Flexible Scheduling of Hyperscale Data Center Loads Under Transmission Constraints

Emmanuel Baskerville[1], Kamar Mann[1], Cyrus Booker[1], Brenden Forrest[1], Moses Garuba[1], Cunzhi Zhao[2], and Xiang Huo[1]*
*Corresponding Author: xiang.huo@hamptonu.edu

[1]Department of Electrical and Computer Engineering
Hampton University
Hampton, VA, USA

[2]Department of Engineering and Computer Science
McNeese State University
Lake Charles, LA, USA

***Abstract*—The rapid expansion of hyperscale data centers (DCs) introduces large, geographically concentrated, and highly uncertain electrical demands that challenge existing transmission operation practices. Unlike traditional load growth, which evolves gradually and is spatially distributed, hyperscale DC deployments exhibit sharp demand increases at specific network locations, significantly impacting power flow patterns, congestion, and system costs. To address these challenges, this paper develops a network-constrained optimization framework to manage the influence of DC load growth on transmission system performance. The objective function incorporates key components of electricity cost minimization, congestion mitigation, and temporal smoothness of DC load variation. Simulations are conducted on the modified IEEE 14-bus system under different DC load control scenarios. The results demonstrate the effectiveness of the proposed approach compared to the conventional greedy (non-coordinated) DC load scheduling.**

***Keywords*—Data centers, large load management, optimization, power transmission system**

I. Introduction

The exponential growth of hyperscale data centers (DCs) is leading to significantly concentrated power demand in modern power systems [1]. Spearheaded by cloud computing, artificial intelligence, and high-performance computing applications, individual DC facilities now can require power capacities comparable to small metropolitan areas [2]. Traditional load growth typically follows gradual demographic and economic trends, while in contrast, DC development is characterized by abrupt increases in demand at geographically concentrated locations [3]. The effective integration and optimal operation of DCs is nontrivial toward the reliable, low-cost, and resilient operation of power transmission networks [4].

Modern DCs have emerged as a rapidly growing component of electrical demand, consuming a significant share of national electricity supply and continuing to expand as digital services scale [5]. These facilities exhibit complex multi-level power architectures, where energy flows from utility substations through building distribution systems to server-level and chip-level loads, introducing new efficiency and reliability challenges. As a result, the increasing scale of DCs places substantial pressure on power system infrastructure, reinforcing the need for advanced planning and operational strategies.

Within power transmission systems, load growth planning includes load forecasting to match certain reliability requirements, such as N-1 contingency, to ensure reliability [6]. However, uncertainty in hyperscale DC deployment complicates traditional forecasting approaches, where a small underestimation of DC demand can lead to thermal overloads, voltage instability, and congestion, and overestimation can result in premature capital investment and stranded infrastructure [7]–[9]. Moreover, DCs have evolved from simple server farms into highly complex, large-scale infrastructures integrating advanced networking, storage, and computational architectures to support cloud computing and data-intensive applications [10]. Consequently, architectural and operational advancements in DCs have been developed to maximize the efficiency of DC resources, such as servers, memory, and network devices that can be manipulated by the DC controller to achieve energy efficiency [11].

Nevertheless, the integration of DCs into transmission networks requires advanced modeling and optimization methodologies that can capture their scale, temporal variability, and impact, particularly on power transmission network constraints. In [12], the authors examine existing designs for the transmission of flows in DC networks, noting that traffic must handle the bandwidth demands generated by a broad number of applications while achieving high utilization of DC infrastructure. Vafamehr *et al.* in [13] develop a framework for the expansion planning of DCs under uncertainty, demonstrating that capacity limits in electricity networks directly reduce the reliability and security of energy supply and that DC growth must be jointly evaluated with transmission network constraints. Complementing this, [14] assesses hyperscale DC interconnections on a benchmark power system, finding that active-power ramp-rate limits must be enforced to prevent oscillations when large DCs connect at high-voltage points of interconnection. Since transmission infrastructure development is time-intensive. Despite the efforts, current transmission planning and operation approaches often rely on single-point load forecasts that do not fully capture the variability of

emerging hyperscale DC demands. As DCs continue to expand, there is an increasing need for scenario-based and risk-informed operation strategies capable of evaluating varying demand trajectories [3], [15].

To this end, this paper develops a flexible network-constrained optimization framework to evaluate the impact of hyperscale DC load growth on the transmission system operation. Section II models the DC-integrated power transmission networks through a constrained optimization problem. Section III studies different scenarios under the proposed DC load scheduling strategy. Section IV concludes the paper.

II. Problem Formulation

A. Power Transmission Networks with Data Centers

Power transmission networks are engineered to facilitate the reliable and economic delivery of electricity from geographically dispersed generation resources to spatially distributed load centers. The operational state of such networks is governed by power flow equations, subject to constraints such as thermal limits, voltage stability requirements, and constraints of other power system components (e.g., DC constraints). Fig. 1 shows a modified IEEE 14-bus system where a DC is connected at Bus 14.

A critical operational challenge in transmission networks is congestion, which arises when line thermal ratings or stability margins limit power transfers. Congestion induces non-uniform marginal costs, resulting in spatial price separation and reduced system efficiency, particularly evident in the emergence of hyperscale DCs. Moreover, the integration of DC loads can largely exacerbate congestion patterns and reduce available transfer capability, resulting in abrupt, localized increases in demand at specific buses. To address these challenges, this paper develops a network-constrained optimization framework that can optimize generation dispatch, network utilization, and DC load scheduling. DCs are modeled as composite loads interconnected at certain buses within the transmission network. Their demand is decomposed into idle, fixed, and variable components, where the variable component represents flexible DC workloads that can be temporally shifted. This integrated framework forms the basis for the optimization problem presented in the following subsection.

B. Optimization Problem Formulation

The goal of the network-constrained optimization problem is to find the optimal operation decisions of generators and DCs, thereby supporting the cost-efficient operation of the power transmission network [16], [17]. To this end, the objective function is defined as:

$$\min \sum_{g\in\mathcal{G},t\in\mathcal{T}} \left(C_g^{\mathrm{nl}} + \sum_{j=1}^{S_g} c_{g,j}^{\mathrm{s}} P_{g,j,t}^{\mathrm{seg}} \right) \Delta t + \lambda_1 \sum_{\ell\in\mathcal{L},t\in\mathcal{T}} \left(\frac{f_{\ell,t}}{F_\ell^{\max}} \right)^2 + \lambda_2 \sum_{k\in\mathcal{K},t\in\mathcal{T}\setminus\{1\}} \left(P_{k,t}^{\mathrm{vl}} - P_{k,t-1}^{\mathrm{vl}} \right)^2 \quad (1)$$

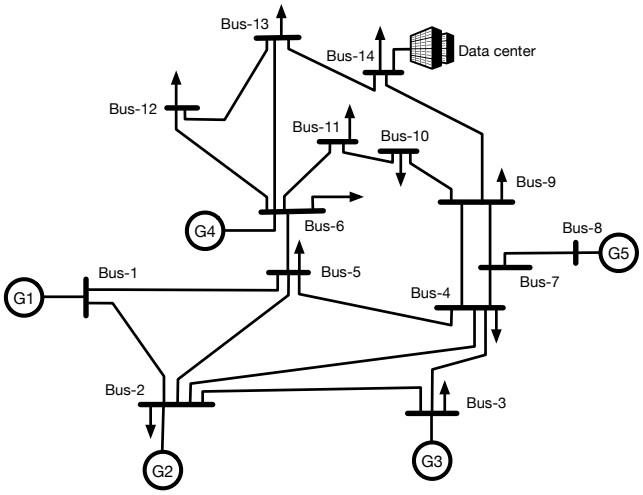

Fig. 1. Example of a modified IEEE 14-bus system, where a data center is connected at Bus 14.

For the objective function (1), the first term $\sum_{g\in\mathcal{G},t\in\mathcal{T}} \left(C_g^{\mathrm{nl}} + \sum_{j=1}^{S_g} c_{g,j}^{\mathrm{s}} P_{g,j,t}^{\mathrm{seg}} \right) \Delta t$ describes the costs associated with generators, where $\mathcal{T}$ and $\mathcal{G}$ denote the sets of time periods and generators, respectively, C_g^{nl} represents the no-load cost of generator g, $P_{g,t}$ is the total power output of generator g at time t, $P_{g,j,t}^{\mathrm{seg}}$ is the power produced in segment j of generator g at time t, S_g is the number of cost segments for generator g, $c_{g,j}^{s}$ is the marginal cost associated with segment j of generator g, Δt is the duration of each time interval.

The second term $\lambda_1 \sum_{\ell\in\mathcal{L},t\in\mathcal{T}} \left(\frac{f_{\ell,t}}{F_\ell^{\max}} \right)^2$ represents a congestion penalty that discourages operating transmission lines close to their thermal limits, where $f_{\ell,t}$ denotes the power flow on line ℓ at time t, $F_\ell^{\max}$ is the corresponding line capacity, and $\mathcal{L}$ denotes the set of lines. The ratio $\frac{f_{\ell,t}}{F_\ell^{\max}}$ represents the normalized loading of line ℓ. The quadratic form assigns a higher penalty to flows approaching the line capacity, thereby encouraging the optimization to distribute power more evenly across the network and reduce congestion. The weighting parameter $\lambda_1 \geq 0$ controls the trade-off between economic efficiency and congestion mitigation.

The last term $\lambda_2 \sum_{k\in\mathcal{K},t\in\mathcal{T}\setminus\{1\}} \left(P_{k,t}^{\mathrm{vl}} - P_{k,t-1}^{\mathrm{vl}} \right)^2$ penalizes the rapid fluctuation of the variable DC load over time [8], [18]. This term encourages smoother scheduling of flexible workloads to reduce operational stress in the transmission network. The parameter $\lambda_2 \geq 0$ controls the trade-off between cost minimization and temporal smoothness, and $\mathcal{K}$ denotes the set of DCs. By limiting the ramping behavior of the DC controllable load, abrupt demand shifts are prevented to avoid voltage fluctuations, frequency deviations, or congestion in the network. It also captures practical limitations in DC workload migration and task scheduling, where instantaneous large-scale power adjustments are infeasible due to computational, thermal, and communication constraints.

Next, we present the constraints for the DC-integrated

transmission operation problem, described by Eqs. (2)-(13).

Constraints (2) and (3) limit the ramp-up and ramp-down rates of generators, ensuring that the change in generation output between consecutive time periods remains within allowable limits:

$$P_{g,t} - P_{g,t-1} \le R_g^{\text{up}}, \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \setminus \{1\}, \tag{2}$$

$$P_{g,t-1} - P_{g,t} \le R_g^{\text{down}}, \quad \forall g \in \mathcal{G},\ t \in \mathcal{T} \setminus \{1\}, \tag{3}$$

where R_g^{up} and R_g^{down} denote the ramp-up and ramp-down limits of the generator g, respectively.

Constraints (4) and (5) define the piecewise linear representation of generator output, where the total generation is expressed as the sum of segment outputs as:

$$P_{g,t} = \sum_{j=1}^{S_g} P_{g,j,t}^{\text{seg}}, \quad \forall g \in \mathcal{G},\ t \in \mathcal{T}, \tag{4}$$

where $P_{g,j,t}^{\text{seg}}$ denotes the power produced in segment j of generator g at time t, S_g is the number of segments, and each segment is bounded by:

$$0 \le P_{g,j,t}^{\text{seg}} \le \bar{P}_{g,j}^{\text{seg}}, \quad \forall g \in \mathcal{G},\ j = 1, \ldots, S_g,\ t \in \mathcal{T}, \tag{5}$$

where $\bar{P}_{g,j}^{\text{seg}}$ is the maximum capacity of generator g for segment j.

The generator power output limits are enforced by:

$$P_g^{\min} \le P_{g,t} \le P_g^{\max}, \quad \forall g \in \mathcal{G},\ t \in \mathcal{T}, \tag{6}$$

where $P_g^{\min}$ and $P_g^{\max}$ denote the minimum and maximum generation capacities, respectively.

Constraint (7) imposes the transmission line thermal limits:

$$F_\ell^{\min} \le F_{\ell,t} \le F_\ell^{\max}, \quad \forall \ell \in \mathcal{L},\ t \in \mathcal{T}, \tag{7}$$

where $F_{\ell,t}$ denotes the power flow on line ℓ at time t, and $F_\ell^{\min}$ and $F_\ell^{\max}$ denote the minimum and maximum line flow limits, respectively.

A DC power flow model is adopted, which simplifies the network constraints using a linear and computationally efficient model, as described below:

$$F_{\ell,t} = -b_\ell S_{\text{base}} \left(\theta_{i,t} - \theta_{j,t}\right), \quad \forall \ell \in \mathcal{L},\ t \in \mathcal{T}, \tag{8}$$

where b_ℓ is the susceptance of line ℓ, S_{base} is the system base power, and $\theta_{i,t}$ and $\theta_{j,t}$ are the voltage angles at the sending and receiving buses i and j of line ℓ, respectively.

Constraint (9) sets the reference bus angle:

$$\theta_{1,t} = 0, \quad \forall t \in \mathcal{T}, \tag{9}$$

where Bus 1 is the slack bus.

The voltage angle limits at all non-reference buses are constrained by:

$$\theta_i^{\min} \le \theta_{i,t} \le \theta_i^{\max}, \quad \forall i \in \mathcal{B} \setminus \{1\},\ t \in \mathcal{T}, \tag{10}$$

where $\theta_i^{\min}$ and $\theta_i^{\max}$ denote the lower and upper bounds of voltage angles, respectively, and $\mathcal{B}$ denotes the set of buses.

Constraint (11) enforces nodal power balance:

$$\sum_{j \in \Omega_i^{\text{in}}} F_{j,t} - \sum_{j \in \Omega_i^{\text{out}}} F_{j,t} + \sum_{g \in \mathcal{G}_i} P_{g,t} = D_{i,t} + \sum_{k \in \mathcal{K}_i} \left(P_{k,t}^{\text{idle}} + P_{k,t}^{\text{fl}} + P_{k,t}^{\text{vl}}\right), \ \forall i \in \mathcal{B},\ t \in \mathcal{T}, \tag{11}$$

where $D_{i,t}$ denotes the base demand of bus i at time t, $P_{k,t}^{\text{idle}}$ denotes the average idle power of DC k at time t, $P_{k,t}^{\text{fl}}$ denotes the fixed load of DC k at time t, $P_{k,t}^{\text{vl}}$ denotes the variable DC load, $\mathcal{G}_i$ denotes the set of generators connected at bus i, $\mathcal{K}_i$ denotes the set of DCs connected at bus i, and Ω_i^{in} and Ω_i^{out} denote the sets of incoming and outgoing lines at bus i, respectively.

Constraint (12) ensures that the total variable energy demand is satisfied:

$$\sum_{t \in \mathcal{T}} P_{k,t}^{\text{vl}}\, \Delta t = E_k^{\text{vl}}, \quad \forall k \in \mathcal{K}, \tag{12}$$

where E_k^{vl} denotes the required total energy for variable load of DC k.

Moreover, the variable load at each time step needs to stay within the limit of:

$$0 \le P_{k,t}^{\text{vl}} \le \bar{P}_k^{\text{DC}} - P_k^{\text{fl}} - P_k^{\text{idle}}, \quad \forall k \in \mathcal{K},\ t \in \mathcal{T}, \tag{13}$$

where $\bar{P}_k^{\text{DC}}$ denotes the maximum capacity of the kth DC.

By combining the aforementioned power transmission system objective and constraints, the flexible DC load scheduling problem is formulated into the following networked optimization problem:

$$\begin{aligned}
\min \ & \sum_{g \in \mathcal{G}, t \in \mathcal{T}} \left(C_g^{\text{nl}} + \sum_{j=1}^{S_g} c_{g,j}^{\text{s}} P_{g,j,t}^{\text{seg}} \right) \Delta t \\
& + \lambda_1 \sum_{\ell \in \mathcal{L}, t \in \mathcal{T}} \left(\frac{f_{\ell,t}}{F_\ell^{\max}} \right)^2 \\
& + \lambda_2 \sum_{k \in \mathcal{K}, t \in \mathcal{T} \setminus \{1\}} \left(P_{k,t}^{\text{vl}} - P_{k,t-1}^{\text{vl}} \right)^2 \\
\text{s.t. } & P_{g,t} - P_{g,t-1} \le R_g^{\text{up}}, \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \setminus \{1\} && (14) \\
& P_{g,t-1} - P_{g,t} \le R_g^{\text{down}}, \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \setminus \{1\} && (15) \\
& P_{g,t} = \sum_{j=1}^{S_g} P_{g,j,t}^{\text{seg}}, \quad \forall g \in \mathcal{G}, t \in \mathcal{T} && (16) \\
& 0 \le P_{g,j,t}^{\text{seg}} \le \bar{P}_{g,j}^{\text{seg}}, \quad \forall g \in \mathcal{G}, j = 1, \ldots, S_g, t \in \mathcal{T} && (17) \\
& P_g^{\min} \le P_{g,t} \le P_g^{\max}, \quad \forall g \in \mathcal{G}, t \in \mathcal{T} && (18) \\
& F_\ell^{\min} \le F_{\ell,t} \le F_\ell^{\max}, \quad \forall \ell \in \mathcal{L}, t \in \mathcal{T} && (19) \\
& F_{\ell,t} = -b_\ell S_{\text{base}} \left(\theta_{i,t} - \theta_{j,t}\right), \quad \forall \ell \in \mathcal{L}, t \in \mathcal{T} && (20) \\
& \theta_{1,t} = 0, \quad \forall t \in \mathcal{T} && (21) \\
& \theta_i^{\min} \le \theta_{i,t} \le \theta_i^{\max}, \quad \forall i \in \mathcal{B} \setminus \{1\}, t \in \mathcal{T} && (22) \\
& \sum_{j \in \Omega_i^{\text{in}}} F_{j,t} - \sum_{j \in \Omega_i^{\text{out}}} F_{j,t} + \sum_{g \in \mathcal{G}_i} P_{g,t} = D_{i,t} && (23)
\end{aligned}$$

$$+ \sum_{k \in \mathcal{K}_i} \left(P^{\text{idle}}_{k,t} + P^{\text{fl}}_{k,t} + P^{\text{vl}}_{k,t} \right), \ \forall i \in \mathcal{B}, \ t \in \mathcal{T}, \tag{24}$$

$$\sum_{t \in \mathcal{T}} P^{\text{vl}}_{k,t} \, \Delta t = E^{\text{vl}}_{k}, \quad \forall k \in \mathcal{K}, \tag{25}$$

$$0 \leq P^{\text{vl}}_{k,t} \leq P^{\text{DC}}_{k} - P^{\text{fl}}_{k} - P^{\text{idle}}_{k}, \quad \forall k \in \mathcal{K}, \ t \in \mathcal{T}. \tag{26}$$

III. Case Studies and Simulation Results

A. Study Setup

To evaluate the impact of hyperscale DC load flexibility on the transmission system operation, simulations are conducted on the modified IEEE 14-bus system shown in Fig. 1. The proposed optimization model is implemented in Python and solved via the Gurobi solver. The machine specifications used to run the code for the test cases are: Apple M2 Max (chip), and 32 GB of memory. A day-ahead scheduling horizon is considered with 24 time slots, i.e., an hourly resolution of $\Delta t = 1$ h. The system base power is set to be 100 MVA. A time-varying demand profile is applied to all buses using a normalized daily load curve to emulate realistic system operation. In total, 5 generators are connected to buses $\mathcal{G} = \{1, 2, 3, 6, 8\}$ in the 14-bus transmission network. The cost coefficients of these generators are set to be $\{c^{\text{s}}_{g,j}\}_{g \in \mathcal{G}} = \{20, 22, 24, 23, 21\}$ \$/MWh^2, $\forall j = 1, \ldots, S_g$, the no-load costs are $\{c^{\text{nl}}_{g}\}_{g \in \mathcal{G}} = \{80, 100, 120, 110, 90\}$ \$/hr.

A DC is connected to Bus 14. The DC load consists of three components:

- *Idle load* (P^{idle}_{1}): fixed demand of 10 MW;
- *Fixed load* ($P^{\text{fl}}_{1,t}$): time-varying demand randomly sampled between 2–3 MW;
- *Variable load* ($P^{\text{vl}}_{1,t}$): flexible demand subject to energy and power transmission system constraints.

The total variable energy requirement is set to $E^{\text{vl}}_{1} = 150$ MWh over the scheduling horizon. The maximum DC capacity is set to be 20 MW, which constrains the instantaneous variable load. In specific, two cases are tested:

- **Case 1 (Greedy scheduling):** The variable load (batch jobs) is processed at the maximum available power rate immediately upon arrival.
- **Case 2 (Optimal flexible scheduling):** The variable load is optimized to minimize system cost, network congestion, and load fluctuation while satisfying system requirements.

In the next section, simulations are conducted for varying DC load control strategies.

B. Simulation Results

Fig. 2 shows the DC load schedules at Bus 14, including the idle DC load, fixed DC load, variable DC loads under both cases, and the variable load upper bound. As can be seen, in Case 1, the variable loads are processed at the maximum rate as soon as they arrive at the DC and finish around 5 PM. Case 2 decreases the electricity cost without considering the line congestion and load fluctuation objectives (by setting penalties $\lambda_1 = 0, \lambda_2 = 0$)

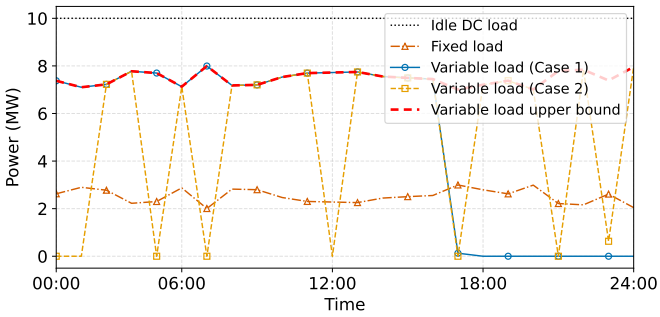

Fig. 2. DC load schedule at Bus 14 ($\lambda_1 = 0, \lambda_2 = 0$).

Table I shows the comparison of total costs for Cases 1 and 2 under three different scenarios, where each scenario represents a different emphasis on the congestion penalty and/or the fluctuation penalty, respectively.

TABLE I
Cost Reduction on the IEEE 14 Bus System

Scenario		Cases	Total cost (\$)	Cost reduction (\$)
S1	$\lambda_1 = 10^6$ $\lambda_2 = 0$	Case 1 Case 2	410643 409788	854
S2	$\lambda_1 = 0$ $\lambda_2 = 10^3$	Case 1 Case 2	198801 142896	55904
S3	$\lambda_1 = 10^6$ $\lambda_2 = 10^3$	Case 1 Case 2	466547 411025	55521

The generators' cost coefficients for both cases are time-invariant. In Scenario 1 (S1), there is no fluctuation penalty, while the congestion penalty is set to be $\lambda_1 = 10^6$, resulting in a total of 854 \$ cost reduction; In contrast, S2 has only the fluctuation penalty as $\lambda_2 = 10^3$, but without the congestion penalty, resulting in a total of 55940 \$ cost reduction; and S3 has both penalties, and similar cost reduction effectiveness as S2. Therefore, the proposed approach alleviates line congestion and/or load fluctuations under different scenarios. Compared with greedy scheduling, the optimal flexible scheduling yields a lower overall objective value owing to improved load shifting and congestion mitigation on peak-demand periods and high-congestion time intervals.

Fig. 3 illustrates the fluctuation of the variable DC load profile at Bus 14 under changing variation penalty λ_2 under Case 2. Compared to Case 1, where the variable load is scheduled greedily, Case 2 allows adjustment of the variable DC load, leading to a flatter curve. Furthermore, the impact of DC load flexibility on transmission congestion is evaluated through line flow analysis. Fig. 4 shows the congestion ratio change between Case 1 and Case 2, i.e., $\left(\frac{\tilde{f}_{\ell,t}}{F^{\max}_{\ell}}\right)^2 - \left(\frac{\hat{f}_{\ell,t}}{F^{\max}_{\ell}}\right)^2$, $\forall \ell \in \mathcal{L}, t \in \mathcal{T}$, where $\tilde{f}_{\ell,t}$ and $\hat{f}_{\ell,t}$ denote the line flows under Case 1 and Case 2, respectively. The congestion rate

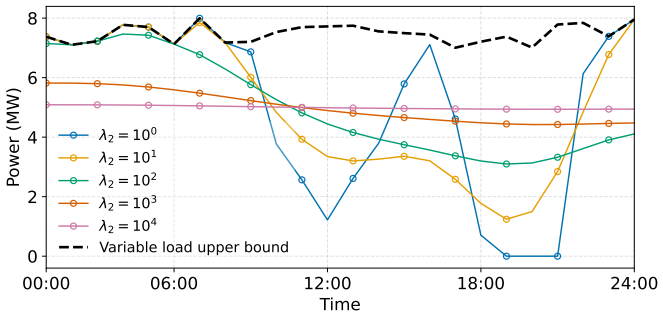

Fig. 3. The fluctuation of variable DC load schedule at Bus 14 ($\lambda_1 = 10^6, \lambda_2 = 10^m, m = 0, 1, 2, 3, 4$).

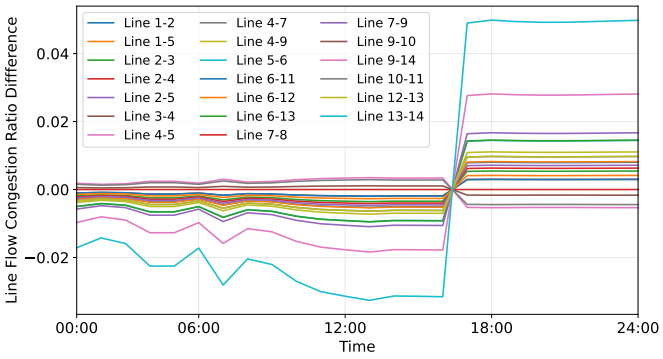

Fig. 4. The fluctuation of variable DC load schedule at Bus 14 ($\lambda_1 = 10^6, \lambda_2 = 10^6$).

deteriorates under Case 1 without optimal flexible scheduling, i.e., between 0 AM and 5 PM, the DCs process variable load in a greedy manner at the maximum available power rate. Compared to Case 1, the congestion penalty term in the objective function of Case 2 effectively discourages excessive loading, leading to improved utilization of transmission capacity and reduced risk of overload. However, after 5 PM, due to the variable load in Case 1 being completed, Case 2 leads to more line congestion.

IV. Conclusions and Future Work

This paper investigates the impact of large DC load integration on transmission system operation through a network-constrained optimization framework. By modeling DC demand as a combination of idle, fixed, and flexible components, the proposed approach enables coordinated scheduling of controllable loads alongside optimal generation dispatch. The results demonstrate that incorporating DC load flexibility can significantly improve transmission system performance. Compared to conventional greedy scheduling, the optimized approach reduces overall operating costs, mitigates transmission congestion, and smooths temporal load variations. These findings highlight the critical role of demand-side flexibility in managing large, concentrated, and uncertain loads introduced by hyperscale DCs. As DC deployment continues to grow, such coordinated optimization strategies will become increasingly important for maintaining grid reliability and economic efficiency. Future work can investigate the risk evaluation of integrating DCs at different geographical locations.

References

[1] Data Centers Standards Needs Analysis and Recommendations Activity, "Review of industry efforts and standards of grid readiness for data center deployment," *Industry Connections Report*, pp. 1–62, 2026.

[2] 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, 2015.

[3] X. Chen, X. Wang, A. Colacelli, M. Lee, and L. Xie, "Electricity demand and grid impacts of AI data centers: Challenges and prospects," *arXiv preprint arXiv:2509.07218*, 2025.

[4] X. Huo, S. Sun, K. A. Haque, L. Al Homoud, A. E. Goulart, and K. R. Davis, "Cooperative optimization of grid-edge cyber and physical resources for resilient power system operation," in *Proceedings of the IEEE Power & Energy Society General Meeting*, Austin, Texas, 27–31 July 2025, pp. 1–5.

[5] P. T. Krein, "Data center challenges and their power electronics," *CPSS Transactions on Power Electronics and Applications*, vol. 2, no. 1, pp. 39–46, 2017.

[6] K. A. Haque, S. Sun, X. Huo, A. E. Goulart, and K. R. Davis, "Scalable discrete event simulation tool for large-scale cyber-physical energy systems: Advancing system efficiency and scalability," *IEEE Access*, pp. 101 900 – 101 921, 2025.

[7] J. Nagy, Ş. Albert, F. Farcaș, C.-A. Lupșe *et al.*, "Experimental control of thermal processes inside the datacenter," in *Proceedings of the IEEE International Conference on Automation, Quality and Testing, Robotics*, Cluj-Napoca, Romania, May 19-21 2022, pp. 1–5.

[8] H. Wan, L. Fang, and X. Li, "Grid operational benefit analysis of data center spatial flexibility: Congestion relief, renewable energy curtailment reduction, and cost saving," *arXiv preprint arXiv:2511.08759*, 2025.

[9] P. Geng, L. Wang, F. C. Alonso, M. Pei, C. F. Wang, J. He, J. Chuang, R. Liu, R. V. Miele, S. K. Saha *et al.*, "Dynamic testing and simulation of chassis attached remote modular heat sink," in *Proceedings of the 23rd IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, Denver, CO, USA, May 28-31 2024, pp. 1–9.

[10] K. Kant, "Data center evolution: A tutorial on state of the art, issues, and challenges," *Computer Networks*, vol. 53, no. 17, pp. 2939–2965, 2009.

[11] J. Shuja, K. Bilal, S. A. Madani, M. Othman, R. Ranjan, P. Balaji, and S. U. Khan, "Survey of techniques and architectures for designing energy-efficient data centers," *IEEE Systems Journal*, vol. 10, no. 2, pp. 507–519, 2016.

[12] R. Rojas-Cessa, Y. Kaymak, and Z. Dong, "Schemes for fast transmission of flows in data center networks," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1391–1422, 2015.

[13] A. Vafamehr, M. E. Khodayar, S. D. Manshadi, I. Ahmad, and J. Lin, "A framework for expansion planning of data centers in electricity and data networks under uncertainty," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 305–316, 2019.

[14] S. Song, J. Kim, S. W. Kim, J. Kim, and M. Yoon, "Oscillation assessment of the kpg-193 test system with hyperscale data center interconnections," *International Journal of Electrical Power Energy Systems*, vol. 177, p. 111811, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014206152600253X

[15] B. Ismalej, M. Smith, and X. Jiang, "Machine learning-based GPU energy prediction for workload management in datacenters," in *Proceedings of the IEEE 15th Annual Computing and Communication Workshop and Conference*, Las Vegas, NV, USA, Jan. 6-8 2025, pp. 01 050–01 054.

[16] V. Chetty and I. E. Davidson, "Challenges of planning future high voltage power systems networks," in *Proceedings of the International SAUPEC/RobMech/PRASA Conference*, Cape Town, South Africa, Jan. 29-31 2020, pp. 1–6.

[17] K. Sheth, D. Patel, and S. K. Sajeev, "Grid transmission evaluation for solar deployment and data center growth," *arXiv preprint arXiv:2509.01778*, 2025.

[18] B. A. Ross, X. Lyu, U. C. Nwaneto, S. M. Mohiuddin, and A. B. Nassif, "Using grid-forming energy storage systems to provide dynamic active power support for hyperscale data center," *IEEE Access*, vol. 4, pp. 1–8, 2026.