# Data Center Model for Transient Stability Analysis of Power Systems

Alberto Jiménez-Ruiz, IEEE Member, and Federico Milano, IEEE Fellow

Abstract—The rising demand of computing power leads to the installation of a large number of Data Centers (DCs). Their Fault-Ride-Through (FRT) behavior and their unique power characteristics, especially for DCs catered to Artificial Intelligence (AI) workloads, pose a threat to the stability of power systems. To ensure its stability, it is required accurate models of the loads involved. Here we propose a dynamic load model that properly captures the behaviour of DCs. Its three most defining features are the use of an Uninterrupted Power Supply (UPS) which sits between the server load and the grid, the cooling load represented by an induction motor, and a pulsing load that represents the transients caused by contemporary DCs with significant AI workloads. The features of the proposed model and its impact on the dynamic performance of transmission systems are illustrated through a model of the all-island Irish transmission system and real-world data of the DCs currently connected to this system.

Index Terms—Power System, Data Center, Load modeling, Dynamic Simulation

#### I. INTRODUCTION

#### A. Motivation

Transmission System Operators (TSOs) around the world are increasingly concerned with the impact of Data Centers (DCs) on the dynamic performance of the system. As a real-world example of the issues the grid with large penetration of DCs faces, Fig. 1 shows the Fault-Ride-Through (FRT) of a DC, where a fault caused a 204 MW drop of DC demand, causing a Rate of Change of Frequency (RoCoF) of  $0.12 \, \text{Hz/s}$  and a zenith of 50.22 Hz in the all-island Irish transmission system [1].

Their large demand also prevents DCs from being installed in arbitrary locations, as the TSO must assess whether they are within a constrained or unconstrained region of the electricity system, if the DC is capable of providing flexibility in their demand by reducing consumption, and if they can bring dispatchable generation equivalent to or greater than their demand [2].

A precise dynamic model of DCs adequate for transient stability analysis is not currently available. TSOs generally use generic load models which do not properly represent the dynamic behavior of DCs. This work addresses this gap and proposes a dynamic DC model that is adequate to reproduce real-world observations and anticipate potential instabilities (e.g., flapping) that the switching logic and protections of DCs can cause to the transmission system.



Fig. 1. Frequency and RoCoF for Kellystown - Woodland fault [1].

#### B. Literature Review

There are in the literature steady-state DC models, which are adequate for operation and economic analyses. For example, [3] focuses on the thermal side of DCs and studies the optimal response of the DCs depending on the electricity market load and the DC load. Reference [4] also studies the optimal grid placement of DCs in the grid. Reference [5] discusses how to optimize the energy consumption of DCs by using multiple workloads and optimize resource usage. Reference [6] suggests using the Uninterrupted Power Supply (UPS) batteries to reduce the energy cost by storing energy in valley hours.

A fair amount of articles cover long-term dynamics of DCs and model exclusively the heat transfer process and thermodynamics aspects such as finding out the best cooling strategy [7]; the new advances in cooling technology [8]; how to distribute the servers to enhance airflow through the DC [9], [10], [11]; and the cooling model for long-term simulations [12]. These studies focus on the DC rather than on the dynamic interaction of the DCs with the grid.

On a shorter time scale, [13] proposes a dynamic load model for voltage and reactive power control suitable for DCs, whereas [14] and [15] consider the DC model the behavior of the server from the power electronics point of view. The authors model the UPS focusing on the internal dynamics of the DC and regarding the load of the servers as dependent exclusively on their instantaneous load. References [16], [17] and [18] approach DC modelling from a bottom-up approach, where storage units, Central Processing Units (CPUs), network elements, and other equipment are considered. Again, these works focus on the device rather than on the grid. As we show in this paper, however, DCs are capable of fast transients and operations that cause concerns to system operators.

In industry, DC's power consumption is usually represented with a standard PQ or voltage dependent load. These generic models do not include the protections of the DCs that protect the servers by disconnecting the entire load from the grid. Disconnection and FRT logics are usually modeled *ad hoc* by means of if-then conditions on voltage and frequency levels.

A. Jiménez-Ruiz and F. Milano are with the School of Electrical and Electronic Engineering, University College Dublin, Belfield Campus, D04V1W8, Ireland. e-mails: alberto.jimenez-ruiz@ucd.ie, federico.milano@ucd.ie

This work is supported by Sustainable Energy Authority of Ireland (SEAI) by funding A. Alberto Jiménez-Ruiz and F. Milano under project FRESLIPS, Grant No. RDD/00681.

However these logic do not include ramps, reconnection delays and internal DC dynamics.

Finally, in [19], the DER\_A model has been utilized to overcome some of the issues above. This model aggregates the behavior of many individual generators (in this case utilized with negative power to emulate the load behavior) that may, after a fault, restart generating back to the grid. After a fault, the DER\_A model reconnects part of the power consumption instantaneously, not considering any delay, which is not how DCs usually operate. Furthermore, this model includes terms for voltage and frequency control, which are services that DCs do not currently provide.

# C. Contributions

We propose in this article a detailed model of a DC that takes into account its dynamic behavior, including both its unique load profile for Artificial Intelligence (AI) loads and their reconnection logic. The specific features of the dynamic behavior of DCs that we consider in this work are as follows.

- A Data Center (DC) model that takes into account its cyclical load consumption for AI DCs.
- A DC model with proper disconnection and reconnection logic that represents their FRT behavior against faults.

The case study considers a real-world grid, namely the allisland Irish transmission system, as well as recent data and observations for EirGrid, the Irish TSO. The Irish system is a particularly interesting case due to its size, islanded nature and the high penetration of DCs, which is currently 800 MW over a peak of 7 GW and is expected to increase to 1.6 GW or, equivalently, 30% of the demand by 2030.

# D. Organization

The remainder of the paper is structured as follows. Section II shows the different power consumption patterns of DCs depending on the workload type. Section III gives a general view of the internal components of DCs from the electrical point of view and introduces their equations, while Section IV shows the DC model in action against certain case studies. Section V draws conclusions and outlines future work.

# II. DATA CENTER CLASSIFICATION

The power usage pattern of each DC depends heavily on the type of computation the devices of the DC carry out. In this article we have identified three usage patterns for the short term load demand of DCs.

# A. Constant Load Pattern

The computing power needs of the society are increasing year after year, but limitations in the characteristics of individual processors and storage units mean that increasing DC capacity involves scaling horizontally, i.e., increasing the number of servers. The global workload of the DC is partitioned and assigned to each individual server so that the computing usage of each individual server is equal. As an example, DCs owned by content delivery network companies replicate clients' data across DCs all over the world for both large reliability and low response time. The load demand depends on the total national or global demand of these services. For this reason, their power consumption is regarded as constant for transient stability analysis. This pattern of consumption can be found in DCs whose load does not vary or whose load vary very slowly throughout the day.

# B. Batched Load Pattern

DCs owned by large research institutions, universities, or any institution which requires to compute large amounts of data in an irregular way, show a batched load pattern. Tasks to be run on DCs are queued and eventually executed if enough computational resources are available.

Fig. 2 plots the load of a DC which follows this pattern. Each task runs for an irregular amount of time, and the load abruptly changes after it finishes and a new task is assigned to the DC. The power consumption of a device that uses switching gates, such as a CPU or a Graphics Processing Unit (GPU) is directly proportional to its frequency and to the square of its voltage [20]. These frequency changes are not instantaneous, as the Operating System (OS) regulates the frequency of the CPUs. Fig. 3 shows, for example, a computer with the Intel(R) Xeon(R) CPU E3-1245 v5 80W CPU that was stressed with the stress-ng utility in order to fully utilize all its CPU cores. The OS then gradually scales up the frequency and voltage of the CPU, thus leading to the consumption spike.



Fig. 2. Batched load.



Fig. 3. Power consumption of transient CPU scaling driver: intel\_pstate.

# C. AI Load Pattern

The surge of AI-based solutions has dramatically changed the power consumption patterns of DCs. The vast majority of AI solutions are based on deep learning with neural networks, where their training is usually performed by feeding training examples into the neural network. This operation is repeated again and again until we achieve the optimal result. A step that covers all the training examples is called an epoch [21]. This training is performed on architectures with large parallelism potential such as GPUs and Tensor Processing Units (TPUs). Before running the epoch training stage, data has to be transferred to the AI accelerators, and after them the results have to be aggregated. Furthermore, epochs are usually fractioned due to the large size of the training data. This regular transition between idle AI accelerators and usage of GPUs and TPUs creates a load pattern similar to the one shown in Fig. 4. This type of load is characterized by quick bursts of power consumption [22], followed by a brief cool down period. Training of neural networks requires that all servers work in lockstep, therefore all GPUs and TPUs activate and deactivate at the same time. As Section II-B mentions, power consumption does not change instantly, and it has been likewise been smoothed out with a low pass filter.



Fig. 4. AI load.

#### III. DATACENTER INTERNAL STRUCTURE

From the electrical point of view, a DC can be represented as shown in Fig. 5. Its components are the server load, the cooling load, the miscellaneous load, and the UPS.



Fig. 5. DC internal structure.

# A. Servers

The majority of the power consumption of the building is done by the servers installed in the facilities (also named as IT load). While the external grid supplies AC power, these electronic loads work with DC power. The conversion is usually performed by the Power Supply Units (PSUs) installed on each individual rack server installed within the DC. New types of electrical power delivery solutions which involve direct DC power delivery to the server racks are being developed that reduce the power conversion stages and increase the overall efficiency and reliability of the DC [23]. Regardless of the PSU type of these servers, they can be modeled as two distinct loads with different behavior: the CPU and GPU. 3

According to [24], the CPU power consumption can be modeled as:

$$p_{\rm CPU}^{\rm raw} = p_{\rm CPU}^{0\%} + \left( p_{\rm CPU}^{100\%} - p_{\rm CPU}^{0\%} \right) \, u_{\rm CPU} + p_{\rm Burst} \,, \qquad (1)$$

where  $u_{\rm CPU}$  is the percentage of the server resources used and  $p_{\rm CPU}^{0\%}$  and  $p_{\rm CPU}^{100\%}$  are the power consumption when the server is idle and at full capacity respectively.  $p_{\rm CPU}^{\rm raw}$  represents the CPU power consumption at  $u_{\rm CPU}$  load, where  $u_{\rm CPU} \in [0, 1]$ . Variable  $p_{\rm Burst}$  was added to model cyclical sudden increases in power consumption in traditional DCs [1]. This behavior can be formally expressed as:

$$p_{\text{Burst}} = \begin{cases} \hat{p}_{\text{Burst}}, & t \in [n\hat{T}_{\text{Burst}}, n\hat{T}_{\text{Burst}} + \tau_{\text{Burst}}], n \in \mathbb{N}, \\ 0, & \text{otherwise}, \end{cases}$$
(2)

where  $\tau_{\text{Burst}}$  is the duration of each burst, generally of the order of few seconds;  $\hat{T}_{\text{Burst}}$  is the period of the observed transients, which are in the range of minutes, and the power increase during the spikes is  $\hat{p}_{\text{Burst}}$ .

Equation (1) represents the power consumption of the CPU due to the fact that the kernel or the internal CPU performance scaling driver can gradually change the frequency and the voltage of the CPU. The value of  $u_{\rm CPU}$  is constant if the DC to be simulated follows a constant load pattern or an AI load pattern. For batched load patterns  $u_{\rm CPU}$  shall be modeled as a piecewise function. The power consumption and length of each interval depends on the processes being run and can be arbitrary. In particular, in the case study, we use the jumps-diffusion process described in [25]:

$$du_{\rm CPU} = c_{\rm CPU} \, dJ_t \,, \tag{3}$$

where d indicates the differential operator,  $c_{\text{CPU}}$  is a constant and  $J_t$  is a compound Poisson process, which generates a sequence of Poisson distributed jumps occurring at times  $t_k$ ,  $k \in \mathbb{N}^+$  with amplitudes that can have an any distribution. For the simulations, we have used uniformly distributed amplitudes. Finally, it should be noted that (1) regards the consumption of the CPU, storage devices, and other auxiliary devices within the server as the aggregated CPU load.

Likewise, if the servers within the DC include GPUs or other AI accelerators, its consumption will be equal to:

$$p_{\rm GPU}^{\rm raw} = p_{\rm GPU}^{0\%} + \left( p_{\rm GPU}^{100\%} - p_{\rm GPU}^{0\%} \right) u_{\rm GPU} \,, \tag{4}$$

where  $p_{\text{GPU}}^{0\%}$ ,  $p_{\text{GPU}}^{100\%}$  and  $p_{\text{GPU}}^{\text{raw}}$  mirror their CPU counterparts.  $u_{\text{GPU}}$  shall be modeled as a train pulse:

$$u_{\rm GPU} = \begin{cases} u_{\rm GPU}^{\rm max}, & t \in [n\hat{T}_{\rm GPU}, n\hat{T}_{\rm GPU} + \tau_{\rm GPU}], n \in \mathbb{N}, \\ u_{\rm GPU}^{\rm min}, & \text{otherwise}. \end{cases}$$
(5)

whose valleys  $u_{\rm GPU}^{\rm min}$  and peaks  $u_{\rm GPU}^{\rm max}$  correspond to the usage of the AI accelerators when they are idle and active, respectively. The duration of each pulse  $\tau_{\rm GPU}$  is equal to the time the AI accelerators are running, while the period  $\hat{T}_{\rm GPU}$  of (5) is equal to the time it takes the results after each epoch to be aggregated with the AI accelerators idle plus  $\tau_{\rm GPU}$ .

To account for transients and the fact that each server's load might be unequal and/or its response not properly synchronized, CPU and GPU loads of the server are smoothed out with fast low-pass filters:

$$T_{\rm CPU} \, p_{\rm CPU}' = p_{\rm CPU}^{\rm raw} - p_{\rm CPU} \,, \tag{6}$$

$$T_{\rm GPU} \, p'_{\rm GPU} = p^{\rm raw}_{\rm GPU} - p_{\rm GPU} \,, \tag{7}$$

where  $p_{\rm CPU}$  represents the power consumption of CPU-related processes (i.e., content distribution, database handling), and  $p_{\rm GPU}$  represents processes which run on AI accelerators, mainly GPUs but also TPUs and other customized devices.  $T_{\rm CPU}$  and  $T_{\rm GPU}$  are time constants that smooth out the sudden variations of power consumption. As Fig. 3 shows, it takes  $\approx 50 {\rm ms}$  for the server to reach the new power consumption state. The transient speed can be adjusted so as to either enhance performance during bursts of CPU usage or improve energy usage.

Finally, the sum of CPU and GPU loads constitutes the total load of the servers, as follows:

$$p_{\rm IT} = p_{\rm CPU} + p_{\rm GPU} + \eta_{\rm IT} , \qquad (8)$$

where  $\eta_{IT}$  is an Ornstein-Uhlenbeck process with zero mean and bounded standard deviation [26]:

$$d\eta_{\rm IT} = -a_{\rm IT} \,\eta_{\rm IT} \,dt + b_{\rm IT} \,dW_t \,, \tag{9}$$

where  $a_{\text{IT}}$  is the mean reversion speed,  $b_{\text{IT}}$  is the diffusion coefficient and  $W_t$  is a Wiener process. Details on the integration of (9) are provided in [26].

## B. Cooling

The processing units of these servers dissipate heat. This thermal energy must be removed from the CPUs and GPUs so that they can perform their tasks at maximum performance without suffering damages. This heat reducing can be performed by the servers themselves by means of fans and chillers and, in addition, DC buildings also offer centralized solutions such as liquid to liquid cooling, where a refrigerant is cooled down by a centralized chiller and delivered to individual server cages; and air containment strategies, which direct cooling air to the equipment. Centrifugal chiller systems perform this process by means of a compression cycle which transports the heat from the inside to the outside of the DC [27].

The cooling load accounts for about 30% of the total power consumption of standalone DCs [28] and is modeled as an induction motor owing to their widespread use in compressors and water pumps [29]. The equations that govern these motors are well known and can be found in books such as [30]. The dq-axis squirrel-cage induction motor model with inclusion of stator and rotor flux dynamics was used to simulate the electro-mechanical transients that manifest after reconnection of the UPS following a cleared fault. The mechanical load of the motor was considered to be constant for studying short-term power system oscillations.

Note that, as the focus on this work is on short-term dynamics and the switching logic of DCs following a system event, we do not consider temperature dynamics nor the thermodynamic model of the DC.

# C. Miscellaneous

There are elements apart from the servers and the cooling load that also consume power. These miscellaneous loads are, for instance, lighting, video surveillance and any auxiliary element used by the staff that attends the DC building. They should only consider the consumption of devices associated, or installed within the facilities of the DCs.

The rest of the loads of the DC are modeled as an aggregated ZIP load, as follows:

$$p_{\rm ZIP} = p_{\rm ZIP}^{o} \left( a_p + b_p \, v_i + c_p \, v_i^2 \right), q_{\rm ZIP} = q_{\rm ZIP}^{o} \left( a_q + b_q \, v_i + c_q \, v_i^2 \right),$$
(10)

where  $p_{\text{ZIP}}^o$  and  $q_{\text{ZIP}}^o$  are the ZIP load consumption at rated voltage; the coefficients satisfy the conditions  $a_p + b_p + c_p = 1$  and  $a_q + b_q + c_q = 1$ ; and  $v_i$  is in the internal voltage magnitude of the DC, i.e., the voltage that the UPS sets or the voltage of the grid depending on its topology (see Section III-E).

In this work, we focus exclusively on DCs as standalone facilities. Mixed-use buildings, such as an office building with a floor dedicated to a server farm might consider for the rest of the building an additional dedicated model that properly features its static and dynamic behavior.

#### D. Back-up Generator

The back-up generator is utilized in emergency conditions, that is, only if the power outage extends through time. During a fault, the cooling load, which is usually not protected by the UPS and connected directly to the grid, is disconnected from the grid as well, effectively removing the entire DC power consumption. If the outage persists, the back-up generator is started up and connected to support the UPS and the cooling motors. As it can be regarded as a power source that seamlessly takes over the UPS battery, the back-up generator is not explicitly modeled here and its contribution is implicitly embedded in the UPS.

# E. UPS

DCs are equipped with an UPS that sits between the IT load loads and the external grid (see Fig. 5). This UPS can be centralized, meaning that a single device protects the entire facility, while decentralized UPS consist of many smaller UPSs that are installed at rack level.

The main function of the UPS is to provide the servers with power in case of a fault and thus ride through voltage disturbances on the electric grid. Fig. 6 shows the main UPS topologies for DCs, where Fig. 6a represents the offline mode of the UPS, where the protected loads are connected under normal operation, and they are quickly disconnected by the static switch if a fault is disconnected, to be then supplied directly by the UPS inverter.

Fig. 6b shows the online mode, where the loads are never connected directly to the grid, and are fed exclusively by the UPS inverter. This topology offers the highest security out of all UPS topologies at the expense of the additional power losses of the inverter and rectifier of the UPS.

The final topology shown corresponds to the DRUPS scheme (see Fig. 6c). A flywheel stores kinetic energy which can be



(c) Diesel Rotary UPS (DRUPS).

Fig. 6. UPS topologies.

transformed by a generator immediately should there a power outage be. Under normal operation, this generator works as a motor and accelerates the flywheel. If more energy is needed, a backup diesel generator can generate the energy required to power the loads of the system. The generator is attached to the motor through a clutch as it is inactive under normal operation.

From the modeling point of view, the active and reactive load that it is being protected by the UPS and back-up generator can be written as:

$$p_{\rm DC} = p_{\rm Cooling} + p_{\rm ZIP} + p_{\rm IT} ,$$
  

$$q_{\rm DC} = q_{\rm Cooling} + q_{\rm ZIP} ,$$
(11)

where the  $p_{\rm Cooling}$  and  $q_{\rm Cooling}$  are the active and reactive power drawn by the cooling motor.

The UPS performs 3 main actions:

- Deliver power to the loads downstream.
- Protect the loads during a fault by isolating them from the grid.
- Charge the energy storage elements of the UPS.

We identify 3 different modes of operation that characterize the power drawn by the UPS from the grid, namely  $p_{\text{grid}}$  and  $q_{\text{grid}}$ , as follows.

• Normal mode: The loads are connected to the UPS, but internally, the loads of UPSs in offline mode are connected directly to the grid, while for UPSs working in online mode the UPS itself draws the power to feed the loads from the grid:

$$p_{\text{grid}} = (1 + \beta) p_{\text{DC}},$$
  

$$q_{\text{grid}} = q_{\text{DC}},$$
(12)

where the coefficient  $\beta$  represents the losses of the UPS. During normal operation, electrical power is transferred through the UPS that work with an online topology. For offline or DRUPS modes,  $\beta = 0$ .

• Emergency mode: The UPS supplies the loads, thus effectively disconnecting the DC from the grid. In this operating condition, the following constraints hold:

$$p_{\text{grid}} = 0, \qquad (13)$$
$$q_{\text{grid}} = 0.$$

• **Charging mode:** After a fault the UPS replenishes the energy supplied to the loads:

$$p_{\text{grid}} = (1+\beta) p_{\text{DC}} + p_{\text{UPS}},$$

$$q_{\text{grid}} = q_{\text{DC}}.$$
(14)

where  $p_{\text{UPS}}$  is the additional power drawn from the grid to charge the battery of the UPS. This parameter is defined based on charging time and/or battery life considerations.

The aggregated energy stored in the UPS and the back-up generator (e) varies depending on the difference between output and input power:

$$e' = p_{\text{grid}} - (1+\beta) p_{\text{DC}}$$
. (15)

During normal mode, the right hand side of (15) is equal to 0, under emergency mode equal to  $-(1 + \beta) p_{\text{DC}}$ , and to  $p_{\text{UPS}}$  under charging mode. In charging mode, when the value of e reaches  $e_{\text{max}}$ , the UPS switches to normal mode.

#### F. UPS Disconnection and Reconnection Logic

The UPS measures two quantities to switch to emergency mode: frequency and voltage. A threshold for each magnitude is defined where  $f_{\min}$ ,  $f_{\max}$ ,  $v_{\min}$  and  $v_{\max}$  are the minimum and maximum values of the frequency and the voltage, respectively. The margins of frequency and voltage usually applied are [0.2, 0.3] Hz and 10% of the nominal voltage. If either of these two values is out of range the UPS should disconnect the loads. The expression that shows this logic is:

Disconnect UPS = 
$$(\Delta f < f_{\min}) \lor (\Delta f > f_{\max}) \lor$$
  
 $(\Delta v < v_{\min}) \lor (\Delta v > v_{\max}).$  (16)

This disconnection should occur as fast as possible with no delays, unlike the reconnection.

After the fault is cleared and the bus voltage magnitude returns to an acceptable value, the UPS can reconnect the DC to the grid, according to any of the following schemes.

- **Instant reconnection:** the load is reconnected immediately after the UPS senses the grid has returned to normal conditions. If the magnitudes are not stabilized, there is a risk of flapping, where the sudden reconnection of the load causes a drop in frequency and/or voltage below the limits of the UPS. For this reason, a delay is usually introduced for reconnecting the loads.
- **Delayed reconnection:** to fix flapping a reconnection delay is introduced. The typical delay ranges from 10 to 50 seconds. During that delay no issue with the grid must be observed.
- Disturbance counting: additionally, a disturbance counting scheme might be applied. This scheme deactivates

the reconnection if a certain number of disturbances are seen within a certain time. The typical number is three disturbances and the typical duration is a minute [31].

• Manual reconnection: certain UPS systems such as the DRUPS topology will not transfer the load back to the grid automatically, but they have to be manually reconnected by an operator [31].

The DC model proposed in this article counts a disturbance when the state of the UPS changes from normal to emergency mode. The time from last fault, however takes as a reference the elapsed time since the voltage and frequency parameters of the grid are under the safety margins. Fig. 7 shows an example of a fault that causes a voltage dip. The shaded areas represent the time intervals that, according to the UPS, warrant a disconnection. The actual reconnection will not occur until the conditions of each reconnection scheme presented in the paragraphs above are met.



Fig. 7. UPS reconnection.

# G. UPS internal voltage and angle

In normal operation, the UPS is connected directly to the grid, thus:

$$v_i \angle \phi_i = v \angle \phi \,, \tag{17}$$

where  $v \angle \phi$  is the external voltage at the point of connection to the grid and  $v_i \angle \phi_i$  the internal voltage of the DC (see Fig. 5).

After a fault, the UPS and the back-up generators of the DC supply power to the loads in isolation. It is thus their sole responsibility to maintain the internal voltage of the DC electrical installation.

During a fault, the online UPS maintains  $v_i$  without variation with respect to the normal operation. However, offline UPS and DRUPS quickly isolate the loads and set  $v_i$ . There are two main schemes, as follows.

- $v_i = 1.0$  pu. This scheme provides the nominal voltage to the loads during the fault. However, after reconnection, the grid voltage might have, or might have had, a different value. This sudden voltage change might cause transients.
- $v_i = v(t_{\text{fault}} \delta)$ . Under this scheme the UPS supplies the measured voltage of the grid after the fault. The voltage right before the UPS disconnect may be outside the safety margins configured by the UPS. Therefore, the voltage used of v is the one measured a small time  $\delta$  before detecting the fault.

The internal voltage angle  $\phi_i$  is immaterial when the DC is disconnected from the grid, as no loads of the DC depend

explicitly on the voltage phase angle. However, to reconnect the DC to the grid the UPS must adjusts  $\phi_i$  to match the phase  $\phi$  of the external grid. Thus, we assume  $\phi_i = \phi(t_{\text{reconnection}} - \delta)$  for the reconnection phase angle condition.

# IV. CASE STUDY

The DC loads are for the most part concentrated in industrial complexes of large population centers. Fig. 8 shows the location of the DCs located in Ireland and Northern Ireland (left panel) and in the Dublin county (right panel) [32]. Out of all the 104 DCs located on the Irish island, 86 of them are located in the Dublin metropolitan area and, as it can be inferred from these figures, DCs' are not only located in industrial complexes, but within these industrial areas their facilities are adjacent to each other. Thus, their power usage strains a select few transmission lines and busbars. Modeling their behavior is useful for TSOs in order to assess the stability of the grid.



Fig. 8. Datacenter locations in Ireland and Northern Ireland (left) and in Dublin county (right) [32]. Source of road map data: OpenStreetMap.

DCs have grown in size dramatically for the last years. Due to economy of scale, it is incentivized to concentrate computational resources in a single facility, giving rise to hyperscale DCs, which are buildings designed to process huge amounts of data. Today's largest hyperscaler sites are multiple buildings with total power of 200 MW to 1 GW [33]. Therefore, the DCs considered for the case studies are large and single loads of hundreds of MWs, and not distributed ones.

As starting point, we have utilized a dynamic model of the all-island Irish power system that includes 1479 buses, 1851 transmission lines and transformers, 22 synchronous generators, along with their appropriate control systems, 169 wind power plants and 245 loads. All wind power plants are assumed to be grid-following converters and not to provide any inertial response nor fast-frequency regulation. All simulations are obtained with the software tool Dome [34].

# A. Fault near a DC

The case studies presented focus on a DC with 300 MW of installed server capacity which is being cooled by a 60 MW chiller. This DC is located in the outskirts of city of Dublin in an industrial area surrounded by generation power plants. It has been set that the internal DC UPS will trip if a voltage

variation of  $\pm 0.1$  pu and a frequency variation of  $\pm 0.3$  Hz is observed. The disconnection time of the DC is equal to 30 s. The DC load profile is considered as constant.



Fig. 9. DC under a fault considering a constant load.

We simulate a fault in a bus near the DC. Fig. 9 shows the magnitude and frequency of the voltage at its terminals, with the DC load demand shown as well in the third subplot. The shaded area in red corresponds to the time the external grid parameters of the grid as measured by the UPS are out of bounds. The UPS detects an undervoltage and quickly disconnects the DC from the grid with no delay. This sudden drop in load causes an imbalance in the generation versus the load that rises the frequency. After the transient is cleared, the UPS starts counting 30 s and finally reconnects the DC to the rest of the grid.

# B. Flapping

In this scenario, we consider a fast reconnection scheme, that is, we consider the case for which the DC is reconnected 10 s after the conditions at the point-of-connection bus of the grid recover normal operating conditions.

Fig. 9 shows that DCs provoke large transients in the grid owing to their large power consumption and sudden disconnection from the grid. This figure also shows that the frequency of grid goes down during the reconnection of the DC close to the disconnection threshold of the DC. If the DC is larger, it can generate flapping [35]. The reconnection triggers, leading to a frequency drop below 49.7 Hz, disconnecting the DC again. Fig. 10 shows this phenomenon. To obtain these results, we have assumed that the DC has double cooling load with respect to the previous scenario. The total power consumption of the DC is thus around 420 MW. Setting

the disconnection time equal to 10 s shows that the sudden disconnection pushes the frequency at the DC terminals below 49.7 Hz, re-disconnecting the DC. The cycle can repeat indefinitely in the worst scenario. In this case, the frequency during the next reconnection is just slightly above 49.7 Hz, exiting the loop.



Fig. 10. DC flapping (disconnection time = 10 s).

Flapping can be avoided in various ways. An obvious option is to increase the time the DC waits before reconnecting to the grid. In the scenario discussed in Section IV-A the reconnection time is 30 s which effectively prevents the occurrence of the flapping. However, in general, it is not possible *a priori* to know what reconnection time will cause the flapping. Moreover, for same reconnection time, the risk of flapping is higher the higher the size of the DCs and the lower the current loading condition of the grid.

Another, more sustainable and safer option from the point of view of the grid is the segmentation of the DC facility into different UPSs with different reconnection times. Fig. 11 has the DC considered for Fig. 10 and splits the load into 10 different UPS. The quickest UPS has a reconnection time of 5 s, while the next one is 1 s slower, and successively longer reconnection times for the remaining ones. This segmentation avoids the flapping issue, enabling a seamless transition from fully disconnection to fully connection that avoids significant voltage and frequency transients in the grid. This approach mirrors the ramp of reconnecting DCs observed in the Irish system [36].

# C. Batched and AI Loads

DCs can cause transients not only due to their behavior during and after a fault, but also during their normal functioning.



Fig. 11. DC gradual reconnection.

As an example, a DC with 300 MW of installed server capacity and 60 MW of cooling that works with a batched load pattern shows the voltage, frequency and power variations presented in Fig. 12. Each task that the DC ought to compute is queued and executed accordingly. It is quite possible that some tasks take advantage of a lower number of individual servers, which explains the power consumption variations. The duration and load percentage of the DC is arbitrary. For this graph, it was plotted a set of tasks whose duration was  $\approx 25$  s and the usage varied from 10 % to 100 % of the DC total capacity.

These variations play a more significant role for the AI DCs. Fig. 13 shows a scenario where the 300 MW of the DC are divided into 150 MW of constant CPU load and 150 MW of GPU load. Each pulse of the GPU load has been set to 10 s, with duty cycle of 0.8, i.e., 80 % of the time of the pulse the GPUs are connected and drawing power, while 20 % of the time they are idle. Results indicate a large frequency variation, while the voltage drops about 0.004 pu. These variations might cause issues in the grid, but they can be dampened by properly adjusting the  $T_{\rm GPU}$  parameter. A larger time constant dampens the oscillations caused by sudden power consumption increases, such as the ones caused by AI DCs.

# D. Periodic Transients

DCs, as noted in [1], have spikes of demand which are regular in nature. Fig. 14 shows the effect on the Irish system of an increase of 10 MW of demand every 5 minutes in a DC. These spikes create micro-transients both for voltage and frequency. These transients are small compared with those caused by faults and sudden reconnections of the DC, but nevertheless concerns the Irish TSO as they can reduce the



Fig. 12. DC CPU batched load.



Fig. 13. DC AI load.

frequency quality (e.g., minutes outside the  $\pm 100$  mHz band). These periodic transients are due to regularly scheduled tasks that the servers within the DC perform. As DCs consist of a myriad of individual devices, de-synchronizing these tasks can effectively remove the demand spikes.



0.98 0.96 v [pu] 0.94 0.92 20 40 60 80 100 50.2₩ 50.0 49.840 60 80 100 20Time [ms  $\phi_{\text{com.}} = 0, v_{\text{pre-f.}} = 0, \text{Flux}_{\text{dyn.}} = 0$  $v_{\rm om} = 0, v_{\rm pre-f_{*}} = 0, Flux_{\rm dyn_{*}} = 1$  $= 0, v_{\text{pre-f.}} = 1, \text{Flux}_{\text{dyn.}} = 0$  $= 0, v_{\text{pre-f.}} = 1, \text{Flux}_{\text{dyn.}} = 1$  $om. = 1, v_{pre-f.} = 0, Flux_{dyn.} = 0$  $= 1, v_{\text{pre-f.}} = 0, \text{Flux}_{\text{dyn.}} = 1$  $\phi_{\rm com.} = 1, v_{\rm pre-f.} = 1, {\rm Flux}_{\rm dyn.} = 0$  $v_{\rm pm.} = 1, v_{\rm pre-f.} = 1, Flux_{\rm dyn.} = 1$ 

Fig. 15. Reconnection of a DC under a fault with constant load.

Fig. 14. DC CPU cyclic transients.

# E. Detailed Reconnection

In this last scenario, we consider a detailed reconnection model, including the effect of the dynamics of the induction motor utilized for the cooling system. Fig. 15 shows the values of voltage and frequency at the DC terminals during the reconnection when considering (or not) angle compensation ( $\phi_{\text{com.}}$ ), using the pre-fault voltage ( $v_{\text{pre-f.}}$ ), and flux dynamics for the cooling load (Flux<sub>dyn.</sub>). As in all previous cases the internal voltage used within the DC while it was disconnected was read 10 ms before the fault. Angle compensation and flux dynamics of the induction motor that cools the DC are also included in the model.

Results indicate that angle compensation can mitigate the fastest transients during the reconnection stage of the DC. On the other hand, setting the internal voltage using the pre-fault voltage or the fixed value 1.0 pu does not much difference for this case study owing to the fact that the previous voltage ( $\approx 0.95$  pu) was already close to 1.0 pu. Fig. 15 also shows the results if flux dynamics are ignored (results shown with dots). For this model angle compensation makes no difference as no internal loads use the voltage angle as an input.

# V. CONCLUSIONS

This paper proposes a DC model that captures the timevarying power consumption and the response following faults and other grid disturbances. The model is tested using a realistic model of the all-island transmission system and information on the DCs that are currently installed in the system. Main conclusions are summarized below.

- Traditional DCs can be regarded as a constant load. Nevertheless, the introduction of AI DCs have changed this dramatically. Due to the way that they make use of AI accelerators, mainly GPUs and TPUs, strong transients due to sudden losses of hundreds of MWs that occur. These transients can be mitigated by adjusting the way the scaling drivers of CPUs and other devices work.
- Reconnection is a sensitive procedure that can destabilize, or push the grid into an unstable condition due to the sudden demand of large amounts of electric power. The article shows that a DC made out of UPSs with different reconnection times allows for a seamless reconnection with only moderate transients.
- Sudden changes in demand are caused by variations of the computing load of the DCs. The paper shows that these transients can be controlled and mitigated by DC by adjusting the speed of the frequency changes. This setting can be controlled in hardware and software and, as TSO requirements tighten, DC might require to tune these parameters for both compliance and maximum performance.
- The cooling load, which is modeled as an induction generator, introduces transients during the reconnection stage if flux dynamics are considered. Results show that matching the voltage grid angle by the UPS dampen these transients, while maintaining the last voltage of the grid results in a much smaller effect.

Modeling of DC for power system analysis is in an early stage. Future work will focus on the modeling of DC long-term behavior and explore control strategies that achieve gradual changes of the consumption of DCs.

#### ACKNOWLEDGMENTS

The authors wish to thank Dr Taulant Kërçi, EirGrid Group, for the providing information on data centers and their effect on the Irish system dynamic performance.

#### REFERENCES

- [1] T. Kërçi, C. Duggan, U. Farooq, S. Tweed, and M. Val Escudero, "Impact of Converter-based Demand on Frequency Quality in the Ireland and Northern Ireland Power Systems," in C4 PS1 - Power System Dynamic Analysis in the Energy Transition: Challenges, Opportunities and Advances, Paris, France, Aug. 2024, p. 11.
- [2] CRU, "CRU Direction to the System Operators related to Data Centre grid connection processing," Commission for Regulation of Utilities (Dublin, Ireland) - An Coimisiún um Rialáil Fóntais (Baile Átha Cliath, Éire), Dublin, Ireland, Decision Paper CRU/21/124, Nov. 2021.
- [3] L. Cupelli, T. Schütz, P. Jahangiri, M. Fuchs, A. Monti, and D. Müller, "Data Center Control Strategy for Participation in Demand Response Programs," *IEEE Trans. on Industrial Informatics*, vol. 14, no. 11, pp. 5087–5099, Nov. 2018.
- [4] P. Luo, X. Wang, and Y. Li, "The impact of datacenter load regulation on the stability of integrated power systems," *Sustainable Energy Technologies and Assessments*, vol. 42, p. 100875, Dec. 2020.
- [5] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *ISCA*, New York, Jun. 2007, pp. 13–23.
- [6] Y. Zhang, H. Tang, H. Li, and S. Wang, "Unlocking the flexibilities of data centers for smart grid services: Optimal dispatch and design of energy storage systems under progressive loading," *Energy*, vol. 316, p. 134511, Feb. 2025.
- [7] A. Habibi Khalaj, T. Scherer, and S. K. Halgamuge, "Energy, environmental and economical saving potential of data centers with various economizers across Australia," *Applied Energy*, vol. 183, pp. 1528–1549, Dec. 2016.
- [8] J. Chu and X. Huang, "Research status and development trends of evaporative cooling air-conditioning technology in data centers," *Energy* and Built Environment, vol. 4, no. 1, pp. 86–110, Feb. 2023.
- [9] A. Beghi, L. Cecchinato, G. D. Mana, M. Lionello, M. Rampazzo, and E. Sisti, "Modelling and control of a free cooling system for Data Centers," *Energy Procedia*, vol. 140, pp. 447–457, Dec. 2017.
- [10] H. An and X. Ma, "Dynamic coupling real-time energy consumption modeling for data centers," *Energy Reports*, vol. 8, pp. 1184–1192, Nov. 2022.
- [11] Y. Fu, W. Zuo, M. Wetter, J. W. VanGilder, and P. Yang, "Equation-based object-oriented modeling and simulation of data center cooling systems," *Energy and Buildings*, vol. 198, pp. 503–519, Sep. 2019.
- [12] S.-W. Ham, M.-H. Kim, B.-N. Choi, and J.-W. Jeong, "Simplified server model to simulate data center cooling energy consumption," *Energy and Buildings*, vol. 86, pp. 328–339, Jan. 2015.
- [13] I. Drovtar, M. Leinakse, K. Tuttelberg, and J. Kilter, "Utilizing Demand Response in Load Modelling for Voltage and Reactive Power Control Studies," *IEEE Trans. on Power Systems*, vol. 40, no. 2, pp. 1389–1400, Mar. 2025.
- [14] H. Suryanarayana *et al.*, "System modeling and fault studies in data center power distribution," in *CIRED*, vol. 2021, Sep. 2021, pp. 1435–1439.
- [15] J. Sun, S. Wang, J. Wang, and L. M. Tolbert, "Dynamic Model and Converter-Based Emulator of a Data Center Power Distribution System," *IEEE Trans. on Power Electronics*, vol. 37, no. 7, pp. 8420–8432, Jul. 2022.
- [16] M. Dayarathna, Y. Wen, and R. Fan, "Data Center Energy Consumption Modeling: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 732–794, 2016.
- [17] H. Cheung, S. Wang, C. Zhuang, and J. Gu, "A simplified power consumption model of information technology (IT) equipment in data centers for energy system real-time dynamic simulation," *Applied Energy*, vol. 222, pp. 329–342, Jul. 2018.
- [18] K. M. U. Ahmed, J. Sutaria, M. H. J. Bollen, and S. K. Rönnberg, "Electrical Energy Consumption Model of Internal Components in Data Centers," in *ISGT-Europe*, Sep. 2019, pp. 1–5.
- [19] P. Pourbeik, J. Weber, D. Ramasubramanian, J. Sanchez-Gasca, J. Senthil, P. Zadkhast, J. Boemer, A. Gaikwad, I. Green, S. Tacke, R. Favela, S. Wang, and S. Zhu, "An aggregate dynamic model for distributed energy resources for power system stability studies," 2019.
- [20] A. Chandrakasan and R. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 498–523, Apr. 1995.

- [22] Q. Ryan, J. Zhao, and K. Thomas, "An Assessment of Large Load Interconnection Risks in the Western Interconnection," Western Electricity Coordinating Council (WECC), Tech. Rep., Feb. 2025.
- [23] Y. Chen, K. Shi, M. Chen, and D. Xu, "Data Center Power Supply Systems: From Grid Edge to Point-of-Load," *IEEE J. of Emerging and Selected Topics in Power Elec.*, vol. 11, no. 3, pp. 2441–2456, Jun. 2023.
- [24] L. A. Barroso, U. Hölzle, and P. Ranganathan, *The Datacenter as a Computer: Designing Warehouse-Scale Machines*. Springer, 2019.
- [25] G. M. Jónsdóttir and F. Milano, "Modeling solar irradiance for short-term dynamic analysis of power systems," in *IEEE PES General Meeting*, 2019, pp. 1–5.
- [26] F. Milano and R. Zárate-Miñano, "A systematic method to model power systems as stochastic differential algebraic equations," *IEEE Trans. on Power Systems*, vol. 28, no. 4, pp. 4537–4544, 2013.
- [27] F. Oliveira and A. Ukil, "Comparative Performance Analysis of Induction and Synchronous Reluctance Motors in Chiller Systems for Energy Efficient Buildings," *IEEE Trans. on Industrial Informatics*, vol. 15, no. 8, pp. 4384–4393, Aug. 2019.
- [28] Q. Zhang et al., "A survey on data center cooling systems: Technology, power consumption modeling and control strategy optimization," J. of Systems Architecture, vol. 119, p. 102253, Oct. 2021.
- [29] M. Hasanuzzaman, N. A. Rahim, R. Saidur, and S. N. Kazi, "Energy savings and emissions reductions for rewinding and replacement of industrial motor," *Energy*, vol. 36, no. 1, pp. 233–240, Jan. 2011.
- [30] P. Kundur, Power System Stability and Control. New York: MHP, 1994.
- [31] North American Electric Reliability Corporation (NERC), "Incident Review - Considering Simultaneous Voltage-Sensitive Load Reductions," NERC, USA, Incidence, Jan. 2025.
- [32] "Data Center Map Colocation, Cloud and Connectivity," https://www.datacentermap.com/.
- [33] L. Lin and A. A. Chien, "Adapting Datacenter Capacity for Greener Datacenters and Grid," in *Proceedings of the 14th ACM International Conference on Future Energy Systems*, ser. E-Energy '23. New York, NY, USA: Association for Computing Machinery, Jun. 2023, pp. 200–213.
- [34] F. Milano, "A Python-based software tool for power system analysis," in IEEE PES General Meeting, 2013, pp. 1–5.
- [35] S. Chatzivasileiadis et al., "Micro-flexibility: Challenges for power system modeling and control," *Electric Power Systems Research*, vol. 216, p. 109002, Mar. 2023.
- [36] T. Kërçi, "Dynamic Response of Inverter-based Resources in the Ireland and Northern Ireland Power Systems," in 2013 IEEE Belgrade PowerTech, Belgrade, Jun. 2023, pp. 1–6.



Alberto Jiménez-Ruiz received from the University of Castilla-La Mancha, Spain, the Ph.D. in Computer Science in 2024. He is currently a postdoctoral researcher with University College Dublin, Ireland. His research interests include renewable energy, parallel computing and heterogeneous architectures.



Federico Milano (F'16) received from the Univ. of Genoa, Italy, the Ph.D. in Electrical Engineering 2003. In 2013, he joined the University College Dublin, Ireland, where he is currently a full professor. He is Chair of the IEEE Power System Stability Controls Subcommittee, IET Fellow, IEEE PES Distinguished Lecturer, Senior Editor of the IEEE Transactions on Power Systems, and Co-Editor in Chief of the IET Generation, Transmission & Distribution. His research interests include power system modeling, control, and stability analysis.