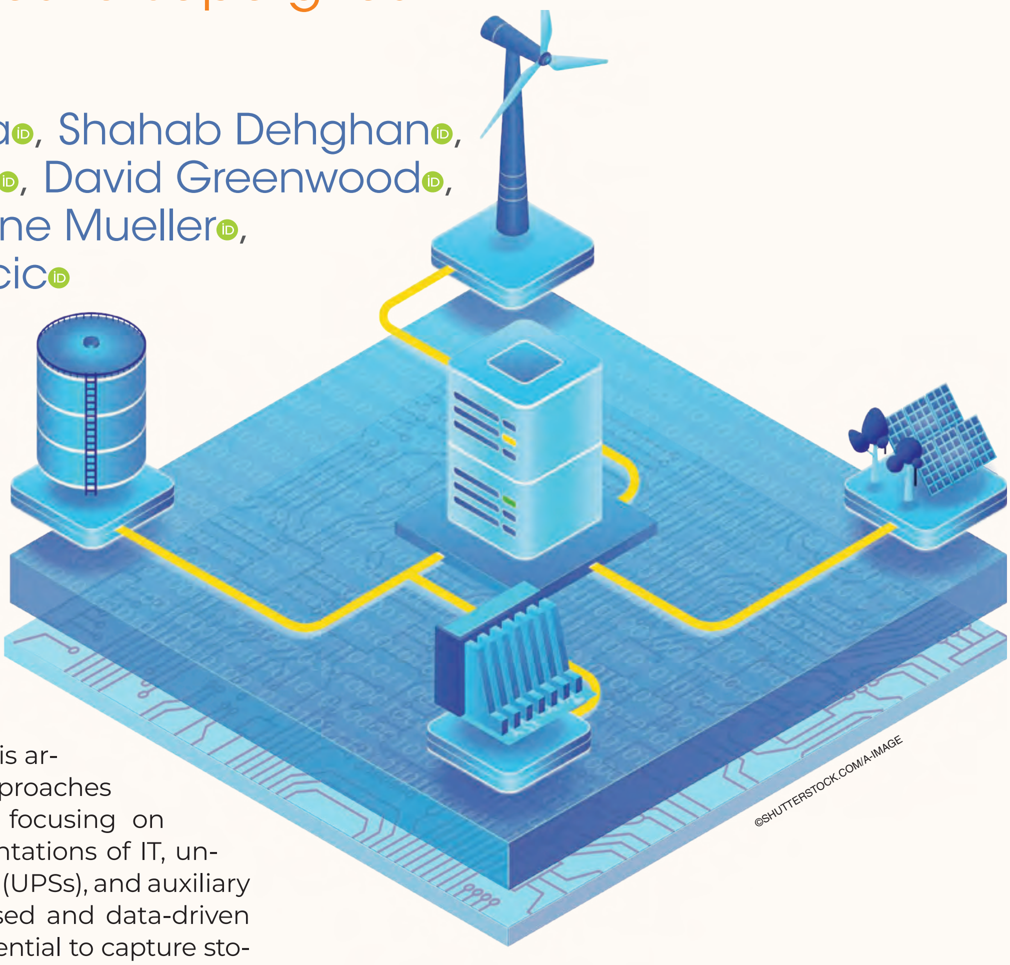
Data Centers for Sustainable Grids

From Microgrids to Supergrids

By Vladimir Terzija, Shahab Dehghan, Damian Giaouris, David Greenwood, Xin Zhang, Juliane Mueller, and Dejan Milojicic



THE EXPANSION OF DATA centers (DCs), accelerated by artificial intelligence (AI) and cloud computing, is redefining their role in power systems. DCs are shifting from being passive loads to active prosumers, with the potential to provide flexibility from microgrids to supergrids. This article explores modeling approaches and integration strategies, focusing on static and dynamic representations of IT, uninterruptible power supplies (UPSs), and auxiliary systems. Hybrid physics-based and data-driven models are identified as essential to capture stochastic workload dynamics. The role of DC energy management systems (EMSs) is analyzed across system levels: optimization within microgrids, various functions within distribution networks, stability support in transmission systems, and balancing in supergrids. Potential contributions to ancillary services, including frequency response, voltage regulation, and black start, are assessed, with attention to novel configurations, such as small modular reactor (SMR) colocation. The study concludes that DCs can evolve into controllable, flexible assets that enhance the resilience, security, and reliability of future electricity networks.

Digital Object Identifier 10.1109/ESM.2025.3634479
Date of current version: 29 January 2026

Motivation

The traditional view defines DCs as specialized facilities that combine IT infrastructure, including servers, storage, and networking, with dedicated power and cooling systems. They enable continuous, large-scale computing but are typically considered purely as concentrated electricity consumers on the grid. In recent years, the rapid growth of generative AI and large language models has driven a surge in demand for computing power, with both advanced training and the rising number of inference tasks requiring far more GPUs (Bash et al., 2024). Consequently, power needs have risen to megawatts for racks, as standardized enclosures that house servers and networking equipment, and gigawatts for full DCs.

This surge in DC integration has already moved faster than grid operators can keep up with. Many hyperscalers (e.g., Microsoft, Google, and Amazon), AI firms (e.g., OpenAI), and large single users (e.g., xAI's Colossus DC) are now looking for alternative supply solutions. Potential alternatives include nuclear and gas-fired behind-the-meter generation and large energy storage systems (ESSs). SMRs are often mentioned as a future possibility (Bash et al., 2024), but none has yet been deployed. In line with the agenda to leverage a broad mix of energy technologies, distributed energy resources (DERs) can partly supply DCs. However, they are unlikely to meet the core demand of hyperscale facilities. This paradigm shift positions DCs as evolving from passive consumers to active prosumers, thereby emphasizing the requirement for robust static and/or dynamic load models that accurately represent their behavior within modern power systems. Therefore, DCs will affect not only the power sector but also integrated multisector energy systems, ranging from short-term operation to long-term investment and even policy design. This evolution, in turn, necessitates closer integration between IT and power/multienergy domains, as outlined in Figure 1. The IT sector has long assumed that reliable electricity would be available at any scale and time. Also, grid operators treated DC facilities as passive electricity load demands under service level agreements. With DC electricity load demands both large and dynamic that are capable of coming on or off the grid at scale, there is an urgent need for new interfaces, services, and grid codes defining new responsibilities for both IT and power domains. In many respects, the one-way relationship from power to IT must evolve into a two-way exchange.

Different national contexts are experiencing DC integration and implication in distinct ways. In the United Kingdom, annual DC consumption is currently estimated at around 7.6 TWh for 2.4 GW of connected DC capacity, equivalent to about 4% of the yearly peak demand (approximately 58 GW) [National Energy System Operator (NESO), 2025]. By 2050, the NESO scenarios project annual DC consumption in the range of 30–71 TWh, corresponding to a peak load of 6.1–14.1 GW, about 6%–15% of the yearly peak demand (NESO, 2025). Also, the U.K. government recently called for 6 GW of new AI-oriented DC integration by 2030 as part of a national initiative. Most of the U.K. DC electricity load demand is concentrated in the south of England, where clusters around London already consume up to one third of the local electricity infrastructure capacity. The clustering of DC facilities in London has resulted in significantly delayed network upgrades and increased reinforcement costs from electricity network congestion. To alleviate southern electricity network bottlenecks, NESO has noted that, with more substantial locational incentives, up to 20% of future DCs could be directed to Scotland (NESO, 2025).

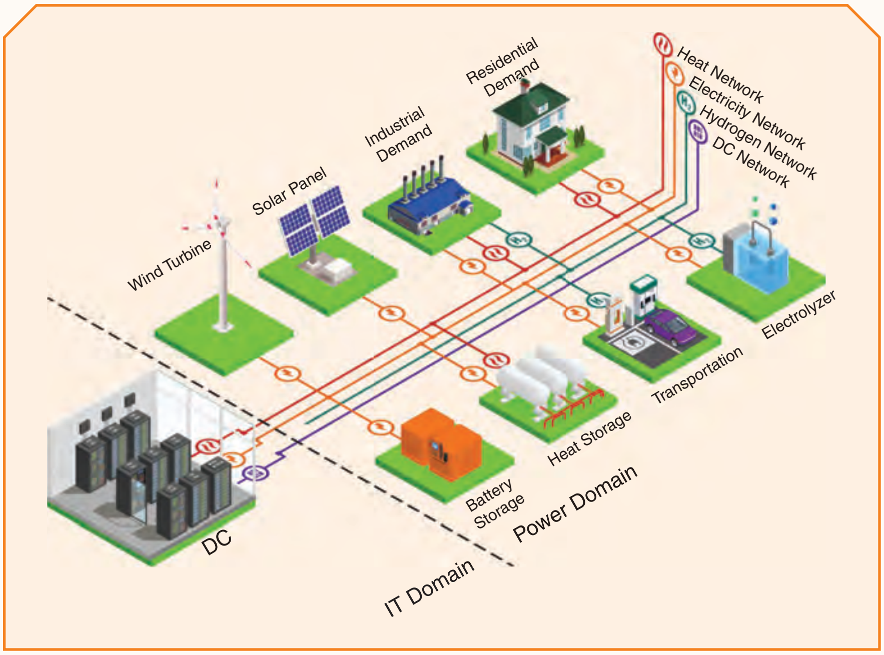


figure 1. The outline of a DC as a prosumer linking IT and power/energy domains.

Furthermore, the FLAP-D markets, including Frankfurt, London, Amsterdam, Paris, and Dublin, have historically dominated European DCs.

FLAP-D hubs account for about 62% of the European DC capacity. In 2023, DCs consumed between 33% and 42% of electricity in Amsterdam, London, and Frankfurt and close to 80% in Dublin. However, their share is expected to fall to 55% by 2030 and 51% by 2035 as developers are forced to look beyond these hubs because of congestion in the electricity network. To remain within the physical limits of electrical infrastructure, Ireland has imposed restrictions on DC integration, with similar measures introduced in The Netherlands and Germany. Only France, with more available headroom in the electricity network, is expected to see steady growth. In Europe, overall DC consumption reached 96 TWh in 2024, accounting for approximately 3% of the total electricity demand, and is expected to reach 236 TWh by 2035. It is worth noting that northern European markets, such as Denmark, Norway, and Sweden, are expected to triple their DC electricity demand by 2030, while southern and eastern European countries are forecast to see threefold-to-fivefold increases by 2035.

The United States hosts the largest concentration of DC facilities globally, with peak demand projected to rise from about 35 GW today, about 4%–5% of the yearly peak demand, to 78 GW by 2035. Clusters in Northern Virginia, Dallas, and Chicago dominate the U.S. DC infrastructure, with Northern Virginia, the world's largest DC market, alone currently hosting more than 7 GW of clustered DC facilities [International Energy Agency (IEA), 2025], as depicted in Figure 2. The clustered DC integration increases the risk of instability within the electricity network. Low-frequency, high-impact unscheduled incidents in clustered DC facilities may result in sudden, simultaneous reductions in local electricity load demand by several gigawatts. Such events have the potential to trigger widespread uncontrolled cascading events and even catastrophic blackouts across the electricity network. A recent near-miss incident in Northern Virginia involved the coincidental loss of 60 DCs with a total load of 1.5 GW. A standard safety mechanism, widely implemented in the DC industry, was activated to prevent damage to electronic DC equipment from voltage fluctuations (North American Electric Reliability Corporation, 2025). PJM Interconnection, the regional grid operator, and Dominion Energy, the local utility provider, implemented immediate corrective actions to prevent a regional blackout. Although a clustered DC load demand can challenge network stability in the event of sudden load demand changes, the large scale of the U.S. DC load demand can also present a potential advantage. If a portion of DC load demand becomes flexible, it may represent the most significant

figure 2. Global distribution of large DC clusters in 2024, utilized and adapted from IEA analysis based on Omdia data (2025), licensed under Creative Commons license BY 4.0 (IEA, 2025).

controllable load demand nationally, supporting grid ancillary services and its overall adequacy, security, and resilience. This DC's ability can provide an opportunity to enhance flexible system operation and mitigate network instability, which is particularly important in power systems with a high share of inverter-based resources (IBRs). Such systems experience reduced grid strength and inertia, both of which are crucial for maintaining frequency stability.

At a global scale, the DC electricity load demand is projected to rise sharply, from about 460 TWh in 2024 (equivalent to approximately 1.5% of the world's DC consumption) to more than 1,000 TWh by 2030 and about 1,300 TWh by 2035. The United States accounts for around 45% of global DC electricity demand in 2024, while China and Europe account for about 25% and 15%, respectively (IEA, 2025). The global distribution of the major DC facilities is depicted in Figure 2. Global investment in DC facilities is projected to reach about US$4.2 trillion by 2030 (IEA, 2025). The power sector investment is expected to reach nearly US$480 billion over the next five years, with about half of that belonging to the United States. Over the same period, the DC-related power sector investment in the United States is expected to contribute more than 15% of the power sector investment (IEA, 2025). Furthermore, the European DC market, worth US$47 billion in 2024, is expected to double to US$97 billion by 2030.

Therefore, it can be concluded that the rapid increase in global electricity consumption by DC facilities is likely to require large-scale investment in infrastructure to keep the power supply secure without pushing the electricity network beyond its physical limits. However, unlike most other large electricity consumers, DCs have a unique capability that supports both the short-term security and the long-term adequacy of electricity systems. This capability lies in their ability to shift substantial loads across both time and space, commonly described as temporal and spatial flexibility. In practice, hyperscale fleets can route computing tasks to different regions or adjust them to different times of day. This flexibility in DC operations can reduce or defer grid upgrades, lower electricity costs, cut energy consumption, and enhance system reliability by shifting load from high-cost peak hours to lower cost off-peak periods. As a practical example, Google has already utilized its carbon-intelligent computing platform to move AI workloads to DC facilities with excess renewable production.

This article addresses the challenges and opportunities of integrating DCs into modern power systems, from microgrids to supergrids. DCs can act as large flexible loads in microgrids, help mitigate congestion and provide grid services at the regional level, and support cross-border balancing at the continental scale. Their spatial and temporal flexibility offers significant potential to enhance EMS platforms, enabling cost-efficient, resilient operation of DC-integrated grids under varying conditions.

DCs as Complex Dynamic Load Models

In recent years, DC load modeling has become a crucial aspect of modern power system studies. Accurate mathematical representation of 1) IT equipment, 2) a UPS that provides backup during outages, and 3) auxiliary systems, such as cooling and pumping, that enable operations of IT equipment, ensures optimal DC integration into microgrids as well as transmission and distribution networks. Well-designed models with accurate parameters not only capture static and dynamic properties but also enable DCs to emerge as active participants in offering spatial and temporal flexibility and, consequently, ensuring grid resilience. With this background, this section outlines the key potential modeling approaches and requirements for the successful integration of DCs into future power systems.

Load modeling describes the behavior of power system loads in computer simulations and stability analysis. It generally involves two stages: 1) selection of the appropriate model, static or dynamic, and 2) determination of model parameters (Regulski et al., 2015). In this context, DC load modeling can be structured by timescale. For steady-state studies, DCs can be modeled as composite static loads, with constant-power IT demand behind UPS systems and auxiliary equipment represented using ZIP models, where Z, I, and P represent impedance, current, and power. For short-term dynamic studies, composite load models must capture UPS-conditioned/protected IT loads and induction-motor cooling systems, including recovery behavior, while also reflecting the pulsing demand/"spiky nature" from AI workloads. For electromagnetic transient studies and power quality assessment, aggregated UPS converter models are required to represent rectifier/inverter behavior and interactions with motor-driven auxiliaries. Therefore, a realistic DC model can be composite, integrating IT load and its characteristics. This can be further explored, both from the perspective of the model itself and the model parameters, which could be dynamically changed over time. An appropriate IT model should be both physics based and data driven: 1) physics based to represent the UPS–IT interface and dynamic response to disturbances and 2) data driven to reflect real workload variability (e.g., pulsing AI demand). Cooling and auxiliary systems can be represented as induction motors with inertia, complemented by a ZIP component for other static loads. In addition, the model may include a flexibility block to capture demand response, workload curtailment, or the capability of UPS–battery systems to provide grid services.

Load models can also be characterized by parameters that must be determined before using the models. Traditionally, this can be achieved through estimation methods that identify unknown load model parameters using different network data, such as voltage, current, and active or reactive power. The nonlinear least-squares (NLS) method has long been used for this purpose, but it is susceptible to the initial parameter guess and may converge to local minima. Other types of estimators, e.g., Kalman filter (KF)-based recursive optimal filters, can also be utilized. AI techniques, surrogate methods, and metaheuristic algorithms have gained attention for their diverse search capabilities, though they can suffer from premature metaheuristic and weak fine-tuning (Regulski et al., 2015). Hybrid methods that combine the search ability of evolutionary algorithms with the speed and accuracy of NLS/KF methods offer further improvements. However, their results highly depend on the robustness of evolutionary algorithms. When assessing IT workloads (e.g., AI/GPU tasks) as well as cooling system dynamics, a significant degree of stochastic behavior can be observed. For power system studies, a hybrid approach that combines physics-based and stochastic/data-driven load models is likely to provide a highly adequate representation of a DC's nature.

In large-scale power networks, composite DC models must be included in stability and frequency response studies, with explicit representation of UPS ride-through and curtailment capability. UPS ride-through capability refers to the ability of a UPS system to maintain continuous power to its connected loads during short disturbances or interruptions in the grid supply. For clusters connected at extra high-voltage or high-voltage levels in transmission networks, electromagnetic transient models are essential to capture converter dynamics. In distribution networks, modeling should focus on hosting capacity, voltage regulation, fault levels, motor starting, and power quality, using feeder-level models that include explicit impedances and detailed UPS transfer behavior. In microgrids, DCs function as priority loads, and models should represent UPS ride-through capability, fast workload shifting or shedding, and interactions with grid-forming/following converters, while also considering DER/ESS limits and converter-to-converter interactions.

SMRs are emerging to reshape DC load modeling by introducing scenarios in which DCs are partially or fully self-powered by on-site nuclear energy (Bash et al., 2024). This shift encourages the development of hybrid load models that reflect on-site production dynamics as well as flexible, reliable consumption profiles. The colocation of SMRs with DCs creates a new supply–demand architecture that changes their representation in power system models. Unlike traditional load-only representations, a colocated SMR allows part or all of a DC's demand to be met by a dedicated on-site baseload source. In modeling terms, this requires treating the DC not just as a composite load but as a hybrid consumption–production entity, or prosumer. In steady-state studies, the SMR block can be represented as a constant production source sized to match contracted demand, reducing or flattening net imports from the grid. For dynamic and transient analyses, the SMR's stability characteristics, including governor and automatic voltage regulator response, must be incorporated alongside the composite DC load to evaluate islanding, fault ride-through, reconnection scenarios, or provision of ancillary services, such as frequency and voltage control. In conclusion, and as expressed in the form of a simplified single-line diagram in Figure 3, the entire DC can be represented by an equivalent dynamic load model, with parameters reflecting its nature and dynamics.

Management Systems for Sustainable Power and IT Domains: From Microgrids to Supergrids

Since DCs have been evolving in recent years from passive consumers to active prosumers,

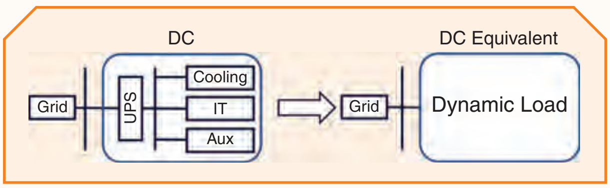


figure 3. DC modeled as a dynamic electrical load.

they will be able to participate in different types of bilateral, retail, and wholesale energy and reserve markets at local, regional, and national levels, depending on their capacity and their connection to microgrids, distribution or transmission networks, or supergrids. DC participation in energy and reserve markets, characterized by extensive information and electricity exchange with power systems at the microgrid, distribution, and transmission levels, necessitates stronger coordination across IT and power domains. Moreover, significant DC integration calls for new standards and protocols governing data exchange between IT and power domains to ensure the cost-efficient resilience of power systems and, more broadly, integrated multienergy systems (e.g., electricity, heating, cooling, gas, and hydrogen networks). This, in turn, necessitates the next generation of coordinated management systems across IT and power domains, as outlined in Figure 4.

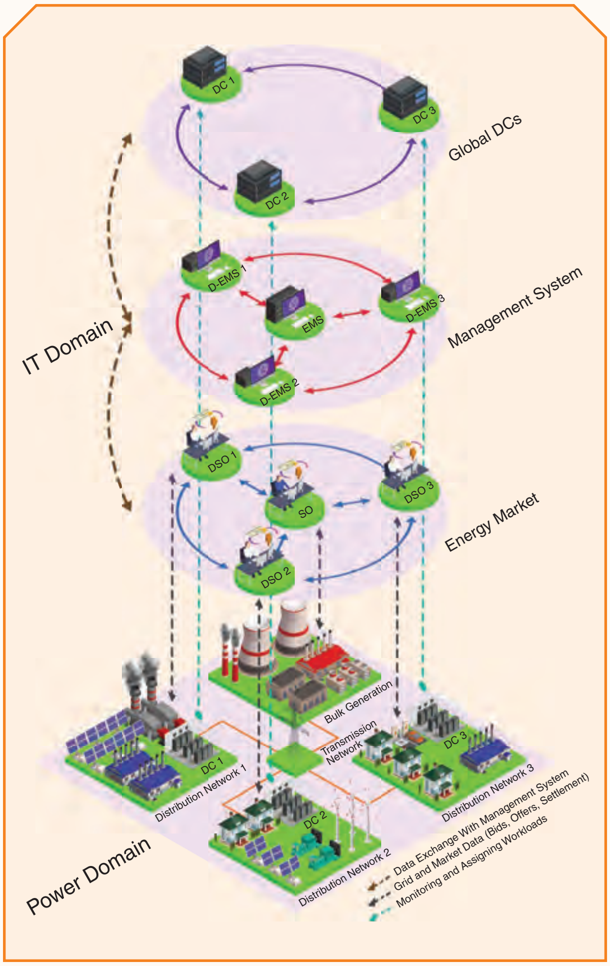


figure 4. Outline of the management systems with data exchange between power and IT domains. DSO: distribution system operator.

Figure 4 aims to conceptually demonstrate the exchange of information and electricity between the IT and power domains. For simplicity, it is assumed that a supergrid comprises three distribution networks, each managed by a distribution system operator (DSO), e.g., DSO 1, DSO 2, and DSO 3, through D-EMS 1, D-EMS 2, and D-EMS 3, respectively. These distribution networks are connected to other bulk producers within the supergrid via a transmission network, which is overseen by a system operator (SO) through the EMS of the supergrid.

DC 1, DC 2, and DC 3 are connected to, and supplied by, distribution networks 1, 2, and 3, respectively, as depicted in Figure 4. Typical consumers, producers, and prosumers that provide *only temporal flexibility* can be managed by DSOs through D-EMSs at the distribution level and by the SO through the EMS at the transmission level. In addition, DC 1, DC 2, and DC 3, as part of global DCs that provide *both spatial and temporal flexibility*, are required not only to exchange information and electricity across D-EMS 1, D-EMS 2, and D-EMS 3, respectively, but also to exchange workload information with the global management system of DCs. This essential exchange of information and electricity across IT and power domains takes place to ensure a resilient, cost-efficient supply–demand balance in response to day-ahead forecasts and real-time changes in renewable production, electricity consumption, and price signals in energy and reserve markets. In the sequel, the management systems of coupled IT and power domains, from microgrids to supergrids, are further discussed in detail.

DC Energy Management in Microgrids

This section focuses on a DC integrated into a single microgrid, which often operates as a standalone network, as outlined in Figure 1, or occasionally connects to the grid. In such scenarios, the DC may interact with the grid by either supporting its operation or requesting support to ensure secure functioning.

The core of a DC at the microgrid level is its EMS, which serves as the central intelligent controller for the local system. At this scale, the EMS is primarily an internally focused system in which the external grid is treated as a backup resource. It operates by collecting high-frequency, real-time data from both 1) multienergy assets, such as on-site renewable energy sources (RESs), combined cooling, power, and heating facilities supplied by electricity, gas, or hydrogen, and ESSs; and 2) the DC-owned IT management systems.

This bidirectional power–IT coupling transfers data to the server GPU, workload queues, and the thermal state of the cooling infrastructure. This information goes into a control and dispatch module whose primary objective is to guarantee a cost-efficient UPS for the IT equipment.

To achieve these objectives, the EMS operates through *offline day-ahead* and *online real-time* stages (Nikkhah et al., 2022). In its offline day-ahead planning stage, the EMS uses forecasts for weather conditions, market prices, and incoming IT workloads to create an optimal day-ahead schedule for all available on-site resources. This phase also includes crucial resilience analysis, in which grid outage scenarios are simulated to preplan islanding responses and establish load-shedding priorities for noncritical DC infrastructure. Following this day-ahead plan, the EMS executes its real-time functions on a seconds-to-minutes basis. This includes 1) dynamic load balancing by adjusting local resources to match the DC's real-time demand and 2) the continuous allocation of computational workloads across servers to cooptimize for energy efficiency and performance. The most critical online function is *emergency management*. Upon detecting a grid failure or undesired perturbation that could lead to instability, the EMS must autonomously manage a sufficiently fast transition to islanded mode. This fast islanding ensures the protection of critical IT load, maintains stable microgrid voltage and frequency, and facilitates a secure reconnection upon grid restoration.

Through the described integrated planning and control functions, the EMS can enhance resilience and economic efficiency at the microgrid level. However, these advantages may be hindered by challenges that are particularly critical in the DC context. Beyond the high capital expenditure for DERs/RESs and the DC itself, the primary challenge is the control complexity. The DC's EMS must solve an interdisciplinary problem that balances the tightly coupled computing, thermal, and electrical areas, which is a task far more complex than ordinary management of a typical power-only microgrid, also requesting a high quantity of data to be processed.

DC Energy Management in Distribution Networks

This section focuses on DCs integrated into 11-kV, 22-kV, or 33-kV distribution networks and discusses operation and integration challenges from the monitoring, protection, and control perspectives.

When a DC is integrated into a medium-voltage distribution network, its EMS evolves from an internal controller into an interactive system to coordinate with the external grid. The EMS's objective function therefore expands, requiring it to balance the DC's internal priorities with the technical and financial constraints of the wider distribution grid. Unlike the internally focused microgrid EMS, its inputs are augmented with critical external data feeds from the DSO, such as real-time locational marginal prices, demand response signals, and local grid conditions. This transforms the control strategy in an EMS into a multiobjective, nonlinear optimization problem, often formulated as a bilevel Stackelberg game, in which the DC and DSO represent the leader and the follower, respectively. Consequently, the EMS's outputs include not only internal control commands but also external communications, providing the DSO with forecasts of the DC's shiftable and nonshiftable load demand, thereby transforming the DC into a dispatchable grid resource.

As in the microgrid-level EMS, offline and online modules coordinate internal DC operations with DSO signals to optimize flexibility provision and maintain stability. However, this level of coordination also presents significant challenges. It requires robust, standardized communication protocols between the DC's and DSO's EMSs, which are often underdeveloped. Moreover, the high concentration of power-electronic-based DC infrastructures can introduce stability risks if not properly managed as an unexpected load change can create severe voltage and frequency deviations on local feeders.

The integration of a flexible DC fundamentally strains the capabilities of the traditional D-EMS. These systems often lack the granular, real-time visibility at the medium- and low-voltage levels necessary to track the rapid voltage fluctuations and harmonic distortions caused by the DC's dynamic, power-electronic-based load. This deficiency necessitates advanced sensing infrastructure, such as microphasor measurement units, for high-fidelity situational awareness. Critically, a DC's power consumption is not a simple, predictable profile but a complex function involving the stochastic nature of the IT workload and the facility's thermal state. Without a direct data feed or an accurate physics-informed model reflecting these internal dynamics, the DSO's state estimation algorithms become unreliable, leading to ineffective grid management and potentially flawed operational decisions.

The temporal flexibility of a distribution-connected DC introduces a profound, dual impact on local grid operations, acting as either a powerful tool for optimization or a significant source of instability, depending on the level of coordination. When actively managed, the DC can leverage its temporal resources to significant effect. By scheduling delay-tolerant workloads or utilizing the thermal inertia of its cooling systems to increase electricity consumption during periods of high renewable production, the DC performs as a controllable load to absorb excess local energy, thereby mitigating overvoltage issues and increasing the grid's RES consumption. Similarly, it can defer noncritical tasks (e.g., backups, archiving, software updates, or nonurgent AI-training) away from network peak hours, alleviating thermal congestion on transformers and feeders and potentially deferring costly infrastructure upgrades. However, this flexibility may also potentially lead to a security threat. An uncoordinated, unexpected disconnection of a multimegawatt DC load, triggered by an internal fault or the sensitive reaction of its UPS to an external disturbance, constitutes a major contingency event for the local distribution network. Such events can cause severe, localized voltage and frequency deviations, threatening the stability of the feeder and risking cascading outages, which ultimately compromise system resilience.

DC Energy Management in Transmission Networks/Supergrids

For hyperscale DCs connected at/to transmission-level networks/supergrids, the role of the EMS evolves into that of a strategic, system-level controller. Often operating as a centralized virtual power plant or a multiagent system coordinator, this advanced EMS manages a cluster of geographically distributed DCs, coordinating them to perform as a highly flexible resource. The focus shifts from local optimization to holistic coordination and market participation.

The architecture of the transmission/supergrid-level EMS is hierarchical and geographically distributed. A central EMS platform processes not only aggregated operational data from each individual DC site but also high-level market and grid data on a national or regional scale, such as wholesale energy and reserve prices, interregional transmission congestion data, and real-time frequency regulation signals. The central processing and control computer solves the large-scale synergistic optimization of power and computing resources. Its objective is to fulfill the total computational demand of the DC portfolio at the minimum possible system-wide cost (Dvorkin, 2025). The central controller then issues high-level dispatch targets to the local EMS at each site, which is subsequently responsible for performing the intrasite optimization to meet that target. For the day-ahead market, the EMS solves an offline optimization problem to formulate its bidding strategy based on price forecasts and the cluster's collective flexibility for the wholesale energy market. In real time, it solves a continuous dispatch problem, whose outputs are the workload reallocation commands that create the "*virtual transmission line*" between geographically distributed DCs by responding to immediate grid congestions and price signals. This capability can also enable system-wide renewable balancing as the coordinated cluster can absorb massive amounts of renewables that would otherwise be curtailed by routing workloads to regions experiencing a surplus of wind or solar power. This coordination can, in turn, provide an effective alternative in which demand-side resources can be used to eliminate the need for reinforcing capital-intensive transmission lines. Furthermore, it can enhance grid resilience against regional events, such as extreme weather conditions, by allowing critical computational services to be migrated away from affected areas. However, the scale of these operations also introduces a significant systemic risk of a simultaneous cluster disconnection, which could potentially threaten regional stability. This capability depends on extremely high-bandwidth and low-latency wide-area communication networks. Finally, it presents novel and complex regulatory questions for market design and grid operation that are still being explored.

In Figure 5, the global DC EMS is outlined, summarizing the microgrid, distribution and transmission network issues, data flow, and applications.

DCs for Power System Ancillary Services

This section further explores the role of DCs in ancillary services, focusing on challenges, opportunities, and potential solutions for integrating hyperscale DC facilities into transmission-level networks. Hyperscale DCs are designed for massive scale and global reach, with siting driven by access to low-cost power, land, renewable hubs, favorable climate, and supportive policy rather than proximity to end users. Interconnection via high-bandwidth, low-latency fiber-optic networks provides the basis for spatiotemporal flexibility. As a result, clusters of hyperscale DCs can act as a virtualized “computing power network” that is tightly coupled with the physical transmission grid, optimizing operations across time and space.

Different Types of Ancillary Services

The unintentional disconnection of a hyperscale DC cluster can represent a *major N-1 contingency*, comparable to the sudden loss of a large power plant. Wide Area Monitoring, Protection, and Control (WAMPAC) schemes must be prepared for such events to prevent cascading failures and potential blackouts. Unlike traditional loads, DCs can migrate critical services to unaffected regions, ensuring digital continuity during grid disturbances. The participation of DCs in ancillary services is becoming increasingly important in weak, low-inertia, and low-fault-level networks with a high penetration of IBRs, such as wind and solar sources. While ancillary services, such as frequency regulation, voltage support, and reserve provision, were traditionally delivered by synchronous generators, their displacement by IBRs necessitates new sources of flexibility, and DCs, with their significant, controllable demand, are suitable to support these services. The key potential ancillary services from DCs are summarized as follows:

- *Frequency response and reserves*: DCs already incorporate UPS systems with batteries to protect IT loads. These batteries can be harnessed for fast frequency response, injecting or absorbing power within seconds to stabilize frequency after disturbances. Additionally, flexible IT workloads (such as AI training or batch processing) can be curtailed or shifted to provide operating reserves, helping to balance the intermittent weather-dependent nature of IBRs.
- *Voltage support and reactive power*: Through their power electronic interfaces, UPS systems can also supply or absorb reactive power, contributing to voltage regulation at the point of connection. This is particularly valuable in distribution grids, where IBRs can create voltage fluctuations and reactive power imbalances.
- *Black-start and grid/system restoration*: With backup systems such as diesel or gas turbines, and, increasingly, potential SMR colocation, DCs could aid black-start services. Their UPS and battery systems can maintain critical operations during restoration, while their on-site production can provide sustained energy support.

As mentioned previously, SMRs can enhance the ancillary service potential of DCs. When colocated, DC–SMR facilities become prosumers, able to both consume and generate power. SMRs provide inertia-like support through turbine-generators, complementing UPS–battery fast response for frequency control. With automatic voltage regulators, they also support voltage stability. Moreover, they enable DCs to operate independently from the grid, easing transmission

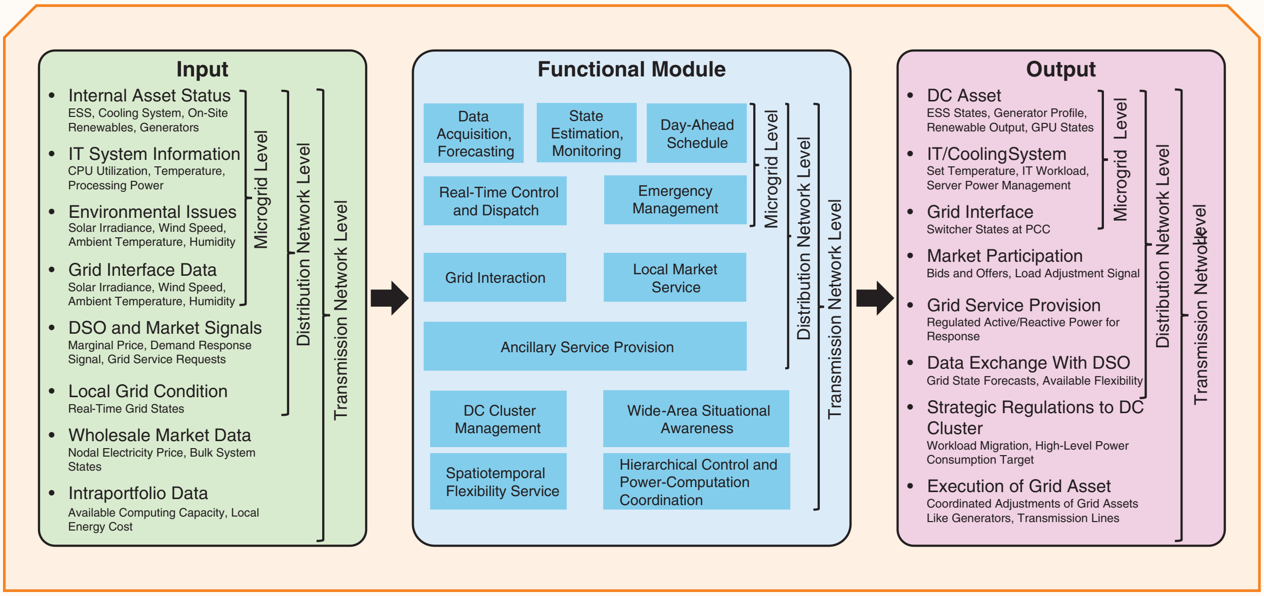


figure 5. Outline of the global DC EMS at different levels. PCC: point of common coupling.

stress and offering dispatchable flexibility. They further allow participation in longer duration services, such as secondary and tertiary reserves, beyond the capabilities of batteries alone.

Real-Life Example: Fast Frequency Containment

The increasing IBR integration has reduced system inertia in recent years, weakening frequency stability and increasing the rate of change of frequency after disturbances. Unlike synchronous machines, IBRs can provide little or no physical inertia, driving the need for fast frequency response services, such as enhanced frequency control capabilities (EFCCs) or fast frequency containment (Hong et al., 2019). DCs can release fast balancing capacity to the grid by reducing their active power demand. This can be achieved by pausing or shifting noncritical IT workloads, using the UPS and ESS to reduce reliance on the grid temporarily. These responses can be activated within seconds, supporting frequency containment in the critical first moments after an event. It is notable to mention that such responses are local, distributed, and do not sacrifice energy yield, unlike renewable curtailment. To capture this critical potential, DCs should be modeled not only as composite loads but also as flexible resources with EFCCs, with parameters defining the magnitude, speed, and recovery of deloading. Coordinated via the WAMPAC system involving phasor measurement units (PMUs) and IT, DCs can operate alongside DERs, RESs, and ESSs to mitigate inertia loss and enhance stability.

In Figure 6, the EFCC scheme (Hong et al., 2019) is presented for the Great Britain (GB) grid. It relies on time-synchronized PMUs monitoring frequency and its rate of change in different areas of the entire power grid. To ensure efficient active power contribution (injection or deloading) after major frequency events, e.g., a sudden disconnection of a large generating unit, the system can be split into regions defined by the network topology and generation mix. By relying on a dynamic GB grid equivalent, expressed through a swing equation representing the system's behavior, together with measured rates of change of frequency from different regions, the instantaneous active power imbalance can be estimated (Hong et al., 2019). Considering the statuses of different sources, also called EFCC service providers, e.g., DERs, RESs, and ESSs, but also DCs, the scheme can create a command for promptly increasing active power injections to the grid, or DC deloading. Through careful design of the scheme and known real-time power system inertia, the scheme can efficiently contribute to the frequency stability. In Figure 7, the power system frequency for two cases 1) without and 2) with the EFCC scheme is presented. Obviously, the EFCC scheme can ensure a frequency response in which the frequency statutory limit (49.5 Hz) has not been violated. Considering the future size of DCs, it is expected that they will be one of the most essential service providers supporting the presented scheme, i.e., ensuring stable system response after major outages and frequency events.

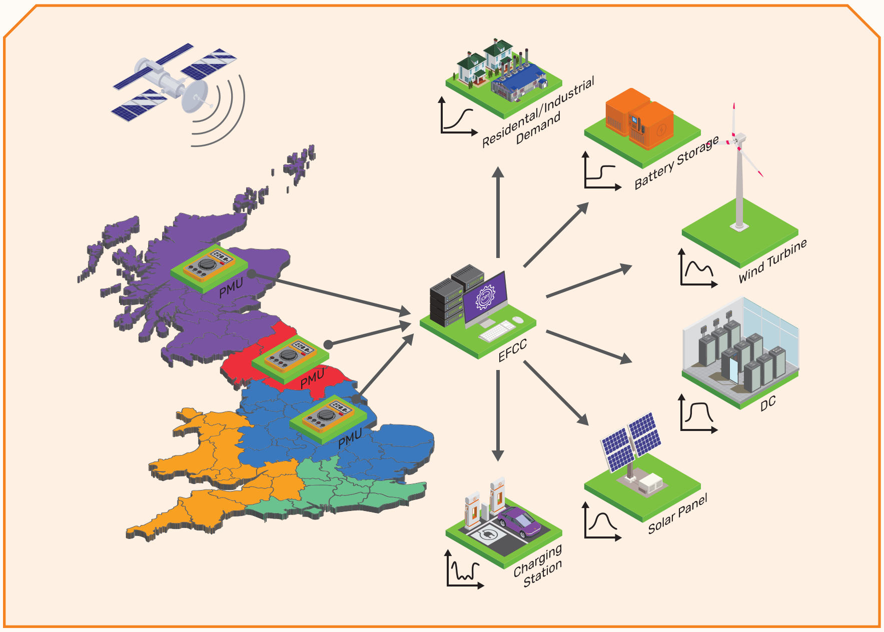


figure 6. The EFCC scheme with DCs actively supporting frequency regulation (Hong et al., 2019). PMU: phasor measurement unit.

Future Outlook and Conclusions

In the context of integrating DCs into power grids, several future challenges remain, some of which will be addressed in this section. Digital twins have been used to monitor and conduct what-if scenarios for DCs, such as in the case of Frontier, the first-ever exa-scale supercomputer. Digital twins can be extended to monitor the power domain in addition to the IT domain. They can also be extended to address reliability and security aspects in DCs (Bash et al., 2025).

To achieve maximum benefits from the increasing and accelerating integration of DCs into the energy ecosystem, multidisciplinary solutions are essential. A key requirement is the development of detailed, predictive models of DC energy consumption, extending beyond aggregate facility-level metrics to include job-level consumption patterns. Such models will likely need to be composite in nature, combining high-fidelity computational fluid dynamics simulations, to capture heat flow, cooling dynamics, and thermal interactions across the physical infrastructure, with data-driven and machine learning (ML) models, which can infer and forecast the energy footprint of workloads at the per-job level while accounting for variability in hardware, scheduling, and AI/ML training demands. By integrating physics-based and data-driven approaches, these composite models can provide actionable insights for real-time energy management and workload scheduling, enabling DCs to operate as intelligent, adaptive energy assets.

In addition, when planning new DCs, there are opportunities to cooptimize hardware and workloads. Hardware choices determine which workloads can be executed efficiently, and workload characteristics, in turn, shape hardware requirements. These interdependencies also affect a DC's ability to participate optimally in the broader distribution network. Moreover, the stochastic nature of workload-driven electricity demand and the need for effective scheduling motivate the development of fast, efficient optimization methods that can handle uncertainty (Vatani et al., 2018).

Finally, fine-grained data collection is essential for advancing the understanding of DC energy consumption. Aggregating usage at the facility or rack level masks the dynamics of workload-driven variability, cooling interactions, and hardware-specific efficiency. Capturing detailed measurements, including per-job energy use, thermal conditions, cooling system response, and power distribution, would enable the development of more accurate models that link workloads to infrastructure behavior. Such data would not only inform improved design and operation of individual DCs but also support their integration as flexible, grid-interactive assets in the broader energy ecosystem.

In summary, DCs are emerging as critical components in the evolution of modern power systems. Once treated as passive, inflexible loads, they are now increasingly recognized as active prosumers capable of contributing to system stability, flexibility, and resilience. Accurate load modeling remains a prerequisite, with hybrid static–dynamic and physics-based data-driven approaches identified as the most appropriate means of capturing the stochastic nature of IT workloads and auxiliary system dynamics. Equally, any model that involves parameters must rely on the reliable selection of those parameters. EMSs for DCs will play a decisive role at multiple levels of the grid: enabling optimization within microgrids, enhancing hosting capacity and power quality in distribution networks, supporting stability and voltage and frequency control in transmission systems, and participating in balancing mechanisms across supergrids. Furthermore, opportunities for ancillary service provision, including frequency response, voltage regulation, and black start, position DCs as valuable service providers in low-inertia, inverter-dominated networks. Looking ahead, novel configurations, such as SMR colocation, underline the potential for DCs to evolve into strong hybrid production–consumption entities. The

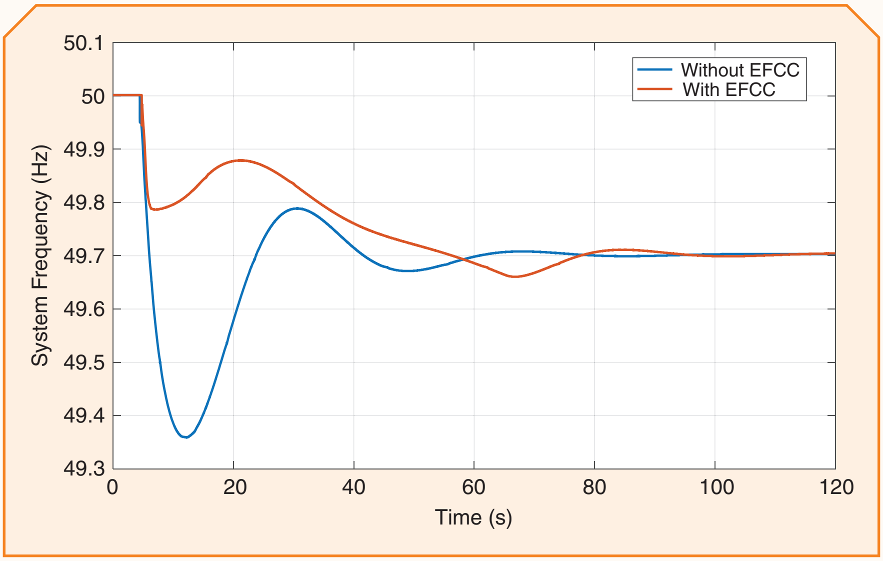


figure 7. Power system frequency response without and with the EFCC scheme (Hong et al., 2019).

evidence reviewed in this article indicates that the integration of DCs into system-level planning and operation is not only desirable but essential for building secure, resilient, and reliable electricity networks.

Acknowledgment

This work was authored in part by the National Renewable Energy Laboratory, operated by the Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes. This work is supported in part by the U.K. Research and Innovation Future Leaders Fellowship, Digitalisation of Electrical Power and Energy Systems Operation, under Grant MR/W011360/2, and by the Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC) through funding provided to the Energy Demand Research Centre (EDRC) under Grant EP/Y010078/1.

For Further Reading

C. Bash, J. Bian, D. Milojicic, C. D. Patel, L. Strezoski, and V. Terzija, "Energy supplies for future data centers," *Computer*, vol. 57, no. 7, pp. 126–134, Jul.2024, doi: 10.1109/MC.2024.3393248.

"Future energy scenarios: Pathways to net zero," Nat. Energy Syst. Operator, Warwick, U.K., Version 4, Jul.2025. [Online]. Available: https://www.neso.energy/document/364541/download

"Energy and AI: World energy outlook special report," IEA, Paris, France, Feb.2025. [Online]. Available: https://iea.blob.core.windows.net/assets/601eaec9-ba91-4623-819b-4ded331ec9e8/EnergyandAI.pdf

"Incident review: Considering simultaneous voltage-sensitive load reductions," NERC, Atlanta, GA, USA, Jan.2025. [Online]. Available: https://www.nerc.com/globalassets/our-work/reports/event-reports/incident_review_large_load_loss.pdf

P. Regulski, D. S. Vilchis-Rodriguez, S. Djurović, and V. Terzija, "Estimation of composite load model parameters using an improved particle swarm optimization method," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 553–560, Apr.2015, doi: 10.1109/TPWRD.2014.2301219.

S. Nikkhah, I. Sarantakos, N. M. Zografou-Barredo, A. Rabiee, A. Allahham, and D. Giaouris, "A joint risk- and security-constrained control framework for real-time energy scheduling of islanded microgrids," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3354–3368, Sep.2022, doi: 10.1109/TSG.2022.3171816.

V. Dvorkin, "Agent coordination via contextual regression (AgentCONCUR) for data center flexibility," *IEEE Trans. Power Syst.*, vol. 40, no. 2, pp. 1832–1842, Mar.2025, doi: 10.1109/TPWRS.2024.3442954.

Q. Hong *et al.*, "Fast frequency response for effective frequency control in power systems with low inertia," *J. Eng.*, vol. 2019, no. 16, pp. 1696–1702, 2019, doi: 10.1049/joe.2018.8599.

C. Bash, J. McDonald, D. Milojicic, J. Mueller, T. Wilson, and Y. Zhang, "Convergence of energy and IT sectors," *Computer*, vol. 58, no. 1, pp. 151–158, Jan.2025, doi: 10.1109/MC.2024.3498572.

B. Vatani, B. Chowdhury, S. Dehghan, and N. Amjady, "A critical review of robust self-scheduling for generation companies under electricity price uncertainty," *Int. J. Electr. Power Energy Syst.*, vol. 97, pp. 428–439, Apr.2018, doi: 10.1016/j.ijepes.2017.10.035.

Biographies

Vladimir Terzija is with the School of Engineering, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K.

Shahab Dehghan is with the School of Engineering, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K. Shahab Dehghan is the corresponding author.

Damian Giaouris is with the School of Engineering, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K.

David Greenwood is with the School of Engineering, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K.

Xin Zhang is with the School of Electrical and Electronic Engineering, The University of Sheffield, S1 3JD Sheffield, U.K.

Juliane Mueller is with the Computational Science Center, National Renewable Energy Laboratory, Golden, CO 80401-3393 USA.

Dejan Milojicic is with HPE Labs, Milpitas, CA 94306 USA.