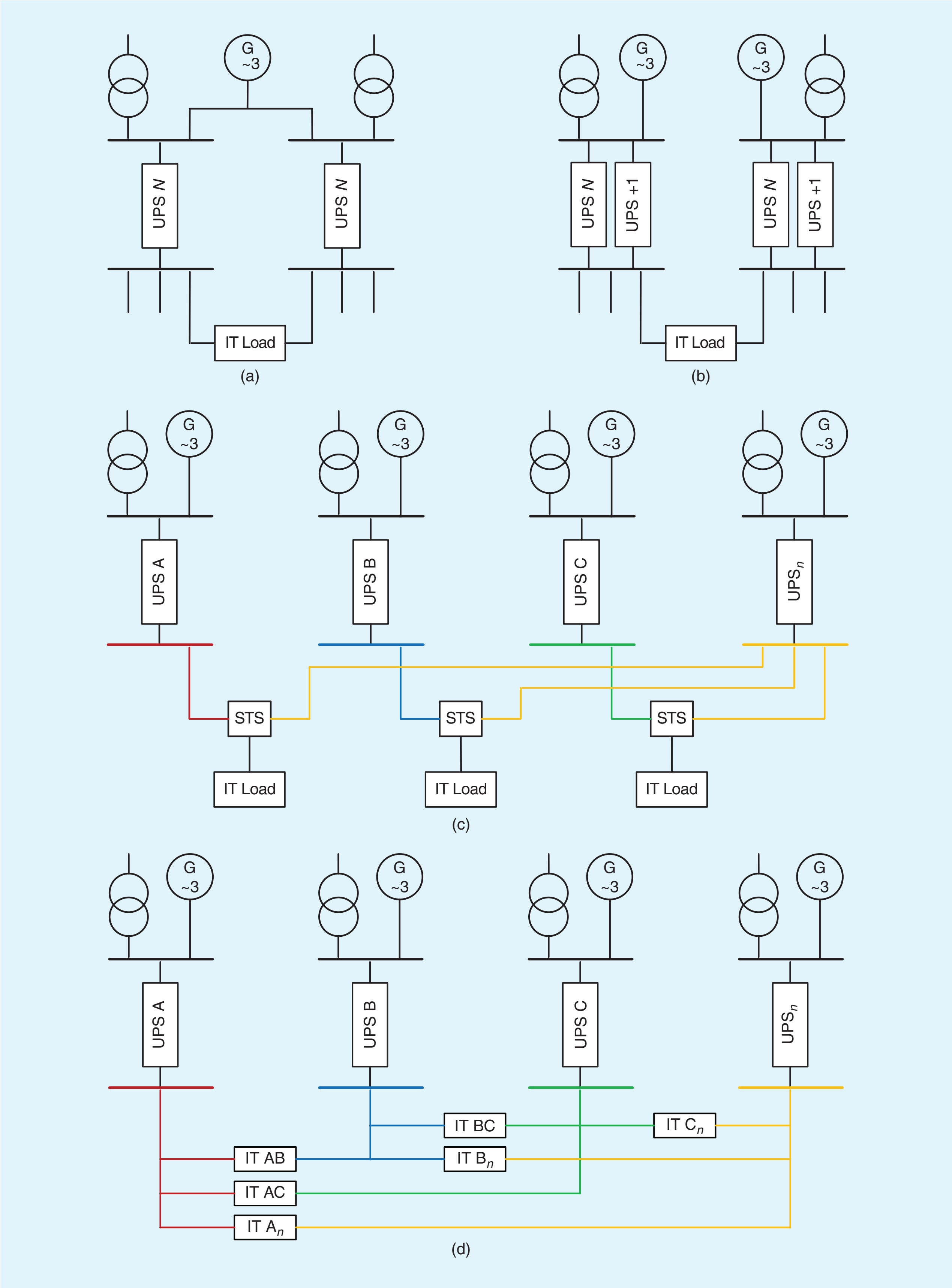

Figure 4. Redundant power distribution topologies using multiple power paths. (a) Tier 3, (b) tier 4, (c) block redundant, and (d) distributed redundant. STS: static transfer switch.

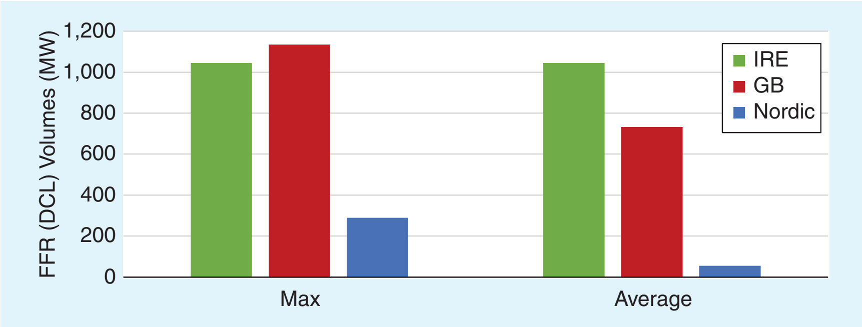

Figure 5. The maximum and average market volumes for FFR and dynamic containment low (DCL) between 1 April and 1 October 2022 in different As. Max: maximum. IRE: Ireland; GB: Great Britain.

need to be used. But, for example, the Central Statistics Office of Ireland publishes quarterly data of metered electricity consumption of data centers within Ireland. From this, the average demand of data centers can be easily derived. Similarly, annual energy consumption data are provided by the National Statistic Office of The Netherlands of the data centers within their country.

With the European Commission-proposed recast of the Energy Efficiency Directive, there will be more stringent requirements for data center operators within the European Union to report their energy consumption and PUE, among many other things, adding transparency to data center energy consumption at the country level. This is a welcome change as it helps to accurately know the data center's power demand and capacities and better understand its role in the power system; and this no longer relies on rough estimates.

Based on the metered data center energy consumption within Ireland, data centers' power demand (connection capacity) was 479 (1,150) MW at the end of 2021. For the United Kingdom, the estimates vary depending on the source, with roughly 950 (2,240) MW as an average, and for the Nordic region, the estimate based on various sources is approximately 330 (740) MW. As there is significant growth in data center markets and energy consumption, the estimates are easily behind actual numbers.

consumption. The results can vary a lot among authors and among studies from the same authors. The common approach has been a bottom-up evaluation, using an assumed number of connected IT devices and their energy consumption within data centers as the starting point. This can lead to some inaccuracies if the initial assumptions are not set correctly.

Another method, and as proposed here, is to evaluate the data center's power demand and energy consumption based on its connection capacity and load utilization rates and average PUE. Estimated connection capacities are available from multiple sources for specific market locations and regions, which can be combined with average utilization rates to estimate the actual power demand of data centers in a specific region.

For most countries and regions, the exact data center energy consumption data are not available and estimates

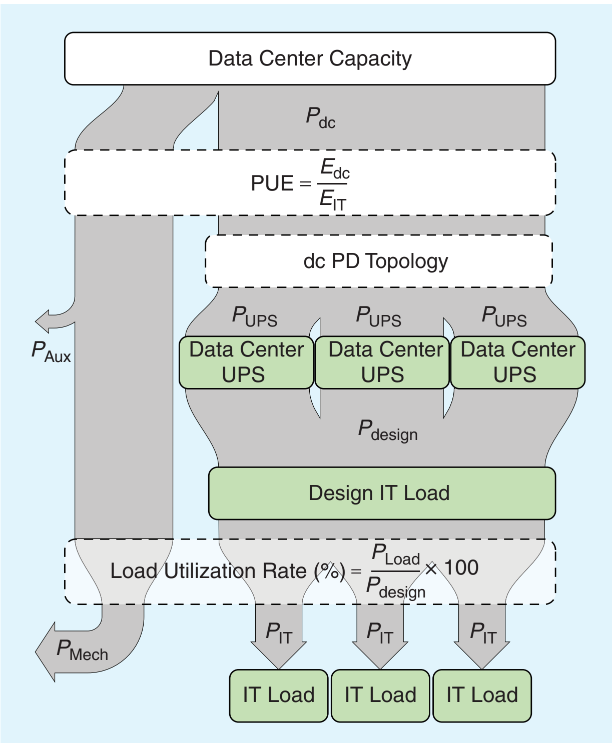

Figure 6. The flow of energy in a data center. PD: power distribution.

FFR Capacities in Selected Markets

System operators acquire ancillary services from the market to meet their specified need. There is variation among markets in the duration of bidding slots and the volatility of reserve volumes. Also, quantities for some of the reserve types are more static, while others can be more dynamic based on actual estimated need for each hour or bidding block.

In the Nordic SA, transmission system operators mainly use FFR only between April and October, when system inertia is low. The highest volumes of FFR are needed during summer weekends and nights when the amount of traditional power generation and system demand is low. During these hours, the share of nonsynchronous sources is potentially highest, and system inertia is at the lowest level. At other times, when industry is running and system demand is higher, system inertia is on a sufficient level and there is no need for FFR. Therefore, the volatility of hourly volumes and market prices in the Nordic FFR market is high.

The U.K. market has less variation in acquired volumes of dynamic containment low (DCL), which is the closest equivalent to FFR, between 4-h Energy Forward Agreement periods (EFA blocks) and seasons, and higher overall volumes are sourced from the market. Currently, in Ireland, the bidding is for six-month slots, but there are plans to move to an hourly market. Due to market design, volumes are static for a six-month period but are steadily increasing between slots as increasing amounts of renewable energy are brought to the system. The FFR volumes are highest in relation to system size in Ireland, which also has the highest SNSP, i.e., the highest share of converter-based power generation in the system (see Figure 5).

As the electricity grid transforms and power generation mix changes, the needed volumes and types of reserves change and vary among systems. What is true today may not be the same tomorrow. Therefore, the assessment of different assets to provide specific services should be updated regularly to reflect actual market conditions.

UPS and Data Center Capacity Versus FFR Volumes

Based on the reviewed 125 data center projects in Europe, the quantity of UPSs used to support mechanical loads is small in the sample base. Therefore, the focus here is on UPS units supporting the IT loads.

The flow of energy in a data center and the principal relationship among nominal (connection) power capacity, IT load design power, actual IT load, and UPS capacity is shown in Figure 6. The overall power demand is split between IT and other loads, and the quantity of UPS units depends on the design IT load and used power distribution topology and resulting overhead.

When designing PUE, the typical UPS-to-IT-load ratio and load utilization rates are known, the quantity of UPS units can be estimated based on system design and connection capacity. Likewise, UPS quantities can be estimated based on actual data center power demand, reported PUE, load utilization rate, and UPS-to-IT-load ratio. These two results can differ from each other as the data center may not be fully operational and part of the power infrastructure has yet to be uninstalled. Therefore, using nominal connection capacity and design PUE may give higher result than a demand-based estimate, as illustrated in Figure 7.

Grid-interactive UPS suitability for frequency regulation has been proven in academic literature, piloting with transmission system operators, and real-life use cases. What has been unclear, however, is the overall potential of grid-interactive UPS systems in data centers to provide ancillary services such as FFR. To evaluate this, data center power demand as well as installed capacities of grid-interactive equipment can be compared to the actual FFR volumes needed and sourced from markets. Based on the earlier presented data, the results are presented in Figure 8.

As can be seen in the figure, the data center connection's capacities, demand, and UPS capacities are significant, if not greatly exceeding, the required FFR or DCL volumes. Therefore, the data center's potential to provide system services like FFR with grid-interactive equipment is significant.

It is important to acknowledge that not all the UPS units installed in data centers are grid interactive or "smart grid ready." Grid-interactive capabilities were recently introduced to the market in the past few years. But as the UPS installation base is renewed and new data centers are built, increasingly higher shares of UPS equipment will be capable of providing ancillary services and FFR. Today, there are at least five major UPS suppliers

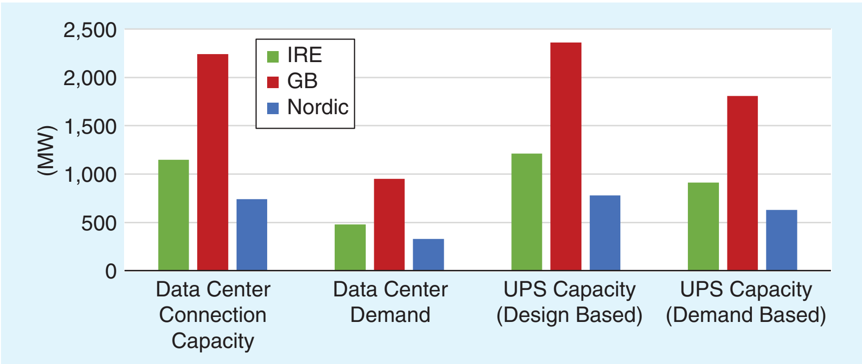

Figure 7. Data center and installed UPS capacities for Ireland (IRE), Great Britain (GB), and Nordic SAs.

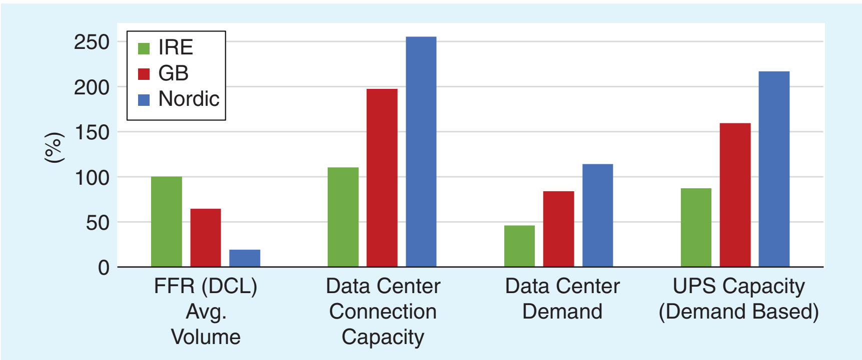

Figure 8. Average FFR (DCL) market volumes, data center connection capacity, data center power demand, and installed UPS capacities in comparison to maximum FFR (DCL) volumes between 1 April and 1 October 2022. IRE: Ireland; GB: Great Britain.

promoting such functionalities for their product(s), and many of these are retrofittable into existing installation bases for the units shipped during the past few years. As such, there already exists a significant potential UPS installation base that could be leveraged for the purpose.

Data Center Motivation

The results from Omdia's smart grid-ready UPS survey in 2022 clearly indicate that money is not the biggest driver for data center operators to use grid-interactive technologies. The biggest motivator is corporate image and sustainability goals, that is, being a good corporate citizen. Nevertheless, providing grid service can be profitable as revenues from FFR markets can be 100 k€/MW/a, or even more. This has importance as one of the top barriers to using grid-interactive technology was the associated potential cost.

Today, data centers are an untapped asset that could provide essential services to enable higher penetration of renewable energy sources, and to reduce the inertia floor.

FFR is particularly well suited for data centers as the typical response needs to be sustained for short times, requiring limited or no additional investments in existing systems and batteries while providing tempting revenues. And at same time, the need for FFR is directly linked to an increasing share of renewable energy in the electricity grid.

Conclusions

If and when a data center wants to support the grid by participating in frequency regulation, it happens through the ancillary services market. Therefore, the method used to evaluate the data center's potential to offer and provide a specific service was approached from a practical point of view based on market volumes. As the markets are continuously changing and new markets and opportunities are emerging, the actual numbers presented are maybe less important than the overall picture and general assessment principle, which can be applied to other markets as well.

A grid-interactive UPS is not just a concept, but a proven technology already in use in a few data centers. The interest grows as increasing numbers of data centers participate in ancillary services markets, and, in the future, data centers can be seen as an active part of the grid, helping to ensure power system reliability while enabling a higher share of renewables. For a data center, this means a more stable and reliable grid, which is important for mission-critical facilities and lower greenhouse gas emissions from their operations.

As presented in this article, the overall potential of data centers to support the grid and help system operators mitigate the issues related to variable renewable energy sources is significant. Today, data centers are an untapped asset that could provide essential services to enable higher penetration of renewable energy sources, and to reduce the inertia floor. This then allows the retirement of fossil fuel-based generating plants from the electrical system, which greatly reduces the greenhouse gas emissions outside a data center. It is also a smarter use of existing assets, resulting in less need for purpose-built reserve units and reducing the embodied carbon in the electricity grid and the associated cost. This carbon handprint created by grid-interactive UPSs and data centers is reducing the socioeconomic impact of the energy transition.

For Further Reading

I. Alaperä, S. Honkapuro, V. Tikka, and J. Paananen, "Dual-purposing UPS batteries for energy storage functions: A business case analysis," *Energy Procedia*, vol. 158, pp. 5061–5066, Feb. 2019, doi: 10.1016/j.egypro.2019.01.645.

I. Alaperä, J. Paananen, K. Dalen, and S. Honkapuro, "Fast frequency response from a UPS system of a data center, background, and pilot results," in *Proc. 16th Int. Conf. Eur. Energy Market (EEM)*, Ljubljana, Slovenia, 2019, pp. 1–5, doi: 10.1109/EEM.2019.8916344.

D. Al Kez, A. M. Foley, F. W. Ahmed, M. O'Malley, and S. M. Muyeen, "Potential of data centers for fast frequency response services in synchronously isolated power systems," *Renewable Sustain. Energy Rev.*, vol. 151, Nov. 2021, Art. no. 111547, doi: 10.1016/j.rser.2021.111547.

"Irish decarbonization and consumer benefits of Grid-Interactive UPS," Baringa, London, U.K., 2022. [Online]. Available: https://www.baringa.com/contentassets/8771b5d7c2ee4f1598d746b5ceafbcd3/baringa_irish_decarbonisation_g-ups_study_report_v4_0.pdf

"Data centers and decarbonization: Unlocking flexibility in Europe's data centers," BloombergNEF, London, U.K., 2021. [Online]. Available: https://www.eaton.com/gb/en-gb/company/news-insights/energy-transition/bnef-data-centres-and-decarbonisation-study.html

J. Paananen and E. Nasr, "Grid-interactive data centers: Enabling decarbonization and system stability," Eaton, Cleveland, OH, USA, Jul. 2021. [Online]. Available: https://www.eaton.com/content/dam/eaton/markets/data-center/eaton-microsoft-grid-interactive-whitepaper-wp153031en.pdf

Biography

Janne Paananen (jannepaananen@eaton.com) is with Critical Power Solutions of Eaton Electrical EMEA, FI-02921 Espoo, Finland.