Utilities Policy 84 (2023) 101636

Contents lists available at ScienceDirect

Utilities Policy

journal homepage: www.elsevier.com/locate/jup

Full-length article

Facilitating the provision of load flexibility to the power system by data centers: A hybrid research method applied to China

Yujie Cao *, Sufang Zhang

School of Economics and Management, North China Electric Power University, China

ARTICLE INFO

Handling editor: Janice A. Beecher

Keywords:
Data center
Load flexibility
Power system

ABSTRACT

With the increasing penetration of renewables, load flexibility is crucial to balance the power system in China. Data centers are rapidly growing energy consumers and their flexible operations could make them a potential source of load flexibility. However, there have been no large-scale attempts to apply this concept in China. This paper utilizes a hybrid research method to explore the flexibility potentials, identifying key technical, financial, policy, and institutional barriers. The study quantifies the flexibility potential of data centers in China and provides solutions for addressing the identified solutions.

1. Introduction

Over the last two decades, China has undergone a swift transition towards low-carbon energy, particularly on the supply side of the power system. Moreover, since China was committed in Sept. 2020 to peak its carbon emissions by 2030 and achieve carbon neutrality by 2060, it is moving towards building a new power system with new energy as the main body, and the proportion of renewable energy is growing steadily. By Feb. 2023, the cumulative installed capacity of wind and solar power in the country has reached 371 GW and 413 GW, respectively, accounting for 14.3% and 15.9% of the national total installed capacity, an increase of approximately three hundred times the level in 2005 (National Energy Administration released January 2023). It is estimated that the ratio of renewable energy capacity will exceed 33% by 2025 (Yin and Yep, 2022).

Given the uncertainty and intermittency of renewable energies, the temporal and spatial mismatch between power demand and supply has become the primary challenge for the operational security of power systems. Therefore, the ability to respond to variability, i.e., power system flexibility, is the most fundamental element and cornerstone to maintaining the real-time system balance (IEA. China, 2019; Ye et al., 2018). Among various flexibility resources available in power systems, demand response is seen as both a cost-effective and climate-friendly option to address integration challenges, as it can reduce the need for new controllable capacities, such as coal and gas-fired power units (Bouckaert et al., 2018). In recent years, the increasing penetration of microgrids and the advent of smart grids have diversified demand response techniques.

Meanwhile, we are currently in an era of the digital economy, and the demand for digital services has seen sustained growth, leading to a subsequent increase in both the scale and energy consumption of data centers - the underlying infrastructure for digital services. As of 2022, China had approximately 5.9 million installed data center racks (Wang, 2022), which are prefabricated with slots to house servers, uninterruptible power supplies (UPS), networks, and cooling systems. As shown in Fig. 1 (Data Center Whitepaper, 2022), these racks are densely located in provinces/municipalities with digitally-driven economies, such as Hebei, Jiangsu, Zhejiang, Guangdong, Guizhou, and Shanghai, exceeding half of the nation's total capacity. As a result, the power consumption of data centers in the country surged to approximately 216.6 TWh by 2021 (Jiang et al., 2022), accounting for 2.6% of China's total power consumption (National Energy Administration, 2022), and is expected to reach 450.5 TWh (Ye et al., 2021), exceeding 3.6% of the country's total by 2035 (Yang, 2022). Fig. 2 presents China's total power consumption and the data centers' portion between 2015 and 2020.

Technically, data centers have the potential to provide load flexibility to the power system while ensuring acceptable service quality for their clients. For example, some workloads within data centers are tolerant of delays, enabling the power load of servers to be accurately adjusted in real time. Moreover, the thermal capacity and inertia of cooling systems allow data centers to shift the power demand for cooling; the UPS can serve as an energy storage system (ESS). Other than that, the power consumption of data centers is continuously monitored and managed by advanced systems in fine granularity. These

* Corresponding author.
E-mail addresses: yujiecao@ncepu.edu.cn (Y. Cao), zsf69826313@sina.com (S. Zhang).

https://doi.org/10.1016/j.jup.2023.101636
Received 14 November 2022; Received in revised form 10 July 2023; Accepted 11 July 2023
Available online 23 July 2023

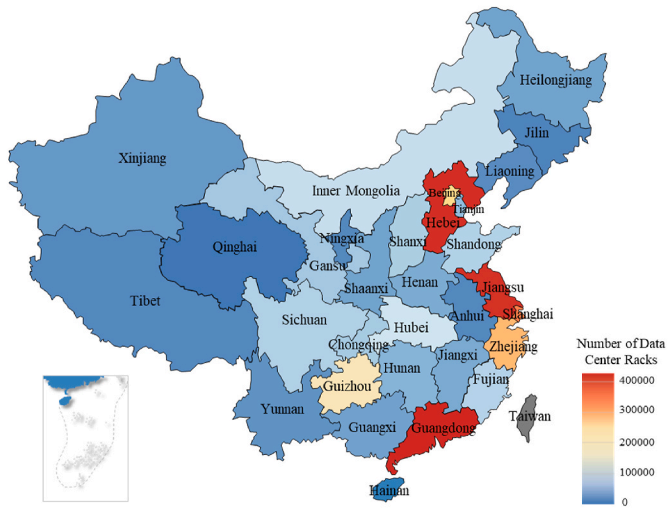

Fig. 1. Distribution of data center racks in China.

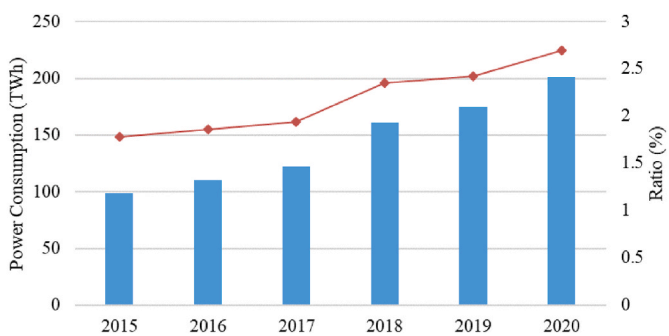

Fig. 2. Total and the ratio of data center power consumption in China: 2015-2020
Source: Annual online news reports on data center power consumption and national power
consumption.

characteristics, coupled with rapidly growing power loads, make data centers promising candidates for demand response in the power system. In addition, their digital connectivity can expand their theoretical demand-size flexibility potentials into a greater share. Given the high-speed transmission, their inherent flexibility enables them to cooperate with the power system and spatial-temporal availability of renewable energies with digital connectivity.

In recent years, the potential of power load flexibility of data centers in the operational scenarios of the power system has been a hot academic issue. Some studies have proposed several frameworks considering various constraints to evaluate the load flexibility of data centers. These include temporal and spatial workload scheduling models (Chen et al., 2019; Yu et al., 2015; Luo et al., 2013), optimal dispatch models for cooling systems, and UPS (Chen et al., 2016; Ali and Ozkasap, 2016; Liu et al., 2012, 2016; Cao et al., 2019) that were developed for a variety of power system services, including frequency regulation (Börjeson and Östbom, 2018; Gao et al., 2019; Li et al., 2013), peak-shaving regulation (Li et al., 2014; Liu et al., 2011a, 2013), and congestion management (Chen et al., 2021; Liu et al., 2018; Wang et al., 2020a).

Regarding practical applications, a team at Lawrence Berkeley National Laboratory conducted experiments to explore demand response opportunities of four data centers, and it was reported that the flexible load in data centers could account for 40% of the overall power load (Ghatikar et al., 2009). Besides, Google deployed a "Carbon-Intelligent Computing" platform that utilized the temporal flexibility of workloads on production data center clusters and shifted workloads according to the availability of hourly carbon-free energy (Radovanovic, 2020). In China, Alibaba has successfully conducted small-scale trials (Alibaba and North China Electric, 2022). For instance, in June 2022, a fraction of the workload was shifted from Nantong in Jiangsu Province in east China, where one of Alibaba's data centers is located, to Zhangbei in Hebei Province in the Northern China power grid, where one of China's biggest wind farms is located. As a result, approximately 150 kWh of electricity is generated in the data center located in Zhangbei, leading to an increase in the utilization of renewable energy in the Northern China power grid during this period and a reduction of 120 kg of carbon dioxide emissions.

We observed that while data centers have been regarded as promising flexible resources on the demand side of the power system in numerous academic models, few are relevant to China. Although, as noted above, Alibaba has had a successful experience in this regard, it must be acknowledged that implementing such an experiment is still difficult from any perspective, let alone widely applying it to data centers. These observations drive three main research questions (RQ) in this study:

RQ1. What is the potential of load flexibility in China's data-center sector?

RQ2. What are the key barriers for data centers to provide load flexibility to the power system in China?

RQ3. What are the possible solutions to the barriers for China's data centers to provide load flexibility to the power system?

This study investigates ways to enable data centers in China to provide load flexibility to the power system to enhance sustainable development in both the data-center and power sectors. The contributions of this work could be summarized as follows:

1. In order to obtain a comprehensive understanding of the situation in China, we employed a hybrid research method, including a thorough literature review, a questionnaire, and semi-structured interviews.
2. In this study, we provide the first quantitative assessment of the power flexibility potentials of data centers in China based on available data and publications, which have not been previously assessed in the existing literature.
3. Due to the multifaceted nature of the research problem, which involves technical, market, policy, and institutional dimensions, the experts interviewed during our research are from diverse fields with sufficient working backgrounds and knowledge, differentiating this work from other research.

The analysis commences with an introduction to our research design, outlined in Section 2, providing a comprehensive description of each of the three research methods noted above. In response to RQ1-RQ3, research findings are detailed in Sections 3-5. Specifically, Section 3 illustrates the theoretical potentials of load flexibility in China's data-center sector; Section 4 identifies the key barriers for China's data centers to provide load flexibility to the power system; Section 5 provides possible solutions to the barriers identified in Section 4. The final section presents the conclusions.

2. Research methods: literature review, questionnaire, and interview

The analytical approach involves a literature review, an offline questionnaire, and semi-structured interviews. To provide a comprehensive perspective on the load-shifting potentials of data centers from theoretical standpoints, we first adopt a literature review approach to analyze flexibility resources available within data centers and relevant policies supporting their applications in the power sector through reviewing relevant documentation. Then, we employ offline questionnaires and semi-structured interviews to address our second and third research questions.

2.1. Literature review

We present an overview of flexibility resources within data centers drawing from relevant academic papers and engineering reports. Furthermore, related government documents concerning data-center load flexibility implementation are analyzed. Supplementing these, information from think-tank publications and press articles is also reviewed to enhance our comprehensive understanding of the latest developments and pressing needs within the data-center sector. Table 1 provides a detailed list of the databases, journals, and websites we used.

2.2. Questionnaire design

We conducted online surveys via questionnaire in November 2020 by directly contacting 13 experts with diverse professional backgrounds, including the energy sector, data-center sector, government administrations, and academic institutes. The purpose of the survey was to complement the analysis of existing research documents that have not been fully explained with practical knowledge and experience and to enrich the research validity from multiple perspectives. The experts interviewed were selected on their experiential relevance; all are working or have worked in related fields. The experts' affiliations and job responsibilities are listed in Table 2. All the surveys were conducted on a voluntary and confidential basis.

The scope of questionnaires was set based on the experts' understandings of the three research questions, and the detailed questions (Q) were summarized, as shown in Table 3, into four main topics: the state-of-art related policies and regulations, potentials of data center load flexibility and risk control, current applications in China, and possible influencing factors and incentives of data center flexibility application and provision.

2.3. Interview design

Semi-structured interviews were adopted to encourage open dialogues with experts to illustrate further their views on in-depth follow-up questions based on the questionnaires. Due to the Covid-19 pandemic, the interviews were conducted individually through video conferences in November 2021, which were recorded and transcribed promptly after the interview. 11 respondents who had worked in the data-center sector for more than five years or as policymakers or think-tank academics were selected from the questionnaire respondents to participate these follow-up conferences.

In order to ensure consistency across different interviews, a transcript for the interview was prepared in advance, covering the background and context of the research, the purpose of the interviews, and a list of key themes to be explored. While the interviews centered on predesigned questions, interviewees were encouraged to provide additional relevant insights based on their professional backgrounds. The interviews were open-ended and avoided directing questions. The main interview themes (I) explored include:

- I1 What is currently being undertaken in data centers in terms of energy management? Based on the need to innovate data-center utilities to provide load flexibility, will these be 'business-as-usual' operation scenarios?
- I2 Is high institutional liquidity on data-center investment project arrangements conducive to integrating data-center flexibility into power systems? If so, what process is involved?
- I3 How does the flexibility of large-scale data center clusters manifest itself under a cross-multi-sector framework with various regulations? To the best of your knowledge, what process/workflow is involved?

The key findings from our hybrid-method survey are presented in sections 3-5. The results of the literature review cite corresponding papers or reports as supporting evidence. The views from anonymous interviewees are directly quoted with minimal editing.

Table 1
Sources for literature search.

Types	Sources
Database	IEEEXplore, Science Direct, ACM
Journal	Energy Policy, Energy Research & Social Science, IEEE Transactions on Smart Grid, Applied Energy; People News, Xinhua News
Website (for report)	Greenpeace, Lawrence Berkley National Laboratory, GIZ, China Academy of Information and Communications Technology (CAICT)
Website (for government document)	Ministry of Industry and Information Technology (MIIT), National Energy Administration (NEA), National Development and Reform Commission (NDRC), Ministry of Natural Resource

Table 2
Overview of interviewees.

Sector	Affiliation	Responsibility	Nos.
Energy utilities	State Grid Integrated Energy Service Co. LTD.	Power transmission, distribution, and grid planning.	2
	Jibei Electric Power Trading Center Co. LTD	Design of electricity market laws, regulations, and membership information. Main platform and management organization for regional power market transactions.	
Data-center utilities	Alibaba Inc.	Internet data center construction and workload management. Providing cloud computing services.	6
	Huawei Inc.	5G infrastructure construction and management.	
	Tencent Inc.	Internet data-center infrastructure construction and management.	
	Global Data Solutions LTD (GDS)	Third-party Data center infrastructure construction, server hosting and management, cloud connectivity services, and internal data center interconnection services.	
Government	China Electricity Council (CEC)	Provide services to enterprises and institutions in the energy sector and assist the National Energy Administration in electricity industry management.	2
	Ministry of Industry and Information Technology (MIIT)	Industrial sector and communications sector management. Providing support for designing policies, strategies, and standards and implementing plans and programs.	
Research institutions	Beijing Institute of Technology	Research on 5G networks and management.	3
	North China Electric Power University	Research on data center flexibility and coordination between data center and power systems.	
	Zhejiang University	Research on electricity market policies.	

3. Potentials of load flexibility in China's data-center sector

Most literature and all the interviewees agreed on the general insight that reducing energy consumption and carbon emissions necessitates efficient and flexible energy use in data centers. Overall, the load

Table 3
Questionnaire design.

RQ #	Topic	Detailed Questions
RQ1	Definition of flexible operation for data center Potentials of data center load flexibility and risk control	Q1.1 Why is it important for data centers to be flexible?
		Q1.2 How much do you know about the flexible power consumption of data centers?
		Q1.3 To the best of your knowledge, are there any strategies for flexibility during the daily operation of data centers?
		Q1.4 Do any domestic data centers ever manage their energy consumption? For what purpose?
RQ2	Key barriers to flexibility integration	Q2.1 How do you see the potential of the coordinative operation between data centers and power systems?
		Q2.2 No domestic data center participates in power system operation or for better renewable energy integration in practice. What are the main barriers to that?
		Q2.3 Are there any data center industries in foreign countries that have managed to participate in the power market and power system operation? In addition, what are their business models?
RQ3	Related policies and regulations	Q3.1 Relevant policies are required to enable data centers to participate in the power market and power system operation. Which parts or what aspects should the policy cover?
		Q3.2 To your best knowledge, are there any other policies that will indirectly affect data center flexibility integration?
		Q3.3 Regarding the ongoing policies/strategies, do you feel that government/regulators would like to provide a positive environment for providing data center load flexibility and its application in power systems? Why/why not?
	Possible incentives	Q3.4 What are those effective incentives that you think could motivate data center industries to provide load flexibility?
		Q3.5 What do you see as possible incentives to integrate data centers' load flexibility into power systems?

flexibility of data centers during operation can be categorized into four groups: flexible shifting workloads, optimizing IT equipment utilization, flexible charging strategies of UPS, and operational robustness of cooling systems.

3.1. Spatial and temporal workload migration

The workloads processed by data centers are programs that execute the computing demand of clients. As widely accepted in research and practice, workload execution can be flexible regarding various time requirements. Since workloads directly affect the power load of data centers, workload execution flexibility also supports the temporal and spatial flexibility of data centers in power consumption.

Generally, workloads could be classified into interactive workloads and batch workloads. Interactive workloads, such as web browsing and online payment transaction, are typically light in computation, requiring a small set of computational resources. Consequently, they can be easily allocated to any server connected to the same internet while guaranteeing their basic service demands. As a result, the execution of interactive workloads offers considerable spatial flexibility. Several papers have suggested utilizing this potential in response to geographically differentiated electricity prices (Dou et al., 2015; Goudarzi and Pedram, 2013; Hogade et al., 2018), renewable energy generation (Liu et al., 2011a, 2011b; Khosravi et al., 2017), and carbon density (Xiang, 2014; Guo et al., 2014) to reduce energy costs and greenhouse gas emissions.

Batch workloads, such as data collection and machine learning, are generally computation- and data-intensive, which are highly resource-demanding. As such, longer execution time is acceptable to clients of batch workloads, enabling data center operators to shift the execution process of batch workload to other time intervals with lower electricity prices (Luo et al., 2013; Chen et al., 2014; Dou et al., 2017; Guo and Fang, 2013) to reduce renewable energy curtailment (Islam et al., 2014; Ren, 2013; Wang et al., 2020b), assist balancing energy (Chen et al., 2021) or resolve grid bottlenecks such as transmission congestions (Liu et al., 2018; Wang et al., 2020a).

Workload scheduling enables resource reallocation. Initially, the location of data centers in China was mainly determined by regulatory compliance, social capital, client proximity, workforce, and tax incentives during the early stage of sector development. As a result, the distribution of data centers in China exhibits highly imbalanced features in geographic locations, as shown in Fig. 1. China has since launched a promising national strategy to relocate data centers hosting non-real-time workloads to Western regions in "Eastern Data and Western Computing" design by 2021 (Implementation plan for computing power, 2021). This strategy is expected to balance regional disparities in the economy and social development while reducing the energy consumption of the data-center sector in the East and improving renewable energy integration in the West.

All professionals who participated in questionnaires and interviews agreed on the flexibility of resources and their feasibility. An academic professional, who has dedicated several years to research in this field and its industrial practice, mentioned their participation in the Alibaba workload migration project reported in (Alibaba and North China Electric, 2022). They stated, "*Thanks to the powerful workload management systems, we successfully experimented with interactive workload migration between data centers in Jiangsu and Hebei with the collaboration of Alibaba Group. The results showed that spatial migration of workloads is practical in production data center clusters, and the power load profiles change accordingly.*"

3.2. Optimizing IT equipment utilization

Another approach to providing load flexibility during operation is optimizing IT equipment utilization through the dynamic voltage frequency scaling (DVFS) method.

The DVFS-based technique can control the operating voltage and frequency of chips, which strongly correlates with server power consumption. By limiting server power load when overheating is detected, DVFS can accurately control the energy density of data centers. This technique can modulate servers' and data centers' real-time power loads in response to power systems' operation signals (Krzywda et al., 2018; Wang et al., 2010, 2017). With online algorithms ensuring time limits for production requirements, studies have shown that DVFS can achieve significant energy savings (Hsu and Feng, 2005; Enhanced Intel Speedstep Technology, 2021; Koronen et al., 2020).

Experts from large-scale internet data-center utilities, representing two-thirds of all data center experts responding to this research, stated that DVFS-based techniques are commonly used in the daily management of large-scale data center clusters to prevent server overheating. "*It can be accomplished through highly automated sensing technologies in computer rooms. Sometimes, operators may also use DVFS to reshape load profiles of servers under circumstances like overloaded.*" In China, DVFS-based techniques have been widely integrated into artificial intelligence (AI) management modules used by large-scale data-center utilities, such as Alibaba, Tencent, Baidu, and Qinhuai, to improve energy efficiency and reduce environmental impact (Data Center Whitepaper, 2022). This technique is also embedded in the Carbon-Intelligent

Computing platform developed by Google, aiming to develop a hyper-scale carbon-aware computing system (Radovanovic et al., 2022).

3.3. Flexible charging strategies of UPS

Energy storage, or UPS, serves as backup power in data centers in case of power outages and generation fluctuations, ensuring the uninterrupted operation of data centers 24/7. It can turn on within 25 ms of a power interruption and power the infrastructure for 15–30 min (Judge, 2021). However, along with improved reliability of power grids and redundant configurations of data centers, most energy storage facilities maintain low efficiency. By optimizing their operational behavior, UPS in data centers can perform load management, reduce the overall energy cost (Liu et al., 2016; Guo and Fang, 2013; Peng et al., 2017), and generate revenue through participation in electricity frequency regulation (Urgaonkar et al., 2011) and demand response.

Along with the increasing scale of data centers, it is estimated that the battery capacity and ride-through time will continue to expand. Moreover, "*Integrating renewable energy with storage is becoming a trend in China's data centers*," as an energy expert stated in the interview, "*Apart from the provision of regulation services, UPS in data centers could be used to absorb a high wind or solar output throughout the year*." Energy storage devices are seen by most as the key facilities for 100% carbon-free energy utilization in data centers. To achieve carbon neutrality targets in data-center sectors, large-scale UPS is one of the promising techniques to integrate renewable energies in data centers, providing even larger opportunities for load flexibility in the future.

3.4. Operational robustness of the cooling system

The cooling system in data centers consists of computer-room air-conditioning units (CRAC) and computer-room air handler units (CRAH) (Alkharabsheh et al., 2015). When the inlet temperature of the computer room rises, the cooling system starts and maintains the temperature within a designed range. During operation, the temperature can be raised to a higher level within the acceptable range, directly affecting the power load. According to (Wang et al., 2013), increasing the outlet air temperature of the cooling system by 1 °C can save 4.3%–9.8% of the electricity demand for cooling. The operational robustness of the cooling system enables data centers to provide load flexibility under operational security constraints. Moreover, the response time of cooling management meets the demand for frequency regulation markets (Li et al., 2013; Salomonsson et al., 2007; Fu et al., 2020).

About 80% of questionnaire respondents acknowledge the load flexibility potentials in cooling systems, stating that "*flexibility in data centers relies on power conservation by controlling cooling equipment, lighting, etc., of which cooling systems are the most promising since they usually contribute the second-highest power consumption in a typical data center other than IT equipment.*" An academic professional noted in the interview that "*cooling systems of large-scale data center clusters could offer much greater demand response capabilities, which possess quite large potentials in flexible energy usage.*" There is no evidence of data centers in China that have proactively changed load profiles for economic purposes through cooling system flexible operation. The current involvement of data center cooling systems in China generally focuses on improving cooling efficiency.

3.5. Flexibility potential evaluation

In this subsection, we would like to estimate the power flexibility of data centers in China based on available data and existing publications. The load flexibility of the resources within data centers is analyzed respectively.

(1) Power flexibility through workload management.

(a) Power flexibility by temporal workload management.

Regarding the flexibility provided by temporal workload management, Cao et al. (2022) assessed the temporal flexibility of periodic batch workloads in 3 production clusters in Alibaba with approximately 8000 servers, which reaches 130 kW during peak hours. Given that aperiodic workloads are less flexible, as introduced in (Cao et al., 2022), and periodic batch workloads comprise two-thirds of all batch workloads in production, we ignored the flexibility of aperiodic workloads in this evaluation. The server scale of China is derived from (Wang, 2022), which is 2×10^7 servers by 2022 in total. The total temporal power flexibility of batch workloads in all data centers in China can be calculated as follows:

130 kW × (National server number/8000 servers) = 130 kW × (2×10^7 / 8000) = 325 MW

This result represents that the maximum power flexibility in China's data-center sector by temporal workload shifting is 325 MW during peak hours.

(b) Power flexibility by spatial workload migration.

As reported in (Alibaba and North China Electric, 2022), interactive workloads shifting across regions drops considerable power load in the original data center clusters. As a result, we would like to assess the spatial flexibility of interactive workloads of data centers in this part.

The computing performance of interactive computing workloads is measured by floating point operations per second (FLOPS). The relationship between computing power and energy performance of about 40 kinds of CPU chips widely used in China has been evaluated by GREEN500 in 2022 (GREEN500. List, 2022), and the results are shown in Fig. 3. As can be seen, the average computing power of them is 1.97 PFLOPS, and the average power consumption is 742.6 kW.

Based on the statistics in (Wang, 2022), the scale of computing power in China exceeds 150 EFLOPS (i.e., 1.5×10^5 PFLOPS) by the end of 2022. We assume that 1% of the total computing power in China can participate in geographical migration. This portion is quite understandable since most interactive workloads require instant computing; the delay intruded on by workload migration will cause service quality degradation.

In sum, the maximum spatial power flexibility provided by spatial workload management can be calculated as $1.5 \times 10^5 \times 1\% \times 742.6/1.97 \approx 565.43$ MW.

(2) Power flexibility through IT equipment

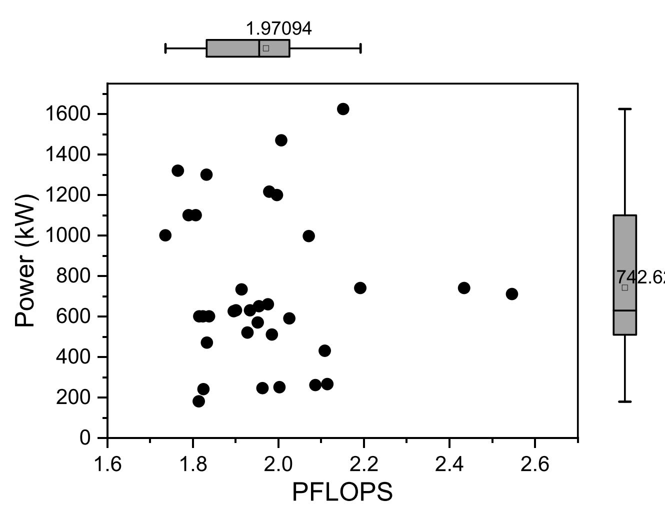

Fig. 3. The correlation between PFLOPS and power consumption (kW) (Source: GREEN500 (GREEN500. List, 2022).).

As previously mentioned, Google has investigated the flexibility potentials of servers powered by DVFS techniques in a carbon- and efficiency-aware manner (Radovanovic et al., 2022). The results showed that their optimal method could help drop 1–2% of the cluster power based on their experiments on production data center clusters. We assume this method could be applied in China in all data centers. Given that the power load of a common server is 500 W, the total power load of data centers in China would be $2 \times 10^7 \times 500 = 10{,}000$ MW. Therefore, the flexibility potential through optimizing IT equipment utilization can be $10{,}000 \times 1\% = 1000$ MW from a conservative perspective.

(3) Power flexibility through UPS.

UPS systems are often deployed with redundancy to meet the high uptime requirements for data centers. There are three main UPS redundancy architectures, i.e., N+1, 2N, 2(N+1), where N is the full UPS capacity required to ensure the total load and cooling demands. Since the UPS system is rather expensive, they are designed to provide enough power for the critical load. We assume the critical load would take up to 32% of the total load (Sawyer), and the batteries run at a sustainable limitation with 20% as the minimum storage capacity. Accordingly, considering a common 2N configuration for UPS systems in data centers, the maximum charging and discharging power would be $10{,}000 \times 32\% \times (1–20\%) \times 2 = 5120$ MW which theoretically could be used as the flexibility that can be provided to power systems.

(4) Power flexibility through cooling systems and IT equipment.

As previously mentioned in Section 3.4, cooling load in data centers is highly related to the outlet temperature, which is often set based on equipment manufacturers' suggestions. According to the design standards for the refrigeration of data centers (ASHRAE, 2022), the recommended temperature range is 15–30 °C. We assume that 23 °C is a generally taken setpoint for cooling systems in China's data centers; thus, the acceptable temperature increase is 7 °C. As evaluated in (Wang et al., 2013), one degree up of cooling temperature could reduce 4.3%–9.8% of the total load. Consequently, cooling systems' maximum flexible power consumption that can be conserved is $7 \times 10{,}000 \times 4.3\% = 3010$ MW.

(5) Total power flexibility potentials in China

As is known, power flexibility should be explained in terms of flexibility capacity, response speed, duration, and direction. Based on our bottom-up analysis illustrated above, the flexibility potentials of the flexible resources within data centers in China are summarized in Table 4. It should be noted that the flexibility capacity is calculated based on the latest literature and advanced cluster management technologies. However, we have to admit that not all data centers in China have the same level of advanced workload management system and volume of workloads as Alibaba or Google, nor do they have the ability to apply edge-cutting theories to practice. Therefore, we believe that considering a 30–50% reduction on top of that would be more reasonable and realistic.

4. Key barriers for data centers to provide load flexibility to the power system in China

Building on the questionnaire and semi-structured interviews described in Section 2, key barriers for China's data-center sector to provide load flexibility to the power system are identified: technical, financial, policy, and institutional barriers.

4.1. Technical barriers

(1) Stability-related concerns and physical constraints on workload shifting

Generally, a high level of system stability and performance is of utmost importance for data centers, specified in Service Level Agreement (SLA) negotiated between Internet service providers (ISPs) and customers. Given this goal, many constraints are incorporated into the current workload scheduling system, including resource overbooking and differentiated workload priority assignments, to match the various demands of workloads with limited computational resource capacity in optimum manners.

As a result, all data center experts mentioned in their questionnaires that implementing flexibility from workload shifting for data center operators would be less interesting. One data center expert explained that "*temporal and spatial workload shifting may violate the time limitations and induce degradation of service quality, which would certainly undermine the reputation of a company that specializes in the provision of high-quality, accurate computing services*." Another data center expert added, "*spatial workload migration is greatly constrained by bandwidth and workloads' demanding data files. Business delay is certainly inevitable as bandwidth is not yet in a state where it can be interconnected*." As for the data constraints, they supplemented that interactive workloads, which possess spatial flexibility, involve little computation but might require retrieving huge data files. Scheduling such workloads requires packaging and transferring the underlying data files, which would burden the limited bandwidth resource.

(2) Lack of successful experience

Data center management system has evolved over generations with highly automated and computerized characteristics, providing optimal scheduling of workloads and immediate control of ancillary facilities, which has been seen as widely-accepted sustainable and optimal operating models. From this perspective, IT sectors in data centers have now stuck into business-as-usual (BAU) cases. This conflict between BAU and flexible operating strategies was reflected in the typical views held by all data center experts interviewed.

Table 4
Flexibility resources in data centers.

Facility	Strategy	Direction	Flexibility Capacity (MW)	Response Speed	Duration	Ref.
IT equipment	Workload temporal scheduling	Up/down load regulation	325	~ seconds	~ hours	Ghatikar (2012)
IT equipment	Workload spatial shifting	Up/down load regulation	565	~ seconds	~ minutes	Ghatikar (2012)
IT equipment	Power management	Load reduction	1000	~ seconds	~ minutes	Wang et al. (2019)
UPS	Charging/discharging power	Up/down load regulation	5120	~ seconds	~ hours	McCluer and Christin (2008)
Cooling system	Adjust the air temperature in the computing room	Load reduction	3010	~ minutes	~ hours	Beil et al. (2016)

Experts from ISPs stated in our questionnaire that

> "It may be worthwhile to form a task force at data centers for periodic work related to real-time information collection, manual intervention on the scheduling results, etc. However, since the importance of flexible power consumption is unclear for data center operators and the successful experience is still needed, few data center operators are willing to make additional efforts on uncertain things against the ongoing performance maximization principle."

Another data center expert contended, "*Manual intervention that gives up the optimal operating solutions would also give rise to other questions, such as who should take the responsibility [for] the performance degradation. This could be alleviated with the experience from a successful implementation of load flexibility application programs on an ongoing basis.*"

The academic professional whose team participated in the spatial workload migration project with Alibaba shared the successful experience in the interviews by stating, "*Our team has made a great effort to persuade several teams inside the Alibaba Group to design the specific scheduling plans. We believe our program went well because we are a group of people who do this full-time. We linked multiple teams within the company to work together based on our multifaceted knowledge and backgrounds. These things don't happen often in normal times.*"

When looking at international experiences, in response to Q2.3 of the questionnaire, it is reported that data center industries are gaining influence in the U.S. electricity market, accounting for about 4.5% of power demand in America (Miller, 2019). Large internet companies like Google, Apple, and Amazon typically use sophisticated commodity risk management operations in the power market to address the complexity of the electric industry. However, these large tech companies generally use economic instruments, such as multi-campus deals, to hedge risks of price fluctuations instead of leveraging the flexibility resources in data centers to minimize costs (Miller, 2018). In addition, some startups, such as *Lancium* in Texas, have made a series of attempts by deploying flexible data centers that can dispatch their power demand in response to market signals, based on the idea proposed in (Chien et al., 2015).

(3) Less load flexibility from UPS

Despite the consensus by a growing number of academic papers on load flexibility through optimal dispatch strategies of UPS in data centers, the feasibility of releasing this flexibility in practice still lacks foundations. In particular, an academic interviewee noted in response to Q2.1 - Q2.3 in the questionnaire: "*Data centers could, indeed, arguably control energy storage to achieve flexible power consumption, reducing energy cost and carbon emissions. However, further research and experiments are needed to realize this measure on a large scale*." One data center expert demonstrated a similar opinion: "*Data centers apply considerable expenditure on UPS installation to achieve desired reliability of system operation and continuity of power supply. As such, they are unwilling to employ this high-priced protective capability for power grid purposes because it would pose a power interruption threat to servers during emergency, resulting in significant financial loss to data center enterprises*." Another data center expert agreed with this opinion in the semi-structured interviews by supplementing that "*since there is limited load flexibility available for individual UPS in data center clusters, it is rather difficult to integrate it into the power systems, despite increasing related proposals. To sum up, individual UPS facilities in data center clusters are not sufficient to be considered in the power system.*"

(4) Co-location data centers

In light of the ownership type of servers and other ancillary facilities, data centers in China fall into two categories: self-built and co-location data centers. The main difference is that the former is purpose-built to serve itself, while the latter hosts IT-related facilities for tenants. In the case of self-built data centers, the ISPs build the data center and own all the facilities within the data center. Therefore, the flexibility actions described above are taken by operators of these data centers with consideration of their agreements with clients. There is no need to consider any third parties.

In contrast, co-location places enterprise-owned computer, storage, and networking assets in a third-party leased facility (Nadeau and Gray, 2013). In other words, several consumers share the physical infrastructure in the co-location data center. Despite their scalability and benefits of saving construction and maintenance costs, it is also true that the service quality commitments to multiple consumers will make the operational requirements stringent, needless to mention the flexibility during operation.

Many respondents in our questionnaire and interviews expressed this concern. One data center expert said, "*Things always get tricky when the value chain is distributed between various stakeholders. The operators of co-location data centers have no right to migrate the workloads or lower the operational frequency of servers. The only sub-system that they could change is the temperature of the computer room, which, however, is also determined in the contracts*." The lack of engagement with customers would make it hard for co-location data centers to provide load flexibility, where "*who is going to have that conversation with all the customers, considering there are tens of them*." Another expert added, "*Perhaps self-built internet data centers, such as Alibaba, Tencent, Huawei, etc., with simple capital ownership could provide more load flexibility during daily operation.*"

4.2. Financial barriers

(1) Insensitive to time-of-use electricity price

China has established a time-of-use (TOU) pricing structure for industrial electricity consumption. In 2021, a policy was issued to improve the TOU price mechanism, which raised the peak-valley difference of the TOU prices (Improvement on TOU Prices Mechanism, 2021). During the semi-structured interviews, interviewees were asked whether electricity prices during peak load hours could motivate data centers to participate in demand response. However, six data center experts and one energy sector expert found little correlation between the two in practice, contrary to what is suggested in academic papers.

> "*Few data centers are sensitive to electricity prices in China,*" said the energy sector expert, "*when choosing a location for a new data center to be built, the typical priority consideration is client proximity rather than energy cost*." One data center expert who works as a strategic investment analyst added, "*Even though energy prices in western China are much cheaper than those in eastern China, data center operators prefer building data centers in eastern for closer distance with more customers, ensuring better computing performance*." Another data-center sector expert had similar views from a different perspective on this by stating that "*western places with cooler climate are favorable for data centers in the long run, but this also represents higher server investment and maintenance cost, which accounts for a large share of the overall cost*." These suggest that saving electricity bills is not enough to make up for the inconvenience of following price signals continuously, whose core business is to provide computing service steadily.

(2) Insufficient incentives from spot electricity price

In response to Q3.1 and Q3.2 in the questionnaire, most respondents held that participation in the power market has never been the sole method for data centers to get a lower electricity price.

Data centers with consistent power demand can secure stable electricity prices through bilateral contracts with power plants, while load flexibility measures, such as workload migration or adjusting frequency and voltage, often result in a reduced computing performance to get an uncertain lower price in the power market, which is not a good deal for data centers. Experts from scientific institutions illustrated that "*electricity prices in China's current power market are less volatile, making it less attractive for data centers to participate in the market bidding.*" One interviewee suggested, "*It shouldn't be the market forces involving data centers to participate in power markets. It is not until more direct administrative regulations, such as requiring a certain amount of flexibility as a condition for interconnection, that data center participation in the power market or other markets will be feasible*."

4.3. Policy barriers

(1) The regulatory system for data centers' energy consumption focuses solely on PUE

As previously noted, the growing trend of power consumption in the data-center sector is expected to continue, placing great pressure on China's energy management and efficiency improvement. Power usage effectiveness (PUE), the ratio of the overall power consumption of a data center to the power consumed by IT devices, is one of the key measures to evaluate the energy efficiency of the infrastructure in the data center. The lower the PUE value of the data center, the more efficient it is.

In response to the increasing demand for electricity in the data-center sector, China has implemented a series of policies to improve its energy efficiency. In 2018, MIIT set a national target for a PUE of 1.50–1.69 for data centers, further narrowed to under 1.50 for all data centers in 2021. Local governments have also introduced several detailed energy efficiency regulations, establishing a data-center sector regulatory system based on the PUE value. These policies encourage data centers to seek higher energy utilization during operation. As a result, the PUE in the data-center sector has decreased from 2.5 to 1.49 since 2010, reducing energy losses by over 60% (Zhang and Chen, 2020).

The current government management approach for power consumption in data centers in China lacks robust and flexible regulations. PUE is the sole de facto metric for assessing energy efficiency, which often requires steady and high energy utilization, hindering flexible operations such as load shifting. This regulatory system focusing only on the PUE values may incentivize data centers to prioritize a low PUE over cleaner energy resources. Furthermore, the static yearly average PUE value is not a flexible metric, hindering the ability of data centers to use electricity flexibly.

(2) Lack of certification standard for data centers' contribution to clean energy consumption

As part of the country's effort to achieve green and sustainable development, China implemented mandatory energy intensity and total energy consumption targets (the so-called Dual Targets of Energy Consumption) in China's 11th and 12th Five-Year Plan. These targets were subsequently incorporated into the performance evaluation system of local governments, leading to strict control of provincial electricity consumption, especially for high energy consumption and energy-intensive industrial and commercial users. Consequently, data centers responsible for non-real-time task computation are the power loads that would be primarily shed if necessary.

A data center expert introduced in the interview, "*In January 2021, some data centers in Zhejiang and Jiangsu provinces experienced power outages due to the mandatory provincial electricity consumption, even though these data centers consumed plenty of distributed renewable power generated by themselves*." In this respect, this expert contended that the government policy should change from the dual control of energy consumption and efficiency to dual control of carbon emission and intensity, incentivizing flexible loads to consume green energy and provide load flexibility to its utmost extent. Interestingly, the Chinese central government officially adopted this suggestion shortly after the interviews in its relevant policy documents.

However, energy and data center experts argued that without sound accounting methods and supporting policies to certify data centers' contribution to renewable energy consumption and carbon emissions reduction, their motivation to exploit the flexibility strategies towards renewable energy consumption is undermined.

4.4. Institutional barriers

(1) The multiplicity of relevant parties in power load shifting

While integrating load flexibility of data centers into power systems has been well explored in some cases, there are complex institutional barriers currently preventing its widespread application on an industrial scale in China. Such barriers arise from the inherent heterogeneity of services and lack of communication between various organizations involved, such as the cooperation of diverse groups of stakeholders, typically the grid utilities and flexible loads. Additionally, data centers providing flexibility through workload shifting require bandwidth support from the telecom sector. The absence of a clear working process framework hinders their cooperation among the multiplicities. An energy expert interviewed illustrated, "*It is because of the involvement of multiple sectors that few projects are implemented spontaneously on [a] large scale. It is rather difficult to drive projects from the bottom up*."

(2) Immature power market

The flexibility of the data centers enables them to be potential participants in the various power markets. In 2020, the North China Energy Regulatory Bureau issued a policy allowing individual power users to participate in peak-shaving markets. In 2020, Chongqing municipality made rules to encourage data centers to provide load flexibility for demand response, the first policy targeting data centers to enter the auxiliary market directly in China. In 2020, Fujian province implemented a power market registration principle for data centers with more than 150 racks. As of now, data centers are not allowed to participate in the spot power market in China, and in provinces and autonomous regions where spot markets are held, such as Fujian, Zhejiang, Inner Mongolia, Sichuan, and Gansu, participants are limited to certain power consumers. Moreover, institutional problems lead to financial issues, as regulations on price variation range in both TOU prices and spot market clearing prices may not provide sufficient motivation to engage in demand response this way.

5. Possible solutions to address the barriers of data centers to provide load flexibility

Drawn on experts' views on RQ3, this section discusses possible solutions for the four categories of barriers to data center flexibility, including technical improvements, financial incentives, policy incentives, and institutional reforms. The barriers and incentives are shown in Fig. 4.

5.1. Technical improvements

As experts raised, addressing the lack of data & information, bandwidth, platform, and pathways is crucial for data centers to adopt load flexibility.

First, information transparency and sufficient data are critical missing parts at the current stage. Therefore, the suggestion, agreed by all experts, was to establish a public or semi-public information platform to ensure transparency in data exchange. Real-time information on spot electricity prices, carbon emission density, and wind and solar power generation is essential for demand-side participation analysis. This information could be collected from provincial grid companies and departments such as MIIT, NEA, and CEC.

In contrast to the principles of the current scheduling system, shifting

Fig. 4. Barriers to and solutions for data centers' load flexibility provision.

workloads for data centers necessitates designated staff for instant manual intervention. In the long run, data centers must integrate price and emission signals into decision-making through an improved management system. The real-time information platform mentioned earlier can be accessed as an additional module embedded in the enhanced data center management system to make well-informed decisions with the help of data-driven methods, making the data center more economically and environmentally intelligent.

Like energy utilities, robust fiber networks are crucial for data centers to provide spatial flexibility. As noted above, the western regions in China have significant advantages for data centers in terms of energy efficiency and renewable integration. However, the insufficient bandwidth and transmission infrastructure in remote areas hinders the utilization of these advantages, indirectly limiting the ability of data centers to migrate workloads geographically to absorb excess energy supplies and reduce peak loads. Therefore, the energy sector and the MIIT should collaborate to improve bandwidth construction and strengthen telecom transmission infrastructure to ensure technical feasibility.

To address the limited flexibility of individual UPS devices, experts emphasize the need for new collaborative pathways, such as aggregation. China is promoting the participation of virtual power plants (VPPs) and other demand-side resources in the ancillary services market. Additional UPS capacity can contribute to frequency regulation and peak shaving while prioritizing computing services. One energy expert noted in the interviews, "*Aggregating UPS from multiple data centers into a VPP is one method to encourage data center participation in the ancillary services market, given the individual capacity limitations of UPS.*" Furthermore, another energy expert emphasized that demonstration projects are needed to evaluate whether they can make significant contributions to the economic operation of the power system and data centers.

5.2. Financial incentives

As illustrated above, all six experts from data-center utilities consistently indicated that data centers are generally hesitant to engage in power markets and provide flexibility at this stage, as data centers sacrificing the optimal options cannot be rewarded properly through flexibility provision and market integration. Thus, additional economic incentives should be provided to further develop data centers' ability to participate in demand response.

Given that electricity price differentials in China's power market are unlikely to be large, some professionals suggested that other financial incentives are needed. Experts from energy utilities stated in the questionnaire, "*For example, fiscal subsidies could be set up to* support *data centers to participate in the power market and compensate for the service degradation caused by the provision of flexibility*." These subsidies could then be passed on to the data center customers willing to accept the risk of service degradation and allow the data center operators to schedule their compute loads. This new business model for data centers could provide more diversified services and motivation for users to achieve their evaluated tradeoff between inconvenience and benefit.

5.3. Policy incentives

The central government should develop clear carbon neutrality targets for energy-intensive sectors as soon as possible. China has committed to peaking carbon emissions by 2030 and achieving carbon neutrality by 2060. However, there is still no comprehensive governance system for collaborative emission reduction progress at the national level or specific quantitative carbon emission requirements for regions and industries. As previously noted, data centers are insensitive to economic incentives and require a more explicit administrative directive to provide load flexibility to the power systems. Implementing a command-and-control policy as a mandatory carbon emission requirement for industrial-level carbon neutrality pathways could provoke data centers to provide load flexibility, thereby achieving higher emission reduction objectives.

Second, it is recommended that multiple metrics should be used in tandem rather than in isolation to obtain a more holistic picture of data center energy usage. One expert suggested that complementary metrics representing clean energy consumption of data centers should be included and implemented as part of a set of green data center measures, both within the data centers themselves and throughout the supply chain. Establishing a more comprehensive regulatory metric system for data center energy consumption could motivate data centers to be more flexible and use cleaner energy resources.

Third, a standardized method for evaluating the renewable energy integration of data centers should be established to provoke their motivation for emission reduction and load flexibility application. The standard should explicitly calculate data centers' contribution to providing load flexibility, including direct and indirect carbon emission reduction, among other factors.

5.4. Institutional reforms

(1) Clear working process of cross-sectoral collaboration

First and foremost, guidelines should be issued to clarify related sectors' responsibilities and working processes concerning the complex organizational reality of data-center load-flexibility provisions. Without such guidelines, it would be troublesome for data centers to interact

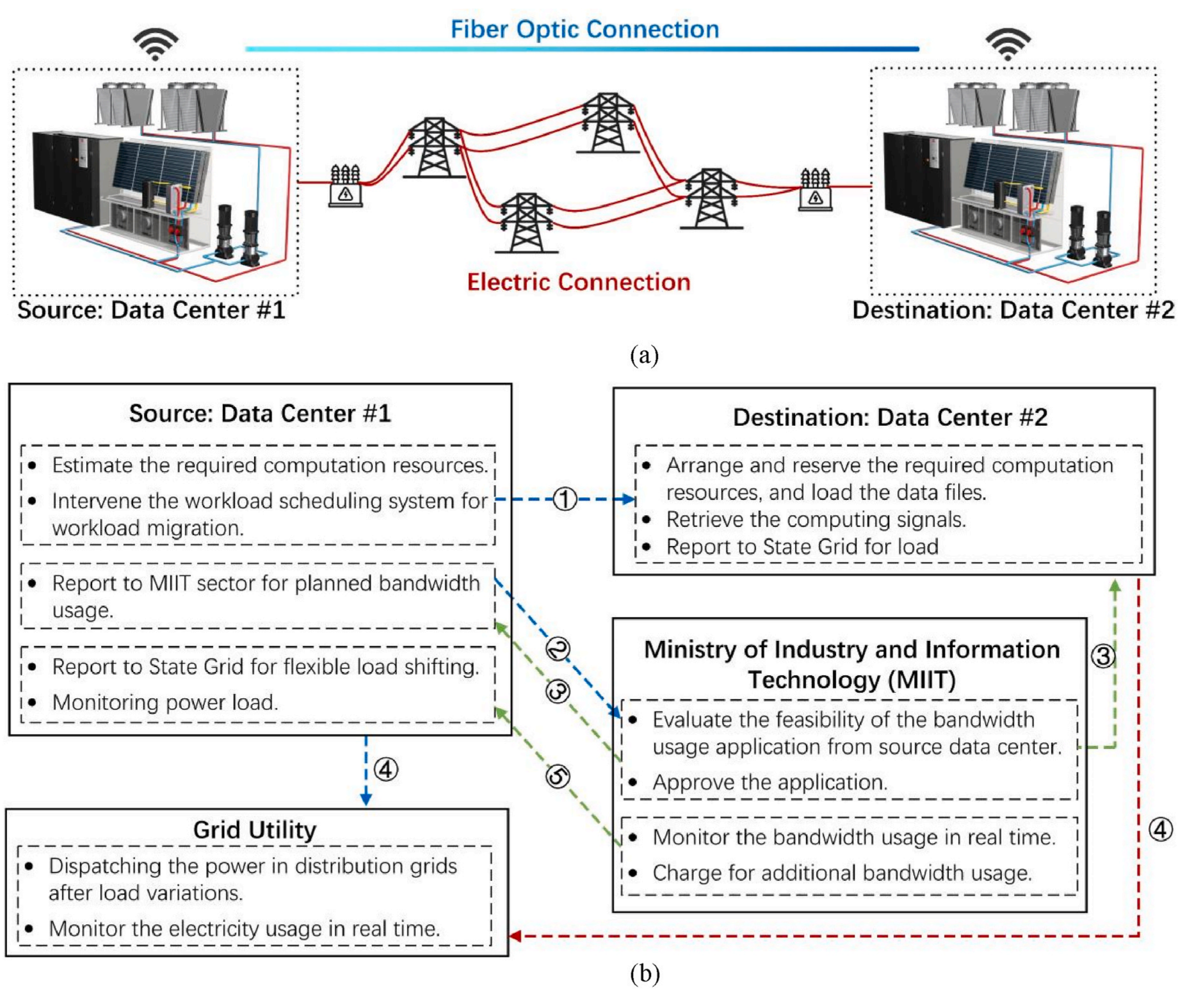

Fig. 5. A conceptual framework of cross-multi-sector collaboration. (a) fiber optic and electric connections; (b) responsibilities between data centers, MIIT, and state grid during workload migration.

with other sectors to retrieve real-time signals for large-scale flexibility provision and implementation. All three experts from scientific institutions stated, "*The working process of different bodies involved in data-center load flexibility integration into power system should be clarified if it were to be implemented on a regular basis*."

A clear working process associated with separate sectors should reflect their roles and needs and, most importantly, the cross-sectoral collaboration workflow. Take spatial scheduling of workloads as an example, as shown in Fig. 5: firstly, data center operators need to estimate the resource requirements of the scheduled workloads and reserve the required capacity in the destination data center in advance. Also, they need to estimate the corresponding power changes and report to the grid utility. Secondly, local MIIT approval is also required to schedule bandwidth resources. Thirdly, after migration, the grid utility accommodates load changes to ensure the operational safety of the distribution grids in both regions.

(2) Electricity market reform

The energy-related governments in China should prioritize improving market liquidity. As previously noted, the peak and valley prices in the TOU mechanism and spot market principles do not provide adequate incentives for data centers to offer flexibility. Generally, high revenues are expected in the short-term electricity market. To address this, the central government should increase the peak-valley ratio of TOU and properly expand the range of clearing prices in spot power markets. Policymakers should focus more on increasing power load flexibility in response to the recent rapid growth of peak load in most provinces in China rather than stabilizing loads.

Understanding data centers' participation based on the current market design is useful for exploring incentives load flexibility. Even though there are several policies concerning new technologies that could provide load flexibility at the national or local level, no policies outline how data centers can participate in electricity or ancillary services markets. A rational and clear market design open to more demand-side players is an important approach to induce more load flexibility.

As mentioned earlier, a major issue with data centers integrating UPS flexibility into the power system is their reluctance to provide expensive and limited energy storage equipment for backup purposes of the power system. In addition, the limited range of electricity price fluctuations in the spot market fails to reflect the value of energy storage capacity. In this context, the capacity market is essential in alleviating data centers' financial concerns regarding UPS capacity installment.

6. Conclusions and policy implications

Concerning the increasing penetration of intermittent and variable renewable energies, flexible loads are considered urgent-needed resources on the demand side of the power system. Recently, data centers have gained attention for their diverse sources of load flexibility. Despite the growing scale of data center industries in China, the provision of power flexibility by data centers lacks industrial-scale practice, and few studies have explored this topic. This paper employs a hybrid research method of literature review, questionnaire, and semi-structured interviews to examine the potentials, barriers, and possible solutions for data centers in China to provide load flexibility. The main conclusions of this research are as follows:

The load flexibility during the daily operation of data centers could be derived from the following four aspects: 1) temporal and spatial workload scheduling; 2) IT equipment utilization optimization; 3) flexible charging strategies of UPS devices; 4) acceptable cooling temperature range. All of them could introduce considerable load flexibility by providing operational robustness to data center operations. Then the flexibility potentials of data centers in China are evaluated with reasonable assumptions.

Even though promising load flexibility potentials exist, data center operators are hesitant to participate in power markets due to several barriers, including technical, financial, policy, and institutional barriers. Technically, the flexibility resources within data centers are difficult to implement on a large scale as data center operators concern that the participation would undermine their core business of stable computing service. Financially, the possible revenue through load flexibility provision is not appealing for data centers, and the risk of core business degradation cannot be sufficiently compensated. In terms of policy, the ongoing related policies only focus on the energy efficiency and stable power load of data centers. Regarding institutional barriers, the immaturity of the power market and unclear working process between cross-sectors undermines the motivation of data centers to provide and explore load-shifting approaches spontaneously.

The following solutions are recommended to address the identified barriers. Technically, data centers should improve their management system and adopt UPS aggregation strategies. The government sectors should focus on improving data transparency and bandwidth construction at the national level. Given the limited incentives from power markets, financial solutions, such as fiscal subsidies, are also important to support data centers. Regarding policy solutions, the central government is suggested to formulate more mandatory requirements, such as carbon neutrality commitment targets for industries, building a joint-up green data center metrics system, and certification standards for renewable energy consumption. In terms of institutional reforms, top-down policies should be issued to clarify the responsibilities of related sectors throughout the process. Moreover, electricity market reforms should be deepened, with real-time spot price signals released to market players to enhance demand-side load flexibility. The solutions and responsible entities are summarized in Table 5.

Two major limitations of this study could be addressed in future research. First, the solutions suggested in the article vary in the difficulty of implementation and the time required to overcome the barriers. Specifically, experts with various backgrounds and knowledge areas could make technical improvements in data center industries through several successful attempts. Financial incentives and policy incentives could be achieved in the short term, starting with launching pilot projects and trials in certain provinces. However, institutional reforms need

Table 5
Summary of the solutions and responsible entities.

Theme	Solution	Responsible Entity
Technical improvements	Bandwidth construction	MIIT, NDRC, telecom companies, Ministry of Natural Resource, and Ministry of Housing and Urban-Rural Development.
	VPP for UPS	NDRC, power trading center, State Administration for Market Regulation, and data center industry.
	Management system improvement	Data center industry.
	Assemble Real-time information platform	NDRC, NEA, MIIT, State Power Corporation, and the data-center industry.
Financial incentives	Fiscal subsidies	NDRC, MIIT, power trading center, and Ministry of Finance.
Policy incentives	Clear carbon neutrality targets for data centers	NDRC, NEA, and the data-center industry.
	More metrics in the regulatory system	Data center industry, universities, and research institutions.
	Evaluation of clean energy usage for data centers	MIIT and data center industry associations.
Institutional reforms	Working process of cross-sectoral collaboration	NDRC, MIIT, NEA, data center industry.
	Electricity market reform	NDRC, NEA, MIIT, State Administration for Market Regulation.

national-level efforts and attention to review the flexibility potential of data centers in the long term, which is more complicated. Second, the evaluation of data center flexibility is not conducted at the provincial level. This perspective is also important and worthy of study because the power grid structures and energy mix vary across provinces, resulting in diverse needs for power flexibility. Thus, assessing data center flexibility by province and evaluating its benefits to provincial power systems can offer a more comprehensive analysis of the situation in China.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

This study was supported by the National Social Science Foundation of China (Grant No.21BJY012).

References

Ali, A., Ozkasap, O., 2016. Price/cooling Aware and Delay Sensitive Scheduling in Geographically Distributed Data Centers. Network Operations and Management Symposium, IEEE, pp. 1025–1030.

Alibaba and North China Electric Power University Achieve the World's First Collaborative Scheduling between Data Centers and Power Systems, 2022. Xinhua, Beijing.

Alkharabsheh, S., Fernandes, J., Gebrehiwot, B., et al., 2015. A brief overview of recent developments in thermal management in data centers. J. Electron. Packag. 137 (4), 040801.

ASHRAE, 2022. Description 2022 ASHRAE Handbook—Refrigeration. https://www.ashrae.org/technical-resources/ashrae-handbook/description-2022-ashrae-handbook-refrigeration.

Beil, I., Hiskens, I., Backhaus, S., 2016. Frequency regulation from commercial building HVAC demand response. Proc. IEEE 104 (4), 745–757.

Börjeson, S., Östbom, L., 2018. Balance the Swedish Transmission System by Using Data Centers: A Study whether UPS-Systems Can Operate as Frequency Regulators. Uppsala Universitet.

Bouckaert, S., Goodson, T., Wanner, B., 2018. The Clean Energy Transition Requires Action on Electricity Demand. IEA, Paris.

Cao, F., Wang, Y., Zhu, F., et al., 2019. UPS node-based workload management for data centers considering flexible service requirements. IEEE Trans. Ind. Appl. 55 (6), 5533–5542.

Cao, Y., Cheng, M., Zhang, S., et al., 2022. Data-driven flexibility assessment for internet data center towards periodic batch workloads. Appl. Energy 324, 119665.

Chen, H., Caramanis, M.C., Coskun, A.K., 2014. Reducing the data center electricity costs through participation in smart grid programs. Int. green comput. conf., IEEE 1–10.

Chen, T., Wang, X., Giannakis, G.B., 2016. Cooling-aware energy and workload management in data centers via stochastic optimization. IEEE J. Sel. Top. Signal Process. 10 (2), 402–415.

Chen, H., Zhang, Y., Caramanis, M.C., et al., 2019. Energyqare: qos-aware data center participation in smart grid regulation service reserve provision. ACM Trans. Model Perform. Eval. Comput. Syst. (TOMPECS) 4 (1), 1–31.

Chen, S., Li, P., Ji, H., et al., 2021. Operational flexibility of active distribution networks with the potential from data centers. Appl. Energy 293, 116935.

Chien, A.A., Wolski, R., Yang, F., 2015. The zero-carbon cloud: high-value, dispatchable demand for renewable power generators. Electr. J. 28 (8), 110–118.

Data Center Whitepaper (2022). Beijing: China Academy of Information and Communications Technology (CAICT);.2022.

Dou, H., Qi, Y., Wei, W., et al., 2015. Minimizing electricity bills for geographically distributed data centers with renewable and cooling aware load balancing. In: 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI). IEEE, pp. 210–214.

Dou, H., Qi, Y., Wei, W., et al., 2017. Carbon-aware electricity cost minimization for sustainable data centers. IEEE Trans. Sustain. Comput. 2 (2), 211–223.

Enhanced Intel Speedstep Technology (EIST), 2021. https://www.techopedia.com/definition/2861/enhanced-intel-speedstep-technology-eist.

Fu, Y., Han, X., Baker, K., et al., 2020. Assessments of data centers for provision of frequency regulation. Appl. Energy 277, 115621.

Gao, C., Wu, G., Chen, S., 2019. A model aimed at reducing power net loss considering frequency scaling of servers in geo-distributed data centers. Proc. CSEE 39 (6), 1673–1681.

Ghatikar, G., 2012. Demand Response Opportunities and Enabling Technologies for Data Centers: Findings from Field Studies.

Ghatikar, G., Piette, M.A., Fujita, S., et al., 2009. Demand Response and Open Automated Demand Response Opportunities for Data Centers. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Goudarzi, H., Pedram, M., 2013. Geographical Load Balancing for Online Service Applications in Distributed Datacenters. IEEE Sixth International Conference on Cloud Computing, IEEE, pp. 351–358.

GREEN500. List-November 2022, 2022. https://www.top500.org/lists/green500/2022/11/.

Guo, Y., Fang, Y., 2013. Electricity cost saving strategy in data centers by using energy storage. IEEE Trans. Parallel Distr. Syst. 24 (6), 1149–1160.

Guo, Y., Gong, Y., Fang, Y., et al., 2014. Energy and network aware workload management for sustainable data centers with thermal storage. IEEE Trans. Parallel Distr. Syst. 25 (8), 2030–2042.

Hogade, N., Pasricha, S., Siegel, H.J., et al., 2018. Minimizing energy costs for geographically distributed heterogeneous data centers. IEEE Trans. Sustain. Comput. 3 (4), 318–331.

Hsu, C.-h., Feng, W.-c., 2005. A power-aware run-time system for high-performance computing. In: SC'05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing. IEEE, 1-1.

IEA. China, 2019. Power System Transformation. IEA, Paris.

Implementation Plan for Computing Power Hub of National Integrated Big Data Center Collaborative Innovation System, 2021. Committee National Development and Reform, Office of the Central Cyberspace Affairs Commission, Technology Ministry of Industry and Information, et al., Beijing.

Improvement on TOU Prices Mechanism, 2021. National Development and Reform Commission (NDRC), Beijing.

Islam, M.A., Ahmed, K., Ren, S., et al., 2014. Exploiting temporal diversity of water efficiency to make data center less "thirsty". Int. Conf. Auton. Comput. 145–152.

Jiang, H., Li, B., Zhang, W., 2022. Promote Green and Low-Carbon Development of Data Centers. China Environment News. http://epaper.cenews.com.cn/html/2022-02/10/content_73778.htm.

Judge, P., 2021. Batteries of the Future. https://www.datacenterdynamics.com/en/analysis/batteries-of-the-future/.

Khosravi, A., Nadjaran Toosi, A., Buyya, R., 2017. Online virtual machine migration for renewable energy usage maximization in geographically distributed cloud data centers. Concurrency Comput. Pract. Ex. 29 (18), e4125.

Koronen, C., Åhman, M., Nilsson, L.J., 2020. Data centres in future European energy systems—energy efficiency, integration and policy. Energy Effic. 13 (1), 129–144.

Krzywda, J., Ali-Eldin, A., Carlson, T.E., et al., 2018. Power-performance tradeoffs in data center servers: DVFS, CPU pinning, horizontal, and vertical scaling. Future Generat. Comput. Syst. 81, 114–128.

Li, S., Brocanelli, M., Zhang, W., et al., 2013. Data Center Power Control for Frequency Regulation, 2013. IEEE Power & Energy Society General Meeting, IEEE, pp. 1–5.

Li, J., Bao, Z., Li, Z., 2014. Modeling demand response capability by internet data centers processing batch computing jobs. IEEE Trans. Smart Grid 6 (2), 737–747.

Liu, Z., Lin, M., Wierman, A., et al., 2011a. Greening geographical load balancing. Perform. Eval. Rev. 23 (2), 657–671.

Liu, Z., Lin, M., Wierman, A., et al., 2011b. Geographical load balancing with renewables. Perform. Eval. Rev. 39 (3), 62–66.

Liu, Z., Chen, Y., Bash, C., et al., 2012. Renewable and cooling aware workload management for sustainable data centers. In: Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems, pp. 175–186.

Liu, Z., Wierman, A., Chen, Y., et al., 2013. Data center demand response: avoiding the coincident peak via workload shifting and local generation. Perform. Eval 70 (10), 770–791.

Liu, L., Sun, H., Li, C., et al., 2016. RE-UPS: an adaptive distributed energy storage system for dynamically managing solar energy in green datacenters. J. Supercomput. 72 (1), 295–316.

Liu, Q., Chen, S., Wu, G., et al., 2018. Congestion management in cross-region grid considering spatially transferable characteristic of data center load. In: 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2). IEEE, pp. 1–6.

Luo, J., Rao, L., Liu, X., 2013. Temporal load balancing with service delay guarantees for data center energy cost optimization. IEEE Trans. Parallel Distr. Syst. 25 (3), 775–784.

McCluer, S., Christin, J.-F., 2008. Comparing data center batteries, flywheels, and ultracapacitors. White paper 65, 202.

Miller, R., 2018. The Power of Scale: Multi-Campus Deals Bring Energy Discounts. https://www.datacenterfrontier.com/energy/article/11430077/the-power-of-scale-multi-campus-deals-bring-energy-discounts.

Miller, R., 2019. Data Centers Becoming Larger Players in Power Markets. https://www.datacenterfrontier.com/energy/article/11429685/data-centers-becoming-larger-players-in-power-markets.

Nadeau, T.D., Gray, K.S.D.N., 2013. Software Defined Networks: an Authoritative Review of Network Programmability Technologies. O'Reilly Media, Inc.

National Energy Administration, 2022. Electricity Consumption of the Whole Society in, 2021. increased by 10.3% year-on-year. http://www.nea.gov.cn/2022-12/02/c_1310681372.htm.

National Energy Administration Released January-February National Electricity Industry Statistics, 2023. National Energy Administration (NEA), Beijing. http://www.nea.gov.cn/2023-03/21/c_1310704448.htm.

Peng, W., Liye, X., Ying, L., et al., 2017. Day-ahead emission-aware resource planning for data center considering energy storage and batch workloads. In: IEEE Conference on

Energy Internet and Energy System Integration (EI2). IEEE, Piscataway, NJ, USA, pp. 1–6.
Radovanovic, A., 2020. Our Data Centers Now Work Harder when the Sun Shines and Wind Blows. https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/.
Radovanovic, A., Koningstein, R., Schneider, I., et al., 2022. Carbon-aware computing for datacenters. IEEE Trans. Power Syst., 2022, 38(2): 1270-1280.
Ren, S., 2013. Batch job scheduling for Reducing Water footprints in data center, 2013. In: 51st Annual Allerton Conference on Communication, Control, and Computing, IEEE, pp. 747–754.
Salomonsson, D., Soder, L., Sannino, A., 2007. An Adaptive Control System for a DC Microgrid for Data Centers, 2007. IEEE Industry Applications Annual Meeting, IEEE, pp. 2414–2421.
Sawyer R. L. Calculating Total Power Requirements for Data Centers - White Paper 3. Schneider Electric's Data Center Science.
Urgaonkar, R., Urgaonkar, B., Neely, M.J., et al., 2011. Optimal power cost management using stored energy in data centers. In: Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems, pp. 221–232.
Wang, Z., 2022. China's Total Computing Power Ranks Among the Top in the World. People's Daily, Beijing in Chinese. http://www.gov.cn/xinwen/2022-08/05/content_5704300.htm.
Wang, L., Von Laszewski, G., Dayal, J., et al., 2010. Towards energy aware scheduling for precedence constrained parallel tasks in a cluster with DVFS. In: 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, IEEE, pp. 368–377.
Wang, N., Zhang, J., Xia, X., 2013. Energy consumption of air conditioners at different temperature set points. Energy Build. 65, 412–418.
Wang, S., Qian, Z., Yuan, J., et al., 2017. A DVFS based energy-efficient tasks scheduling in a data center. IEEE Access 5, 13090–13102.
Wang, W., Abdolrashidi, A., Yu, N., et al., 2019. Frequency regulation service provision in data center with computational flexibility. Appl. Energy 251.
Wang, Q., Liu, Y., Huang, Y., et al., 2020a. Congestion management of urban power grid considering demand response of data center. Power Syst. Technol. 44 (8), 3129–3138.
Wang, P., Cao, Y., Ding, Z., 2020b. Flexible multi-energy scheduling scheme for data center to facilitate wind power integration. IEEE Access 8, 88876–88891.
Xiang, X., 2014. Greening geo-distributed data centers by joint optimization of request routing and virtual machine scheduling. IEEE/ACM Int. Conf. Util. Cloud Comput., IEEE 1–10.
Yang, K., 2022. Achievements and Prospects of China's Electric Power Development in the New Era. https://cec.org.cn/detail/index.html?3-314599.
Ye, Q., Jiaqi, L., Mengye, Z., 2018. Wind Curtailment in China and Lessons from the United States. China's Energy in Transition.
Ye, R., Yuan, Y., Wei, J., et al., 2021. Decarbonization of Digital Infrastructure in China: Data Centers and 5G Carbon Reduction Potential and Challenges (2020-2035). Greenpeace, Beijing.
Yin, I., Yep, E., 2022. China Could Exceed Renewables Generation Target of 33% by 2025. S&P Global. https://www.spglobal.com/commodityinsights/en/market-insights/latest-news/agriculture/042023-indias-palm-oil-imports-to-taper-down-in-april-june-quarter-traders.
Yu, L., Jiang, T., Cao, Y., et al., 2015. Joint workload and battery scheduling with heterogeneous service delay guaranteesfor data center energy cost minimization. IEEE Trans. Parallel Distr. Syst. 26 (7), 1937–1947.
Zhang, E., Chen, M., 2020. Five Things to Know about China's Mega East-West Data Center Plan. https://www.sixthtone.com/news/1010207.