Long-Term Joint Decarbonization Planning of Power Systems and Data Centers: A Case Study in PJM

Zhentong Shao[a], Nanpeng Yu[a,*] and Daniel Wong[a]

[a]Department of Electrical and Computer Engineering, University of California, Riverside, Riverside, California, USA

ARTICLE INFO

Keywords:
Data center
Joint planning
Embodied carbon emissions
Stochastic optimization
Benders decomposition

ABSTRACT

With the accelerating growth of artificial intelligence (AI) and cloud-based services, data centers have become critical infrastructures driving digital economies. Their surging energy demand has elevated concerns over electricity consumption and carbon emissions, underscoring the need for integrated, carbon-aware infrastructure planning. However, most existing studies adopt static power system assumptions, focus solely on operational emissions, and neglect the potential for co-optimization. This paper proposes a dynamic joint planning framework that co-optimizes the long-term development of data centers and power systems over a 15-year horizon. The model determines investment decisions for data center siting, capacity, and type, as well as power generation expansion, storage deployment, and generator retirements, while incorporating both operational and embodied carbon emissions. To address multi-scale uncertainty, a large-scale two-stage stochastic programming formulation is developed and efficiently solved via an enhanced Benders decomposition algorithm. The framework is applied to the PJM Interconnection, where all system parameters and processed datasets are curated and made publicly available on GitHub to support future research. Case study results show that the PJM system can accommodate up to 55 GW of peak data center demand, with Virginia (DOM) and Northern Illinois (ComEd) identified as optimal hosting regions. Compared to non-joint planning approaches, the proposed framework reduces total investment cost by 12.6%, operational cost by 8.25%, and carbon emissions by 5.63%. Relative to the case without embodied carbon considerations, incorporating lifecycle emissions increases renewable deployment by 25.5%, underscoring the critical role of embodied carbon in driving deeper decarbonization.

1. Introduction

1.1. Motivation

Data centers have emerged as critical infrastructures driving digital economies and enabling rapid advancements in technologies such as artificial intelligence (AI) and the Internet of Things [1]. As the energy consumption of these facilities surges, so does the urgency to mitigate their associated carbon emissions [2]. Traditional energy management efforts have primarily focused on operational improvements, such as enhancing power usage effectiveness (PUE) [3], but these approaches fall short of meeting the ambitious decarbonization goals set by policymakers and environmental agencies [4].

Moreover, data centers do not operate in isolation, they interact with power systems that are simultaneously challenged by renewable intermittency, shifting load profiles, and aging infrastructure [5]. However, a majority of existing planning studies treat the power system as static and optimize data center deployment independently, neglecting the opportunities for integrated planning. As shown in Table 1, prior works typically adopt one-shot, short-term planning framework and focus solely on operational emissions, thereby overlooking embodied emissions and long-term co-optimization.

To address these limitations, this paper proposes a dynamic joint planning framework that co-optimizes data center and power system investments over a 15-year horizon. Unlike previous studies, our approach is implemented on the real-world PJM system and integrates data center siting, capacity expansion, and workload dispatch with long-term energy resource planning of power system. Both operational and embodied carbon emissions [6] are included to enable a comprehensive lifecycle assessment. By capturing spatial and temporal dynamics, this framework provides a more realistic and carbon-aware pathway for sustainable infrastructure development.

*Corresponding Author.
nyu@ece.ucr.edu (N. Yu)

1.2. Literature Review

Most existing studies have examined expansion planning and energy management issues primarily from the perspective of a single data center [7–12]. Consequently, they have not addressed the coordination between data centers and power systems. By contrast, our paper adopts the perspective of regional government agencies, such as public utilities commissions or Independent System Operators (ISOs). The primary goal of this paper is to investigate how the joint planning of data centers and power systems can enhance the long-term societal benefits through improved energy utilization and carbon reduction.

Compared to single-site data center planning, joint planning for power systems and data centers is more complex. In particular, multi-site data center planning must incorporate system-wide power balance and power flow representations [13], while also considering issues such as renewable energy supply uncertainties and transmission congestion [14]. As a result, the joint planning problem takes the form of a non-linear combinatorial optimization problem, which poses significant computational challenges for practical transmission networks [15]. In the case of single-site data centers, power network complexities can often be neglected, leading to a much simpler planning formulation [9].

Another critical issue for the joint planning problem is the management of geographical impacts on both data centers and generation resources, which affects both energy efficiency and corresponding carbon emissions for the integrated energy system [16]. This is because, for data centers, different regions exhibit varying ambient temperatures, which in turn affect the energy efficiency of cooling systems [17], whose energy consumption constitutes a significant portion of the data center's total energy use. Furthermore, the embodied carbon emissions related to constructing data centers and generation resources differ by location [18]. The assessment of these embodied carbon emissions is important for fully understanding the sustainability of planned infrastructure. In addition, variations in resource endowment across regions also influence the feasibility of renewable energy installations and, consequently, the degree of carbon reduction achievable by data centers. Although references [16, 19–21] have considered power system factors in data center planning, several important aspects remain underexplored. For instance, references [19–21] do not account for key geographical energy efficiency and carbon emission factors in data center planning, nor do they consider power system expansion planning. While reference [16] incorporates comprehensive geographical factors, the power system is simplified into preset locational marginal prices (LMP) and locational marginal carbon emissions (LMCE). Additionally, most of the above studies account only for operational carbon emissions, thus overlooking the embodied carbon emissions essential for long-term planning.

The modeling granularity of data centers is a key factor that significantly influences computational efficiency and model accuracy in joint planning. Appropriate consideration of data center components, such as the number of server racks, cooling system technology (water, air, or liquid cooling), and workload transfer, is important for ensuring realistic representation of energy consumption in operations. However, detailed data center models, such as those that explicitly determine CPU utilization and control Quality of Service (QoS) [12, 19, 21, 22], can be prohibitively complex for long-term planning. Conversely, over-simplified data center models can lead to inaccuracies. For example, [23] treats data centers as a simple load curves without considering cooling system choices, while [20, 24] approximate data centers as shiftable loads, reducing data centers to demand response resources. In contrast, rack-level modular modeling of data centers, as studied in [16] and [25], is more appropriate. As it effectively accounts for major components of data center planning without distortion and facilitates long-term planning by considering carbon emissions and energy efficiency.

In addition to the aforementioned challenges, the multi-source uncertainties arising from the joint planning of power systems and data centers have not yet been fully addressed. Many existing studies, such as [19–25], have focused on short-term uncertainties (e.g., intermittent renewable generation and variable demands). For instance, reference [20] uses a scenario-based stochastic optimization (SO) method to address renewable energy and load uncertainties in the expansion planning of data centers and data transmission facilities, whereas reference [21] considers the planning of data centers by incorporating the demand-side flexibility of data centers under electricity-carbon markets. A scenario-based SO method is adopted to capture the uncertainties in the operational stage of the system. Reference [22] proposes an optimal planning method for regional integrated electricity–heat systems with data centers, where wind power uncertainty is considered and captured using a three-stage distributionally robust optimization (DRO) planning model. Nonetheless, the uncertainties of long-term demand and computing workload growth, which are crucial for joint planning results, are often neglected in the existing literature [16, 24]. Moreover, practical limitations in handling numerous stochastic scenarios often lead to reduced scenario sets, which can compromise solution accuracy and system reliability. These gaps in the literature underscore the necessity of advanced modeling and computational acceleration techniques to tackle the joint planning of power systems and data centers more effectively.

Table 1
Comparison of Representative Literature on Data Center Planning

Reference	Joint Planning	Retirement Planning	Practical Case Study	Carbon Emissions	Planning Scope	Planning Structure	Planning Horizon
[12]	✗	✗	✗	OP	Servers	One-shot	Short-term (1 yr)
[16]	✗	✗	✗	OP	Data Center	One-shot	Mid-term (5 yrs)
[23]	✗	✗	✗	OP	Microgrid	One-shot	Short-term (1 yr)
[19]	✗	✗	✗	✗	Data Center	One-shot	Short-term (1 yr)
[25]	✗	✗	✗	✗	Microgrid	One-shot	Mid-term (4 yrs)
[20]	✗	✗	✗	✗	Data Center	Dynamic	Long-term (20 yrs)
[21]	✗	✗	✗	OP	Data Center	One-shot	Short-term (1 yr)
[22]	✗	✗	✗	✗	Microgrid	One-shot	Short-term (1 yr)
[24]	✗	✗	✗	OP	Building System	One-shot	Short-term (1 yr)
This paper	✓	✓	✓	OP & EM	Power System & DC	Dynamic	Long-term (15 yrs)

Notes:
⋄ *Joint planning makes investment decisions for both the power system and data centers, enabling coordinated development. In contrast, existing studies assume a fixed power system and plan data centers accordingly.*
⋄⋄ *OP refers to operational carbon emissions, OP & EM refers to both operational and embodied carbon emissions.*
⋄ ⋄ ⋄ *Dynamic planning involves annual decisions over time, while One-shot determines all decisions at once.*

Table 1 compares representative studies on data center planning, highlighting critical gaps that this work addresses. First, most existing studies assume a fixed power system and optimize data center deployment accordingly, thereby overlooking the benefits of joint planning. As shown in the table, none of the prior works co-optimize investments in both power systems and data centers, missing the opportunity for coordinated infrastructure development under growing energy demands. Second, while operational (OP) carbon emissions are occasionally considered, embodied emissions (EM), which account for lifecycle carbon impacts and are essential for long-term dynamic planning, are rarely incorporated, limiting the comprehensiveness of the environmental assessment. Third, most studies adopt one-shot planning approaches that fail to capture the evolving nature of infrastructure development. Although a few works, such as [20], implement dynamic planning, they still lack integration with both embodied carbon considerations and joint infrastructure expansion. In contrast, our work proposes a dynamic joint planning framework over a 15-year horizon for the practical PJM system, co-optimizing investments in both power systems and data centers while considering operational and embodied carbon emissions. This comprehensive approach bridges key methodological gaps in the literature and provides a more realistic and carbon-aware pathway for long-term infrastructure planning.

1.3. Contributions

This paper addresses key methodological gaps in the existing literature by proposing a dynamic joint planning framework that integrates the long-term development of power systems and data centers. The main contributions are summarized as follows:

1. **Dynamic joint planning of coupled infrastructures:** Unlike prior studies that optimize data center deployment based on a static power system, this work co-optimizes the investment trajectories of both power generation assets and data centers over a 15-year horizon. The proposed model captures the spatial and temporal interdependencies between power supply and digital demand, enabling coordinated infrastructure development.

2. **Lifecycle carbon integration:** This work is among the first to incorporate both operational and embodied carbon emissions into joint infrastructure planning. By accounting for emissions from facility construction, equipment manufacturing, and long-term operations, the framework supports a more comprehensive and carbon-conscious planning pathway. The consideration of embodied carbon effectively promotes the development of renewable energy and thereby enables deeper decarbonization.

3. **Stochastic optimization under multi-scale uncertainty:** A large-scale stochastic programming model is developed to simultaneously address long-term demand growth uncertainty and short-term renewable intermittency. To enhance tractability, a customized Benders decomposition algorithm is implemented to achieve interactive

iteration between the investment stage and the operation stage, and finally achieve an effective joint planning solution.

4. **Application to a real-world regional system:** The proposed framework is applied to the PJM interconnection, one of the largest power systems in North America. To support this application, we systematically compiled and harmonized fragmented public data sources to construct a consistent and comprehensive dataset for PJM system planning. All processed data and system parameters have been made publicly available via GitHub [26]. This open dataset is expected to serve as a valuable resource for future research on integrated infrastructure planning.

In addition to methodological contributions, the case study conducted on the PJM system yields several empirical insights that demonstrate the practical value of the proposed framework: (a) The results show that the PJM system can accommodate up to 55 GW of peak data center demand, with the DOM (Virginia) and ComEd (Northern Illinois) zones identified as the most suitable hosting regions. (b) Joint planning of power systems and data centers leads to notable system-wide benefits. Compared to independent data center planning, it reduces total investment costs by 12.63%, operational costs by 8.25%, and total carbon emissions by 5.63%. (c) Compared to the case where embodied carbon emissions are not considered, incorporating embodied carbon into the planning framework leads to a 25.5% increase in renewable capacity deployment and a 16.9% reduction in operational carbon emissions.

1.4. Paper Organization

The remainder of this paper is organized as below. Section II provides the planning framework and the mathematical formulation of the joint planning model. Section III derives the solution strategy and illustrates the solution approach of the strengthened Benders decomposition. Section IV shows the results of numerical studies on the PJM interconnection. Finally, Section V concludes the paper.

2. Mathematical Formulation

The joint data center and power system planning model is constructed from the perspective of a regional government agency, such as public utility commission, whose objective is to satisfy computing and electric load in a way that reduces both costs and greenhouse gas emissions. The optimal planning outcome can guide both data center and energy resource developers to strategically locate data centers and power plants in a manner that maximizes system-wide benefits.

2.1. Two-Stage Joint Planning Framework

This paper proposes a two-stage joint planning framework to determine the optimal siting, sizing, and cooling system selection of datacenter and power system resources to achieve the decarbonization and cost minimization goal. The proposed framework models a two-stage decision-making process. In the first stage, long-term investment decisions are made to determine the placement, capacity, and type of resources. The second stage focuses on verifying the operational feasibility of the system by analyzing short-term daily operations. The overall structure of the joint planning framework is shown in Fig. 1.

The multi-timescale stochastic planning timeline is illustrated in Fig. 2, indicating that the proposed model incorporates a large-scale framework by considering multiple years and expanding to a large number of scenarios. This presents a significant computational challenge, which will be addressed in the next section by introducing a customized Benders decomposition method.

The key planning assumptions for data centers and power systems are explained below.

a) *Datacenter expansion planning:* For data centers, the planning stage involves determining the location, the size of the data center (reflected by the number of server racks), and the cooling equipment options (represented by the power utilization efficiency under different cooling technologies). These decisions are made on an annual timescale, aiming to account for embodied carbon emissions from construction while minimizing investment costs. In the operation stage, the objective is to minimize O&M costs and operational carbon emissions by considering location-specific differences and distributing the total power load across data centers in various locations. This is achieved by dispatching active servers and facilitating power load response. This process accounts for locational variations in cooling systems, energy prices, and carbon emissions.

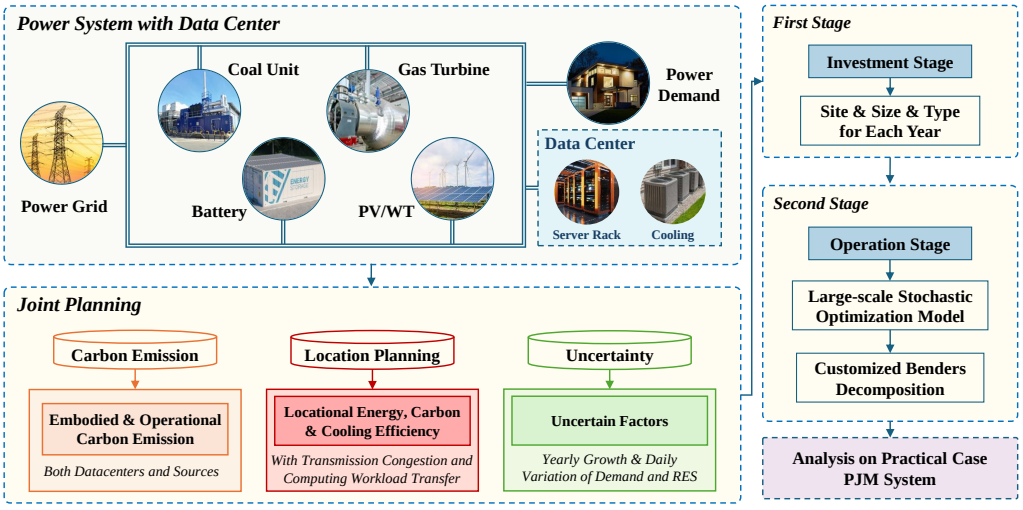

Figure 1: The structure of joint planning of power systems with data centers.

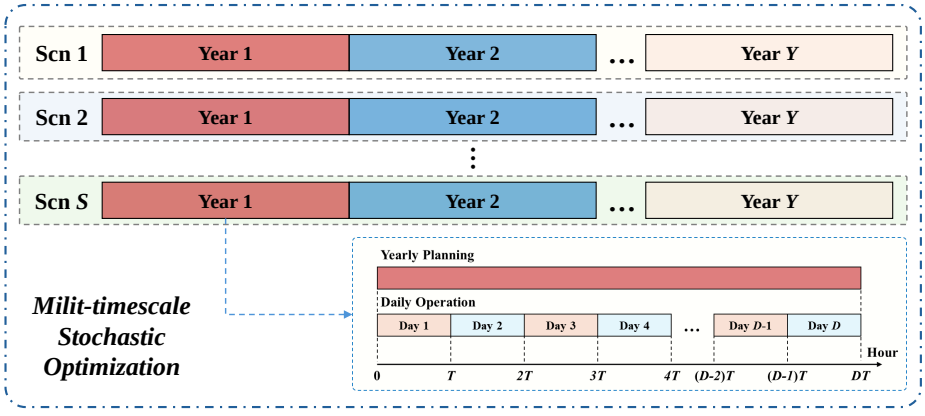

Figure 2: The illustration of multi-timescale stochastic planning optimization.

b) *Power system expansion planning:* For the power system modeling, the planning stage focuses on developing RES and natural gas generation (NG) units on an annual basis and determining the annual retirement of CG units to achieve carbon reduction targets. These planning decisions for power generation resources encompass three key aspects: site selection, capacity determination, and construction scheduling. In the operation stage, the power system is modeled using a multi-period optimal power flow with transmission constraints. This approach captures the locational differences in energy distribution while accounting for transmission congestion. The model explicitly incorporates the annual growth of system electric load and data center computing load. Additionally, it evaluates the reliability of system operations under short-term load and renewable energy uncertainties, as well as long-term uncertainties for demand growth.

Nomenclature

	Sets
d	Index of days, $d \in \mathcal{N}_d$.
i, j	Index of power system nodes/locations, $i, j \in \mathcal{N}_i$.

ij	Index of transmission lines from node i to node j, $ij \in \mathcal{N}_l$.
k	Index of candidate generation resources $k \in \mathcal{N}_k$.
l	Index of transmission lines, $l \in \mathcal{N}_l$.
m	Index of server types, $m \in \mathcal{N}_m$.
t	Index of hours, $t \in \mathcal{N}_t$.
y	Index of years, $y \in \mathcal{N}_y$.
	Decision Variables
C^y_{curt}	Total curtailment penalty cost for PV and WT of year y (\$).
C^y_{DC}	Investment cost of data centers of year y (\$).
C^y_{ex}	Total energy exchange cost of the power system of year y (\$).
C^y_{gen}	Total generation cost of all resources of year y (\$).
C^y_{GR}	Investment cost of all generation resources of year y (\$).
C^y_{om}	Total maintenance cost of all resources of year y (\$).
C^y_{TC}	Investment cost for data transmission facilities of year y (\$).
C^y_{TL}	Investment cost of transmission lines of year y (\$).
C^y_{RT}	Salvage value of retired equipments of year y (\$).
$E^{y,d,t,i}_{ES}$	State of charge of energy storage devices on location i in year y, day d and hour t (MWh).
$N^{y,i}_k$	Number of installed units of resources $k \in \{NG,PV,WT,ES,TL,CG\}$ in year y at location i.
$N^{y,i}_k$	Number of retired NG units in year y at location i.
$P^{y,d,t,i}_k$	Power output of resources $k \in \{NG,PV,WT,ESC,ESD,CG\}$ at location i in year y, day d and hour t (MW).
$P^{y,d,t,i}_{BUY}$	Purchased energy of PJM from other ISOs at location i in year y, day d and hour t (MW).
$P^{y,d,t,i}_{DC}$	Power demand of data center at location i in year y, day d and hour t (MW).
$P^{y,d,t,i}_{k,cap}$	Available power output capability for resource $k \in \{PV,WT\}$ of location i in year y, day d and hour t (MW).
$P^{y,d,t,i}_{SELL}$	Sold power of PJM to other ISOs from location i in year y, day d, and hour t (MW).
$P^{y,d,t,l}_{TL}$	Power flow on transmission line l in year y, day d and hour t (MW).
$X^{y,d,t,i}_{SVR,m}$	Number of active servers in a data center (rack).
$Y^{y,d,t,i}_{DC}$	Computing workload served by data centers at location i in year y, day d, hour t (Requests).
$Y^{y,d,t,ij}_{TC}$	Computing workload transferred from location location i to location j in year y, day d, hour t (Requests).
$\bar{Y}^y_{ij}$	Computing workload transfer capability from location i to location j in year y (Requests).
$\mathbf{x}$	Abstract vector representing decision variables for the operational stage, including power outputs of all generation resources and data centers, as well as data workload decisions for servers.
$\mathbf{z}$	Abstract vector representing decision variables for the planning stage, encompassing all installed capacity decisions, as well as newly added and retired capacity decisions.
$BN^{y,i}_{SRV,m}$	Newly added number of m-type servers in year y at location i (rack).
CE^y_{EMB}	Total embodied carbon emissions of the system in year y (short ton)
CE^y_{OPR}	Total operational carbon emissions of the system in year y (short ton)
$IC^{y,i}_k$	Installed capacity of system resources $k \in \{NG,PV,WT,TL,CG\}$ in year y at location i (MW).
$IC^{y,i}_{ES}$	Installed capacity of energy storage devices in year y on location i (MWh).
$IN^{y,i}_{DC}$	Installed number of all server types in the data center at location i in year y (rack).
$IN^{y,i}_{SRV,m}$	Installed number of m-type server in the data center at location i in year y (rack).
$RN^{y,i}_{SRV,m}$	Number of m-type servers retired in year y at location i (rack).
RV	Total residual value of all resources at the end of the planning horizon (\$).
	Parameters

$\beta_k^{\min}$	Capacity scaling factor representing the minimum output of resource $k \in \{NG,CG,PV,WT,ES,ESC,ESD\}$ (%).
$\beta_k^{\max}$	Capacity scaling factor representing the maximum output of resource $k \in \{NG,CG,PV,WT,ES,ESC,ESD\}$ (%).
χ_k^{emb}	Embedded carbon emission coefficient for resources $k \in \{NG,PV,WT\}$ (short ton/MW).
χ_{ES}^{emb}	Embodied carbon emission coefficient for energy storage devices (short ton/MWh).
χ_{Fix}^{emb}	Embodied carbon emission coefficient for fixed resources in data centers (short ton/rack).
$\chi_{SRV,m}^{emb}$	Embodied carbon emission coefficient for m-type servers in data center (short ton/rack).
χ_{TC}^{emb}	Embodied carbon emission coefficient for data transmission facilities (short ton/km · Requests).
χ_{TL}^{emb}	Embodied carbon emission coefficient for transmission lines (short ton/MW · km).
χ_k^{gen}	Generation carbon emission coefficient for resources $k \in \{NG,CG\}$ (short ton/MWh).
$\delta_{d,t}^{data}$	Daily variation factor for computing workload demand in day d and hour t (%).
$\delta_{d,t}^{load}$	Daily variation factor for electricity demand in day d and hour t (%).
$\delta_{d,t}^{pv/wt}$	Daily variation factor for solar power and wind power outputs in day d and hour t (%).
Δt	Time interval (hour).
η_{esc}/η_{esd}	Charging/Discharging efficiency for energy storage devices (%).
γ_{tech}^{y}	Investment cost discount factor, representing the reduction in investment costs for GPU and other server equipments due to technological advancements (%).
$\Gamma_{l,i}$	Power transfer distribution factor (PTDF) from node i to line l.
$\hbar$	Weighting factor that reflects the decision maker's preference toward carbon emission reduction ($/short ton).
λ_y^{data}	Yearly growth factor for computing workload demand (%).
λ_y^{load}	Yearly growth factor for electricity demand (%).
λ_{tech}^{y}	Embodied carbon emission discount factor for GPU and other server equipments representing technological advancements (%).
$\phi_{y,d,t,i}^{temp}$	Temperature-dependent cooling efficiency scaling factor (%).
$\psi_{SVR,m}$	The equivalent processing capability of the m-type server (Requests/hour · rack).
σ	Discount factor (%).
φ_i^{data}	Locational distribution factor of computing workload load for location i (%).
φ_i^{load}	Locational distribution factor of electric load for location i (%).
$\xi_{i,t}^{buy/sell}$	Energy buying/selling price at node i and hour t ($/MWh).
ξ_k^{curt}	Curtailment cost coefficient for resource $k \in \{PV,WT\}$ ($/MWh).
ξ_k^{gen}	Generation cost coefficient for resource $k \in \{NG,CG\}$ ($/MWh).
ξ_k^{om}	Maintenance cost coefficient for resources $k \in \{NG,PV,WT,ES,CG\}$ ($/MWh).
ξ_{ES}^{inv}	Unit investment cost for energy storage devices, including integrated charging/discharging capacity costs. ($/MWh).
ξ_{Fix}^{inv}	Unit investment cost for fixed resources of a data center, e.g., land and material costs ($/rack).
$\xi_{SRV,m}^{inv}$	Unit investment cost for m-type server ($/rack).
ξ_{TC}^{inv}	Unit investment cost for data transmission facilities ($/km · Requests).
ξ_{TL}^{inv}	Unit investment cost for transmission lines ($/km · MW).
ξ_k^{inv}	Unit investment cost for generation resources $k \in \{NG,PV,WT,CG\}$ ($/MW).
ζ_s	Stochastic scenario.
E_{ES}^{unit}	Unit energy storage capacity (MWh).
H_l	Length of transmission line l (km).
H_{ij}	Physical distance between location i and location j (km).
L_k	Lifetime of system resources $k \in \{NG,PV,WT,ES,CG,Fix,TC,TL\}$ (year).
$L_{SRV,m}$	Lifetime of m-type servers (year).
R_k^{y}	Residual value factor for system resource $k \in \{NG,PV,WT,ES,Fix,TC,TL\}$ installed in year y (%).
R_{CG}^{y}	Residual value factor for coal-fired units retired in year y (%).

R_k^0	Scrap value of system resources $k \in \{NG,PV,WT,ES,CG,Fix,TC,TL\}$ at the end of its life (%).
$R_{SRV,m}^y$	Residual value factor of m-type servers in year y (%).
$R_{SRV,m}^0$	Scrap value of m-type servers at the end of its life (%).
$R_k^{up/down}$	Up/Down ramping rate for resources $k \in \{NG,CG\}$ (%).
$\bar{RT}_{CG}^y$	Retired capacity limitation for CG units in year y (MW).
P_k^{unit}	Unit generation capacity of resources $k \in \{NG,PV,WT,CG,TL\}$ (MW).
$P_{LOAD}^{y,d,t,i}$	Electric load of the system without data centers at location i in year y, day d, hour t (MW).
$P_{EX}^{\max}$	Maximum allowable power exchange (MW).
$P_{SVR,m}^{rate}$	Rated power consumption of m-type server (MW/rack).
P_{LOAD}^{peak}	Peak power demand of the system (MW).
PUE_m	Power usage effectiveness for m-type server (%).
$\bar{IC}_{CG}^y$	Upper limit of capacity retirement of coal-fired generation units of year y (MW).
$\bar{IC}_k^{y,i}$	Maximum installed capacity limitation of resources $k \in \{PV,WT\}$ in year y at location i (MW).
$\underline{IC}_k^{y,i}$	Minimum installed capacity requirement of resources $k \in \{PV,WT\}$ in year y at location i (MW).
$\bar{IC}_{ES}^{y,i}$	Maximum installed capacity limitation of energy storage devices in year y at location i (MWh).
$\underline{IC}_{ES}^{y,i}$	Minimum installed capacity requirement of energy storage devices in year y on location i (MWh).
$Y_{LOAD}^{y,d,t,i}$	Computing workload demand on location i in year y, day d, hour t (Requests).
Y_{LOAD}^{peak}	Peak computing workload of the system (Requests).
Y_{ij}^0	Existing data transmission resources from location i to location j (Requests).
z_{ij}	Binary parameter indicating whether data transmission facilities should be built between location i and location j.

	Abbreviations
CG	Coal-fired generation unit.
DC	Data center.
ES	Energy storage.
ESC	Energy storage charging.
ESD	Energy storage discharging.
ISO	Independent system operator.
NG	Natural-gas generation unit.
PUE	Power usage effectiveness.
PV	Photovoltaic generation unit.
RES	Renewable energy resource.
RPS	Renewable portfolio standards.
SOC	State of charge.
SRV	Servers in data center.
TC	Data transmission facility.
TL	Transmission lines.
WT	Wind turbine generation unit.
Fix	Fixed facilities of data center (e.g., building, land and materials).

2.2. Formulation of the Joint Planning Model

The joint planning model aims at minimizing the total costs (TC) and total carbon emissions (TE) of the whole system for both investment and operational stage.

2.2.1. The economic objective

The economic objective is defined in (1), encompassing the investment cost in the planning stage, represented by $F(\mathbf{z})$, as well as the operational cost in the operational stage represented by $G(\mathbf{x})$.

$$TC = \min_{\mathbf{z}\in\mathbb{Z},\ \mathbf{x}\in\mathbb{X}} \left[\underbrace{F(\mathbf{z})}_{Planning\ Stage} + \underbrace{G(\mathbf{x})}_{Operational\ Stage} \right] \tag{1}$$

Where $\mathbf{z}$ and $\mathbf{x}$ denote the sets of decision variables for the planning and operational stages. The economic objective for planning is calculated as follows.

$$F(\mathbf{z}) = \sum_{y\in\mathcal{N}_y} \left[\frac{C^y_{GR} + C^y_{DC} + C^y_{TL} + C^y_{TC} - C^y_{RT}}{(1+\sigma)^y} \right] - \frac{RV}{(1+\sigma)^{|\mathcal{N}_y|}} \tag{2}$$

$$C^y_{GR} = \sum_{i\in\mathcal{N}_i} \sum_{k\in\{NG,PV,WT,ES\}} \xi^{inv}_k \left(IC^{y,i}_k - IC^{y-1,i}_k \right) \tag{3}$$

$$C^y_{DC} = \sum_{i\in\mathcal{N}_i} \left[\sum_{m\in\mathcal{N}_m} \left(\gamma^y_{tech} \xi^{inv}_{SRV,m} BN^{y,i}_{SRV,m} \right) + \xi^{inv}_{Fix} \left(IN^{y,i}_{DC} - IN^{y-1,i}_{DC} \right) \right] \tag{4}$$

$$C^y_{TL} = \sum_{l\in\mathcal{N}_l} H_l\, \xi^{inv}_{TL} \left(IC^{y,l}_{TL} - IC^{y-1,l}_{TL} \right) \tag{5}$$

$$C^y_{TC} = \sum_{i\in\mathcal{N}_i} \sum_{j\in\mathcal{N}_i, j\neq i} z_{ij} H_{ij} \xi^{inv}_{TC} \left(\bar{Y}^y_{ij} - \bar{Y}^{y-1}_{ij} \right) \tag{6}$$

$$C^y_{RT} = \sum_{i\in\mathcal{N}_i} \left[R^y_{CG} \xi^{inv}_{CG} \left(IC^{y-1,i}_{CG} - IC^{y,i}_{CG} \right) + \sum_{m\in\mathcal{N}_m} \left(R^0_{SRV,m} \gamma^y_{tech} \xi^{inv}_{SRV,m} RN^{y,i}_{SRV,m} \right) \right] \tag{7}$$

$$RV = \sum_{y\in\mathcal{N}_y} \left[\begin{array}{l} \displaystyle\sum_{i\in\mathcal{N}_i} \sum_{k\in\{NG,PV,WT,ES\}} R^y_k \xi^{inv}_k \left(IC^{y,i}_k - IC^{y-1,i}_k \right) + \sum_{i\in\mathcal{N}_i} R^y_{Fix} \xi^{inv}_{Fix} \left(IN^{y,i}_{DC} - IN^{y-1,i}_{DC} \right) + \\ \displaystyle\sum_{l\in\mathcal{N}_l} R^y_{TL} H_l \xi^{inv}_{TL} \left(IC^{y,l}_{TL} - IC^{y-1,l}_{TL} \right) + \sum_{i\in\mathcal{N}_i} \sum_{j\in\mathcal{N}_i, j\neq i} R^y_{TC} \xi^{inv}_{TC} z_{ij} H_{ij} \left(\bar{Y}^y_{ij} - \bar{Y}^{y-1}_{ij} \right) \end{array} \right] \tag{8}$$

$$R^y_k = \frac{L_k - (|\mathcal{N}_y| - y + 1)}{L_k} (1 - R^0_k) + R^0_k, \forall k \in \{NG,PV,WT,ES,Fix,TC,TL\} \tag{9}$$

Equation (2) evaluates the net present value of total investment costs using discount factor σ for generation resources (C_{GR}), data centers (C_{DC}), transmission line expansion (C_{TL}), and data transmission facilities (C_{TC}). Additionally, the disposal value of retired devices (C_{RT}) and the residual value (RV) of all resources at the end of the planning horizon are incorporated in (2). Equation (3) represents the construction costs associated with generation resources $k \in \{NG,PV,WT,ES\}$. Equation (4) captures the investment costs of servers in data centers and the fixed investments in data centers, including buildings, land, and materials, which are proportional to the number of total installed server racks (IN_{DC}) in the data center. Note that the index m in equation (4) encodes the server type selection for data center planning. The concept of server types is illustrated in Fig. 3, where each type corresponds to servers with uniform processors and cooling systems.

Equation (5) quantifies the investment costs associated with transmission line expansion. Equation (6) models the investment costs for the expansion of data transmission facilities, where z_{ij} is a binary parameter indicating whether location i can provide data services to location j based on propagation delay, as defined by constraints (48)–(49). Equation (7) calculates the disposal value of CG units and servers in data centers. CG units are restricted to retirement, with no new installations allowed to align with decarbonization objectives. Note that the factor R^y_{CG} in (7) evaluates the residual value of the retired CG units. In this paper, we set $R^y_{CG} = 80\%$ in the first year, and then linearly decrease it by 4.5% each year thereafter. Equation (8) computes the residual value of all devices at the end of the planing horizon, and the residual value rate of devices installed in y-th year (R^y_k) is computed in equation (9).

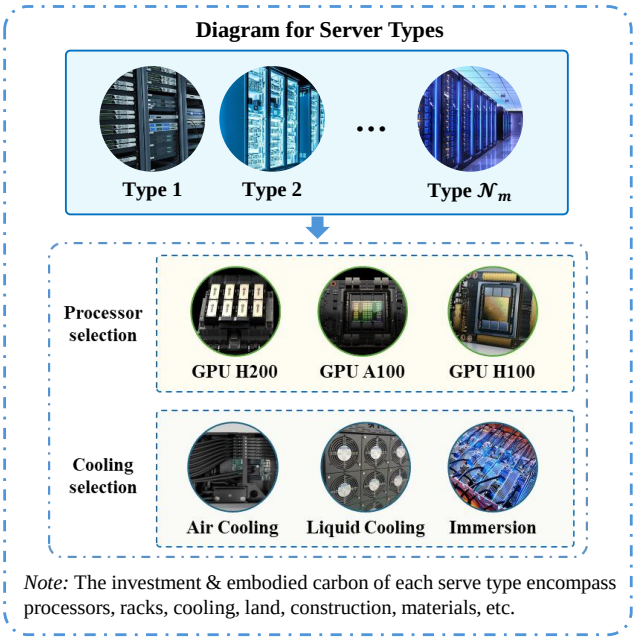

Figure 3: The illustration of types of server racks.

The economical objective for operational costs are given as below:

$$G(\mathbf{x}) = \sum_{y\in\mathcal{N}_y} \frac{C^y_{gen} + C^y_{om} + C^y_{ex} + C^y_{curt}}{(1+\sigma)^y} \tag{10}$$

$$C^y_{gen} = \sum_{d\in\mathcal{N}_d}\sum_{t\in\mathcal{N}_t}\sum_{i\in\mathcal{N}_i}\left(\xi^{gen}_{NG}P^{y,d,t,i}_{NG}\Delta t + \xi^{gen}_{CG}P^{y,d,t,i}_{CG}\Delta t\right) \tag{11}$$

$$C^y_{om} = \sum_{d\in\mathcal{N}_d}\sum_{t\in\mathcal{N}_t}\sum_{i\in\mathcal{N}_i}\left[\sum_{k\in\{NG,CG,PV,WT\}}\xi^{om}_k P^{y,d,t,i}_k\Delta t + \xi^{om}_{ES}\left(P^{y,d,t,i}_{ESD}\Delta t + P^{y,d,t,i}_{ESC}\Delta t\right)\right] \tag{12}$$

$$C^y_{ex} = \sum_{d\in\mathcal{N}_d}\sum_{t\in\mathcal{N}_t}\sum_{i\in\mathcal{N}_i}\left(\xi^{buy}_{i,t}P^{y,d,t,i}_{BUY}\Delta t - \xi^{sell}_{i,t}P^{y,d,t,i}_{SELL}\Delta t\right) \tag{13}$$

$$C^y_{curt} = \sum_{d\in\mathcal{N}_d}\sum_{t\in\mathcal{N}_t}\sum_{i\in\mathcal{N}_i}\left[\xi^{curt}_{WT}\left(P^{y,d,t,i}_{WT,cap} - P^{y,d,t,i}_{WT}\right)\Delta t + \xi^{curt}_{PV}\left(P^{y,d,t,i}_{PV,cap} - P^{y,d,t,i}_{PV}\right)\Delta t\right] \tag{14}$$

Equation (10) summarizes the total system operational costs. Specifically, electricity generation costs (C_{gen}) are calculated in equation (11) for NG and CG units. The maintenance costs (C_{om}) for all generation resources are given in (12). The power transactions (C_{ex}) between the PJM and other ISOs are specified in (13). The RES curtailments (C_{curt}) are punished in (14).

2.2.2. Environmental objective

To minimize the system-wide carbon emissions, including embodied carbon emissions CE_{EMB} from the planning stage and operational carbon emissions CE_{OPR} from the operational stage.

$$TE = \min_{\mathbf{z}\in\mathbb{Z},\,\mathbf{x}\in\mathbb{X}} \sum_{y\in\mathcal{N}_y}\left[\underbrace{CE^y_{EMB}}_{\textit{Planning Stage}} + \underbrace{CE^y_{OPR}}_{\textit{Operational Stage}}\right] \tag{15}$$

$$CE^{y}_{EMB} = \left[\begin{array}{l} \sum\limits_{i\in\mathcal{N}_i} \sum\limits_{k\in\{NG,PV,WT,ES\}} \chi^{emb}_{k} \left(IC^{y,i}_{k} - IC^{y-1,i}_{k} \right) + \sum\limits_{i\in\mathcal{N}_i} \sum\limits_{j\in\mathcal{N}_i, j\neq i} \chi^{emb}_{TC} z_{ij} H_{ij} \left(\bar{Y}^{y}_{ij} - \bar{Y}^{y-1}_{ij} \right) + \\ \sum\limits_{l\in\mathcal{N}_l} \chi^{emb}_{TL} H_l \left(IC^{y,l}_{TL} - IC^{y-1,l}_{TL} \right) + \sum\limits_{i\in\mathcal{N}_i} \sum\limits_{m\in\mathcal{N}_m} \lambda^{y}_{tech} \chi^{emb}_{SRV,m} \left(BN^{y,i}_{SRV,m} \right) + \sum\limits_{i\in\mathcal{N}_i} \chi^{emb}_{Fix} \left(IN^{y,i}_{DC} - IN^{y-1,i}_{DC} \right) \end{array} \right] \tag{16}$$

$$CE^{y}_{OPR} = \sum_{d\in\mathcal{N}_d} \sum_{t\in\mathcal{N}_t} \sum_{i\in\mathcal{N}_i} \left(\chi^{gen}_{NG} P^{y,d,t,i}_{NG} + \chi^{gen}_{CG} P^{y,d,t,i}_{CG} \right) \Delta t \tag{17}$$

Equation (15) summarizes the carbon emissions. Equation (16) calculates the total embodied carbon emissions during system construction, and equation (17) represents the operational carbon emissions from electricity generation.

2.2.3. Planning constraints for power system

The constraints for the planning stage of the power system are as follows ($\forall y \in \mathcal{N}_y, \forall i \in \mathcal{N}_i$).

$$IC^{y,i}_{k} = N^{y,i}_{k} P^{unit}_{k} + IC^{y-1,i}_{k}, \ \forall k \in \{NG,PV,WT\} \tag{18}$$

$$IC^{y,i}_{ES} = N^{y,i}_{ES} E^{unit}_{ES} + IC^{y-1,i}_{ES} \tag{19}$$

$$IC^{y,l}_{TL} = N^{y,l}_{TL} P^{unit}_{TL} + IC^{y-1,l}_{TL} \tag{20}$$

$$IC^{y-1,i}_{CG} = N^{y,i}_{CG} P^{unit}_{CG} + IC^{y,i}_{CG} \tag{21}$$

$$0 \leq IC^{y,i}_{CG}, \ \sum\nolimits_{i\in\mathcal{N}_i} N^{y,i}_{CG} P^{unit}_{CG} \leq \bar{RT}^{y}_{CG} \tag{22}$$

$$\underline{IC}^{y,i}_{k} \leq IC^{y,i}_{k} \leq \bar{IC}^{y,i}_{k}, \ \forall k \in \{NG,PV,WT,ES\} \tag{23}$$

$$N^{y,i}_{k} \in \mathbb{Z}_{+}, \forall k \in \{NG,PV,WT,ES,TL\}; N^{y,i}_{CG} \in \mathbb{Z}_{+} \tag{24}$$

Equations (18)-(21) indicate that the newly installed capacities of generation resources and transmission lines must follow discrete unit sizes. Notably, (21) specifies that coal-fired generators can only be retired. Thus, we have $IC^{y-1,i}_{CG} \geq IC^{y,i}_{CG}$. Constraints (22) limits and ensures the maximum retired capacity of CG units according to the scheduled retirement plan. Constraint (23) limits the installed capacity of power generation resources. Meanwhile, the renewable portfolio standards (RPS), which require regions to install RES to a certain degree by specific years [27], are also reflected in this constraint. Constraints (24) ensure the non-negativity and discreteness of the planning variables.

2.2.4. Planning constraints for data centers

The constraints for data centers' planning are given as follows ($\forall y \in \mathcal{N}_y, \forall i \in \mathcal{N}_i, m \in \mathcal{N}_m$).

$$IN^{y,i}_{SRV,m} - IN^{y-1,i}_{SRV,m} = BN^{y,i}_{SRV,m} - RN^{y,i}_{SRV,m} \tag{25}$$

$$IN^{y,i}_{SRV,m} - IN^{y-1,i}_{SRV,m} \geq 0, \tag{26}$$

$$IN^{y,i}_{DC} = \sum\nolimits_{m\in\mathcal{N}_m} IN^{y,i}_{SRV.m} \tag{27}$$

$$RN^{(y+L_{SVR,m}),i}_{SRV,m} = \begin{cases} BN^{y,i}_{SRV,m}, & \text{if } y + L_{SVR,m} \leq |N_y| \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

$$RN^{y,i}_{SRV,m}, BN^{y,i}_{SRV,m} \in \mathbb{Z}_{+} \tag{29}$$

$$z_{ij,i\neq j} = \begin{cases} 1, \ H_{ij} \leq H_0 \\ 0, \ H_{ij} > H_0 \end{cases} \tag{30}$$

$$H_0 = T_{req}/\tau_0 \tag{31}$$

Constraint (25) calculates the number of installed servers based on the number of newly added servers (*BN*) and the number of retired servers (*RN*). Constraint (26) ensures that the number of installed servers in data centers do

not decrease over time. Constraints (27) calculates the total number of installed servers in the data center at location i. Constraint (28) ensures that the servers will retire once they reach the expected life. Constraints (29) limits the variables representing the number of installed and retired servers to be discrete. Constraints (30) calculate the binary variables z_{ij} to determine whether location i can provide data service for location j. If $H_{ij} \leq H_0$, location i can provide service for location j while meeting the required propagation delay; otherwise, it cannot. Equation (31) calculates the maximum acceptable distance H_0, where T_{req} represents the permitted propagation delay for data center location, and τ_0 denotes the propagation delay per unit distance. The communication networks' parameters are assumed to be fixed, with a typical propagation delay of approximately 0.82 ms per 100 miles [28].

2.2.5. Operational constraints for power system

The power system constraints for the operational stage are as follows ($\forall y \in \mathcal{N}_y, d \in \mathcal{N}_d, t \in \mathcal{N}_t, \forall i \in \mathcal{N}_i$).

$$P_{GEN}^{y,d,t,i} = \left[P_{NG}^{y,d,t,i} + P_{CG}^{y,d,t,i} + P_{WT}^{y,d,t,i} + P_{PV}^{y,d,t,i} + P_{ESD}^{y,d,t,i} - P_{ESC}^{y,d,t,i} + P_{BUY}^{y,d,t,i} - P_{SELL}^{y,d,t,i} \right] \tag{32}$$

$$\sum_{i \in \mathcal{N}_i} \left(P_{GEN}^{y,d,t,i} - P_{DC}^{y,d,t,i} - P_{LOAD}^{y,d,t,i} \right) = 0 \tag{33}$$

$$P_{TL}^{y,d,t,l} = \sum_{i \in \mathcal{N}_i} \Gamma_{l,i} (P_{GEN}^{y,d,t,i} - P_{DC}^{y,d,t,i} - P_{LOAD}^{y,d,t,i}) \tag{34}$$

$$-IC_{TL}^{y,l} \leq P_{TL}^{y,d,t,l} \leq IC_{TL}^{y,l}, \ \forall l \in \mathcal{N}_l \tag{35}$$

$$0 \leq P_{BUY}^{y,d,t,i}, \ P_{SELL}^{y,d,t,i} \leq P_{EX}^{\max} \tag{36}$$

Equation (32) defines the nodal power generation. Equation (33) enforces the system-wide power balance. Equation (34) calculates the transmission line power flows using power transfer distribution factors ($\Gamma_{l,i}$). Constraints (35) enforce power flow limits on transmission lines. Constraint (36) imposes capacity limitations on power transactions through the system interface.

The operational constraints for NG and CG units are as follows ($\forall y \in \mathcal{N}_y, d \in \mathcal{N}_d, t \in \mathcal{N}_t, \forall i \in \mathcal{N}_i$)

$$\beta_{NG}^{\min} IC_{NG}^{y,i} \leq P_{NG}^{y,d,t,i} \leq \beta_{NG}^{\max} IC_{NG}^{y,i} \tag{37}$$

$$\beta_{CG}^{\min} IC_{CG}^{y,i} \leq P_{CG}^{y,d,t,i} \leq \beta_{CG}^{\max} IC_{CG}^{y,i} \tag{38}$$

$$-R_{NG}^{down} IC_{NG}^{y,i} \leq P_{NG}^{y,d,t,i} - P_{NG}^{y,d,t-1,i} \leq R_{NG}^{up} IC_{NG}^{y,i} \tag{39}$$

$$-R_{CG}^{down} IC_{CG}^{y,i} \leq P_{CG}^{y,d,t,i} - P_{CG}^{y,d,t-1,i} \leq R_{CG}^{up} IC_{CG}^{y,i} \tag{40}$$

$$P_{NG}^{y,d,0,i} = P_{NG}^{y,d,T,i}, \ P_{CG}^{y,d,0,i} = P_{CG}^{y,d,T,i} \tag{41}$$

Constraints (37)-(38) specify the capacity limitations for NG and CG units. Constraints (39)-(40) restrict the ramping capabilities accordingly. Equation (41) ensures the initial output of the day equals the end output of the day with $T = |\mathcal{N}_t|$.

The RES and ES operational constraints are formulated as follows ($\forall y \in \mathcal{N}_y, d \in \mathcal{N}_d, t \in \mathcal{N}_t, \forall i \in \mathcal{N}_i$)

$$0 \leq \left[P_{WT}^{y,d,t,i}, \ P_{PV}^{y,d,t,i} \right] \leq \left[P_{WT,cap}^{y,d,t,i}, \ P_{PV,cap}^{y,d,t,i} \right] \tag{42}$$

$$E_{ES}^{y,d,t,i} = E_{ES}^{y,d,t-1,i} + (P_{ESC}^{y,d,t,i} \eta_{esc} - P_{ESD}^{y,d,t,i} / \eta_{esd}) \Delta t \tag{43}$$

$$\beta_{ES}^{\min} IC_{ES}^{y,i} \leq E_{ES}^{y,d,t,i} \leq \beta_{ES}^{\max} IC_{ES}^{y,i} \tag{44}$$

$$\left[P_{ESC}^{y,d,t,i}, P_{ESD}^{y,d,t,i} \right] \begin{cases} \geq \left[\beta_{ESC}^{\min}, \beta_{ESD}^{\min} \right] IC_{ES}^{y,i} \\ \leq \left[\beta_{ESC}^{\max}, \beta_{ESD}^{\max} \right] IC_{ES}^{y,i} \end{cases} \tag{45}$$

$$E_{ES}^{y,d,0,i} = E_{ES}^{y,d,T,i} \tag{46}$$

Constraints (42) enforce the RES output limit. Equation (43) describes the energy balance of the ES system. Constraints (44) limit the energy capacity. Constraints (45) limit the charging and discharging power. Equations (46) enforce the starting and ending energy levels of energy storage systems to be equal.

2.2.6. Operational constraints for data center

The following constraints describe the operation of data center ($\forall y \in \mathcal{N}_y, d \in \mathcal{N}_d, t \in \mathcal{N}_t, \forall i \in \mathcal{N}_i, \forall m \in \mathcal{N}_m$).

$$Y_{DC}^{y,d,t,i} + \sum_{\forall j, j \neq i} Y_{TC}^{y,d,t,ij} - Y_{LOAD}^{y,d,t,i} = 0 \tag{47}$$

$$-z_{ij}\bar{Y}_{ij}^{y} - Y_{ij}^{0} \leq Y_{TC}^{y,d,t,ij} \leq z_{ij}\bar{Y}_{ij}^{y} + Y_{ij}^{0} \tag{48}$$

$$Y_{DC}^{y,d,t,i} = \sum_{m \in \mathcal{N}_m} \psi_{SVR,m} X_{SVR,m}^{y,d,t,i} \Delta t \tag{49}$$

$$P_{DC}^{y,d,t,i} = \sum_{m \in \mathcal{N}_m} \phi_{y,d,t,i}^{temp} PUE_m X_{SVR,m}^{y,d,t,i} P_{SVR,m}^{rate} \tag{50}$$

$$0 \leq X_{SVR,m}^{y,d,t,i} \leq IN_{SVR,m}^{y,i} \tag{51}$$

Equation (47) describes the nodal computing workload balance, where $Y_{DC}^{y,d,t,i}$ denotes the computing service provided by the data center at location i, $Y_{TC}^{y,d,t,ij}$ represents the computing workload transferred location j to location i, and $Y_{LOAD}^{y,d,t,i}$ denotes the nodal computing workload demand. Constraints (48) represent computing workload transfer limitation. Equation (49) calculates the computing workload provision at location i based on the active number of rack servers $X_{SVR,m}$. Equation (50) calculates the power consumption of the data center at location i, considering the power utilization efficiency PUE_m, a locational and seasonal cooling efficiency correction factor $\phi_{y,d,t,i}^{temp}$. Constraint (51) ensures that the active number of rack servers does not exceed the number of installed servers.

3. Methodology

The proposed joint planning model is formulated as a large-scale, mixed-integer, stochastic optimization problem with multiple objectives. To enhance computational efficiency, we employ several solution techniques and acceleration strategies, including the single-objective reformulation, candidate location selection, long- and short-term uncertainty sampling, and a customized Benders decomposition approach. These techniques collectively enhance computational efficiency, transforming the problem into a tractable optimization framework.

3.1. Single-objective Reformulation

The joint planning model simultaneously considers both economic and environmental objectives. Given the complexity associated with solving multi-objective optimization problems, we employ the weighted sum method [29] to convert the multi-objective formulation into an equivalent single-objective optimization problem, as expressed in (52). In this formulation, $\hbar$ represents the weighting factor that reflects the decision maker's preference toward carbon emission reduction.

$$\begin{aligned} \boldsymbol{J} = & \min_{\mathbf{z} \in \mathbb{Z},\, \mathbf{x} \in \mathbb{X}} [TC(\mathbf{z}, \mathbf{x}) + \hbar \cdot TE(\mathbf{z}, \mathbf{x})] \\ \text{s.t.} \quad & \text{Constraints (1)} - \text{(51)} \end{aligned} \tag{52}$$

This approach is practical, as $\hbar$ is a quantifiable parameter, which can be interpreted as the price for carbon emission. Specifically, we obtain the carbon emission auction price data from the Regional Greenhouse Gas Initiative (RGGI) and use an initial price of \$22 /short ton with a 2.3% yearly increase, as estimated in the exploratory policy scenario of the technical report [30].

3.2. Long-Term and Short-Term Uncertainty Modeling

In this paper, we consider two types of system uncertainties. The first type pertains to long-term uncertainties, such as the annual increase in power demand and computing workload demand, which exhibit predictable patterns to some extent. The second type encompasses short-term uncertainties, including the daily variations in both power and computing workload demand, as well as the inherent variability of solar and wind power generation. To model these uncertainties, we define ζ_s as the stochastic scenario representation and incorporate them into the planning model through (53)-(55) ($\forall y \in \mathcal{N}_y, d \in \mathcal{N}_d, t \in \mathcal{N}_t, \forall i \in \mathcal{N}_i$).

$$P_{LOAD}^{y,d,t,i} = \varphi_i^{load} \lambda_y^{load}(\zeta_s)\, \delta_{d,t}^{load}(\zeta_s)\, P_{LOAD}^{peak} \tag{53}$$

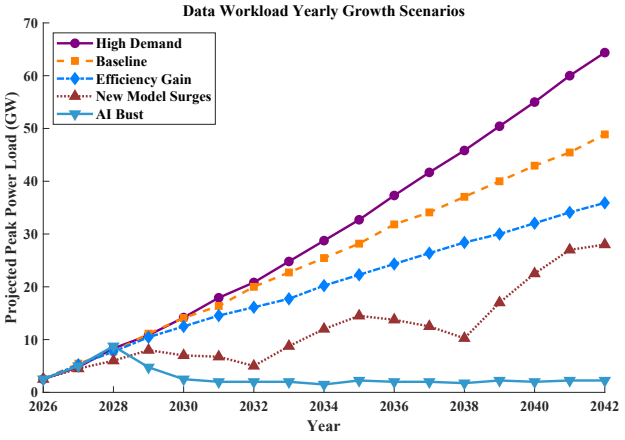

Figure 4: Scenarios for AI demand growth (i.e., $\lambda_y^{data} Y_{LOAD}^{peak}$ converted to GW).

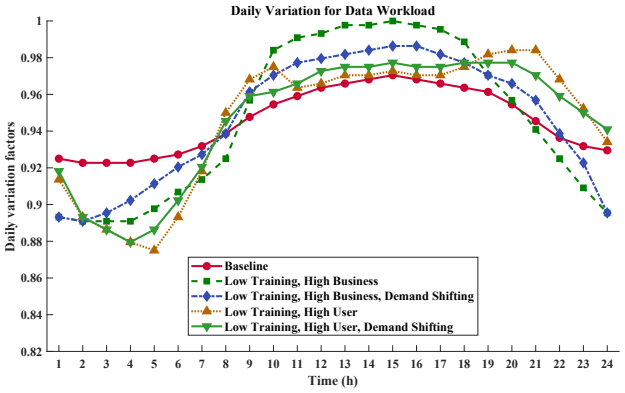

Figure 5: Scenarios for daily variation factors of computing workload (i.e., $\delta_{d,t}^{data}$).

$$Y_{LOAD}^{y,d,t,i} = \varphi_i^{data} \lambda_y^{data}(\zeta_s) \delta_{d,t}^{data}(\zeta_s) Y_{LOAD}^{peak} \tag{54}$$

$$\left[P_{WT,cap}^{y,d,t,i},\ P_{PV,cap}^{y,d,t,i}\right] = \left[\delta_{d,t}^{wt}(\zeta_s) IC_{WT}^{y,i},\ \delta_{d,t}^{pv}(\zeta_s) IC_{PV}^{y,i}\right] \tag{55}$$

Equation (53) models the total power demand by multiplying an annual increase factor, λ_y^{load}, a daily variation factor, $\delta_{d,t}^{load}$, and the base peak load, P_{LOAD}^{peak}. The annual increase factor represents long-term uncertainties, whereas the daily variation factor accounts for short-term fluctuations. In this study, the annual increase in power demand is assumed to be 1% per year on average. The daily variation is derived from historical data using a sampling method similar to that employed for the daily variation factor of RESs, as detailed in later sections.

Equation (54) characterizes the computing workload demand of the system, which can be estimated by dividing the total power consumption of data centers in the PJM system by the average energy consumption per request. The computing workload is assumed to be AI-dominated. The annual increase factor, λ_y^{dc}, is classified into five scenarios—high demand, baseline, efficiency gains, new model surges, and AI bust—to capture various potential trajectories of AI-driven data center growth. The values of the peak computing workload for each year (i.e., $\lambda_y^{dc} Y_{LOAD^{peak}}$) under each growth pattern are presented in Fig. 4 [31]. Note that we consider AI-dominated data centers. Thus, a more general measurement for AI computation workload is adopted, that is GPU · h, which is convenient to measure the processing capability of different types of servers. A GPU·h represents the amount of computing workload that can be performed using a benchmark GPU at full load for one hour. This paper uses the NVIDIA V100 GPU as the benchmark, and it assumes that 1 GPU h = 10^4 requests. The daily computing workload demand is assumed to follow five representative scenarios, as shown in Fig. 5 [31].

Equation (55) determines the available power output from RES by applying time-varying factors, $\delta_{d,t}^{wt/pv}$, to scale the installed capacity. To account for daily demand-supply uncertainties, a hierarchical sampling strategy is employed for stochastic scenario generation [32]. First, seasonal patterns and short-term fluctuations of random variables (e.g., RES generation and power demand) are characterized based on historical data or their probability distributions. For instance, the Weibull distribution is commonly used to model the stochastic nature of wind power [33], while the random sequences of electric demand are synthesized based on hourly variations following a normal distribution [34]. Subsequently, stochastic scenarios are generated using the Latin Hypercube Sampling (LHS) method, which preserves correlations among different random variables [35]. To enhance computational efficiency while maintaining accuracy, the K-means clustering technique is applied to reduce the scenario set to a manageable size.

3.3. Enhanced Benders Decomposition

3.3.1. Compact Formulation

The joint planning model is presented in (52). For the convenience of expressing the customized Benders decomposition method, its compact matrix formulation is given in (56).

$$J = \min \sum_{y\in\mathcal{N}_y}\left(c_y^T z_y + \sum_{d\in\mathcal{N}_d}\sum_{s\in\mathcal{N}_s}\pi_s q^T x_{y,d,s}\right) \tag{56a}$$

$$\text{s.t. } A_y z_y \le b_y,\ \forall y \tag{56b}$$

$$E_{y,d}\ x_{y,d,s} = f_{y,d,s},\ \forall y,d,\forall s\in\mathcal{N}_s \tag{56c}$$

$$F_{y,d,s}\ z_y + G_{y,d}\ x_{y,d,s} \le h_{y,d,s},\ \forall y,d,\forall s\in\mathcal{N}_s \tag{56d}$$

$$z_y \in \mathbb{Z}_+,\ x_{y,d,s}\in\mathbb{R}, \tag{56e}$$

Where z_y represents the planning decision variables for year y, and $x_{y,d,s}$ denotes the operational decision variables for year y, day d, and scenario s. Relaxing the daily cycling constraints (41) and (46) enables the separation of operational constraints for each day. Consequently, system operations can be modeled as independent daily operational subproblems, which forms the basis of the customized Benders decomposition method. Additionally, π_s in (56a) denotes the probability of scenario s, constraint (56b) represents the planning constraints, where equalities are reformulated as equivalent inequality pairs for brevity. Constraints (56c) and (56d) define the operational constraints for each day of each year.

3.3.2. Feasibility-check subproblem

We first define a subproblem ($FP_{y,d,s}$) to check the feasibility of planning-stage solution (denoted by $\hat{z}_y$) with respect to the operational constraints under each daily stochastic scenarios. The formulation of $FP_{y,d,s}$ is presented as follows:

$$FP_{y,d,s}:\ \tilde{\Upsilon}_{y,d,s} = \min\left[\mathbf{1}^T(\delta_{y,d,s}^+ + \delta_{y,d,s}^-) + \mathbf{1}^T\epsilon_{y,d,s}\right] \tag{57a}$$

$$\text{s.t. } E_{y,d}\ x_{y,d,s} = f_{y,d,s} + \delta_{y,d,s}^+ - \delta_{y,d,s}^- \tag{57b}$$

$$F_{y,d,s}\ z_y + G_{y,d}\ x_{y,d,s} \le h_{y,d,s} + \epsilon_{y,d,s} \tag{57c}$$

$$z_y = \hat{z}_y \quad (\text{dual vars: } \lambda_{y,d,s}) \tag{57d}$$

$$\delta_{y,d,s}^+, \delta_{y,d,s}^-, \epsilon_{y,d,s} \ge 0 \tag{57e}$$

The objective function (57a) minimizes the sum of all slack variables, which are non-negative variables introduced in (57e). Constraints (57b) and (57c) represent the operational constraints, incorporating slack variables $\delta_{y,d,s}^+$, $\delta_{y,d,s}^-$, and $\epsilon_{y,d,s}$. Equation (57d) enforces the planning decision variables, z_y, to be equal to the obtained planning-stage solution, while $\mu_{y,d,s}$ denotes the dual variables associated with this constraint.

This formulation employs a practical technique to streamline the solution process in Benders decomposition. Since planning-stage decisions are the only link between the planning and operational stages, the Benders cuts are constructed based solely on the subproblem's derivatives with respect to planning decisions. Notably, state-of-the-art solvers, such as Gurobi, can efficiently compute dual solutions for specific constraints. Equation (57d) directly facilitates the identification of dual solutions, thereby accelerating the formulation of Benders cuts.

3.3.3. Optimality subproblem

Next, the optimality subproblem, denoted as ($OP_{y,d,s}$), is defined based on the operational constraints. The subproblem $OP_{y,d,s}$ is solved only when the feasibility-check problem, $FP_{y,d,s}$, attains a zero objective value, indicating that the planning-stage solution satisfies all operational constraints without violating feasibility conditions. The formulation of $OP_{y,d,s}$ is presented below.

$$OP_{y,d,s}: \tilde{\Phi}_{y,d,s} = \min q_y^T x_{y,d,s} \tag{60a}$$

$$\text{s.t. } E_{y,d}\, x_{y,d,s} = f_{y,d,s} \tag{60b}$$

$$F_{y,d,s}\, z_y + G_{y,d}\, x_{y,d,s} \le h_{y,d,s} \tag{60c}$$

$$z_y = \hat{z}_y \quad (\text{dual vars: } \mu_{y,d,s}) \tag{60d}$$

The objective function (60a) minimizes the recourse cost for each operational day. Constraints (60b)–(60c) define the operational constraints of the power system with data centers, given the planning decision $\hat{z}_y$. Equation (60d) follows the proposed technique to capture the dual variables of $OP_{y,d,s}$ with respect to the planning-stage solution. It

Algorithm 1 Customized Benders Decomposition Method

Step 1. Set $LB = -\infty$, $UB = \infty$, and $l = 0$.

Step 2. Solve the master problem (MP) as below.

$$MP_l: O^* = \min \sum_{y\in\mathcal{N}_y} \left(c_y^T z_y + \sum_{d\in\mathcal{N}_d} \sum_{s\in\mathcal{N}_s} \pi_s \tilde{\Phi}_{y,d,s} \right) \tag{58a}$$

$$\text{s.t. } A_y z_y \le b_y,\ \forall y \tag{58b}$$

$$\tilde{\Phi}_{y,d,s} \ge \tilde{\Phi}^{j,*}_{y,d,s} - \mu^{j,*}_{y,d,s}(z_y - z_y^{j,*}),\ \forall y, d, s,\ 1 \le j \le l \tag{58c}$$

$$0 \ge \tilde{\Upsilon}^{j,*}_{y,d,s} - \lambda^{j,*}_{y,d,s}(z_y - z_y^{j,*}),\ \forall y, d, s,\ 1 \le j \le l \tag{58d}$$

$$z_y \in \mathbb{Z}_+, \tilde{\Phi}_{y,d,s} \in \mathbb{R}_+, \quad \forall y, d, s \tag{58e}$$

- If MP_l is infeasible, then terminate the algorithm and report the infeasibility.
- Otherwise, let $l \leftarrow l + 1$. Get solution $z_y^{l,*}, \forall y$ and the optimal value O^*. Then, update $LB = O^*$.

Step 3. Solve the subproblems *FP* and *OP* for each scenario $s \in \mathcal{N}_s$ as below.

for $\forall y \in \mathcal{N}_y$ **do**

  for $\forall d \in \mathcal{N}_d$ **do**

    Solve the feasibility-check problem $FP_{y,d,s}$.

    - If $\tilde{\Upsilon}^{l,*}_{y,d,s} = 0$, then set $\lambda^{l,*}_{y,d,s} = 0$, and solve the $OP_{y,d,s}$ to derive optimal solution $\tilde{\Phi}^{l,*}_{y,d,s}$, $\mu^{l,*}_{y,d,s}$.
    - Otherwise, derive the optimal solution $\tilde{\Upsilon}^{l,*}_{y,d,s}$ and $\lambda^{l,*}_{y,d,s}$. Set $\tilde{\Phi}^{l,*}_{y,d,s} = 0$, $\mu^{l,*}_{y,d,s} = 0$.

  end for

end for

Step 4. Update $UB = \min\{UB, UB^{l,*}\}$, where,

$$UB^{l,*} = \sum_{y\in\mathcal{N}_y} \left(c_y^T z_y^{l,*} + \sum_{d\in\mathcal{N}_d} \sum_{s\in\mathcal{N}_s} \pi_s \tilde{\Phi}^{l,*}_{y,d,s} \right) \tag{59}$$

Step 5. Check the tolerance and generate Benders cuts.

- If $|UB - LB| \le$ tolerance, then terminate and report the optimal solution.
- Otherwise, generate and add Benders cuts in (58c) and (58d). Go to the **Step 2.**

is mentioned that with such formulation, there is no need to solve the dual problems explicitly when formulating the Benders cuts.

3.3.4. Customized Benders decomposition

With the above defined subproblems, the customized Benders decomposition method is presented in Algorithm 1.

Remark: Several enhancement techniques for the customized Benders decomposition algorithm are provided here:

- The lower-level subproblems in Algorithm 1 are decomposed based on scenarios, which provides a well-structured framework to leverage the advantages of parallel computing, thereby improving computational efficiency.
- In the current framework in **Step 3**, subproblems are divided by individual uncertain scenarios. In practical coding, grouping several scenarios into a single subproblem can yield better computational efficiency. The number of scenarios included in each subproblem is determined heuristically. In this study, grouping four scenarios per subproblem yields the best performance.
- The master problem can include an initial operating scenario to expedite the feasibility check of the subproblems. In this study, we heuristically incorporate the scenario corresponding to the peak load into the master problem as the initial scenario, significantly enhancing the overall computational efficiency.

4. Case Study

4.1. The PJM Interconnection Test Case

This paper evaluates the proposed joint planning model using the test case of PJM interconnection, one of the most prominent ISOs in the United States, covering a significant portion of the nation's data centers (approximately one-third). The system data of PJM interconnection is obtained from the U.S. Energy Information Administration (EIA), with updates through 2022 [36]. The transmission zones of PJM are defined in [37], and each zone is modeled as a bus. The transmission lines rated above 161 kV between the zones are modeled. The resulting simplified PJM interconnections consists of 21 buses and 196 transmission lines. The dataset for the 21-bus PJM interconnection has been compiled and is available at [26]. The annual peak non-datacenter power demand prior to the planning horizon is assumed to be 173 GW [38]. The data center power demand at the beginning of the planning period is set to 2.5 GW [31, 39].

The data for capital and operational cost of generation resources are collected from [40, 41]. The embodied carbon emissions data for generation resources are obtained from [42–45]. The data for embodied carbon emissions and operational parameters of data center racks are derived from [31, 46]. The capital cost of data center servers is estimated based on publicly available data of NVIDIA. Other system parameters are borrowed from [15, 32]. The basic operational scenarios for power demand and RES are obtained from the PJM API tools [38]. The data center demands are sourced from [31]. The summary of the collected data is given in Table 3 – Table 6. In this paper, a 15-year long-term planning horizon is considered, with four representative days per year selected to characterize each season. A total of 50 operating scenarios are generated, incorporating both long-term demand growth and short-term RES variations, based on the method described in Section 3.2, which also provides the daily and yearly operational patterns of data centers. The representative scenarios for each season are illustrated in Fig. 6, providing a clear overview of their characteristics. Besides, the following assumptions are considered during the planning.

4.1.1. Energy Resource Installation Capacity Constraints

The area of transmission zones varies significantly, resulting in differences in the installation capacity of RES and thermal units across zones. To address this, we estimate each zone's energy resource installation capacity based on its peak power demand at the start of the planning horizon. Accordingly, the maximum installation capacity of each generation resource in each transmission zone is constrained to be less than five times its peak power demand in the initial year of planning.

4.1.2. Limitations for Coal-Fired Power Plant Retirement

If there were no retirement capacity limit, the system would rapidly retire all CG units and replace them with RES within two years to secure long-term retirement compensation and benefit from the low generation cost of RES.

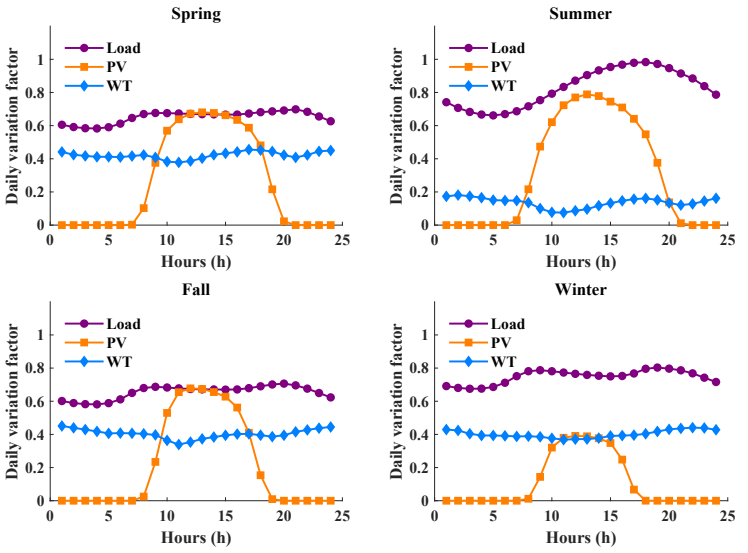

Figure 6: Seasonal patterns in non-data center power demand, solar, and wind power.

Table 3
Generation Resources and Energy Storage Devices

Generation Resources							
Label	Capacity (MW)	Gen. Cost ($/MWh)	Inv. ($/MW)	O&M ($/MWh)	Emb CO_2 (lbs CO_2/MW)	Opr CO_2 (lbs CO_2/MWh)	Life Time (yr)
CG	100	37.43	4,103,000	6.4	169,091	1400	30
NG	100	39.59	921,000	3.3	152,654	1200	30
PV	10	/	1,302,000	2.8	1,485,900	/	25
WT	10	/	1,686,000	9.6	503,360	/	25
Energy Storage Devices							
Label	Energy Rating (MWh)	Power Rating (MW)	Energy Inv. ($/MWh)	Power Inv. ($/MW)	O&M ($/MWh)	Emb CO_2 (lbs CO_2/MWh)	Life Time (yr)
ES	1	0.35	436,000	1,245,000	4.2	240,000	18

Table 4
Data Center Server Module Specifications

Label	Description	Unit (Servers/Rack)	Inv. ($/Rack)	Emb CO_2 (lbs CO_2/Rack)	Power (kW/Rack)	Capability (GPU·h/Rack)	PUE	Life Time (yr)
Fix Inv.	Buildings & Racks	/	850,000	150,000	/	/	/	25
Type-I	V100 & Air Cooling	150	1,500,000	900,000	50	150	1.5	5
Type-II	H100 & Water Cooling	150	4,800,000	750,000	130	2,370	1.3	5
Type-III	H200 & Liquid Cooling	150	5,650,000	650,000	135	3,090	1.2	5

However, this does not accurately reflect real-world conditions. Therefore, we limit the annual retirement of CG power plant capacity to a maximum of 3 GW to account for reliability and other technical factors.

4.1.3. Renewable Portfolio Standards

The state authorities in PJM have established RPS or clean energy laws that have been enacted or issued. Many states operating in PJM have had RPS laws in effect for nearly 20 years. These ambitious carbon reduction targets set by the executive orders have significant implications for long-term planning. Therefore, the RPS information is incorporated into our long-term planning model. The relevant RPS constraints are summarized below in Table 7, with data sourced from Table 12 in Reference [27].

Table 5
Electricity and Data Transmission

Transmission Lines				
Label	Rating (MW)	Inv. ($/km · MW)	Emb CO_2 (lbs CO_2/km · MW)	Life Time (yr)
TL	10	21,940	21,340	30
Data Transfer Facilities				
Label	Bandwidth (Gbps)	Inv. ($/km)	Emb CO_2 (lbs CO_2/km)	Life Time (yr)
TC	50	60,000	17,637	30

Table 6
Economic and Technical Parameters

Parameter	Description	Value
σ	Discount rate	0.04
R^0	Residual factor at retirement	0.15
	On-peak Electricity price ($/MWh)	156
$\xi^{buy/sell}$	Mid-peak Electricity price ($/MWh)	115
	Off-peak Electricity price ($/MWh)	82
$\xi^{curt}_{PV/WT}$	Curtailment cost for PV and WT ($/MWh)	30
τ_0	Propagation delay per unit distance (ms/100 km)	0.51
T_{req}	Maximum propagation delay (ms)	10
λ_{tech}	Embodied carbon equivalent rate from server tech advancements	0.95
γ_{tech}	Investment equivalent rate from server tech advancements	0.85
$\hbar$	Initial carbon price ($/short ton)	22

Table 7
Renewable Portfolio Standards for States in PJM

State	Bus Index	Percentage of load served by RES
New Jersey	7, 8, 11, 12	35% by 2025; 50% by 2030; 100% by 2035
Maryland	1, 6, 9	50% by 2030; 100% by 2045
Illinois	16	40% by 2030; 50% by 2040
Delaware	9	25% by 2026; 40% by 2035
Indiana	21	10% by 2026
Michigan	21	60% by 2035; 100% by 2040
North Carolina	13	70% by 2035
Ohio	17, 18, 19	10% by 2026
Pennsylvania	2, 3, 4, 5, 10	10% by 2026
Virginia	13	100% by 2045
Washington DC	6	100% by 2032

4.2. Comparison Tests

This section presents a comparative evaluation of the proposed joint planning method against a conventional non-joint planning approach. The objective is to demonstrate the effectiveness of co-optimizing power systems and data centers from both economic and decarbonization perspectives. In addition, we examine the impact of including embodied carbon emissions in the planning process to highlight its importance in achieving long-term sustainability goals. Subsequent sections will provide an in-depth analysis of the PJM case study based on the proposed framework.

4.2.1. Comparison with Non-Joint Planning Approach

Non-joint planning approaches are commonly adopted in the literature, such as [20, 21, 23], where data centers determine their siting and capacity decisions based on locational carbon intensity and electricity prices, while power

Table 8
Comparison of Proposed Joint and Conventional Non-Joint Planning Approaches

Metric	Unit	Non-Joint	Joint (Proposed)	Improvement (%)
Total Investment Cost	B$/yr	48.76	**42.60**	**−12.63%**
▷ Power System Investment	B$/yr	28.43	**25.10**	**−11.72%**
▷ Data Center Investment	B$/yr	20.33	**17.50**	**−13.93%**
Operational Cost	B$/yr	62.19	**57.05**	**−8.25%**
Operational Carbon Emissions	$MtCO_2$/yr	154.7	**145.9**	**−5.63%**
Embodied Carbon Emissions	$MtCO_2$/yr	85.1	**72.6**	**−14.69%**
Renewable Capacity Share	%	36.2%	**45.4%**	**+25.41%**

Table 9
Comparison of Joint Planning With and Without Embodied Carbon Consideration on the 15-Year PJM Case

Metric	Joint-OP (Only OP in Objective)	Joint-EM (OP & EM)	Change (%)
Total Investment Cost (B$)	967.6	**1035.3**	**+7.0%**
Operational Carbon Emissions ($MtCO_2$)	2929.2	**2434.0**	**−16.9%**
Embodied Carbon Emissions ($MtCO_2$)	364.9	**326.6**	**−10.5%**
Renewable Generation Capacity (GW)	229.6	**288.2**	**+25.5%**
Renewable Share (% of total capacity)	47.6%	**59.6%**	**+25.2%**
Storage Deployment (GWh)	68.8	**101.2**	**+47.1%**
RES Curtailment (% of RES gen)	13.2%	**8.4%**	**−36.4%**

system planning is conducted independently, accounting for projected power demand growth. In contrast, our proposed method adopts a system-level co-optimization framework that jointly plans both infrastructures.

Table 8 summarizes the comparison results. Both methods are designed to meet the same total power demand and data center service levels, ensuring a fair comparison. The comparison is conducted using a one-year static planning instance based on the PJM system, while the subsequent sections explore the full 15-year dynamic planning outcomes.

As shown in Table 8, the proposed joint planning framework significantly outperforms the non-joint approach. It achieves a 12.6% reduction in total investment cost, an 8.3% reduction in operational cost, and a 5.6% decrease in operational carbon emissions. Furthermore, the coordinated development of power systems and data centers increases the renewable capacity share by over 25%, emphasizing the effectiveness of joint planning in advancing decarbonization objectives.

4.2.2. Impact of Embodied Carbon Consideration in Joint Planning

To evaluate the role of embodied carbon emissions in long-term planning, we compare two variants of the joint planning model: (i) **Joint-OP**, which includes only operational carbon emissions in the objective function; and (ii) **Joint-EM**, which accounts for both operational and embodied emissions. Both scenarios are evaluated under the same 15-year planning horizon and High-Demand settings for the PJM system.

Results in Table 9 highlight that incorporating embodied carbon emissions results in a modest 7.0% increase in total investment cost, primarily due to the preference for low-carbon technologies and infrastructure. However, this investment leads to substantial environmental benefits, including a 16.9% reduction in operational carbon emissions and a 10.5% reduction in embodied emissions.

Moreover, the proposed embodied carbon-aware planning approach promotes higher deployment of renewable energy (up by 25.5%) and storage capacity (up by 47.1%), while simultaneously reducing renewable energy curtailment by over one-third. These results underscore the critical importance of considering embodied emissions in long-term planning to avoid suboptimal investment strategies and to support deep decarbonization goals.

4.3. Analysis of Capacity Expansion Results

In this paper, we consider the planning decisions for the NG units, PV units, WT units, and ES devices, as well as the retirement of CG power plants. In PJM, hydroelectric and nuclear power are also present and are treated as base-load

units in the planning process. We present the construction plan for generation resources across five computing workload demand scenarios: the high-demand scenario, baseline scenario, efficiency-gain scenario, new model surge scenario, and AI bust scenario, reflecting possible AI development trends. The planning results are shown in Fig. 7.

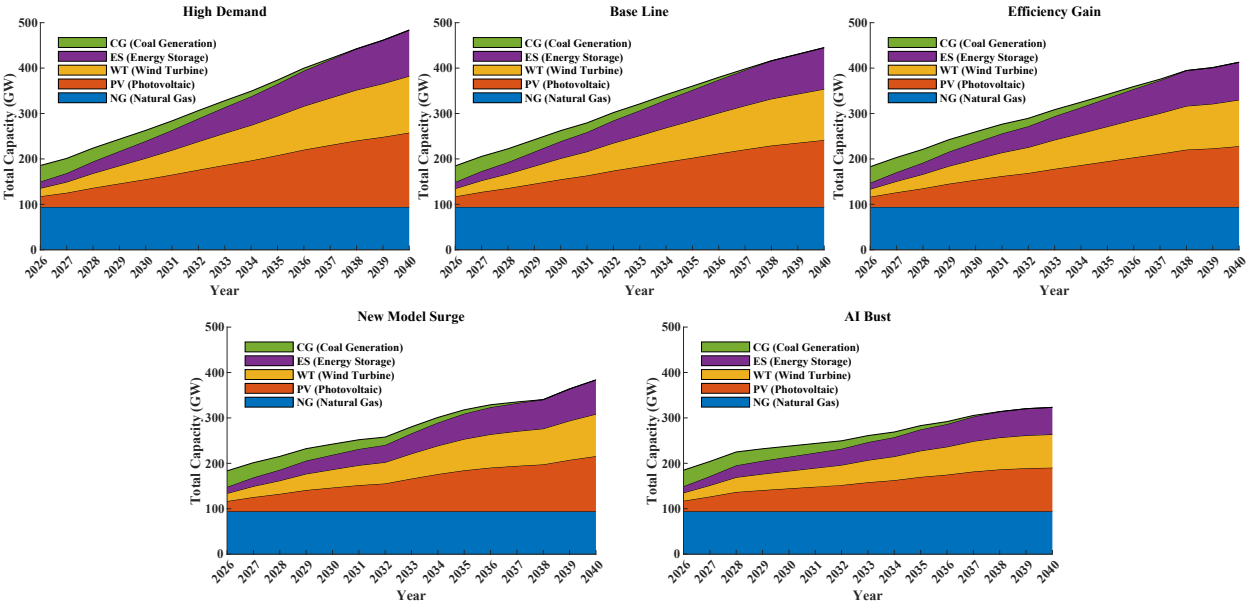

Figure 7: The long-term planning results for generation resources under five computing workload demand conditions.

As illustrated in the Fig. 7, the total installed capacity of RES continues to increase as the planning process progresses, thereby facilitating carbon reduction. Prior to the start of the planning horizon, the total installed capacity of RES (including PV, WT, and ES) was 21.25 GW, accounting for 10.55% of PJM's total generation capacity. In the planning result, under the High-Demand condition, the RES installed capacity reaches 74.36% of the total installed capacity in 2040, with the value of 389.4 GW. This indicates that the carbon reduction objective effectively encourages the installation of RES. Meanwhile, the RES planning results for the AI-bust scenario show an installed RES capacity of 228.92 GW, which is 58.79% of the installed capacity in the High-Demand condition. Majority of the increased computing workload in the High-Demand scenario can be supported by RES in the planning results. This is because the daily profile of computing workload has high correlation of PV generation profile.

The capacity expansion trends for the High-Demand, Baseline, and Efficiency-Gain scenarios exhibit similar patterns due to their comparable load growth trajectories, differing primarily in their growth rates and final peak values. Specifically, in the Baseline scenario, 350 GW of RES resources are installed by 2040, while the Efficiency-Gain scenario leads to 318 GW RES installation by 2040. These two installed capacities represent 89.9% and 81.7% of the RES capacity in the High-Demand scenario, respectively. This discrepancy is primarily driven by variations in AI-related load growth. In contrast, the planning results from the New Model Surge scenario is different from the others. In this scenario, the emergence of advanced AI models leads to a substantial improvement in computing energy efficiency, resulting in temporary reductions in load demand around 2032 and 2038. These demand troughs, in turn, delay the expansion of power generation capacity, culminating in a total RES installation of 289 GW by 2040 in this scenario. Furthermore, the results reveal that among various types of RES, namely PV, WT and ES, the installation capacity ratio is approximately 0.4:0.3:0.3. This distribution is influenced by multiple factors, for example, PV exhibits the highest installation rate due to its economic advantages. WT deployment is largely driven by RPS established by local governments, and the widespread adoption of ES is crucial for mitigating the variability of renewable energy generation.

Another key observation from Fig. 7 is that the installed capacity of NG units remains unchanged across different AI development scenarios. This is primarily due to the fact that, under decarbonization objectives, the addition of new NG units is not environmentally sustainable. While NG units offer operational flexibility and contribute to system resilience,

their long-term viability remains limited within a low-carbon framework. Moreover, the retirement trajectories of CG units exhibit consistent patterns across various AI-driven development scenarios, with annual retirements reaching the maximum threshold of 3 GW. This underscores the urgent need to accelerate the phase-out of high-carbon-emission CG units to achieve decarbonization targets effectively.

4.4. Analysis of Investment Cost

This section presents the annual investment costs associated with generation resources and data centers, along with the operational costs assessed using the Levelized Cost of Electricity (LCOE). The corresponding results are depicted in Fig. 8.

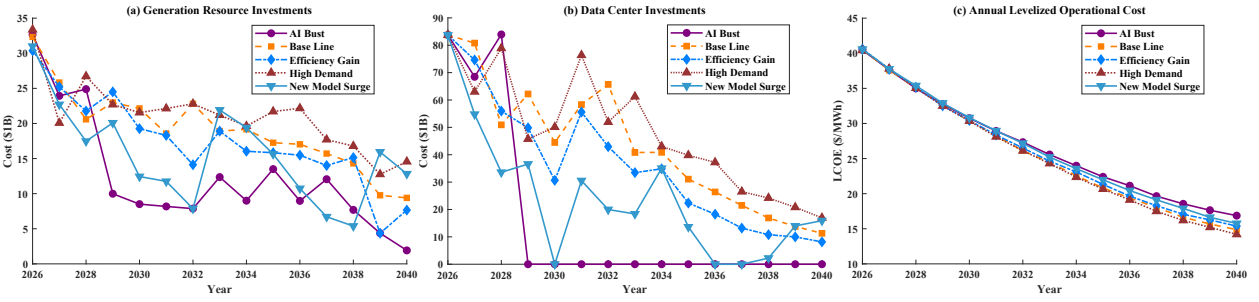

Figure 8: Cost analysis for joint planning, including (a) annual investments in power generation resources, (b) annual investments in data centers, and (c) annual operational costs, evaluated using the LCOE.

The annual investment costs for installing generation resources are illustrated in Fig. 8(a). Since no new NG units are constructed, these costs exclusively represent the investments associated with expanding RESs. Observing the overall trend, it is evident that the annual investment exhibits a declining trajectory across various scenarios. For instance, in the High-Demand scenario, annual investment decreases from 33.36 billion USD in 2026 to 14.56 billion USD in 2040. This decline can be attributed to two primary factors. First, the cost of constructing RES of the same capacity decreases over time due to the consideration of discount factors. Second, the marginal benefits of RES investments diminish as RES capacity expands. Specifically, in the early stages of deployment, the power system possesses sufficient NG units to provide the necessary flexibility to mitigate the uncertainty of RES. During this phase, only a limited amount of ES is required in conjunction with RES expansion, leading to relatively low marginal construction costs. However, in the later stages, a substantial amount of ES must be deployed to accommodate the increasing uncertainty associated with higher RES penetration. Consequently, the marginal construction cost of RES increases, thereby limiting further investment. Furthermore, a comparison across different scenarios reveals that the amount of investment has a clear positive correlation with the development trend of AI demand. For example, the total investment in the High-Demand scenario amounts to 389.4 billion USD, whereas in the AI-Bust scenario, it is 185.9 billion USD, which is approximately half that of the High-Demand case. This highlights the significant impact of AI-driven electricity demand on power generation investments, particularly in renewable-dominated power systems. To reliably meet AI-related computing loads, a substantial expansion of renewable generation capacity is required.

Fig. 8(b) illustrates the investment trends for data centers under various scenarios. It is observed that data center investments exhibit a declining trend, with peak investment occurring at the initial stage. This pattern is primarily attributed to the application of a technological discount on equipment costs, which assumes that as technology advances, the cost of equipment with equivalent computing power decreases over time. Specifically, a discount factor of 0.85 per year is applied. Across different AI development scenarios, investment levels remain relatively stable between 2026 and 2028. However, in the New Model Surge scenario, a significant decline in investment is observed in 2028, likely due to the introduction of advanced AI models that improve resource utilization, thereby reducing the demand for additional computing equipment. Also, the New Model Surge scenario exhibits pronounced investment fluctuations, underscoring the substantial impact of AI model innovations on computing infrastructure investments. In the AI Bust scenario, beginning in 2029, the burst of the AI technology bubble results in a complete cessation of AI-related investments, leading to zero investment in subsequent years. The High-Demand scenario represents a sustained strong demand for AI, where investments in AI computing assets remain consistently high. Under the Baseline scenario, data

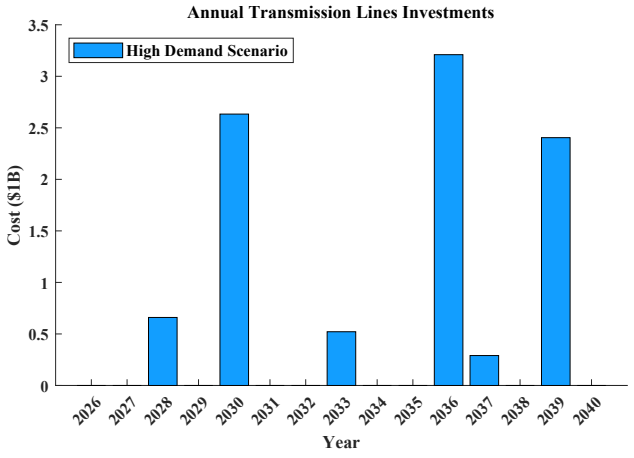

Figure 9: Annual investments for power transmission lines under AI High-Demand Scenario.

center investments begin to decelerate after 2030, while in the Efficiency Gain scenario, the slowdown is even more pronounced. This trend highlights the influence of AI development uncertainty on infrastructure investments.

The LCOE for the entire system is illustrated in Fig. 8(c), providing an assessment of operational costs under various scenarios. A significant reduction in system operating costs is observed over the planning horizon. For instance, under the High-Demand scenario, the LCOE is projected to be 40.42$/MWh in 2026. By the end of the planning period, it declines to 14.2$/MWh, representing a 64.87% reduction. This substantial decrease is primarily attributed to the large-scale deployment of RESs, which supply low-cost, clean electricity. However, it should be noted that PV, WT and ES technologies still incur maintenance costs and are not entirely zero-marginal-cost resources.

The comparison of operating costs across different scenarios reveals relatively minor variations between 2026 and 2030. However, from 2040 onward, the differences become more pronounced. Among the scenarios, the High-Demand case exhibits the lowest operating cost, driven by substantial AI-related electricity demand and the associated carbon emissions from AI infrastructure development, which incentivize the expansion of renewable energy capacity. This expansion, in turn, reduces the system-wide LCOE. In contrast, the AI-Bust scenario experiences the highest operating cost due to the limited deployment of RES. Overall, the substantial decline in system operating costs underscores the effectiveness of integrated energy planning strategies in achieving carbon reduction targets. Furthermore, the results highlight the profound impact of future high AI-related electricity demand on the overall energy system structure.

The growth of both data center and non-data center electric loads has contributed to increased congestion in the transmission network. To analyze this congestion, this study employs a simplified transmission line planning model based on DC power flow. Among all scenarios, the High-Demand scenario exhibits the highest load, resulting in the most significant transmission congestion. Consequently, the corresponding transmission line investment costs for this scenario are presented in Fig. 9 for further analysis. The results in Fig. 9 indicate that transmission line expansion occurs in discrete stages rather than on an annual basis. This pattern suggests that as load demand increases, the transmission network remains congested until an economic equilibrium is reached, at which point large-scale transmission expansions are undertaken. The entire transmission expansion process can be categorized into three distinct stages. In the early stage, a small-scale investment of 0.66 billion USD is made in 2028, followed by a substantial investment of 2.63 billion USD in 2030. In the mid-term stage, after a period of stable system operation, another small-scale expansion of 0.52 billion USD is implemented in 2033, followed by a major expansion of 3.21 billion USD in 2036, marking the final year of this stage. Similarly, in the final stage, an initial small-scale investment of 0.29 billion USD occurs in 2037, followed by a large-scale investment of 2.4 billion USD in 2039. This investment pattern aligns with transmission expansion trends observed in the PJM region and highlights the effectiveness of the joint planning approach proposed in this study. It should be noted that the joint planning model presented here incorporates multiple sectors: electricity production, power transmission, data center, and communication network. Consequently, our simplified DC power flow model and transmission line construction cost model, may not fully capture the complexities of real-world transmission system expansion planning. Nonetheless, the model provides a reasonable representation of transmission planning trends in the PJM interconnection.

4.5. Analysis of Carbon Emissions

This section analyzes the carbon emissions associated with the proposed long-term planning for PJM interconnection, considering both embodied carbon emissions from the construction phase and operational carbon emissions from power generation. Notably, operational carbon emissions in this study are limited to those arising from CG and NG power plants. The results are illustrated in Fig. 10. We first examine the embodied carbon emissions from the construction of power generation resources, as shown in Fig. 10(a). The data is presented in a cumulative annual format. In the early phase (2026–2029), differences across AI development scenarios remain minimal. By 2029, total embodied carbon emissions reach approximately 50 million short tons of CO_2 across all scenarios.

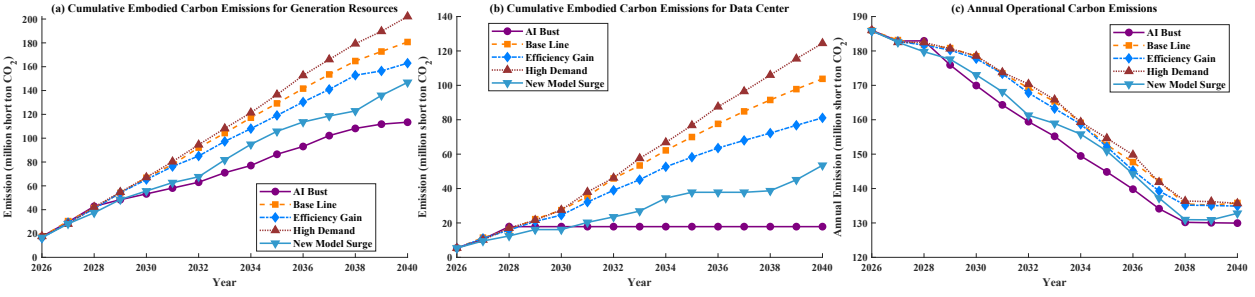

Figure 10: Carbon emission analysis for joint planning, including (a) cumulated embodied carbon emission for power generation resources, (b) cumulated embodied carbon emission for data centers, and (c) annual operational carbon emission of the whole system.

From 2030 onward, the trajectory of embodied carbon emissions diverges across various AI development scenarios. The High-Demand scenario consistently exhibits the highest emissions, driven by substantial growth in AI-related power consumption. This increasing demand necessitates a significant expansion of power generation capacity, particularly from RESs. By 2040, emissions in this scenario peak at 203.6 million short tons of CO_2. In the Baseline scenario, emissions follow a slightly lower trajectory, reaching 178.2 million short tons of CO_2 in 2040. The Efficiency Gain scenario, which assumes reduced RES expansion alongside moderate AI-driven energy consumption growth, results in emissions of 156.7 million short tons of CO_2. Similarly, the New Model Surge scenario, characterized by a more constrained expansion of RES, leads to emissions of 143.8 million short tons of CO_2. In contrast, the AI-Bust scenario follows a distinct trend due to minimal AI-driven power demand growth. Consequently, this scenario exhibits the lowest embodied carbon emissions, reaching 108.6 million short tons of CO_2 by 2040. Here, emissions are primarily driven by the expansion of non-data center power loads rather than AI-related energy consumption.

A comparison between the AI-Bust and High-Demand scenarios highlights the substantial impact of AI-driven load growth on embodied carbon emissions. Specifically, emissions increase from 108.6 million short tons of CO_2 in the AI-Bust scenario to 203.6 million short tons of CO_2 in the High-Demand scenario—an increase of 95 million short tons of CO_2, which corresponds to 87.48% of the total embodied carbon emissions in the AI-Bust scenario. This finding underscores the significant influence of AI-induced load growth on power generation infrastructure planning, leading to considerable embodied carbon emissions. *Notably, this factor is often overlooked in conventional planning approaches, which primarily focus on power system operations rather than infrastructure-related emissions.* Furthermore, this result highlights the effectiveness of the proposed method in addressing such challenges. A detailed analysis of embodied carbon emissions further reveals that approximately 62.32% of these emissions originate from the construction of solar energy resources. This finding provides a critical insight: while PV substantially reduce operational carbon emissions, their deployment introduces significant embodied carbon emissions. Thus, careful consideration of these emissions is essential for achieving truly sustainable energy planning.

The cumulative embodied carbon emissions associated with data center construction are illustrated in Fig. 10(b). To account for technological advancements, a discount factor is introduced to reflect the anticipated reduction in embodied carbon emissions from AI equipment in data centers due to improvements in GPU manufacturing technology. As depicted in Fig. 10(b), the High-Demand, Baseline, and Efficiency Gain scenarios demonstrate a clear upward trend in embodied carbon emissions. This trend is primarily driven by the continuous expansion of data centers to meet increasing AI demand, necessitating the deployment of a large number of GPUs, which have high embodied carbon emissions from manufacturing. In contrast, the New Model Surge scenario follows a three-stage trajectory.

This pattern arises because the development of more efficient AI models reduces the immediate need for computing resource expansion, leading to a phased increase in emissions. Meanwhile, in the AI-Bust scenario, no further data center expansion is expected after 2028, resulting in a stabilization of embodied carbon emissions.

It is crucial to highlight the substantial absolute scale of data center-related embodied carbon emissions. Under the High-Demand scenario, these emissions reached 128.6 million short tons of CO_2 in 2040, accounting for 63.2% of the total accumulate embodied carbon emissions from all power generation resources in the same year. This underscores the significant carbon footprint associated with data center construction, primarily driven by the manufacturing of GPU chips. As a result, the rapid expansion of AI infrastructure poses considerable environmental challenges, emphasizing the need for increased awareness and proactive mitigation strategies.

The results of operational carbon emissions are presented in Fig. 10(c), illustrating an overall downward trend driven by the large-scale integration of RES into the power system. The evolution of operational carbon emissions can be categorized into three distinct stages. In the first stage (2026–2028), operational carbon emissions decline at a slow rate due to the initially low share of RES in the energy mix. In the second stage (2029–2037), emissions experience a significant reduction, following distinct trajectories across different scenarios. The High-Demand scenario exhibits the highest emissions, whereas the AI-Bust scenario records the lowest. This overall decline is primarily attributed to the continuous expansion of RES, while variations across scenarios stem from differences in AI-driven load demand. In the third stage (2038–2040), the rate of reduction in operational carbon emissions slows considerably, as the marginal benefit of additional RES deployment diminishes. This suggests that careful planning of RES expansion is necessary to maximize carbon reduction efficiency. Across all scenarios, the average annual operational carbon emissions amount to 158.06 million short tons of CO_2 per year. This substantial figure underscores the formidable challenge of balancing energy conservation and decarbonization amid the rapid growth of AI-driven demand.

4.6. Data Center Construction Plan

This section examines the geographical distribution of data center construction under different AI development scenarios. The final year of the planning horizon (i.e., 2040) is selected as a representative year for analysis, with the planning results depicted in Fig. 11. For analytical purposes, the size of data centers is expressed in terms of their equivalent electrical demand capacity. Additionally, data centers are assumed to be located at the geometric centers of the PJM transmission zones, as illustrated in the figure.

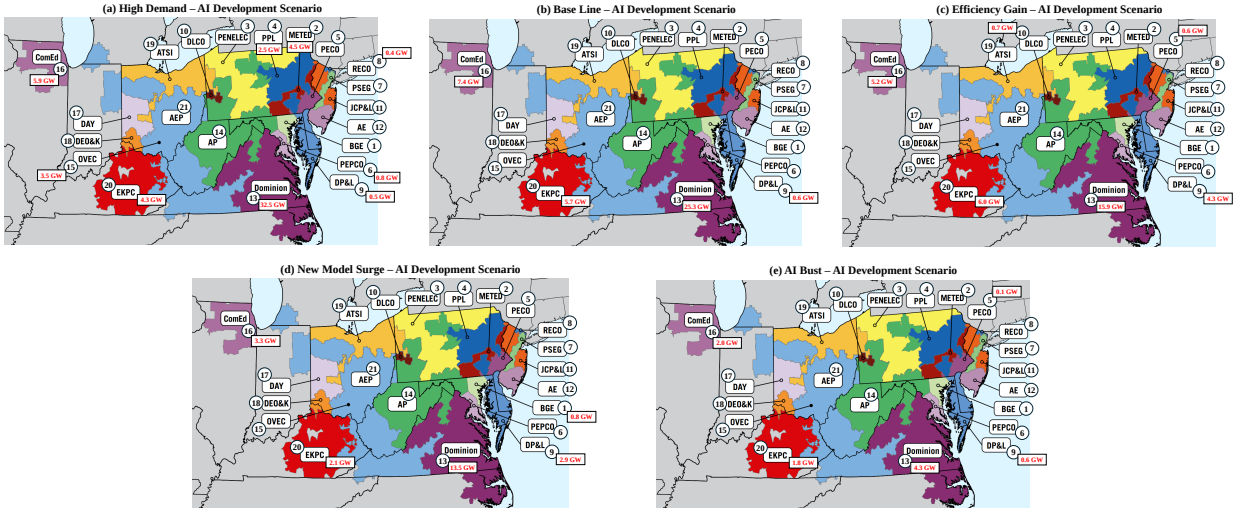

Figure 11: The locational planning results for data centers in 2040 under five AI demand scenarios, the map is obtained from PJM [37].

Examining the High-Demand scenario in Fig. 11(a), data centers are distributed across nine PJM transmission zones, with Dominion (DOM) exhibiting the highest concentration at 32.5 GW. This is followed by ComEd (5.9 GW), METED (4.5 GW), EKPC (4.3 GW), and OVEC (3.5 GW), while the remaining five zones each accommodate less than 3 GW. A notable observation from these planning results is the significantly higher data center capacity in DOM

compared to other regions. Notably, the geometric center of DOM is near Richmond, Virginia's capital, which is already one of the largest data center hubs in the United States. This strong alignment between the model's projections and real-world data center distribution supports the validity and effectiveness of the proposed planning framework.

More specifically, the High-Demand scenario represents the upper bound of AI-related load growth, suggesting that the projected 32.5 GW data center load in the DOM region can be interpreted as an upper limit of its carrying capacity. In comparison, the current data center load in DOM is estimated to be between 2 and 5 GW [39], indicating significant potential for further expansion. Three key factors enable DOM to support large-scale data center deployment:

1. *Abundant Energy Resources:* According to the planning results, DOM has a substantial supply of RES and NG generation units, which contribute to lower local electricity prices and reduced carbon emissions.
2. *Robust Communication Infrastructure:* The DOM region serves as a critical hub for multiple submarine cables and optical fiber networks, providing high-capacity, low-latency connectivity essential for AI-driven applications.
3. *Well-Integrated Power Transmission Network:* The DOM region is well-connected to surrounding transmission lines, ensuring that large-scale data center development does not result in significant power system congestion.

These factors collectively reinforce DOM's strategic advantage as a major data center hub, aligning with both the model's projections and real-world trends. Furthermore, ComEd (north Illinois), which ranks second in data center capacity according to the planning results, is also a major data center hub within the current PJM region. This alignment between the model's projections and real-world data center distribution further validates the effectiveness of the proposed planning approach.

Examining the Base-Line scenario in Fig. 11(b), data centers remain primarily concentrated in the DOM region, reaching a capacity of 25.3 GW, followed by ComEd (7.4 GW), EKPC (5.7 GW), and DP&L (0.6 GW). Compared to the High-Demand scenario, the Base-Line scenario exhibits a more concentrated distribution, with data center capacity predominantly allocated within the three transmission zones of DOM, ComEd, and EKPC. Similarly, analyzing the planning results for the Efficiency Gain, New Model Surge, and AI Bust scenarios in Fig. 11(c)–(e), it is evident that data center development remains concentrated in DOM, ComEd, and EKPC. Among these, DOM consistently demonstrates the highest construction capacity, while ComEd and EKPC exhibit comparable data center capacities. These findings highlight the strategic importance of DOM, ComEd, and EKPC in data center planning, suggesting that in practical investment decisions, these three locations should be prioritized for large-scale data center development.

5. Conclusion and Discussion

This paper presents a long-term, dynamic joint planning framework from the perspective of system operators (e.g., ISOs), aiming to coordinate the development of power systems and data centers to achieve deep carbon emission reductions. The proposed framework encompasses comprehensive planning decisions across both infrastructure types. On the power system side, it determines the location, capacity, and retirement of thermal units, RESs and storage batteries, as well as the sizing of power transmission lines. On the data center side, it optimizes the location, capacity, and type of data centers, taking into account locational carbon intensity, electricity prices, and cooling efficiency. In addition, communication network investments are incorporated to ensure connectivity across facilities.

From a decarbonization standpoint, the model accounts for both operational and embodied carbon emissions, including those associated with infrastructure construction and GPU manufacturing. The resulting large-scale stochastic optimization problem is efficiently solved using an enhanced Benders decomposition algorithm. The framework is applied to a real-world case study on the PJM system, with all datasets curated and publicly released via GitHub [26] to support related research. The case study yields several key empirical insights. First, the PJM system can support up to 55 GW of peak data center demand, with the DOM (Virginia) and ComEd (Northern Illinois) zones identified as the most suitable hosting regions. Second, compared to independent data center planning, joint planning reduces total investment costs by 12.63%, operational costs by 8.25%, and total carbon emissions by 5.63%. Third, when embodied carbon emissions are considered, renewable capacity deployment increases by 25.5%, and operational carbon emissions are reduced by 16.9%, highlighting the importance of life-cycle carbon accounting in planning for deep decarbonization.

While this work provides a comprehensive planning framework, several avenues remain open for future research. First, although this study adopts a system operator-centric perspective, effective coordination mechanisms between

system operators and data center investors are needed. Future work may explore market-based mechanisms to incentivize or enforce alignment with the joint planning strategy. Second, the modeling of communication network infrastructure is simplified by allocating fixed investment shares based on data center size, which can be extended by developing a more detailed and realistic model of data communication planning. Third, the generation resources in this study are modeled as continuous units. In practice, generators operate with discrete output levels governed by unit commitment decisions. Directly integrating unit commitment into the planning model would improve realism but impose significant computational burden. Future work could explore more efficient formulations to incorporate such operational constraints.

Acknowledgment

This work was supported by the National Science Foundation under award #2324940.

References

[1] Sirui Chen, Peng Li, Haoran Ji, Hao Yu, Jinyue Yan, Jianzhong Wu, and Chengshan Wang. Operational flexibility of active distribution networks with the potential from data centers. *Applied Energy*, 293:116935, 2021.

[2] Osten Anderson, Mikhail A Bragin, and Nanpeng Yu. Optimizing deep decarbonization pathways in California with power system planning using surrogate level-based lagrangian relaxation. *Applied Energy*, 377:124348, 2025.

[3] Dongxiang Yan, Mo-Yuen Chow, and Yue Chen. Low-carbon operation of data centers with joint workload sharing and carbon allowance trading. *IEEE Transactions on Cloud Computing*, 2024.

[4] Ana Radovanović, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, et al. Carbon-aware computing for datacenters. *IEEE Transactions on Power Systems*, 38(2):1270–1280, 2022.

[5] Hao Wang, Jianwei Huang, Xiaojun Lin, and Hamed Mohsenian-Rad. Exploring smart grid and data center interactions for electric power load balancing. *ACM SIGMETRICS Performance Evaluation Review*, 41(3):89–94, 2014.

[6] Eirik Resch, Carine Lausselet, Helge Brattebø, and Inger Andresen. An analytical method for evaluating and visualizing embodied carbon emissions of buildings. *Building and Environment*, 168:106476, 2020.

[7] Fang Cao, Yajing Wang, Feng Zhu, Yujie Cao, and Zhaohao Ding. UPS node-based workload management for data centers considering flexible service requirements. *IEEE Transactions on Industry Applications*, 55(6):5533–5542, 2019.

[8] Liang Yu, Tao Jiang, and Yulong Zou. Distributed real-time energy management in data center microgrids. *IEEE Transactions on Smart Grid*, 9(4):3748–3762, 2016.

[9] Zhaohao Ding, Liye Xie, Ying Lu, Peng Wang, and Shiwei Xia. Emission-aware stochastic resource planning scheme for data center microgrid considering batch workload scheduling and risk management. *IEEE Transactions on Industry Applications*, 54(6):5599–5608, 2018.

[10] Soongeol Kwon, Lewis Ntaimo, and Natarajan Gautam. Demand response in data centers: Integration of server provisioning and power procurement. *IEEE transactions on Smart Grid*, 10(5):4928–4938, 2018.

[11] Shahab Bahrami, Vincent WS Wong, and Jianwei Huang. Data center demand response in deregulated electricity markets. *IEEE Transactions on Smart Grid*, 10(3):2820–2832, 2018.

[12] Ana Radovanovic, Bokan Chen, Saurav Talukdar, Binz Roy, Alexandre Duarte, and Mahya Shahbazi. Power modeling for effective datacenter planning and compute management. *IEEE Transactions on Smart Grid*, 13(2):1611–1621, 2021.

[13] Zhentong Shao, Qiaozhu Zhai, Zhihan Han, and Xiaohong Guan. A linear AC unit commitment formulation: An application of data-driven linear power flow model. *International Journal of Electrical Power & Energy Systems*, 145:108673, 2023.

[14] Zhentong Shao, Qiaozhu Zhai, Yingming Mao, and Xiaohong Guan. A method for evaluating and improving linear power flow models in system with large fluctuations. *International Journal of Electrical Power & Energy Systems*, 145:108635, 2023.

[15] Xiaoyu Cao, Jianxue Wang, and Bo Zeng. Networked microgrids planning through chance constrained stochastic conic programming. *IEEE Transactions on Smart Grid*, 10(6):6619–6628, 2019.

[16] Fengjuan Wang, Chengwei Lv, and Jiuping Xu. Carbon awareness oriented data center location and configuration: An integrated optimization method. *Energy*, 278:127744, 2023.

[17] Victor Depoorter, Eduard Oró, and Jaume Salom. The location as an energy efficiency and renewable energy supply measure for data centres in Europe. *Applied Energy*, 140:338–349, 2015.

[18] Tamar Eilam, Pradip Bose, Luca P Carloni, and et al. Reducing datacenter compute carbon footprint by harnessing the power of specialization: Principles, metrics, challenges and opportunities. *IEEE Transactions on Semiconductor Manufacturing*, 2024.

[19] Caishan Guo, Fengji Luo, Zexiang Cai, Zhao Yang Dong, and Rui Zhang. Integrated planning of internet data centers and battery energy storage systems in smart grids. *Applied Energy*, 281:116093, 2021.

[20] Ali Vafamehr, Mohammad E Khodayar, Saeed D Manshadi, Ishfaq Ahmad, and Jeremy Lin. A framework for expansion planning of data centers in electricity and data networks under uncertainty. *IEEE Transactions on Smart Grid*, 10(1):305–316, 2017.

[21] Bo Zeng, Yinyu Zhou, Xinzhu Xu, and Danting Cai. Bi-level planning approach for incorporating the demand-side flexibility of cloud data centers under electricity-carbon markets. *Applied Energy*, 357:122406, 2024.

[22] Weiwei Li, Tong Qian, Yin Zhang, Yueqing Shen, Chenghu Wu, and Wenhu Tang. Distributionally robust chance-constrained planning for regional integrated electricity–heat systems with data centers considering wind power uncertainty. *Applied energy*, 336:120787, 2023.

[23] Jinhui Liu, Zhanbo Xu, Jiang Wu, Kun Liu, Xunhang Sun, and Xiaohong Guan. Optimal planning of internet data centers decarbonized by hydrogen-water-based energy systems. *IEEE Transactions on Automation Science and Engineering*, 20(3):1577–1590, 2022.

[24] Da Xu, Shizhe Xiang, Ziyi Bai, Juan Wei, and Menglu Gao. Optimal multi-energy portfolio towards zero carbon data center buildings in the presence of proactive demand response programs. *Applied Energy*, 350:121806, 2023.

[25] Wenbo Qi, Jie Li, Yaoqing Liu, and Chen Liu. Planning of distributed internet data center microgrids. *IEEE Transactions on Smart Grid*, 10(1):762–771, 2017.

[26] Zhentong Shao and Nanpeng Yu. A 21-bus system for PJM. GitHub repository, 2025. Available: https://github.com/ZTSshao123/PJM-21-bus-system.

[27] Zachary Zimmerman, Dinos Gonatas, Anjali Patel, and Rob Gramlich. Transmission planning for PJM's future load and generation: Version 1, May 2024. Available: https://cleanenergygrid.org/portfolio/transmission-planning-for-pjms-future-load-and-generation/.

[28] Tasnim Ibn Faiz and Md Noor-E-Alam. Data center supply chain configuration design: A two-stage decision approach. *Socio-Economic Planning Sciences*, 66:119–135, 2019.

[29] R Timothy Marler and Jasbir S Arora. The weighted sum method for multi-objective optimization: New insights. *Structural and multidisciplinary optimization*, 41:853–862, 2010.

[30] RGGI. Regional Greenhouse Gas Initiative Program Review: Notes on Modeling Materials, September 2024. Available: https://www.rggi.org/sites/default/files/Uploads/Program-Review/2024/Third_Program_Review_Update_9-23-2024.pdf.

[31] Isabelle Riu, Dieter Smiley, Stephen Bessasparis, and Kushal Patel. Load Growth is Here to Stay, but Are Data Centers?: Strategically Managing the Challenges and Opportunities of Load Growth. *Energy and Environmental Economics, Inc.*, July 2024.

[32] Zhentong Shao, Xiaoyu Cao, Qiaozhu Zhai, and Xiaohong Guan. Risk-constrained planning of rural-area hydrogen-based microgrid considering multiscale and multi-energy storage systems. *Applied Energy*, 334:120682, 2023.

[33] Xiaoyu Cao, Jianxue Wang, and Zhong Zhang. Multi-objective optimization of preplanned microgrid islanding based on stochastic short-term simulation. *International Transactions on Electrical Energy Systems*, 27(1):e2238, 2017.

[34] *HOMER Pro 3.14 User Manual*, 2020. Available: https://www.homerenergy.com/pdf/HOMERHelpManual.pdf.

[35] Han Yu, Chi Yung Chung, Kit Po Wong, Heung-Wing Joseph Lee, and Jianhua Zhang. Probabilistic load flow evaluation with hybrid latin hypercube sampling and Cholesky decomposition. *IEEE Transactions on Power Systems*, 24(2):661–667, 2009.

[36] U.S. Energy Information Administration. Energy Infrastructure and Resources Maps, 2022. Available: "https://atlas.eia.gov/pages/energy-maps".

[37] PJM Interconnection. PJM Maps Library, 2025. Available: https://www.pjm.com/library/maps.

[38] PJM Interconnection. PJM Data Miner 2, 2025. Available: https://dataminer2.pjm.com/list.

[39] CBRE. North America Data Center Trends H1 2023, September 2023. Available: https://www.cbre.com/insights/reports/north-america-data-center-trends-h1-2023.

[40] U.S. Energy Information Administration (EIA). Capital cost and performance characteristics for utility-scale electric power generating technologies. January 2024. Available: https://www.eia.gov/analysis/studies/powerplants/capitalcost/.

[41] Daniel DeSantis, Brian D James, Cassidy Houchins, Genevieve Saur, and Maxim Lyubovsky. Cost of long-distance energy transmission by different carriers. *IScience*, 24(12), 2021.

[42] XD Wu, JL Guo, and GQ Chen. The striking amount of carbon emissions by the construction stage of coal-fired power generation system in China. *Energy Policy*, 117:358–369, 2018.

[43] Laura Daniels, Phil Coker, and Ben Potter. Embodied carbon dioxide of network assets in a decarbonised electricity grid. *Applied Energy*, 180:142–154, 2016.

[44] Antonio Augusto Morini, Dachamir Hotza, and Manuel J Ribeiro. Embodied energy and carbon footprint comparison in wind and photovoltaic power plants. *International Journal of Energy and Environmental Engineering*, pages 1–11, 2022.

[45] Thomas Le Varlet, Oliver Schmidt, Ajay Gambhir, Sheridan Few, and Iain Staffell. Comparative life cycle assessment of lithium-ion battery chemistries for residential storage. *Journal of Energy storage*, 28:101230, 2020.

[46] Ahmad Faiz, Sotaro Kaneda, Ruhan Wang, Rita Osi, Prateek Sharma, Fan Chen, and Lei Jiang. LLMCarbon: Modeling the end-to-end carbon footprint of large language models. *arXiv preprint arXiv:2309.14393*, 2023.