SMALL-SCALE DATA CENTERS IN THE DAY-AHEAD ELECTRICITY MARKET

Enea Figini[1] and Mario Paolone[1]

[1]Distributed Electrical Systems Laboratory, École polytechnique fédérale de Lausanne (EPFL)

1. Context

- Total data center capacity is projected to grow from approximately $100\,\mathrm{GW}$ today to over $225\,\mathrm{GW}$ by 2035 [4].
- Transmission System Operators (TSOs) in Ireland, the Netherlands and Germany **refuse connection requests** of new large-scale (e.g., $> 100\,\mathrm{MW}$ data centers to the grid [2]. Some TSOs are issuing **data center-specific grid codes** to regulate connections [5].
- 90% of AI workload (WL) is expected to become inference-based by 2030 [3]. Inference-based WL requires low-latency processing close to end-users, making small-scale edge data centers (e.g., $< 20\,\mathrm{MW}$) the preferred solution.
- Small-scale data centers often lack market access and demand side response incentives (fixed pricing structures).

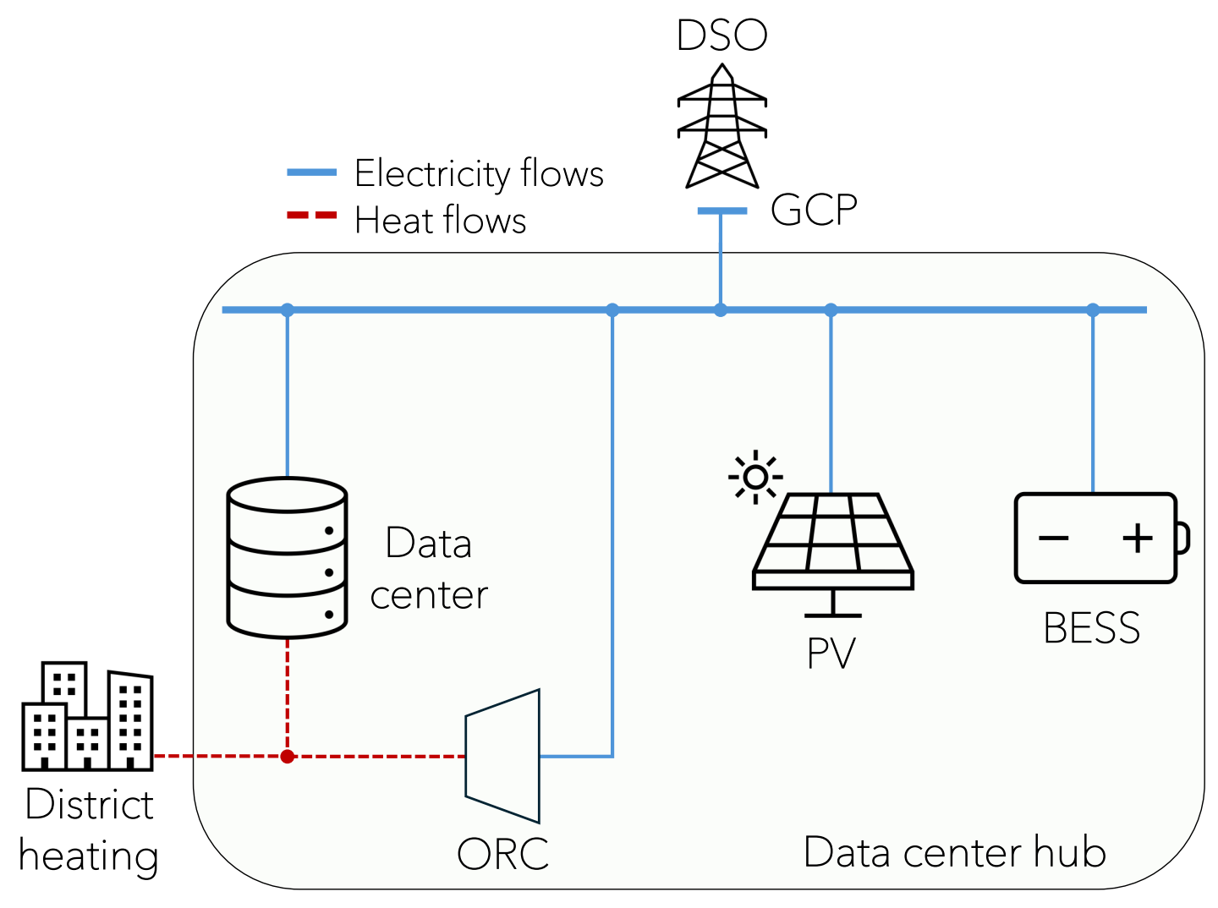

Figure 1. Schematic view of data center microgrid.

2. Bilateral agreement with the DSO

- **Data Center Operator (DCO) market access**
  - The data center is granted access to the day-ahead electricity market.
  - The data center operates as a price-taker.
  - The Distribution System Operator (DSO) acts as an interface with the market platform.
  - The DCO is financially accountable for deviations from the day-ahead schedule.
- **Dynamic rating of the Grid Connection Point (GCP)**
  - The DSO retains the right to apply a day-ahead virtual de-rating of the data center GCP.
  - The de-rating signal is communicated day-ahead and must be incorporated into the DCO's bidding process.
  - A minimum import capacity is guaranteed (2).
  - Cumulative daily and weekly de-rating are limited (3 and 4).

$$\Delta P^{t}_{\text{gcp}} = P^{\text{rated}}_{\text{gcp}} - P^{t}_{\text{gcp,cap}} \tag{1}$$

$$P^{t}_{\text{gcp,cap}} \geq P_{\text{gcp,min}},\ \forall t \tag{2}$$

$$\sum_{t\in\mathcal{D}} \Delta P^{t}_{\text{gcp}} \Delta T \leq (P^{\text{rated}}_{\text{gcp}} - P_{\text{gcp,min}}) t_{\text{daily,lim}} \tag{3}$$

$$\sum_{t\in\mathcal{W}} \Delta P^{t}_{\text{gcp}} \Delta T \leq (P^{\text{rated}}_{\text{gcp}} - P_{\text{gcp,min}}) t_{\text{weekly,lim}} \tag{4}$$

3. Stochastic optimization MILP bidding strategy

The data center **participates in the day-ahead electricity market** and leverages its **flexibility** to minimize expected costs. Flexibility sources include Distributed Energy Resources (DERs), such as PhotoVoltaic (PV), Battery Energy Storage System (BESS), and Organic Rankine Cycle (ORC), as well as the data center WL and waste heat utilization.
When part of the WL is flexible (i.e., its execution can be shifted in time), **Virtual Capacity Curves (VCCs) are used to virtually constrain the compute capacity** of the data center and steer WL scheduling [6].

- **Objective**

$$\min \quad (1-\beta)\,\mathbb{E}(\boldsymbol{\Pi}^{\omega}_{\text{op}}) + \beta\,\text{CVaR}_{\alpha}(\boldsymbol{\Pi}^{\omega}_{\text{op}}) \tag{5}$$

$$\begin{aligned}\boldsymbol{\Pi}^{\omega,t}_{\text{op}} =& \boldsymbol{\Pi}^{\omega,t}_{\text{d-a}} + \boldsymbol{\Pi}^{\omega,t}_{\text{imb}} + \boldsymbol{\Pi}^{\omega,\,t}_{\text{op, DERs}} - \boldsymbol{\Pi}^{\omega,t}_{\text{heat}} \\ &+ \pi_{\text{carbon}}\left(\boldsymbol{C}^{\omega,t}_{\text{gcp}} + \boldsymbol{C}^{\omega,\,t}_{\text{op, DERs}}\right)\end{aligned} \tag{6}$$

- **Constraints: data center model**
  - *Rigid WL*: offers no flexibility. Time-series forecasts of resource usage due to such workload must be satisfied according to Quality of Service (QoS) agreement. Chance Constraints (CCs) are formulated with violation indicator variables to ensure that the share of unserved rigid WL remains below the allowed quota.
  - *Flexible WL*: can be scheduled within $24\,\mathrm{h}$. Forecasts of cumulated flexible demand must be satisfied according to QoS agreement.
  - *Power consumption*: fit the coefficients of (7) using resource usage and power consumption of IT equipment measurements [1].

$$\boldsymbol{P}^{\omega,t}_{\text{dc}} = \eta_{\text{PUE}} \sum_{\text{c}\in\mathcal{C}} \Big(\rho^{\text{c}}_{\text{inter}} + \sum_{\text{res}\in\mathcal{R}\cup\mathcal{R}_{\text{mem}}} \rho^{\text{c, res}}_{\text{coeff}} \boldsymbol{u}^{\omega,\,t,\,\text{c, res}}\Big) \tag{7}$$

  - *Recovered heat*: recovered heat is fitted to power consumption of compute units (Direct Liquid Cooling (DLC)).
- **Constraints: energy resources models**
  - *ORC*: non-linear relationship between input heat and output electrical power modeled through Special Ordered Set of type 2 (SOS2) piecewise constraints.
  - *BESS*: bucket model and enforce power and energy operational constraints. Binary activation constraints used to recover charge and discharge power to model asymmetric system efficiency. Cycling-based aging (8) is tracked and converted to operational costs and carbon emissions using the asset's cost and life cycle assessment, respectively.
  - *PV generation*: modeled as a linear function of local irradiance conditions. It can be curtailed.

$$\boldsymbol{a}^{\omega,\,t}_{\text{bess}} = \frac{|\boldsymbol{P}^{\omega,\,t}_{\text{bess}}|\Delta T}{2L^{\text{cycles}}_{\text{bess}} E^{\text{rated}}_{\text{bess}}} \tag{8}$$

- **Constraints: markets model**
  - *Electricity market*: forecast spot prices (9) and model imbalance costs by splitting GCP imbalances into a long and a short component (10). Short and long imbalance price are assumed to be proportional to the spot price and are calibrated such that this approximation results in an underestimation of imbalance costs in 40% of historical prices. Forecasts of dynamic grid carbon intensity are generated and carbon emissions due to grid imports are tracked (11).

$$\boldsymbol{\Pi}^{\omega,t}_{\text{d-a}} = \pi^{\omega,t}_{\text{spot}} \boldsymbol{P}^{t}_{\text{gcp, d-a}} \Delta T, \tag{9}$$

$$\boldsymbol{\Pi}^{\omega,t}_{\text{imb}} = \Delta T(\pi^{\omega,t}_{\text{short}} \boldsymbol{P}^{\omega,\,t}_{\text{gcp,+}} - \pi^{\omega,t}_{\text{long}} \boldsymbol{P}^{\omega,\,t}_{\text{gcp,-}}) \tag{10}$$

$$\boldsymbol{C}^{\omega,\,t}_{\text{gcp}} = c^{\omega,t}_{\text{gcp}} \boldsymbol{P}^{\omega,\,t}_{\text{gcp}} \Delta T \tag{11}$$

  - *Heat market*: the recovered heat can be sold to the local utility provider at a fixed tariff.

4. Results

The case of a $200\,\mathrm{kW}$ rated academic data center of EPFL ($200\,\mathrm{kW}$ PV, $250\,\mathrm{kW/kW\,h}$ BESS, $100\,\mathrm{kW}$ ORC) is studied. The bidding strategy is used in all study cases, even those with fixed price Power Purchase Agreement (PPA).

- **Financial advantages of market access**
  - 22% reduction of ex-post expected daily costs
  - 6% increase in carbon footprint
  - Long tails due to imbalance costs
- **Low value of flexible WL on day-ahead market**
  - Expected value on day-ahead market is $<$ 2.5% of the value of the shifted workload
  - Little incentive for users to participate in flexible programs
- **Constrained operation under virtual de-rating**
  - ORC activates under constrained operation
  - WL flex. reduces ORC dispatch
  - Constrained operation increases costs by about 20%. The PPA must be carefully designed (3, 4)

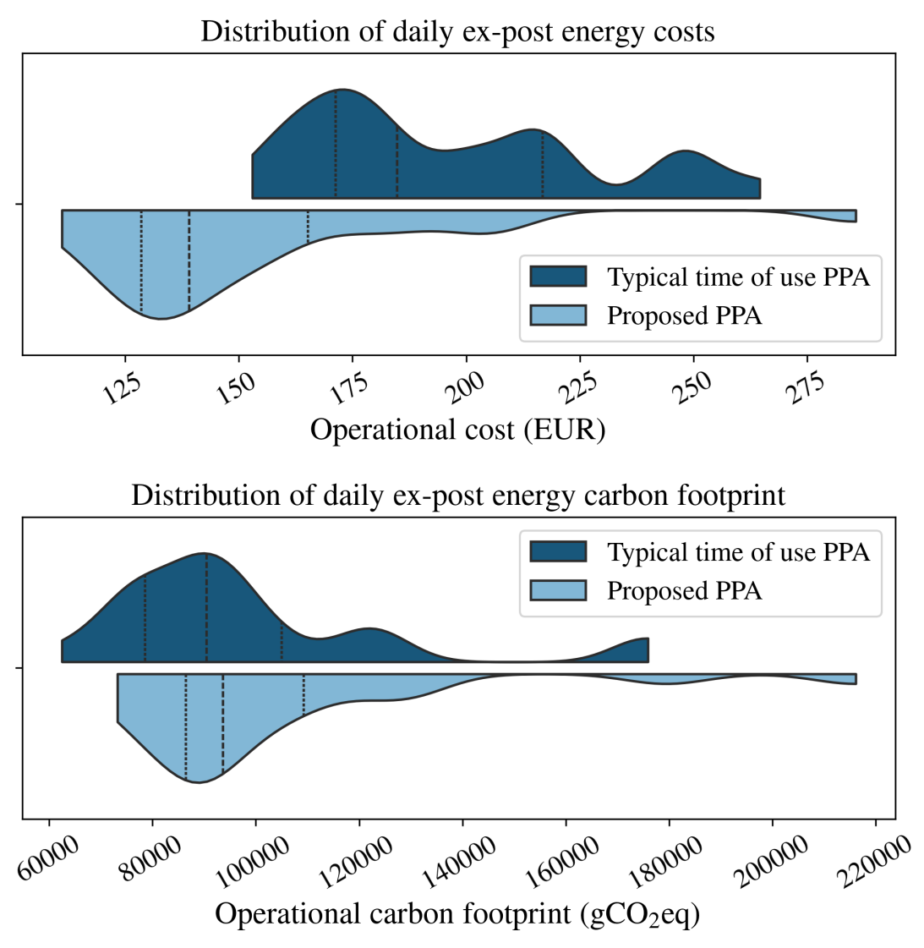

Figure 2. Ex-post operational costs under ToU and custom PPA.

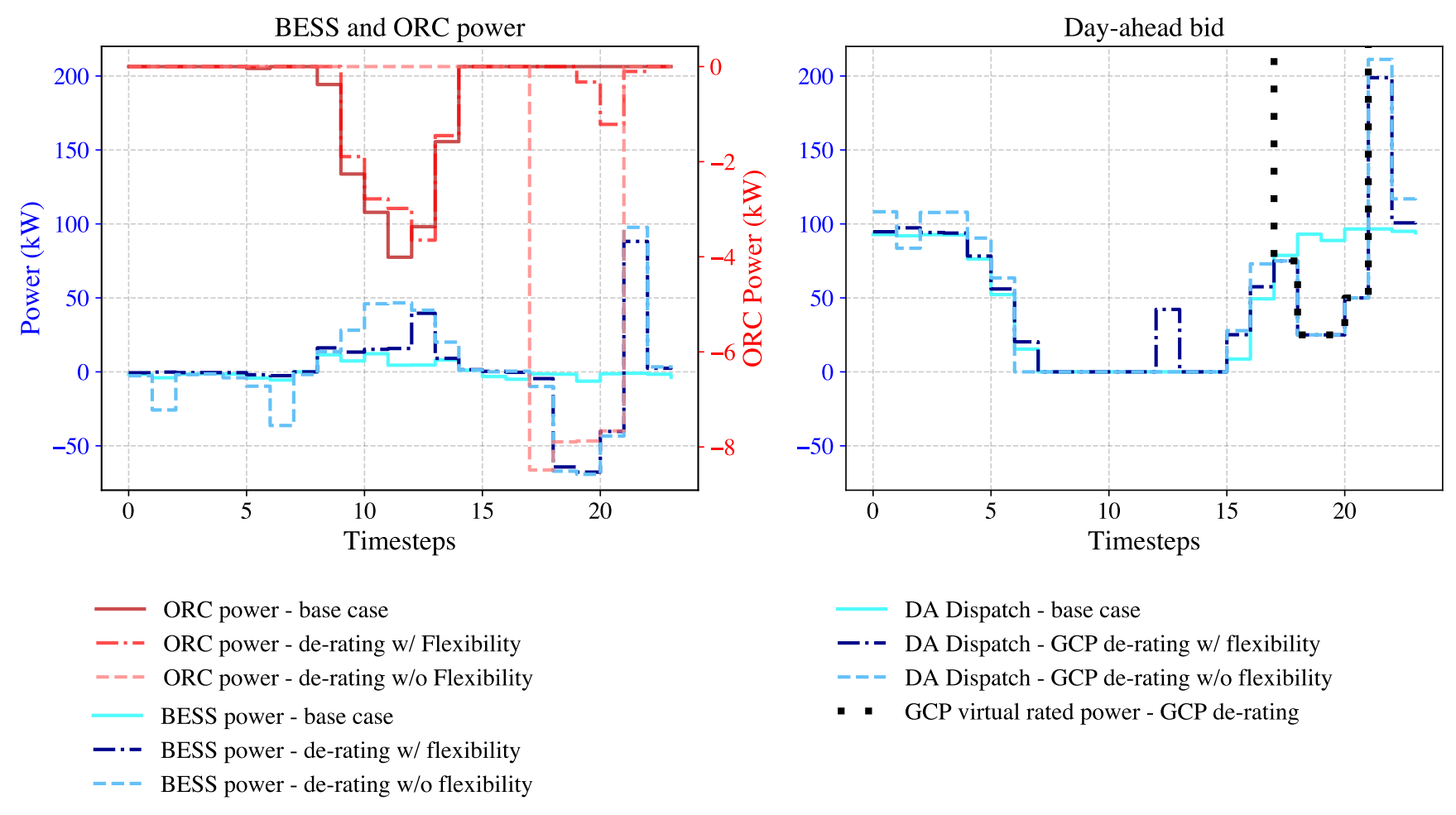

Figure 3. Dispatch of resources and bids under virtual de-rating.

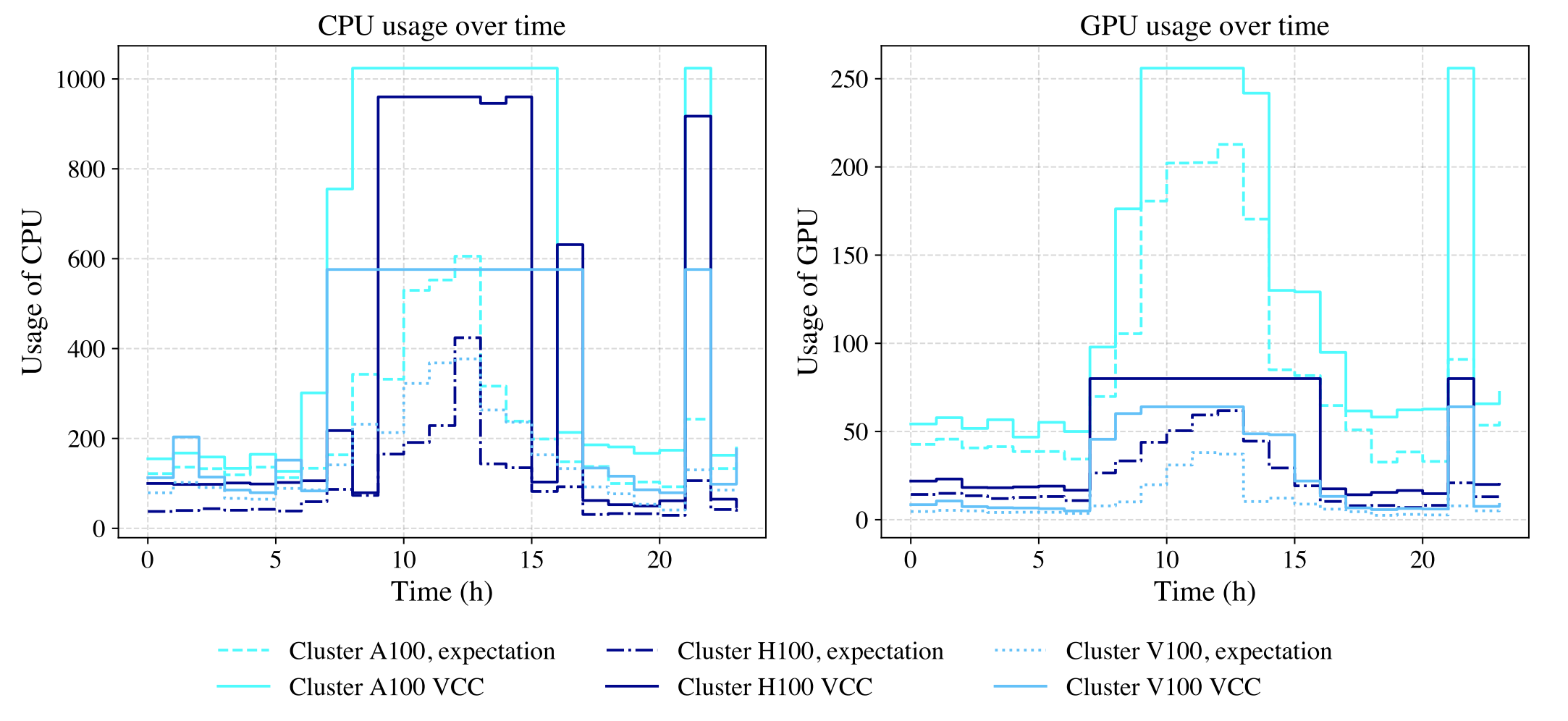

Figure 4. Compute usage and VCCs under virtual de-rating operation.

5. Highlights

- Small-scale data centers **gain access to the spot electricity market** through a custom PPA, the DSO gets some **demand side flexibility** via simple virtual de-rating mechanism.
- The risk-averse carbon-aware day-ahead bidding strategy combined with the PPA is expected to **reduce energy supply costs by 22%**.
- The value of WL flexibility in the day-ahead market is marginal and **unlikely to motivate data centers to propose billing schemes** for flexible WL (from the energy supply perspective only).
- Virtual GCP transfer capacity de-rating requests by the DSO **should not significantly impact operational costs** if the supply agreement is well-designed, showing benefits for both actors.
- **Future work** will extend this approach toward real-time control strategies for data centers, building on the bidding framework proposed here.

References

[1] Miyuru Dayarathna, Yonggang Wen, and Rui Fan. "Data Center Energy Consumption Modeling: A Survey". In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 732–794. ISSN: 1553-877X. DOI: `10.1109/COMST.2015.2481183`. URL: `http://ieeexplore.ieee.org/document/7279063/` (visited on 04/09/2025).

[2] Ember. *Europe's AI ambitions at risk of gridlock*. URL: `https://ember-energy.org/chapter/chapter-1-grids-for-data-centres/` (visited on 10/13/2025).

[3] Aditie Garg et al. "Considerations for Distributed Edge Data Centers and Use of Building Loads to Support Large Interconnections". en. In: *Renewable Energy* (2025).

[4] IEA. *Grid congestion is posing challenges for energy security and transitions*. 2025. URL: `https://www.iea.org/commentaries/grid-congestion-is-posing-challenges-for-energy-security-and-transitions`.

[5] IEEE SA. "Review of Industry Efforts and Standards of Grid Readiness For Data Center Deployment". en. In: (Jan. 2026).

[6] Ana Radovanović et al. "Carbon-Aware Computing for Datacenters". en. In: *IEEE Transactions on Power Systems* 38.2 (Mar. 2023), pp. 1270–1280. ISSN: 0885-8950, 1558-0679. DOI: `10.1109/TPWRS.2022.3173250`. URL: `https://ieeexplore.ieee.org/document/9770383/` (visited on 07/04/2025).

READ OUR PAPER