

分类号\_\_\_\_\_

密级\_\_\_\_\_

U D C\_\_\_\_\_

编号\_\_\_\_\_

清 华 大 学

## 博 士 后 研 究 工 作 报 告

能源互联网需求侧分布式灵活资源聚合机制及优化调节策略研究

刘 迪

工作完成日期 2020年9月—2022年9月

报告提交日期 2022年9月

清 华 大 学 （北京）

2022年9月

# 能源互联网需求侧分布式灵活资源聚合机制及优化调节策略研究

## RESEARCH ON AGGREGATION MECHANISM AND OPTIMAL ADJUSTMENT STRATEGY OF DISTRIBUTED FLEXIBLE RESOURCES ON DEMAND SIDE IN ENERGY INTERNET

博 士 后 姓 名： 刘 迪

流动站（一级学科）名称： 控制科学与工程

专 业（二级学科）名称： 控制理论与控制工程

合 作 导 师： 吉吟东 教授

研究工作起始时间 2020年9月30日

研究工作期满时间 2022年9月30日

清华大学自动化系（北京）

2022年9月

## 内 容 摘 要

伴随着我国以新能源为主体的新型电力系统的建设，我国的可再生能源装机容量快速上升。可再生能源的大量渗透，使电力系统供给侧的灵活性迅速下降，对电网的安全稳定运行造成了极大的威胁，亟需调动需求侧的灵活性资源，支撑电网的安全稳定运行。

目前而言，需求侧大型工商业用户的优化调节已经被许多国家和地区广泛实施，但居民、商业等分布式的中小型用户尚未能很好地参与电网的调节。海量的中小型用户聚合后，能够形成相当可观的调节能力，但需求侧用户具有主体多、数量大、差异性大、不确定性强等特点，因此需求侧用户的聚合调节需要从机制设计和算法优化两个方面实现。

基于上述分析，本报告关注需求侧分布式的用户群体，从不同的角度设计灵活聚合机制及相应的优化调节策略，实现需求侧的高效用能。具体的工作点如下：

(1) 针对用户差异化的特点，提出了一种需求响应增量激励机制及异步双交互深度强化学习算法，提高消费者的响应深度，同时降低单位激励成本。

(2) 针对分布式储能归属于不同主体的现状，提出了两种分布式储能的共享机制和优化策略：1) 通过负荷服务商的激励实现储能共享；2) 通过多主体的主动协同实现储能共享，两种共享机制均能有效实现提升储能利用效率，提升用户收益的同时更多地消纳本地的可再生能源。

(3) 面向分布式数据中心，构建了分布式数据中心协同优化调节架构，以及基于动态微分方程的信息层和能量层耦合优化问题的统一调节模型，引入最优控制理论实现数据中心信息能量的秒级协同优化控制，利用数据中心负荷地理迁移的能力提升可再生能源的消纳率。

(4) 针对电力系统碳排放问题，提出了需求侧的电-碳耦合交易框架，利用“虚拟储碳”实现需求侧碳排放的灵活流转，进而从用户视角出发，提出优化竞价策略，通过需求侧的调节实现碳减排。

关键词：需求响应，可再生能源消纳，深度强化学习，分布式储能

## Abstract

With the construction of a new power system with new energy as the main body, the installed capacity of renewable energy in China has risen rapidly. The massive penetration of renewable energy makes the flexibility on the supply side of the power system drop rapidly, posing a great threat to the safe and stable operation of the grid. It is urgent to mobilize the flexibility resources on the demand side to support the safe and stable operation of the grid.

At present, the optimal regulation of large industrial and commercial users on the demand side has been widely implemented in many countries and regions, but residential, commercial, and other distributed small and medium-sized users have not yet been well involved in the regulation of the power grid. After the aggregation of many small and medium-sized users, considerable adjustment capacity can be formed. However, demand side users are characterized by many subjects, large number, great diversity, and strong uncertainty. Therefore, the aggregation adjustment of demand side users needs to be realized from two aspects: mechanism design and algorithm optimization.

Based on the above analysis, this report focuses on the demand side distributed user groups, and designs flexible aggregation mechanisms and corresponding optimization and adjustment strategies from different perspectives to achieve efficient energy use on the demand side. The specific work points are as follows:

(1) According to the characteristics of user differentiation, a demand response incremental incentive mechanism and an asynchronous dual-interaction deep reinforcement learning algorithm are proposed to improve the depth of consumer response and reduce the unit incentive cost.

(2) In view of the fact that distributed energy storage belongs to different entities, two sharing mechanisms and optimization strategies of distributed energy storage are proposed: 1) energy storage sharing is realized through the incentive of load service providers; 2) Through the active collaboration of multiple agents, energy storage can be shared. Both sharing mechanisms can effectively improve the efficiency of energy storage utilization, increase user income and consume more local renewable energy.

(3) For the distributed data center, the collaborative optimization and regulation architecture of the distributed data center is constructed, as well as the unified regulation model of the coupling optimization problem between the information layer and the

energy layer based on the dynamic differential equation. The optimal control theory is introduced to achieve the second level collaborative optimization control of the information energy of the data center, and the capacity of the data center to geographically migrate the load is used to improve the consumption rate of renewable energy.

(4) Aiming at the problem of carbon emissions in the power system, a demand side electricity carbon coupling trading framework is proposed, which uses "virtual carbon storage" to achieve the flexible flow of carbon emissions on the demand side. From the perspective of users, an optimized bidding strategy is proposed to achieve carbon emission reduction through demand side regulation.

**Keywords:** demand response, renewable energy consumption, deep reinforcement learning, distributed energy storage

## 目 录

1	能源互联网需求侧资源优化概述.....	1
1.1	面向分布式用户的需求响应研究现状.....	1
1.2	分布式储能共享优化研究现状.....	3
1.3	数据中心优化调节研究现状.....	4
1.4	需求侧碳电耦合市场机制研究现状.....	5
1.5	基于深度强化学习的优化求解方法研究现状.....	5
1.6	本报告工作概要.....	6
2	面向差异化用户的增量激励机制研究.....	8
2.1	系统和模型.....	9
2.1.1	增量激励机制制定.....	9
2.1.2	消费者建模.....	10
2.1.3	LSE 建模.....	11
2.2	增量激励机制分析.....	13
2.2.1	优越性分析.....	13
2.2.2	效率和公平性分析.....	16
2.3	问题制定和解决方案.....	18
2.3.1	决策动作时序及求解方法.....	18
2.3.2	马尔可夫决策过程.....	20
2.3.3	异步双交互 DRL 算法.....	21
2.4	仿真分析.....	24
2.4.1	仿真参数设置.....	24
2.4.2	所提机制的效果分析.....	25
2.4.3	所提算法的可靠性分析.....	27
2.5	本章小结.....	29
3	面向分布式储能的共享机制及优化策略.....	30
3.1	基于主从博弈的储能共享方式.....	30

---

3.1.1	分布式储能系统架构与共享机制 .....	31
3.1.2	产消者和微电网运营商的建模 .....	33
3.1.3	优化问题制定和解决方法 .....	36
3.1.4	仿真实验 .....	41
3.2	基于合作博弈的储能共享方式.....	50
3.2.1	微电网联盟架构 .....	50
3.2.2	区域电力市场 .....	50
3.2.3	模型构建 .....	52
3.2.4	基于 soft actor critic 的竞价策略 .....	53
3.2.5	仿真分析 .....	56
3.3	本章小结.....	58
4	分布式数据中心信息能量协同优化策略 .....	60
4.1	分布式数据中心架构.....	61
4.2	分布式数据中心信息能量模型.....	62
4.2.1	数据中心能耗控制模型 .....	62
4.2.2	多数据中心间计算任务分配模型 .....	64
4.2.3	考虑延迟的信息能量耦合模型 .....	64
4.3	问题构建及求解.....	65
4.3.1	问题构建 .....	65
4.3.2	问题求解 .....	68
4.4	仿真实验.....	69
4.4.1	仿真环境设置 .....	69
4.4.2	控制频率影响分析 .....	71
4.4.3	经济性与延迟优化分析 .....	75
4.5	本章小结.....	77
5	需求侧电-碳耦合交易机制.....	78
5.1	电-碳耦合市场模型 .....	79

---

5.1.1	系统结构 .....	79
5.1.2	虚拟碳存储 .....	81
5.1.3	结算价格的确定 .....	82
5.1.4	电量和碳强度的确定 .....	82
5.2	问题构建.....	83
5.2.1	通过虚拟碳储存调整碳强度 .....	83
5.2.2	用户模型 .....	84
5.2.3	目标函数 .....	85
5.3	优化问题的求解.....	86
5.3.1	基于 MDP 的强化学习.....	86
5.3.2	近端策略优化 .....	88
5.4	仿真分析.....	90
5.4.1	仿真设置 .....	90
5.4.2	成本分析 .....	91
5.4.3	碳减排分析 .....	93
5.4.4	不确定性分析 .....	94
5.5	本章小结.....	95
6	总结与展望.....	96
	参考文献.....	97
	项目及研究成果.....	108
	个人简介.....	109

# 1 能源互联网需求侧资源优化概述

## 1.1 面向分布式用户的需求响应研究现状

在许多国家，由于电力市场管理大量参与交易的实体的能力有限，相对较小的消费者无法直接参与电力批发市场。独立消费者必须通过负荷服务实体（Load service entity, LSE）参与电力批发市场<sup>[1]</sup>。一般情况下，为了在零售电力市场中向消费者提供能源供应服务，LSE 通过投标和竞价从电力批发市场中获得电力。LSE 的经济效益可以通过需求响应得到提升。

需求响应被认为是利用需求侧可调整资源提高微电网灵活性的有效手段，它主要包括两种类型：一种是基于价格的需求响应（例如，[2]-[5]），另一种是基于激励的需求响应（例如，[6]-[8]）。在基于价格的需求响应领域，由于分时电价较为稳定，便于实施，因而被广泛应用。在实践中通过分时电价释放用户的时间互补特性，以降低社会成本<sup>[2][3]</sup>。具有更高灵活性的实时价格也是许多研究人员关注的问题，并且通过适当的机制和在线优化方法，引导用户改变电力负荷的消费行为以达到理想的状态<sup>[9]</sup>。

基于激励的需求响应可以为系统运营商提供灵活的调度资源，有利于需求侧不同类型的灵活资源的协同优化<sup>[7]</sup>。基于激励的需求响应通常在高峰负荷期间实施，消费者根据其负荷的响应情况直接获得补贴<sup>[10]</sup>。由于在需求响应中被削减的负荷，包括电动汽车等可转移负荷以及温控负荷等可延迟的负荷，因此在需求响应中，负荷的削减可能会引起后续时段的负荷反弹，从而可能在更长时间内造成负荷曲线的劣化。因此在激励型需求响应的研究中，不但要考虑本时段的供需平衡，也要考虑其对后续时段的影响，目前已经有一些研究关注到了此问题，例如文献[9],[11]-[13]。

与基于价格的需求响应相比，消费者可以直接获得基于激励的需求响应的收入，因此具有更高的主动性<sup>[14]</sup>。此外，基于激励的需求响应可以为电网提供更灵活的可调度资源，例如，美国的激励型需求响应贡献了高峰时期 93% 的负荷削减量<sup>[15]</sup>。因此，面向家庭负荷的激励型需求响应需要被重点关注。

在激励型需求响应中，LSE 通过给予不同的激励信号，引导用户改变消费行为。当用户认为其舒适度损失可以被弥补时，就会参与需求响应。根据微观经济学中的消费者行为理论，消费者的边际成本随着响应深度的累积而增加，并且不

同消费者之间的响应行为具有很大的差异性，但是在许多关于激励型需求响应的现有研究中（例如，[16][17]），消费者获得的激励均相同，并没有根据其差异性获得不同的激励，这就限制了消费者的响应深度，并进一步降低了 LSE 从激励型需求响应中获得的利润。

为了满足消费者的用电需求，LSE 还需要制定在电力市场中购电的定价策略。在实践中，电力市场中的竞价通常在日前进行，而激励型需求响应则在日内实施。在电力市场中的竞价策略和日内的激励型需求响应相互影响。但是在日前的竞价中，LSE 并不能提前获知需求响应策略，从而其需要购买的电量，也不能被提前获知。因此 LSE 在日前电力市场中需要在信息不完整的环境下做出竞价决策。

具体而言，有两个关键问题需要被解决，第一，是如何在保证公平性的前提下，充分利用消费者差异化的特点，以提升整体的效益；第二，是综合考虑日前竞价和日内激励型需求响应的最优决策方法。

根据现有的研究成果，消费者的响应特征受到许多因素的影响，并且不同消费者之间有着很大的差异性<sup>[18][19]</sup>。文献[20]中考虑了消费者类型的差异，并分别针对工业用户和住宅用户制定不同的激励措施，以引导不同类型用户改变电力消费行为，从而最大限度地降低系统整体的用电成本。而针对统一类消费者，文献[21]也进一步将其分类，并根据每类用户的响应深度给予不同的激励，文献[22]中也有类似的研究成果。但是，每一类内的用户还是获得一样的激励，即虽然上述研究考虑了消费者类间的差异性，但类内消费者的差异性被忽略了，从而限制了消费者响应潜力的释放。

为了解决这个问题，文献[23]提出了一种更加灵活的激励机制，首先根据类型设置消费者不同的权重，然后根据每个消费者在激励型需求响应中的贡献来计算补贴。类似的，文献[24]中提出了一个面向住宅用户的激励型需求响应的框架，其中聚合商根据消费者的响应行为基于不同的补贴。在文献[25]的研究中，消费者也能够根据自己在需求响应业务中的贡献情况获得不同的补贴。

虽然上述激励型需求响应机制充分利用了消费者的差异性，但仍存在一些不足之处。当消费者数量庞大时，激励型需求响应的实施面临两难境地。为每位消费者制定差异化激励，虽然可以最大限度地发挥激励型需求响应的效果，但在实践中几乎是无法实施的，因为很难获得每个消费者的详细特征信息，计算效率也

面临很大的挑战。并且，对消费者的差异化激励可能会导致公平问题。

此外，消费者的舒适性损失函数是非线性和凹陷的<sup>[18]</sup>，即单位响应引起的舒适性损失随着响应深度的增加而增加。在现有机制中，消费者每单位响应得到相同的激励，因此现有机制无法跟踪消费者舒适度损失的变化，导致用户响应的初始阶段，消费者剩余过高<sup>[26]</sup>，从而降低了激励型需求响应的效率和 LSE 的利润。

## 1.2 分布式储能共享优化研究现状

储能被认为是应对可再生能源发电波动的有效手段，其安装量在全球范围内迅速增加<sup>[27]</sup>。由于储能可以归微电网运营商或产消者所有，为了提高储能的利用效率，现有的研究中提出了储能容量共享的机制，主要分为集中式储能共享和分布式储能共享两种模式。

在集中式模式下，储能由微电网运营商或独立的储能运营商投资和运营，产消者购买所需的储能容量。使用 Stackelberg 博弈论来分析参与者之间的关系<sup>[28]</sup>，市场框架旨在最大化系统的收入<sup>[29]</sup>。文献[30]提出了一种装备了储能的单微电网离线优化方法，将常规电网的能量成本降至最低。文献[31]中提出了一种针对微电网的双层能源管理系统，储能用于最小化总运营成本，并处理可再生能源的不确定性。还有许多其他的文献在做类似的研究，例如文献[32]-[35]。

在分布式模式下，储能由每个产消者拥有，容量可以由产消者通过激励或交易共享<sup>[36]-[38]</sup>。受电力供需变化的影响，微电网运营商在每个时隙中所需的共享储能容量各不相同。过剩的共享储能容量会导致容量资源的浪费，而共享储能容量不足会影响微电网运营商的调整。然而，由于不确定性的影响，微电网运营商所需的共享储能容量和产消者可以共享的储能容量都有波动，如何在每个时间段获得适当的共享储能容量在现有研究中尚未得到很好的解决。

储能装置通常被装在微电网中，以满足微电网中负荷的需求，同时消纳微电网中装设的光伏出力<sup>[39]</sup>。文献[40]中考虑了储能使用中的折旧成本，提出了一种双层能量管理方法，以最大限度的降低微电网的运营成本。为了应对可再生能源波动的影响，在文献[41]中提出了一种多储能系统的能量管理方法，在保证微电网内部功率平衡约束的同时，实现成本的最优化。文献[42]考虑了多个微电网之间的能量交互，通过设计交易机制来进一步提升微电网社区的运行效率。文献[43]则考虑了系统的鲁棒性，通过鲁棒优化实现了系统安全性和经济性之间的平衡。

由于电力负荷的波动以及光伏出力的波动，每个微电网所装设的储能并不能被充分的利用，如果微电网闲置的储能容量可以与其他微电网共享，那么微电网群体的总体运营成本可以被进一步降低<sup>[44]</sup>。一个经济储能的共享框架在文献[45]中被提出，参与者通过竞价实现储能的共享。在文献[46]中，博弈论被用于建立储能的共享机制，每个代理需要同时确定容量交易和储能的充放电调度。为了反映终端用户之间的差异，在文献[47]中提出了基于信用的容量共享方法，实现了系统总体效益的提升。

虽然储能的利用可以提高光伏的本地消纳率，但不能完全满足微电网的负荷需求。因此，微电网还需要参与电力市场以购买所需电力。通常，一个微电网的投标不会影响市场的出清价格，但如果多个微电网组成联盟作为独立实体参与电力市场，其负荷需求足够大，其投标就可以影响市场的结算价格，从而微电网联盟被称为价格制定者。市场参与者作为价格制定者的投标策略也受到一些研究的关注。文献[48]中开发了一个规划模型，用于确定价格制定者-零售商在市场中的竞价策略。同样，在文献[49]中提出了一个基于情景的随机优化框架，用于价格制定者的竞价，并在研究中考虑了日前市场和实时市场。文献[50]研究了实时电力市场中价格制定者能源负荷的多阶段需求响应，提高了能源负荷的利润。由于其他参与者的竞价和报价不能事先知道，因此文献[51]中引入了概率分布函数来描述需求价格配额曲线，而文献[52]中使用鲁棒优化来处理电力市场中的不确定性。

### 1.3 数据中心优化调节研究现状

目前的研究中，主要从数据中心本身的能效提升<sup>[53]</sup>以及数据中心与电网协同优化<sup>[54]</sup>两方面展开。针对高能耗问题，除了从数据中心建设本身提升数据中心能源利用率外，在满足时延要求的前提下适当降低数据中心的服务器率，能够在一定程度上降低数据中心的能耗<sup>[55][56]</sup>。针对能源供给中可再生能源占比低的问题，可以通过计算任务在不同数据中心之间的灵活分配实现负荷的迁移<sup>[57]</sup>，以更多地消纳可再生能源。具体而言，可以通过直接负荷控制、电价信号等手段<sup>[58]</sup>，引导数据中心参与电网需求响应，以平抑新能源波动、缓解电网阻塞，提升系统运行效率<sup>[59]</sup>。

文[60]针对混合供电的数据中心的能耗优化问题，分别从数据中心本身能效的提升以及多个数据中心负载均衡等角度出发，综述了目前的最新研究成果，并

对数据中心能耗与可再生能源出力的联合调节提出了研究展望，但并未深入地研究。文[61]提出了一种多数据中心流量分配的实时算法，通过计算任务的优化分配，提升数据中心的长期效益。文[62]在此基础上，进一步考虑了时延等因素，提出了分布式的多数据中心流量分配实时算法，以最小化多个数据中心的长期运营成本。文[63]建立了数据网络与电力网络混合运行模型，将数据中心任务分配纳入电网经济运行的优化调度中。

#### 1.4 需求侧碳电耦合市场机制研究现状

现有的许多研究表明，市场的交易可以引导供需双方匹配得到合适的出清价格，以提高市场的总体效益。文献[64]中提出了批发电力市场的动态市场机制，通过交易可以实现预期的系统效率。为了有效利用小规模住宅消费者的需求侧灵活性，在文献[65]中提出了一个两阶段的能源交易市场模型，其中需求侧灵活性可以在本地进行交易。在文献[66]中也可以找到类似的研究成果，其中提出了一个分层交易能源市场，以实现需求侧灵活性资源的协作。

然而，上述研究忽略了碳排放成本。尽管碳排放直接由发电厂所产生，但发电厂的出力情况取决于负荷的消费情况，因此碳排放成本最终应由电力消费的用户所承担<sup>[67]</sup>。目前已经有一些研究，试图将碳排放成本与电价成本相结合。文献[68]中提出了能源-碳综合价格机制，其中碳价格和电价分别建模，两者之和为能源消费的总价格。文献[69]中比较了电力碳价格模型的不同公式，并证明基于碳排放流量的电力碳价格模式具有更好的性能，因为它能够反映用户的实际排放贡献。碳排放流量是电力系统中从供应侧到需求侧的碳排放责任的量，根据基于比例共享定理，按照电力系统的实际潮流进行计算<sup>[69]-[72]</sup>。需要指出的是，尽管现有的研究关注了碳排放责任从供给侧到需求侧的转移过程，但是目前很少有研究关注需求侧之间碳排放责任的交易与转移。

#### 1.5 基于深度强化学习的优化求解方法研究现状

在需求响应中，分布式的用户需要通过 LSE 参与电力市场，因此 LSE 一方面需要制定面向消费者的需求响应激励，另一方面需要从电力市场中购电以满足消费者的电力消费需求。随着储能的发展，分布式储能的共享与使用与需求响应的相互耦合，给优化问题的求解带来了更大的挑战。从 LSE 的视角看，耦合环境中的策略优化也是需要解决的重要问题之一。在目前的许多研究中，主要是将该

耦合问题转换为具有平衡约束（MPAC）的数学规划问题<sup>[73][74]</sup>。为了实现信息不完整的竞价决策，目前的一些研究中引入了迭代方法<sup>[75][76]</sup>。

然而，以住宅消费者为代表的分布式用户的多样性使得在实践中，难以选择合适的模型并识别相应的参数来准确描述每个消费者的响应行为，此外电力市场价格、可再生能源出力等多重不确定性的叠加，使得优化问题的求解显得更加困难。因此，在模型不明确以及多重不确定性叠加的环境下，传统的求解方法难以实施。无模型的深度强化学习，算法可以利用历史数据，而不需要显性的模型对优化问题进行求解，很适合用来解决不确定性环境下耦合问题的优化求解问题。深度强化学习算法已经在电力系统中得到了深度强化学习算法已经在电力系统中得到了应用并取得了良好的效果，如电池充电控制、电机控制、电压控制等<sup>[77]-[79]</sup>。在需求侧调节的耦合问题优化求解中也有了一些尝试并被证明是有效的<sup>[80][81]</sup>。但是上述优化问题或多或少对求解问题进行了简化，因此基于深度强化学习的优化方法，在需求侧管理中的应用，还需要被进一步深入研究。

## 1.6 本报告工作概要

本报告面向需求侧的灵活性资源开展研究，第一章对国内外现状进行综述，分别从分布式用户需求响应研究现状、分布式储能共享优化研究现状、数据中心优化调节研究现状、需求侧碳电耦合市场机制研究现状，以及基于深度强化学习的优化求解方法研究现状几方面展开论述。

第二章面向需求侧差异化的用户，提出了一种基于激励的需求响应增量激励机制，其中消费者根据响应增量获得不同的激励，从而使激励能够跟随消费者边际成本的变化。我们从理论上说明，与其他现有激励机制相比，所提出的增量激励机制可以有效提高负荷服务实体的利润，以及高度灵活的消费者的利益。在实践中，负荷服务实体在日前市场的竞价策略受到日内激励型需求响应策略的影响，该策略无法提前知道。为了解决日前市场中信息不完全的竞价问题，提出了一种异步双交互深度强化学习算法，以最大化负荷服务实体在全天多个时隙的累积利润。仿真结果表明，所提出的机制可以提高消费者的响应深度，同时降低单位激励成本，并且所提出的深度强化学习算法即使在高度不确定的环境中也具有相对稳定和令人满意的性能。

第三章提出了两种储能共享模式，集中激励模式和分布协作模式。在集中激

励模式下，提出了一种两阶段储能共享机制，其中售电商在日前通过激励集合空闲储能容量，为实时优化提供可靠资源。为了处理多重不确定性的影响，采用蒙特卡罗抽样，以确保在任何情况下共享储能容量足够。在分布式协作模式下，以微电网为例，提出了一种基于中央协调器的协作框架。由于微电网联盟大到足以影响市场中的清算价格，应用了深度强化学习算法，实现市场竞价和储能控制的协同优化。

第四章面向地理分布式的数据中心，构建了分布式数据中心协同优化调节架构，并分析了多数据中心计算任务分配与单数据中心功率优化的动态特性。其次，构建了基于动态微分方程的信息层和能量层耦合优化问题的统一调节模型。最后，综合考虑系统运营成本及计算时延构建目标函数，引入最优控制理论对该问题求解，实现数据中心信息能量的秒级协同优化控制。仿真结果表明，相比分钟级的控制，基于该策略的快速控制能够较好的追踪可再生能源出力以及计算任务的波动，从而有效提升系统的经济效益及可再生能源就地消纳率。

第五章考虑需求侧碳减排的问题，设计了一个联合电力碳交易框架，通过交易和需求侧管理减少碳排放。在所提机制内，提出了“虚拟碳储存”的概念，以提高用户之间碳排放交易的灵活性。为了在所提出的机制内制定交易和需求侧管理的最佳策略，提出了基于蒙特卡罗的近端策略优化（PPO）算法。仿真研究表明，与没有碳交易的电力市场和没有虚拟碳存储的用户相比，所提出的机制可以分别实现 27%和 31.7%的碳减排，同时将经济成本保持在可接受的低水平。仿真还表明，所提出的蒙特卡罗采样方法可以显著提高算法的收敛速度。

## 2 面向差异化用户的增量激励机制研究

需求响应已成为智能电网的关键部分，旨在鼓励消费者调整用电行为以提高能源利用效率<sup>[82]</sup>。一般来说，大型工商消费者被认为是需求响应业务的较好的参与者，因为每个消费者都能提供可观的响应，而且消费者的数量不是特别大，有利于需求响应的实施<sup>[83]</sup>。然而，在许多国家，具有巨大需求侧灵活性的住宅消费者约占峰值负荷的 50%，而其灵活性潜力的利用不到 2%<sup>[84]</sup>。由于每个住宅消费者只能提供有限的响应，因此需要聚集大量消费者以形成足够大的响应能力。

目前的研究在激励机制和优化策略方面都存在不足。在激励机制方面，激励型需求响应中对消费者差异的充分利用和计算效率之间的矛盾尚未得到很好的解决。由于在现有激励机制中，增量激励成本与消费者的总响应呈正相关，因此高度灵活的消费者的响应深度有限。此外，消费者剩余的不均衡分布也降低了负荷服务实体（Load service entity, LSE）的利润。在优化策略方面，由于日内激励型需求响应策略和日前投标策略相互影响，并且在实践中在不同的时间段执行，需要提出优化策略，以在日前投标中，未知日内需求响应决策的情况下，实现优化决策。

本文的主要贡献是提出一种激励机制，以提高差异化住宅消费者的激励型需求响应效率，同时确保公平性和计算效率。通过数学分析和仿真验证了所提出的增量激励机制的优越性、效率和公平性。同时，提出了一种异步双交互深度强化学习算法来解决无激励型需求响应信息的投标优化问题。主要的重要性和贡献如下：

(1) 提出了一种增量激励机制，其中将激励型需求响应中的增量单位激励成本与总响应量解耦。同时从理论上证明，现有研究中的统一激励机制是本文所提出的增量激励机制的一种特殊形式。

(2) 与现有研究中的统一激励机制相比，所提出的机制可以提高激励型需求响应中高度灵活的消费者的响应深度，并可以通过跟踪消费者响应弹性的变化来平衡每单位增量响应的消费者剩余，以避免过度的消费者剩余，从而提高激励型需求响应中的 LSE 利润。

(3) 提出了一种基于深度确定性策略梯度（deep deterministic policy gradient, DDPG）的异步双交互深度强化学习算法，以解决不完全信息下的竞价优化问题，

以及在实践中无法准确识别多样化消费者模型参数的困难，从而最大化 LSE 全天多个时段的累积利润。

## 2.1 系统和模型

本文针对 LSE 的优化决策问题展开研究。其在电力零售市场中为多个消费者提供能源供应服务，同时实施激励型需求响应以提升其自身效益。假设一天被分解为  $T$  个时段，如图 1 所示，LSE 向消费者提供电力供应服务，并在电力批发市场中参与投标，以竞得所需的电力。当电力批发市场中的出清价格较高时，LSE 可以实施激励型需求响应，引导消费者降低该时段的负荷需求，从而提高 LSE 全天的利润。需要指出的是，LSE 作为能源供应商，在任何情况下都需要满足消费者的电力需求，否则将会面临因电力不平衡而造成的高额罚款。

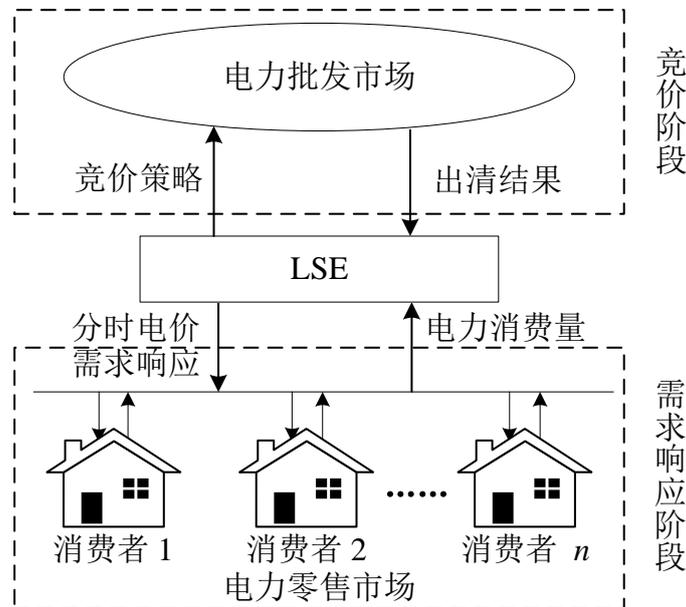


图 2-1 竞价及激励型需求响应业务架构

### 2.1.1 增量激励机制制定

根据经济学中的需求弹性理论，激励型需求响应中消费者的响应弹性  $E$  可以用如下公式来表示<sup>[85]</sup>：

$$E = \frac{dR/R}{dI/I}, \quad (2-1)$$

其中  $R$  是消费者的响应量， $I$  是激励价格，由于消费者舒适度的损失，消费者的响应弹性  $E$  随着响应量的上升而降低。同时，由于消费者的多样性， $E$  在消费者中的变化趋势各不相同。然而，在现有的激励机制中，消费者在激励型需求响应

事件中的所有累积响应都以相同的激励价格结算，这意味着 $E$ 的变化无法被完全反映。

为了在充分利用消费者差异的同时，对消费者响应弹性的变化进行追踪，本文提出一种增量激励机制，即消费者根据响应增量获得收益：

$$U_{i,t}(R_{i,t}) = \begin{cases} \int_0^{R_{i,t}} f(x)dx, & R_{i,t} > 0, \\ 0, & \text{others,} \end{cases} \quad (2-2)$$

其中 $f$ 增量激励函数，并且单调递增，其使得消费者获得的单位收入随着响应深度的增加而增加。需要指出的是，任何满足上述要求的函数都可以在所提出的机制中用作增量激励函数，例如线性函数，多项式函数，指数函数等。

### 2.1.2 消费者建模

消费者根据其电力需求在每个时段从 LSE 购买所需电力，并参与激励型需求响应以提高其自身效益。假设时隙  $t$  中第  $i$  个消费者的基本负荷为  $L_{i,t}^a$ ，即激励型需求响应中消费者响应的基线。基于此，时隙  $t$  中第  $i$  个消费者的响应  $R_{i,t}$  为：

$$R_{i,t} = L_{i,t}^b - L_{i,t}^a, \quad (2-3)$$

其中， $L_{i,t}^a$  是时隙中第  $i$  个消费者在时隙  $t$  中的真实负荷。

当消费者在时隙  $t$  中参与激励型需求响应时，由于可转移负载（如温度控制负载、电动车辆等）的影响，部分负载被转移到后续时隙<sup>[11]</sup>。由于消费者的负荷消费行为是周期性的，他们的负荷反弹在同一时间段内具有相似的特征<sup>[86]</sup>，可以表示为：

$$L_{i,t}^r = \sum_{j=0}^{t-1} \xi_{i,j} R_{i,j}, \quad (2-4)$$

式中， $L_{i,t}^r$  是第  $i$  个耗电设备的负荷反弹量， $\xi_{i,j} \in [0,1]$  是第  $i$  个耗电设备的负荷相关系数，表示前一个时隙的负荷减少与后一时隙负荷反弹之间的关系。应注意的是，反弹的负荷量在实践中无法被直接测量，只能测量到包含了负荷反弹后的真实用电量。

参与激励型需求响应造成的舒适度损失与消费者的响应量和灵活性有关，可以表示为：

$$C_{i,t}^c = \int_0^{R_{i,t}} l_i(\tilde{R}_{i,t}) d\tilde{R}_{i,t}, \quad (2-5)$$

其中,  $l_i(\tilde{R}_{i,t})$  是第  $i$  个消费者的舒适度损失函数。

根据经济学原理, 当负荷需求减少时, 边际成本, 即消费者每单位响应的舒适度损失增加, 因此  $l_i(\tilde{R}_{i,t})$  是单调递增的函数, 其形式和参数根据消费者的特征而变化。

第  $i$  个消费者在  $t$  时间段内的购电成本为:

$$C_{i,t}^p = \lambda_t^{TOU} L_{i,t}^a, \quad (2-6)$$

其中  $\lambda_t^{TOU}$  为分时电价, 由 LSE 事先确定。

消费者通过解决以下成本最小化问题来获得每个时隙  $t$  中的最佳决策:

$$\min_{L_{i,t}^a} [C_{i,t}^p + C_{i,t}^c - U_{i,t}], \quad s.t. L_{i,t}^a \geq 0, \quad (2-7)$$

其中  $U_{i,t}$  是在每个时隙中参与激励型需求响应的第  $i$  个消费者的收入, 并且可以根据 (2-2) 获得。

### 2.1.3 LSE 建模

LSE 需要优化电力批发市场中的出价策略和电力零售市场中的激励型需求响应策略。在电力批发市场中, LSE 通过竞价购买电量, 竞价策略  $\pi_b(t)$  可以用单调递增函数表示<sup>[87]</sup>:

$$p_t = \alpha_t + \beta_t L_t, \quad (2-8)$$

其中  $L_t$  是 LSE 预计在时隙  $t$  内以价格  $p_t$  购买的电力,  $\alpha_t$  确定 LSE 的最高购买价格,  $\beta_t \leq 0$  决定投标曲线的趋势。电力批发市场出清后, 交易单价高于或等于电力批发市场中出清价格的投标电量获得成交。

根据 (2-8), 成交的电量  $L_t^W$  可以表示为:

$$L_t^W = (\lambda_t^W - \alpha_t) / \beta_t, \quad (2-9)$$

其中  $\lambda_t^W$  是电力批发市场中的出清价格, 在竞争性电力市场中, 其结果不受单个 LSE 投标的影响。从而可以计算时隙  $t$  中电力批发市场中 LSE 的成本为:

$$C_{W,t} = \lambda_t^W (\lambda_t^W - \alpha_t) / \beta_t. \quad (2-10)$$

消费者的电力需求应始终得到满足,如果竞价获得的电量 $L_t^W$ 不能满足消费者的电力需求, LSE 必须根据负荷不平衡的程度向独立系统运营商 (independent system operator, ISO) 支付罚款。由于负荷不平衡造成的罚款如下<sup>[80]</sup>:

$$C_{S,t} = \phi \left( \sum_{i=1}^n L_{i,t}^a - L_t^W \right), \quad (2-11)$$

其中 $C_{S,t}$ 是 LSE 向 ISO 支付的不平衡罚款, 其与电力的不平衡程度成正相关关系。从而 LSE 需要满足的功率平衡约束为:

$$L_t^W + L_t^S = \sum_{i=1}^M L_{i,t}^a, \quad (2-12)$$

在 LSE 向 ISO 支付不平衡罚款后, 负荷的不平衡量由 ISO 负责弥补。此外, 任何时隙中用电的总负载不应超过输电线路的最大容量 $L_{max}$ , 即

$$0 \leq \sum_{i=1}^M L_{i,t}^a \leq L_{max}. \quad (2-13)$$

LSE 所从电力批发市场和电力零售市场获得的综合利润是:

$$U_{LSE,t} = \sum_{i=1}^n C_{i,t}^p - \sum_{i=1}^n U_{i,t} - C_{W,t} - C_{S,t}, \quad (2-14)$$

其中,  $U_{LSE,t}$ 是 LSE 在时隙  $t$  中的利润,  $U_{i,t}$ 为第  $i$  个消费者在时隙  $t$  中通过参与激励型需求响应获得的收入。

LSE 的优化目标是最大化全天多个时隙的累积利润, 即

$$\max_{f(R), \alpha, \beta} \sum_{t=0}^{T-1} U_{LSE,t}. \quad (2-15)$$

为了保证电力零售市场中没有负激励,  $f(R)$ 应为正且单调递增, 并且 $f(R) \geq 0$   $df(R)/dR \geq 0$ 应被满足。在电力批发市场中, LSE 提交的竞价曲线应该单调递减, 因此 $\alpha_t \geq 0$ ,  $\beta_t \leq 0$ 应该被满足。

## 2.2 增量激励机制分析

本节从理论上分析所提出的增量激励机制的优越性，并引入斯塔克尔伯格博弈理论来分析所提出机制的公平性和帕累托效率。

### 2.2.1 优越性分析

在这一节中，本文通过两个命题从理论上说明了所提出的增量激励机制的优点。本节首先证明，目前的统一激励机制是本文所提的增量激励机制的一种特殊的形式。当增量激励函数与响应量不相关时，即当 $f(R)$ 为一个常数时，使 $f(R) = I$ ，则消费者参与需求响应获得的收益为：

$$\int_0^{R_{i,t}} f(x)dx = \int_0^{R_{i,t}} I dx = IR_{i,t}, \quad (2-16)$$

其中 $I$ 是一个常量，代表激励型需求响应中的激励价格。从而消费者的收入是单位激励价格 $I$ 和总响应量 $R_{i,t}$ 的乘积，即为现有研究中的统一激励机制的形式。

下面本文提出两个命题，分别从响应积累过程中个体消费者的消费者剩余变化，以及消费者剩余在不同柔性的消费者中的分布两个方面探讨所提出的增量激励机制的优势。

**命题 1:** 在激励型需求响应事件中，使 $R_{1,f}^*$ 和 $R_{1,g}^*$ 表示消费者 1 在所提的增量激励机制 $f$ 中，以及现有统一激励机制 $g$ 中的最优响应量。LSE 需要根据消费者的响应支付相应的激励，在两种激励机制下分别为 $C_f(R_{1,f}^*)$ 和 $C_g(R_{1,g}^*)$ ，进而可以得到以下结论：

$$C_f(R_{1,f}^*) < C_g(R_{1,g}^*), \quad \forall R_{1,f}^* = R_{1,g}^*. \quad (2-17)$$

**证明:** 假设 $f(R)$ 是本文提出的增量激励函数， $g(R)$ 为现有研究中的统一激励函数，其值在特定的激励型需求响应事件中保持不变。现有的研究表明，随着消费者负荷的削减，其舒适度损失在加速上升<sup>[20]</sup>。从而与经济学中的供应曲线类似，消费者的负荷削减函数可以表示为 $I = z(R)$ ，该公式代表了消费者在不同激励水平下的负荷削减量。在激励型需求响应中，消费者激励函数 $f(R)$ 获得响应补贴。假设消费者是理性的，则他们在负荷削减量的决策中，目标是最大化经济效益和舒适度损失之间的差异，如下所示：

$$R^* = \arg \max_R \int_0^R [f(x) - z(x)] dx \quad (2-18)$$

其中 $R^*$ 是消费者在激励函数 $f(R)$ 中的最优响应量， $\int_0^R [f(x) - z(x)] dx$ 是具有响应量 $R$ 时的消费者剩余<sup>[26]</sup>。当消费者剩余为正时，消费者可以通过响应获得利益，从而选择参与激励型需求响应。由于消费者的舒适度损失与响应正相关，因此 $z(R)$ 是单调递增函数。考虑到消费者的特征存在巨大差异，他们的响应行为受多种因素影响， $z(R)$ 可能有多种形式。

如图 2-2 所示，假设在特定的激励型需求响应事件中， $f(R)$ 、 $g(R)$ 和 $I_1(R)$ 在同一点相交。在这种情况下，消费者的响应量在两种激励机制中是相同的，即 $R_{1,f}^* = R_{1,g}^*$ 。

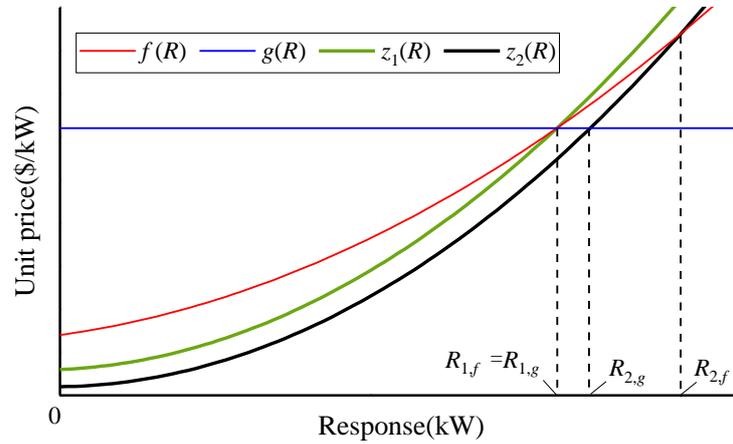


图 2-2 说明增量激励机制的示意图

两种激励机制中 LSE 的成本分别为

$$C_f(R_{1,f}^*) = \int_0^{R_{1,f}^*} f(x) dx, \quad (2-19a)$$

$$C_g(R_{1,g}^*) = g(R_{1,g}^*) R_{1,g}^*. \quad (2-19b)$$

通过推导 (2-19)，可以得到两种激励机制的增量成本如下：

$$\Delta C_f(R) = \frac{dC_f(R)}{dR} \Delta R = f(R) \Delta R, \quad (2-20a)$$

$$\Delta C_g(R) = \frac{dC_g(R)}{dR} = g(R) \Delta R, \quad (2-20b)$$

其中 $\Delta R$ 是响应的增量。由于 $f(R)$ 是单调递增函数，因此以下不等式始终成立：

$$f(R) < g(R), \quad \forall R < R_{1,f}^* \quad (2-21)$$

根据 (2-18) - (2-21), 很容易证明 (2-17) 成立。

命题 1 描述了激励型需求响应中 LSE 与消费者之间的关系。结果表明, 当消费者有相同的响应量时, LSE 的增量激励机制的成本总是低于现有的统一激励机制中的成本。增量激励机制可以更好地跟踪消费者响应弹性的变化, 使消费者每单位响应量获得的消费者剩余保持在合理的范围内, 从而降低 LSE 的激励成本。

**命题 2:** 假设存在两个具有不同灵活性的消费者, 低弹性消费者和高灵活性消费者的响应函数分别用  $I_1 = z_1(R)$  和  $I_2 = z_2(R)$  表示。使  $R_{2,f}^*$  和  $R_{2,g}^*$  分别表示在激励函数  $f(R)$  和  $g(R)$  中, 高柔性消费者的响应量, 则以下两个式子始终成立:

$$R_{2,f}^*/R_{1,f}^* > R_{2,g}^*/R_{1,g}^*, \quad \forall R_{1,f}^* = R_{1,g}^* \quad (2-22)$$

$$\frac{\int_0^{R_{2,f}^*} [f(x) - z(x)] dx}{\int_0^{R_{1,f}^*} [f(x) - z(x)] dx} > \frac{\int_0^{R_{2,g}^*} [g(x) - z(x)] dx}{\int_0^{R_{1,g}^*} [g(x) - z(x)] dx}, \quad \forall R_{1,f}^* = R_{1,g}^* \quad (2-23)$$

**证明:** 如图 2-2 所示, 低灵活性消费者在两种机制中有相同的响应量。根据 (2-18), 可以通过求解来  $g(R) = z_2(R)$  获得  $R_{2,g}^*$ 。由于  $f(R)$  是一个单调递增函数, 并且  $f(R_{1,f}^*) = g(R_{1,g}^*)$ , 从而

$$f(R) > g(R), \quad \forall R > R_{1,g}^* \quad (2-24)$$

将  $R_{2,g}^*$  和 (2-24) 代入 (2-18), 可以获得增量响应产生的消费者剩余变化:

$$\left. \frac{d \int_0^R [f(x) - z(x)] dx}{dR} \right|_{R=R_{2,g}^*} = f(R_{2,g}^*) - z(R_{2,g}^*) > 0. \quad (2-25)$$

对于高柔性消费者而言, 在  $R_{2,g}^*$  处增加响应量会增加其收益, 因此其会选择进一步加深响应, 直至消费者剩余为 0, 即  $f(R) = z(R)$ 。因此  $R_{2,f}^*$  可以通过求解  $f(R) = z(R)$  获得, 并且总是会大于  $R_{2,g}^*$ , 从而 (2-22) 成立。此外, 根据 (2-22) 和 (2-24), 可以得到如下方程

$$\begin{cases} \int_{R_{1,f}^*}^{R_{2,f}^*} [f(x) - z(x)]dx > \int_{R_{1,g}^*}^{R_{2,g}^*} [f(x) - z(x)]dx, \\ \int_0^{R_{1,f}^*} [f(x) - z(x)]dx < \int_0^{R_{1,g}^*} [g(x) - z(x)]dx, \end{cases} \quad (2-26)$$

从而 (2-23) 成立。

命题 2 描述了激励型需求响应中消费者之间的关系。它说明了增量激励机制可以增加高灵活性消费者的响应深度，并将更多的消费者剩余转移给高灵活性消费者。通过增量激励机制，积极参与响应的消费者可以获得更高比例的收益，从而增强消费者参与激励型需求响应的积极性。

### 2.2.2 效率和公平性分析

公平是激励型需求响应机制的基本要求。根据文献[88]，公平性需要从以下 4 个方面来证明：(i) 共享分配，(ii) 帕累托效率，(iii) 策略一致和 (iv) 无嫉妒性。所提机制的公平性分析如下：

**(i) 激励分配：**在激励型需求响应中，激励根据消费者的响应情况进行分配，而不是简单的平均分配。在具有公平性的激励机制下，如果消费者  $i$  在时隙  $t$  中的响应大于消费者  $j$  的响应，则消费者  $i$  必须获得更高的收入。在式 (2-2) 中，由于  $f(\tilde{R}_{i,t})$  是单调递增函数，消费者的收入相对于响应  $R$  单调递增，即，如果  $R_{i,t} > R_{j,t}$ ，则  $\int_0^{R_{i,t}} f(x)dx > \int_0^{R_{j,t}} f(x)dx$ 。因此， $U_{i,t} > U_{j,t}$  始终成立。

**(ii) 帕累托效率：**Stackelberg 博弈模型可用于描述 LSE 与消费者之间的互动关系，其中 LSE 是领导者，制定激励策略，消费者是跟随者，在不同时间段根据激励做出响应决策。当 Stackelberg 博弈满足以下条件时，存在唯一的最优解 [20][89]。(a) 一旦得知 LSE 的激励策略，消费者的收入具有唯一的最大值。(b) 对于给定的消费者策略，LSE 的利润具有唯一的最大值。根据反向归纳原理，所提出机制的帕累托效率分析如下 [20]：

首先分析 (a) 中消费者的最优决策过程。如上所述，只有当消费者获得的收入大于其舒适度损失时，消费者才参与激励型需求响应。参与激励型需求响应的消费者获得的收益为：

$$U_{i,t}(R_{i,t}) = \int_0^{R_{i,t}} [f_t(x) - z_{i,t}(x)] dx. \quad (2-27)$$

$U_{i,t}(R_{i,t})$ 关于 $R_{i,t}$ 的一阶导数为:

$$\frac{dU_i}{dR_{i,t}} = f(R_{i,t}) - z_i(R_{i,t}), \quad (2-28)$$

其中,  $f(R_{i,t})$ 和 $z_i(R_{i,t})$ 都是连续且可微的递增函数。

如果 $f(0) > z_i(0)$ 并且存在 $R_{i,c,t} \in [0, R_{i,max}]$ , 使 $f(R_{i,c,t}) = z_i(R_{i,c,t})$ , 则可以得到如下结果:

$$\begin{cases} f(R_{i,t}) - z_i(R_{i,t}) \geq 0, \forall R_{i,t} \in [0, R_{i,c,t}], \\ f(R_{i,t}) - z_i(R_{i,t}) \leq 0, \forall R_{i,t} \in [R_{i,c,t}, R_{i,max}]. \end{cases} \quad (2-29)$$

$U_{i,t}$ 关于 $R_{i,t}$ 在 $R_{i,c,t}$ 处的二阶导数为:

$$\frac{d^2U_i}{dR_i^2} = f'(R_i) - z'_i(R_i), \quad (2-30)$$

根据导数的定义, (2-30) 可以被写为

$$\frac{d^2U_i}{dR_i^2} = \frac{f(R_{i,c,t}) - z_i(R_{i,c,t})}{\Delta R} - \frac{f(R_{i,c,t} - \Delta R) - z_i(R_{i,c,t} - \Delta R)}{\Delta R}, \quad (2-31)$$

根据(2-31),  $d^2U_i/dR_i^2 < 0$ 成立, 并且 $U_{i,t}$ 在 $R_{i,c,t}$ 处达到最大值, 即 $R_{i,t}^* = R_{i,c,t}$ 。

如果 $f(0) > z_i(0)$ 并且 $R_{i,c,t} \notin [0, R_{i,max}]$ , 令 $f(R_{i,c,t}) = z_i(R_{i,c,t})$ , 第 $i$ 个消费者参与激励型需求响应总能获得的收益, 因此在这种情况下, 消费者会尽可能多地参与响应, 即 $R_{i,t}^* = R_{i,t,max}$ 。

如果 $f(0) \leq z(0)$ , 则消费者 $i$ 不参与激励型需求响应, 即 $R_{i,t}^* = 0$ 。

接下来分析 (b) 中 LSE 的决策行为。令 $\chi_t$ 表示增量激励, 则式 (2-14) 关于 $\chi_t$ 的一阶导数为:

$$\frac{\partial U_{LSE}}{\partial \chi_t} = \sum_{i=1}^n \left( (\lambda_t^W - \lambda_t^{TOU} - f(R_{i,t}^*)) \frac{\partial R_{i,t}^*}{\partial \chi_t} \right). \quad (2-32)$$

消费者的响应 $R_{i,t}^*$ 随着获得的单位激励价格而增加, 即 $\partial R_{i,t}^*/\partial \chi_t \geq 0$ 。

如果 $\lambda_t^W \leq \lambda_t^{TOU}$ ,  $U_{LSE,t}$ 随 $\chi_t$ 的增加单调递减, 从而 LSE 不执行激励型需求响

应。

如果  $\lambda_t^W > \lambda_t^{TOU}$ ，有

$$\begin{cases} \frac{\partial U_{LSE}}{\partial \lambda_t} > 0, \lambda_t^W > \lambda_t^{TOU} + f(R_{i,t}^*), \\ \frac{\partial U_{LSE}}{\partial \lambda_t} \leq 0, \lambda_t^W \leq \lambda_t^{TOU} + f(R_{i,t}^*), \end{cases} \quad (2-33)$$

$U_{LSE,t}$  在  $\lambda_t^W = \lambda_t^{TOU} + f(R_{i,t}^*)$  处达到最大值。

综上所述，所提出的机制中存在独立均衡状态，能够实现帕累托优化。

**(iii) 策略一致：** 在一个策略一致的机制中，消费者无法从报告虚假信息中获得额外的好处。由于消费者需要报告的唯一信息是他们的响应，因此任何其它响应都将偏离其帕累托最优值，即：

$$(C_{i,t}^p + C_{i,t}^c - U_{i,t})|_{R_{i,t}=R_{i,t}^*} < (C_{i,t}^p + C_{i,t}^c - U_{i,t})|_{R_{i,t}=R_{i,t}^-}, \quad (2-34)$$

其中  $R_{i,t}^-$  表示除了  $R_{i,t}^*$  之外的任何可能值，从而本文所提机制是策略一致的。

**(iv) 无嫉妒性：** 在无嫉妒性机制中，没有消费者嫉妒另一个消费者的分配。具体来说，消费者在相同的响应下获得相同的收入。根据式 (2-11)，以下等式始终成立：

$$\int_0^{R_{i,t}} f(x) dx = \int_0^{R_{j,t}} f(x) dx, \forall R_{i,t} = R_{j,t}, \quad (2-35)$$

因此，本文所提出的机制是无嫉妒性的。

## 2.3 问题制定和解决方案

本节首先阐明了电力市场竞争和激励型需求响应的决策顺序。然后，将电力批发市场和电力零售市场中的耦合问题转化为马尔可夫决策过程（Markov Decision Process, MDP）。最后，提出了一种异步双交互深度强化学习（Deep Reinforcement Learning, DRL）算法来解决优化问题。

### 2.3.1 决策动作时序及求解方法

尽管电力市场竞争和激励型需求响应决策相互关联，但实际上，它们是在不同的时隙中执行的。电力市场竞争在日前市场中进行，而激励型需求响应在日内

执行，如图 2-3 所示。

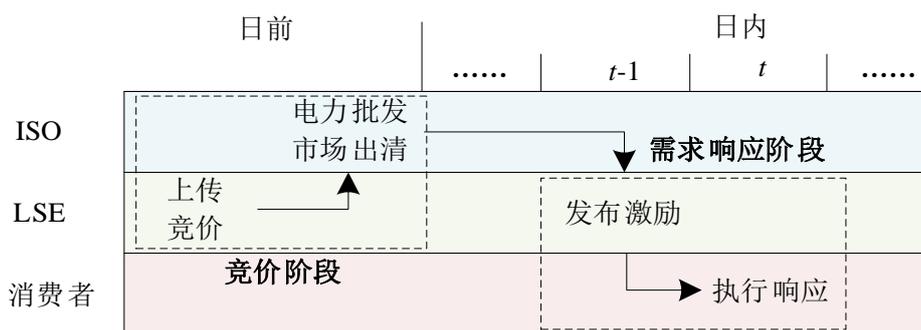


图 2-3 动作时序示意图

在日前市场中，LSE 向 ISO 提交第二天每个时段的投标，然后 ISO 根据买方和卖方的投标/报价出清电力批发市场，以产生批发电价以及每个参与者成交的电量。在日内时隙  $t$  之前，LSE 向消费者发布下一时隙的需求响应激励，消费者根据激励执行响应。

理论上，这一耦合问题可以转化为双层优化问题，其中上层是投标问题，而市场出清价格和电量被用作下层激励型需求响应问题的输入。此外，下层的优化结果也影响上层的决策。然而，每个消费者的舒适度损失函数的参数难以获得，因此无法直接计算优化决策。

DRL 和启发式算法，如遗传算法、粒子群算法等，可以在无模型环境中找到最优解。如图 2-3 所示，投标和激励型需求响应是异步决策过程。对于在线算法，解决问题需要完整的信息。然而，在解决投标问题时，激励型需求响应的行动和优化结果是未知的。因此，在线启发式算法难以解决本文中的耦合问题。

DRL 算法的训练过程是离线的，在投标和激励型需求响应都完成后，可以使用完整的历史信息对网络进行训练。此外，得益于神经网络的学习和记忆能力，DRL 可以在信息不完全的情况下推断激励型需求响应的可能优化结果，并据此做出最优投标决策。考虑到投标和激励型需求响应的行动空间是连续的，有必要应用基于无模型策略的 DRL 算法来解决耦合问题。根据现有研究，在基于无模型策略的 DRL 算法中，DDPG 算法在预测精度和收敛速度方面具有相对较好的性能，但它需要更多的状态转移样本<sup>[90][91]</sup>。由于智能电表可以方便地采集电力消耗数据，本文提出了基于 DDPG 的异步双交互 DRL 算法来解决耦合问题。

### 2.3.2 马尔可夫决策过程

耦合问题可以表述为 MDP，它被定义为五元组  $M = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$ ，其中  $\mathcal{S}$  是状态空间， $\mathcal{A}$  是动作空间， $\mathcal{T}$  是状态之间的转移概率， $\mathcal{R}$  是奖励函数， $\gamma$  是折扣因子。

由于电力批发市场是日前的市场，因此 LSE 可用的电力批发市场相关的信息包括电力批发市场的预测价格  $\hat{\lambda}_t^W$ ，分时电价（Time-of-use, TOU） $\lambda^{TOU}$ ，以及第二天没有激励型需求响应的预测电力需求  $\hat{L}_t^o$ 。令  $s_{W,t}$  表示在电力批发市场中执行出价动作  $a_{W,t}$  之前 LSE 可用的信息集，其可以表示为：

$$s_{W,t} = (\hat{\lambda}_t^W, \lambda_t^{TOU}, \hat{L}_t^o), \quad \tau = 1, 2, \dots, T. \quad (2-36)$$

激励型需求响应在日内实施，LSE 可用的与电力零售市场相关的信息包括当前时隙前电力批发市场中的实际负荷和价格、预测负荷需求  $\hat{L}_t$ 、电力批发市场中出清价格  $\lambda_t^W$  以及分时电价  $\lambda_t^{TOU}$ 。令  $s_{R,t}$  表示在电力零售市场中执行激励型需求响应动作  $a_{R,t}$  之前 LSE 可用的信息集，其可以表示为：

$$s_{R,t} = (\lambda_t^W, \lambda_t^{TOU}, L_{\kappa}^a, \hat{L}_t^o), \quad \tau = 1, \dots, T, \kappa = 1, \dots, t - 1. \quad (2-37)$$

在所考虑的场景中，LSE 需要决定电力批发市场中的投标策略，即决定  $\alpha_t$  和  $\beta_t$ 。同时，还需要确定增量激励函数  $f(R)$ 。由于增量激励函数的形式不同，需要确定的参数数量也不同。设  $\chi_t$  表示增量激励函数的所有参数，动作  $a_t$  定义为

$$a_t = (\alpha_t, \beta_t, \chi_t). \quad (2-38)$$

由于优化目标是增加 LSE 的利润，因此将  $t$  时隙中的奖励  $r_t$  定义为 LSE 获得的利润：

$$r_t = \sum_{i=1}^n C_{i,t}^p - \sum_{i=1}^n U_{i,t} - C_{w,t} - C_{S,t}. \quad (2-39)$$

受信用分配机制<sup>[92]</sup>的启发，重新分配竞价和激励型需求响应的奖励，而不是直接使用总体奖励  $r_t$ 。由于激励型需求响应的成本不影响竞价，竞价的奖励  $r_{W,t}$  定义为：

$$r_{W,t} = \sum_{i=1}^n C_{i,t}^p - C_{w,t} - C_{S,t}. \quad (2-40)$$

类似地，由于负载不平衡引起的惩罚与激励型需求响应无关，激励型需求响应的奖励 $r_{R,t}$ 定义为：

$$r_{R,t} = \sum_{i=1}^n C_{i,t}^p - \sum_{i=1}^n U_{i,t} - C_{w,t}. \quad (2-41)$$

从时隙  $t$  起的累计折扣奖励由 $D_t$ 表示，并可表示为：

$$D_t = \sum_{k=t}^{T-1} \gamma^{k-t} r_t, \quad (2-42)$$

其中 $\gamma \in (0,1]$ 是折扣系数。

### 2.3.3 异步双交互 DRL 算法

LSE 需要在电力批发市场和电力零售市场中依次做出最优决策，即，在观测值 $s_t^{BID}$ 中的投标策略 $\pi_w(s_{w,t}) = (\alpha_t, \beta_t)$ 。虽然投标和激励型需求响应是顺序执行的，但这两个问题是耦合的。投标策略的制定需要激励型需求响应的优化结果，而这些优化结果无法提前知道。此外，奖励由投标方和激励型需求响应共同计算，即当投标行动完成时，无法立即获得奖励。为了解决上述问题，我们提出了一种异步双交互 DRL 算法，在竞价和激励型需求响应动作都完成后，网络离线训练。算法架构图如图（2-4）所示。

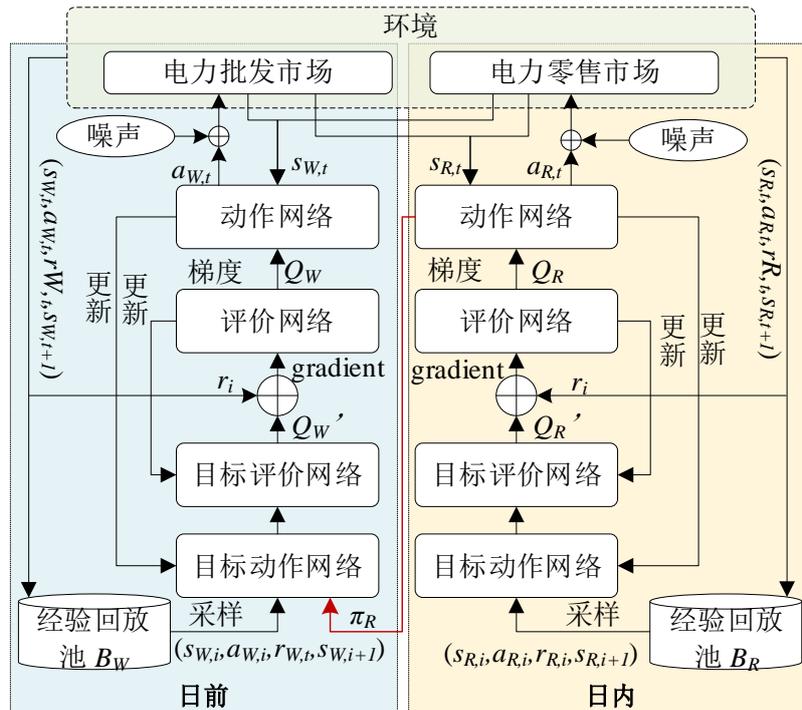


图 2-4 异步双交互 DRL 算法

使 $G_W$ 和 $G_R$ 分别表示电力批发市场和电力零售市场中的子代理。在日前， $G_W$ 首先根据 $\pi_W(s_{W,t})$ 在电力批发市场中执行动作 $a_{W,t}$ 。而后，在时隙 $t-1$ 中， $G_R$ 根据 $\pi_R(s_{R,t})$ ，在电力零售市场中执行动作 $a_{R,t}$ ，并且可以相应地计算奖励值 $r_t$ 。

在投标和激励型需求响应完成后，电力批发市场和电力零售市场中的两组状态转换 $(s_{W,t}, a_{W,t}, r_{W,t}, s_{W,t+1})$ 和 $(s_{R,t}, a_{R,t}, r_{R,t}, s_{R,t+1})$ 分别存储在经验回放池 $B_W$ 和 $B_R$ 中。每个经验回放池可以存储 $K$ 个样本，当存储到达上限时，将滚动消除最早的样本。由于用于训练网络的数据需要是独立的和相同分布的，所以从经验回放池随机采样数据以训练网络。在所提出的 DRL 算法中，两个子代理与环境之间的交互是耦合的，但网络训练是独立的。由于实际负载受到激励型需求响应的影响，因此需要先训练激励型需求响应代理，然后使用激励型需求响应代理作为参与招标代理训练的环境。根据电力零售市场中的优化策略 $\pi_R$ ，可以预测激励型需求响应后的实际负荷需求：

$$\hat{L}_t^o \xrightarrow{\pi_R} \hat{L}_t^a, \quad (2-43)$$

即，通过激励型需求响应行动策略 $\pi_R$ ，可以预测消费者参与激励型需求响应后的实际负荷。

LSE 希望最大化整个周期的总利润，而所提出的 DRL 算法的目标是学习一种策略 $\pi$ ，以最大化初始时段中动作的预期回报，即， $J[\pi] = \mathbb{E}_{a_t \sim \pi}[D_0]$ ，其中 $J[\pi]$ 是目标函数， $\mathbb{E}$ 是数学期望， $D_0$ 是从时隙 0 开始的累积回报。

在基于策略的 DRL 算法中，动作值函数用于衡量策略的性能 $\pi$ ，

$$Q^\pi(s_t, a_t) = \mathbb{E}_{a_{i>t} \sim \pi}[D_t | s_t, a_t], \quad (2-44)$$

其中 $Q^\pi(s_t, a_t)$ ，表示时隙后每个操作的总预期回报 $t$ 。计算每个步骤的回报显然需要大量的时间。为了提高算法的训练效率，(2-44)可以根据贝尔曼方程用递归形式替换

$$Q(s_t, a_t) = \mathbb{E}[r_t + \gamma Q(s_{t+1}, \mu(s_{t+1}))], \quad (2-45)$$

其中 $\mu(s_{t+1})$ 是动作函数，它通过将状态映射到特定操作来指定当前策略。

设 $\theta^Q$ 和 $\theta^\mu$ 分别表示评价网络和动作网络的参数向量。此外，分别设 $\theta^{Q'}$ 和 $\theta^{\mu'}$ 表示目标评价网络和目标动作网络的参数向量。 $\theta^Q$ 通过最小化以下损失函数来更

新评价网络的参数向量:

$$L(\theta^Q) = \frac{1}{Z} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2, \quad (2-46)$$

其中 $Z$ 是从经验回放池获取的样本数,  $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{t+1} | \theta^{\mu'}) | \theta^Q)$ 。  
使用采样策略梯度更新动作网络 $\theta^Q$ 的参数向量:

$$\nabla_{\theta^Q} J \approx \frac{1}{Z} \sum_i \nabla_a Q(s, a; \theta^Q |_{s=s_i, a=\mu(s_i)}) \nabla_{\theta^{\mu}} \mu(s; \theta^{\mu} |_{s_i}), \quad (2-47)$$

其中 $\nabla_{\theta^{\mu}}$ 是 $\theta^{\mu}$ 的梯度, 并且 $\nabla_a$ 是 $a$ 的梯度。

为了提高动作空间的探索效率, 探索策略 $\mu'$ 通过增加噪声来构建, 其中噪声由过程 $\mathcal{N}$ 来描述:

$$\mu'(s_t) = \mu(s_t | \theta_t^{\mu}) + \mathcal{N}_t, \quad (2-48)$$

其中, Ornstein-Uhlenbeck 过程被用来构建噪声<sup>[93]</sup>:

$$\mathcal{N}_{t+1} = (1 - \theta) \mathcal{N}_t + \sigma dW_t \mathcal{N}_t, \quad (2-49)$$

其中 $\theta > 0$ 和 $\sigma > 0$ 分别是漂移项和扩散项中的参数,  $W_t$ 表示维纳过程。

算法 1 中给出了详细的异步双交互 DRL 算法的伪代码。

---

#### 算法 1 异步双交互 DRL 算法

---

随机初始化参数向量  $\theta_W^Q, \theta_W^{\mu}, \theta_R^Q, \theta_R^{\mu}$ .

初始化目标网络参数  $\theta_W^{Q'} \leftarrow \theta_W^Q, \theta_W^{\mu'} \leftarrow \theta_W^{\mu}, \theta_R^{Q'} \leftarrow \theta_R^Q, \theta_R^{\mu'} \leftarrow \theta_R^{\mu}$ .

初始化经验回放池  $B_W$  and  $B_R$ .

**For** episode=1, ...,  $M$  **do**

    分别为动作 $a_W$ 和动作 $a_R$ 初始化随机过程 $\mathcal{N}_W$ 和 $\mathcal{N}_R$

    接收初始的观测状态 $s_{W,t}$ 和 $s_{R,t}$

**For**  $t = 0, \dots, T - 1$  **do**

        选择 $a_{W,t} = \mu(s_{W,t} | \theta_{W,t}^{\mu}) + \mathcal{N}_{W,t}$

        在电力批发市场中执行 $a_{W,t}$ 并接收 $L_{i,t}^b$

        在电力零售市场中执行 $a_{R,t}$ 并接收奖励 $r_t$ 和下一时段状态 $s_{t+1}$ 。

        将状态迁移 $(s_{W,t}, a_{W,t}, r_t, s_{W,t+1})$ 和 $(s_{R,t}, a_{R,t}, r_t, s_{R,t+1})$ 存入经验回放池 $B_W$ 和 $B_R$ 。

**If** 激励型需求响应的训练过程没有结束

        随机从经验回放池 $B_R$ 中抽取 $Z_R$ 组样本

        设置 $y_i = r_i + \gamma Q_R'(s_{R,i+1}, \mu_R'(s_{R,i+1}; \theta_R^{\mu'}) | \theta_R^{Q'})$

        通过最小化 (2-46) 中的  $L$  更新评价网络 $G_R$ :

$$\theta_R^{Q'} = \theta_R^{Q'} - \nabla_{\theta_R^{Q'}} L(\theta_R^{Q'})$$

        使用 (2-47) 中的梯度更新动作网络 $G_R$ :

$$\theta_R^{\mu'} = \theta_R^{\mu'} + \nabla_{\theta_R^{\mu'}} J(\theta_R^{Q'})$$

        更新目标网络 $G_R$ :

$$\theta_R^{Q'} \leftarrow \tau_R \theta_R^{Q'} + (1 - \tau_R) \theta_R^Q, \quad \theta_R^{\mu'} \leftarrow \tau_R \theta_R^{\mu'} + (1 - \tau_R) \theta_R^{\mu}$$

**Else**

        随机从经验回放池 $B_W$ 中抽取 $Z_W$ 组样本

---

根据 $\pi_R$ 获得激励性需求响应的结果，并将其添加至观测： $s_{W,t}: \hat{L}_t^o \rightarrow \hat{L}_t^a$

设置  $y_i = r_i + \gamma_W Q'_W \left( s_{W,i+1}, \mu'_W \left( s_{W,i+1}; \theta_W^\mu \right); \theta_W^o \right)$

通过最小化 (2-46) 中的 L 更新评价网络 $G_W$ :

$$\theta_W^o = \theta_W^o - \nabla_{\theta_W^o} L(\theta_W^o)$$

使用 (2-47) 中的梯度更新动作网络 $G_W$ :

$$\theta_W^\mu = \theta_W^\mu + \nabla_{\theta_W^\mu} J(\theta_W^o)$$

更新目标网络 $G_W$ :

$$\theta_W^o \leftarrow \tau_W \theta_W^o + (1 - \tau_W) \theta_W^{o'}, \theta_W^\mu \leftarrow \tau_W \theta_W^\mu + (1 - \tau_W) \theta_W^{\mu'}$$

End for

End for

## 2.4 仿真分析

### 2.4.1 仿真参数设置

电力批发市场中的实时电价 $\lambda_t$ 取自 PJM 电力市场。消费者的数量设置为 25, Pecan Street<sup>[94]</sup>中的实际负载数据被用作消费者的基线负载，其中 92 天被选为训练集，10 天作为验证集，3 天作为测试集。消费者响应函数设置为 $z_i(R) = a_i R^2 + b_i R + c_i$ ，其中 $a_i$ 、 $b_i$ 和 $c_i$ 分别以均匀分布在 $[0,0.3]$ 、 $[0.0,1]$ 和 $[0,1.02]$ 范围内随机生成。本文建立了增量激励函数 $f(R) = \alpha_1 + \beta R$ 和统一价格 $f(R) = \alpha_2$ 。反弹系数 $\xi_{i,t}$ 在 $[0,1]$ 范围内以均匀分布随机生成，惩罚函数设置为电力批发市场中出清价格的两倍，即， $\phi = 2\lambda_t^{DA} L_{t,s}$ ，式中， $L_{t,s}$ 是时隙 t 中的电力不平衡量。为了模拟预测误差引起的不确定性，将干扰添加到实际值中以模拟预测值，即 $\hat{\lambda}_t^{DA} = (1 + \varepsilon)\lambda_t^{DA}$ 和 $\hat{L}_t^a = (1 + \varepsilon)L_t^a$ ，其中 $\varepsilon$ 是正态分布产生的随机误差，其期望值和方差如表 2-1 所示。分时电价如表 2-2 所示：

表 2-1 均值和方差

低不确定性场景		高不确定性场景	
期望	方差	期望	方差
0	0.05	0	0.25

表 2-2 各时段的 TOU 资费

0.03\$/千瓦时 (峰值)	0.025\$/千瓦时		0.02 美元/千瓦时 (谷)	
11:00-19:00	6:00-10:00	20:00-22:00	0:00-5:00	23:00

DRL 中的所有网络采用具有 3 个隐藏层的全连接网络，其中前两个具有 256 个神经元，第三个具有 128 个神经元。两个子代理的动作网络和评价网络的学习率分别设置为 0.000001 和 0.00001。该算法在 Python 中使用 PyTorch 实现。仿真实验在配备 Intel (R) Core (TM) i7-9750H 处理器和单个 NVIDIA GeForce GTX

1660 Ti GPU 的笔记本电脑上进行。

## 2.4.2 所提机制的效果分析

在这一部分，本文测试了所提出的增量激励机制的性能。由于电力市场的 RTP 波动较大，且各天之间差异很大，因此我们选择了连续三天进行模拟。

在电力零售市场中，LSE 以固定的分时电价向消费者提供能源供应服务。在大多数时段中，电力零售市场中的分时电价高于电力批发市场中的 RTP，因此 LSE 可以通过电力交易获得利润。因此，为了提高总利润，LSE 根据电力批发市场中的价格波动实施激励型需求响应。每个时段在不同机制下的响应量如图 2-5 所示。

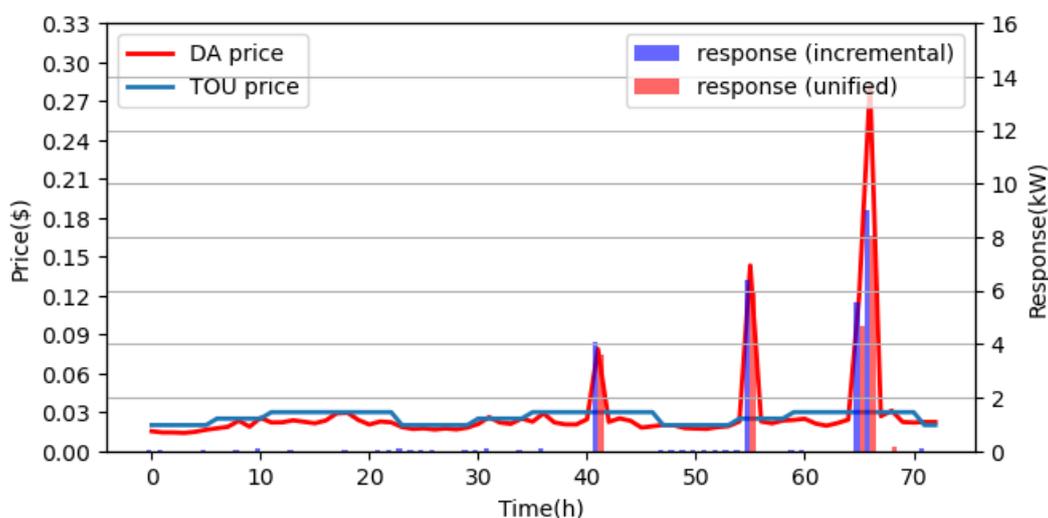


图 2-5 每个时间段不同机制下的响应

从图 5 可以看出，本文提出的算法可以跟踪电力批发市场中 RTP 的波动。在大多数时段，TOU 价格高于 RTP，激励型需求响应未实施。在一些 RTP 较高的时段，例如时段 41、55、65、66 等，实施激励型需求响应以减少亏损。由于激励型需求响应的激励成本在增加，不同时段的激励型需求响应深度也随着 RTP 的变化而变化。例如，时段 66 的 RTP 高达 0.29\$/kW，而 TOU 价格仅为 0.03\$/kW，这意味着 LSE 向消费者提供的电力损失了 0.26\$/kW。LSE 需要通过激励型需求响应尽可能降低消费者的功耗，消费者的响应超过 8kW。相比之下，时段 41 的 RTP 为 0.08\$/kW，比 TOU 价格高 0.05\$/kW，响应小于 4kW。

对不同机制的激励型需求响应的单位成本和响应的进一步分析如图 2-6 所示。

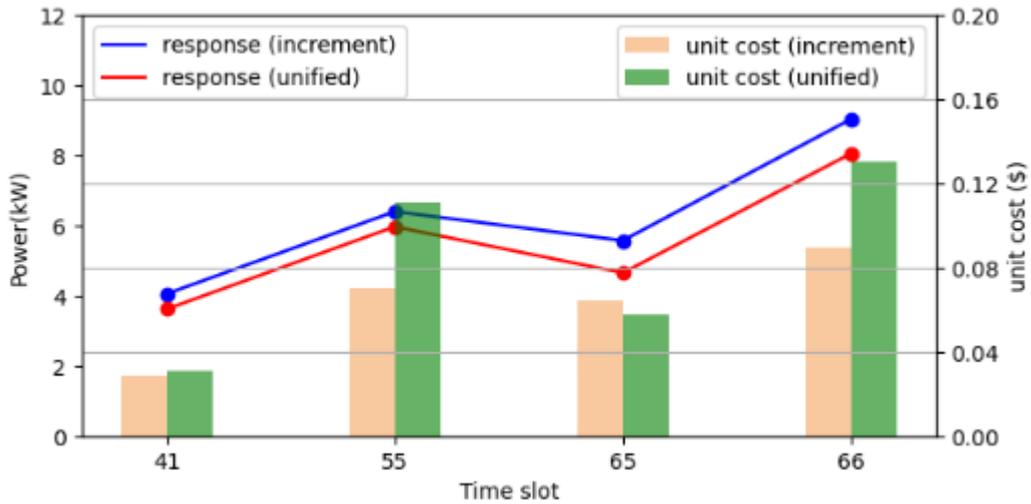
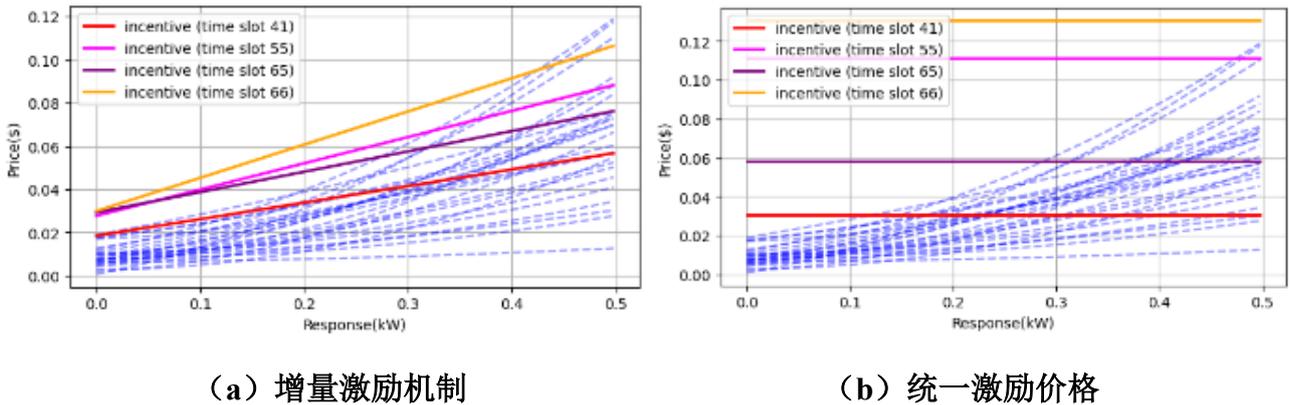


图 2-6 不同机制的单位成本和响应

从图 2-6 可以看出，在四个典型时段内，使用增量激励函数的消费者响应始终高于使用统一激励价格的消费者响应。同时，在 41、55、66 时段，采用增量激励函数的单位激励成本也低于采用统一激励价格的成本，说明增量激励函数可以有效降低 LSE 的激励成本，充分释放 LSE 的减载潜力。同时消费者。在 65 时段，采用增量激励函数的成本略高于采用统一激励价格的成本，这是由于消费者响应差距较大所致。时段 65 和时段 66 的消费者响应差异几乎相同，但单位成本存在显著差异，说明采用增量激励机制的单位成本与统一激励价格之间的差异随着消费者响应的提高而增大。这种现象是由消费者的响应特性引起的。



(a) 增量激励机制

(b) 统一激励价格

图 2-7 激励和消费者响应曲线

在图 2-7 中，每条蓝色虚线代表一个消费者的负荷减少函数  $z(R)$ ，它表示单位激励和期望的负荷减少之间的关系。实线分别表示时间段 41、55、65 和 66 中的激励函数。激励函数和消费者响应函数的差值为消费者剩余。

从图 2-7 (b) 可以看出, 统一激励价格在负荷降低的初始阶段造成大量消费者剩余, 并随着响应的增加而迅速减少。当 LSE 所希望获得更高的减载时, 它只能将横向激励价格曲线向上移动, 这加剧了消费者剩余分配不均的问题。例如, 时隙 65 的消费者剩余在负载减少的初始阶段约为  $0.04\$/kW$ , 而在时隙 66 中高达  $0.1\$/kW$ 。

本文提出的增量激励函数可以有效缓解上述问题。使用增量激励函数, 时隙 65 中最大的消费者剩余仅为  $0.01\$/kW$ , 而时间段 66 中的最大消费者盈余仅为  $0.02\$/kW$ 。消费者剩余的再分配提高了 LSE 的激励效率, 使其能够以更低的单位激励成本获得更多的负荷削减。

### 2.4.3 所提算法的可靠性分析

由于 LSE 需要为负载不平衡支付罚款, 因此电力批发市场中的投标在很大程度上影响了 LSE 的利润。LSE 从电力批发市场获得的功率和每个时隙中消费者的实际用电需求如图 2-8 所示。

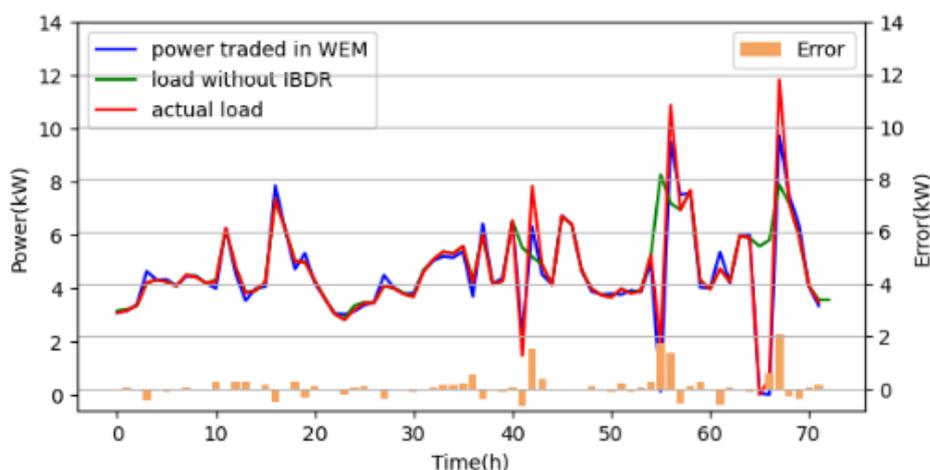


图 2-8 电力批发市场中的交易电力和消费者的电力消耗

从图 2-8 可以看出, 所提出的异步双交互 DRL 算法即使在受激励型需求响应和负载反弹影响的时隙内, 也能保证交易电量接近消费者的实际用电需求。例如, 在时间段 55, 交易电量在激励型需求响应之后随着消费者的实际负荷需求而减少, 而在随后的时间段 56, 交易电量随着负荷反弹而增加。在竞价策略的优化中, 只能观察到时隙  $t$  之前的信息和时隙  $t$  的预测值, 无法观察到后续时隙的负载反弹。从仿真结果可以看出, 该算法可以通过有限的观测信息推断负载反弹, 同时很好地跟踪负载变化。算法可以将所有时隙内的功率偏差保持在一个很小的

范围内，从而最大限度地减少因功率不平衡而造成的利润损失。

LSE 在决策中观察到的价格和消费者负荷是预测值，预测误差影响决策效果。因此，我们验证了算法在高不确定性场景中的性能，如图 2-9 所示。

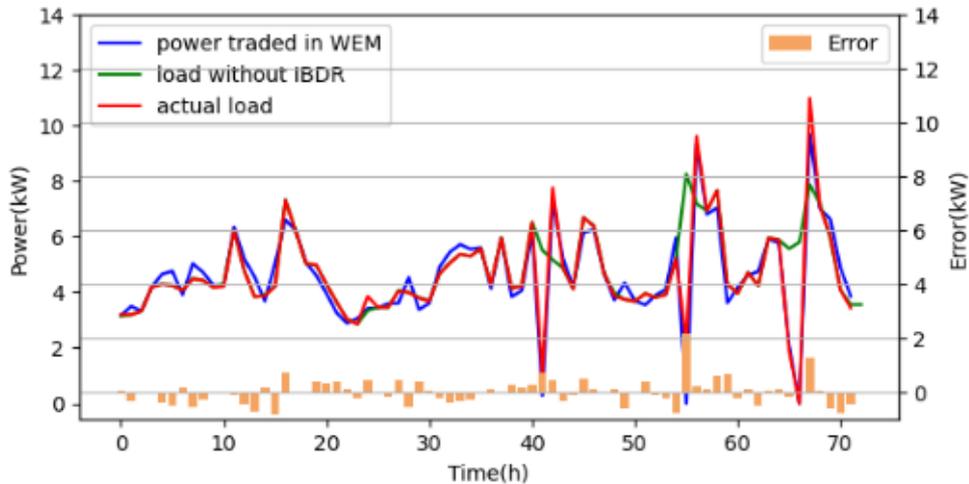


图 2-9 电力批发市场中的交易电力和消费者的电力消耗具有高度的不确定性

在高不确定性场景下，最大预测误差超过 20%。从图 2-9 可以看出，竞价策略仍然可以很好地跟踪具有很高的不确定性的功率曲线的变化。尽管功率不平衡增加，但仍保持在可接受的范围内。

此外，由于 LSE 经常需要与激励型需求响应中的大量消费者交互，因此算法的可扩展性也需要验证。我们设置了 500 个消费者来验证算法在相同场景下的性能，如图 2-10 所示。

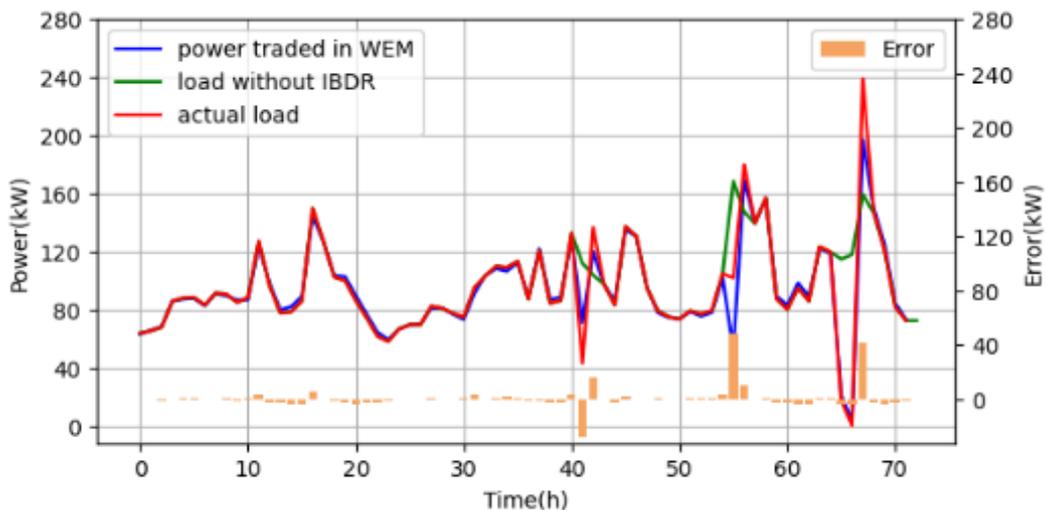


图 2-10 电力批发市场和功耗中的交易电力（500 名消费者）

在所提出的算法中，所有的观测值都是消费者的总价值，例如，总功率、总响应等。因此，消费者数量的增加不会对算法造成额外的负担。从图 2-10 可以看出，该算法仍然可以保持相对稳定的良好性能，说明所提出的算法可以适应大量消费者的场景。

## 2.5 本章小结

本章提出了一种激励型需求响应的增量激励机制，其中消费者根据他们的增量响应获得激励。通过数学分析，我们发现，现有的统一激励机制是增量激励机制的一种特殊形式。通过将统一激励价格扩展到增量激励机制，消费者剩余可以更合理地分配。累积响应中过度的消费者剩余得到缓解，高弹性消费者通过提高响应可以获得更多的利润。LSE 的单位激励成本也降低了，同时增加了消费者的响应。此外，异步双交互 DRL 算法在仿真中的不同场景下具有相对稳定和令人满意的性能，验证了所提算法具有令人满意的稳定性和适应性，能够很好地处理耦合环境下的决策问题。

在未来的工作中，可以进一步研究不同形式的增量激励机制对多元化消费群体的表现，从而实现针对消费群体特点的增量激励机制形式的自适应优化。此外，为了提高系统效率，可以扩展增量激励机制的应用，例如基于价格的需求响应、点对点交易等。

### 3 面向分布式储能的共享机制及优化策略

微电网 (microgrid, MG) 是具有自控能力的电力系统的一个子集, 通常由分布式发电机、负载、储能设施等组成<sup>[95]</sup>。与配电网不同, MG 可以在孤岛模式或并网模式下运行, 从上层系统的角度来看, MG 是电力系统中的独立实体<sup>[96]</sup>。为应对环境污染问题, 可再生能源近年来发展迅速。在可再生能源占比较高的 MG 中, 分布式可再生发电设备, 例如屋顶光伏 (PV), 往往安装在需求侧, 将消费者转变为具有供需双重属性的产消者。可再生能源输出的不确定性给 MG 的高效运行带来了挑战。储能 (ES) 在减少微电网中可再生能源的不确定性和周期性造成的供需时空功率失衡方面发挥着越来越重要的作用。通过共享可以提高属于不同实体的分布式 ES 的利用效率, 通过 ES 共享和需求响应的协调, 可以为微电网提供相当大的灵活性资源。

#### 3.1 基于主从博弈的储能共享方式

本节研究提出了一种两阶段 ES 共享机制, 其中日前对闲置的 ES 容量进行聚合, 为实时优化提供可靠的资源。然后, 提出一种基于 DDPG 算法的双层半耦合优化策略, 以解决日前共享和日内优化的异步决策问题。为了应对多重不确定性的影响, 应用蒙特卡罗采样来确保共享的 ES 容量在任何情况下都是足够的。仿真验证了可再生能源的本地消费率有效提高了 12.9%, 微电网运营商和产消者都可以通过 ES 共享和需求响应的联合优化来提高收入。

为了提高分布式 ES 的利用效率, 提出一种基于激励的两阶段 ES 共享机制, 其中微电网运营商可以获得所需的 ES 容量以降低运营成本, 而产消者可以从闲置 ES 容量的共享中获得收益。然后, 提出一种基于 DDPG 的双层半耦合优化策略, 以解决信息不完整的决策问题, 并应用蒙特卡罗抽样处理不确定性的影响。主要重要性和贡献总结如下。

(1) 提出一个两阶段优化框架, 实现需求响应 (Demand Response, DR) 和 ES 共享的合作。与现有仅关注 DR 的研究或仅关注 ES 共享的研究相比, 联合优化可以更充分地释放需求侧资源的可调潜力, 从而提高微电网运营商的收入和可再生能源的本地消费。

(2) 由于日前 ES 共享所需的 ES 容量是由实时优化确定的, 无法事先知道, 因此提出一种基于 DDPG 的两层半耦合优化策略, 实现对分布在不同时隙的耦合

决策问题的异步优化。

(3) 尽可能充分地考虑由预测误差、产消者行为等引起的多重不确定性。与现有的研究忽略了 DR 中负荷反弹的不确定性不同，引入了贝叶斯转移概率来描述产消者行为引起的不确定性。

(4) 为了应对多重不确定性对 ES 容量共享的影响，在所提出算法的网络训练中应用了蒙特卡罗采样。与现有忽视不确定性对 ES 共享影响的研究相比，本文提出的优化策略可以确保在任何情况下都能以最低的成本获得足够的共享 ES 容量进行实时优化。

### 3.1.1 分布式储能系统架构与共享机制

本节首先介绍 MG 中 ES 共享和 DR 的系统架构，然后解释协同调度中动作的时序。

#### 3.1.1.1 系统架构

如图 3-1 所示，本文关注住宅社区 MG，考虑的系统由微电网运营商(microgrid operator, MGO)和不同类型的产消者组成，例如住宅、公寓或商业建筑等，假设产消者配备了不同容量的屋顶光伏和 ES。同时，产消者的负荷包括可调负荷（如电动汽车、热水器、空调等）和非可调负荷（如照明、冰箱等）。所有需求侧设备均由智能终端控制，即产消者可以根据外部信号的变化做出优化决策。住宅建筑的日用电量通常为数十千瓦时<sup>[97]</sup>，因此包含多个建筑物的 MG 的日用电量通常为数百千瓦时。住宅小区的可再生能源主要是屋顶光伏，在高普及率的住宅小区，光伏的最大输出可以达到几百千瓦甚至更高。通过适当的光伏阵列布置，屋顶光伏可以聚合成稳定的光伏电源<sup>[98]</sup>。为了存储多余的光伏输出，ES 设备安装在建筑物内的家庭中。由于家庭 ES 设备的容量约为 10kWh<sup>[99]</sup>，假设每个家庭都配备 ES 设备，MG 中的总 ES 容量可以达到数千 kWh。

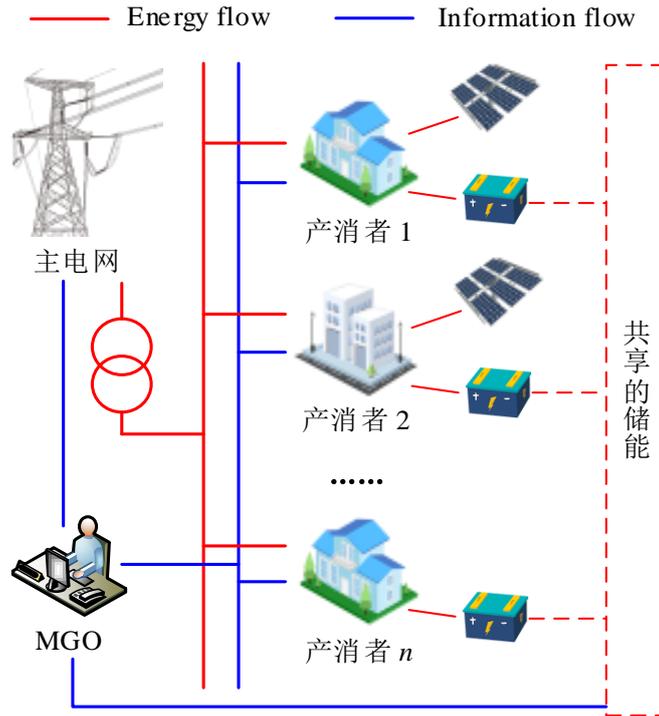


图 3-1 储能共享架构

MGO 是一个独立实体，通过向产消者提供能源供应服务来获得收入。通常，MGO 以实时价格(RTP)从主电网购买电力，并以 TOU 电价将其出售给产消者。当 RTP 高于 TOU 时，向产消者售电会导致 MGO 出现赤字，这可以通过 DR 减少负载需求来缓解。此外，MGO 可以通过激励聚合产消者闲置的 ES 产能，并通过适当的充放电操作提高自身收入。

### 3.1.1.2 共享机制和动作顺序

微电网运营商需要优化的变量包括 ES 共享的激励价格，ES 的充电/放电功率以及 DR 的激励价格。操作的时间序列如图 3-2 所示。

假设一天被划分为 $T$ 个时隙，每个时隙由 $t \in \{0, 1, \dots, T\}$ 索引。如图 3-2 所示，日前，MGO 向产消者公布 ES 共享的激励价格，每个产消者相应地反馈共享的 ES 容量。需要注意的是，每个时隙的激励价格随所需的 ES 容量而变化。在时间段  $t-1$ ，MGO 根据产消者的净负荷需求和主电网的 RTP 公布下一个时间段的 DR 激励价格。随后，在时隙  $t$ ，MGO 控制共享 ES 的充放电功率，并与主电网交互购买或出售功率，以维持 MG 的功率平衡。

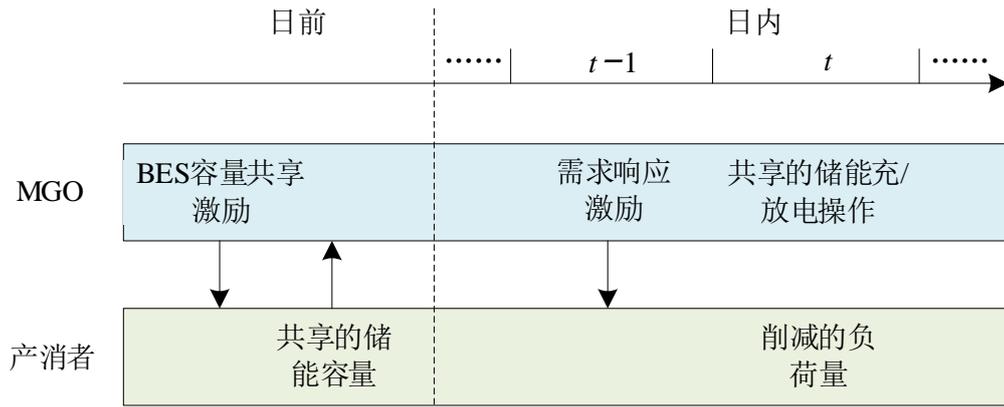


图 3-2 储能共享优化及需求响应的动作时间序列

### 3.1.2 产消者和微电网运营商的建模

#### 3.1.2.1 产消者建模

我们考虑一组产消者  $i=\{1,\dots,n\}$ ，其负载由 MGO 提供。产消者可以通过分享他们闲置的 ES 容量来获得收入。充电/放电会降低 ES 的循环寿命，并且单位容量的成本随着充电/放电深度的增加而增加<sup>[100]</sup>，即生产者必须承担共享 ES 容量

$$C_{i,t}^S = f_{i,t}(O_{i,t})O_{i,t}, \quad (3-1)$$

其中， $C_{i,t}^S$ 是 ES 共享中第  $t$  个时隙内第  $i$  个产消者的总成本， $O_{i,t}$ 是共享 ES 的容量， $f_{i,t}()$ 是单调递增函数，表示 ES 共享中的单位成本。

参与 ES 共享的产消者的目标是最大化他们自己的收入。因此，激励 $\lambda_t^{CS}$ 下产消者共享的 ES 容量为：

$$O_{i,t}^* = \operatorname{argmax}(f_{i,t}(O_{i,t})O_{i,t} - \lambda_t^{CS}O_{i,t}), \quad s.t. O_{i,t}^* \leq O_{i,t}^I, \quad (3-2)$$

其中 $\lambda_t^{CS}$ 是时隙  $t$  中 ES 共享的激励价格， $O_{i,t}^*$ 是时隙  $t$  第  $i$  个产消者共享的 ES 容量， $O_{i,t}^I$ 是第  $i$  个产消者安装的 ES 容量。

在 RTP 高于 TOU 的时段，MGO 必须承担亏损向产消者提供电力，并且可以通过 DR 降低产消者的负荷需求来缓解赤字。减少负荷为产消者带来了经济收入，但消费者不得不遭受舒适度的损失。根据经济学理论，产消者负荷减少与需求响应激励价格之间的关系可以用需求函数表示：

$$\Delta P_{i,t}^L = g_{i,t}(\lambda_t^L), \quad (3-3)$$

其中 $\Delta P_{i,t}^L$ 是负载减少量， $\lambda_t^L$ 是时间段 t 的 DR 激励价格， $g_{i,t}(\cdot)$ 是单调递增函数。需要指出的是，由于产消者的不同特点， $g_{i,t}(\cdot)$ 可能有不同的表现形式。需要指出的是，本文中使用的无模型 DRL 算法可以通过学习适应不同的需求函数。

由于产消者的可调负载由可还原负载和可转移负载组成，因此负载的降低引起后续时间段负载需求的变化<sup>[86]</sup>，因此每个时间段的产消者负载为

$$P_{i,t}^{co,L} = P_{i,t}^L - P_{i,t}^{PV} + \sum_{j=0}^{t-1} \eta_{i,j,t} \Delta P_{i,j,t}^L, \quad (3-4)$$

其中 $P_{i,t}^{PV}$ 是第 i 个产消者在时隙 t 配备的 PV 的输出功率，是考虑负载反弹的修正负载， $P_{i,t}^{co,L}$ 是时隙 j 减载对时隙 t 负载的影响系数，反映了负载减少和负载反弹之间的关系。负载反弹是 DR 后一段时间内负载需求会高于正常负载的现象，主要是由温控负载、电动汽车等可转移负载引起的。负载的回弹功率与 DR 的深度相关<sup>[9][11]</sup>。考虑到产消者行为的不确定性会影响负载反弹的力量，引入贝叶斯转移概率来描述负载反弹与 DR 的关系：

$$\eta_{i,j,t} = \frac{\sum_{j=0}^{t-1} \sum_{m=0}^M p \left( P_{i,t}^L = P_{i,t-1}^L + \Delta P_{m,i,j}^L \mid \Delta P_{m,i,j}^L \right) \Delta P_{m,i,j}^L}{\sum_{j=0}^{t-1} \sum_{m=0}^M \Delta P_{m,i,j}^L}, \quad (3-5)$$

假设在时隙 j 有 M 个负载参与减载，则 $\Delta P_{m,i,j}^L$ 为第 m 个负载减载的功率，第 m 个减载转移到时隙 t 的概率为 $p \left( P_{i,t}^L = P_{i,t-1}^L + \Delta P_{m,i,j}^L \mid \Delta P_{m,i,j}^L \right)$ 。

### 3.1.2.2 微电网运营商建模

在每个时隙中，微电网运营商可以控制共享 ES 的充放电功率，ES 的动态变化过程如下<sup>[10]</sup>：

$$SOC_t^E = SOC_{t-1}^E + \eta_t^c P_t^c \Delta t - \frac{1}{\eta_t^d} P_t^d \Delta t, \quad (3-6)$$

$$0 \leq SOC_t^E \leq \sum_{i=1}^N O_{i,t}^*, P_t^c \leq P_t^{c,max}, P_t^d \leq P_t^{d,max}, \quad (3-7)$$

其中 $\eta_t^c$ 和 $\eta_t^d$ 分别为 ES 的充放电效率， $P_t^c$ 和 $P_t^d$ 分别为时隙 t 的充放电功率，

$SOC_t^E$  为共享 ES 中存储的能量, 且不应超过每个时隙中共享 ES 的总容量  $\sum_{n=1}^N S_{i,t}^*$ 。

由于产消者行为的不确定性, MGO 通过公布每个时隙的 ES 共享激励价格可以实现的预期共享 ES 容量为:

$$E(O_{i,t}^{sum}) = E\left(\sum_{i=1}^N (O_{i,t}^* + \varepsilon_{i,t}^{CS})\right), \quad (3-8)$$

其中  $E$  是函数的期望值,  $\varepsilon_{i,t}^{CS}$  是第  $i$  个产消者共享 ES 容量的实际值与理论值之间的误差。

同理, DR 激励价格下产消者的实际反应也存在误差, 产消者在时隙  $t$  的预期响应可表示为:

$$E(\Delta P_t^{L,sum}) = E\left(\sum_{i=1}^N (\Delta P_{i,t}^L + \varepsilon_{i,t}^{DR})\right), \quad (3-9)$$

其中  $\varepsilon_{i,t}^{DR}$  是第  $i$  个产消者在时隙  $t$  的响应的实际值与理论值之间的误差。从而, 每个时隙产消者的总预期实际负载需求为:

$$E(P_t^{L,sum}) = E\left(\sum_{i=1}^N P_{i,t}^{co,L}\right) - E(\Delta P_t^{L,sum}), \quad (3-10)$$

在每个时隙中, 必须满足功率平衡约束:

$$P_t^{L,sum} + P_t^S + P_t^c = P_t^P + P_t^d, \quad (3-11)$$

其中  $P_t^P$  是 MGO 在高峰时段从外部电网购买的电力,  $P_t^S$  是 MGO 在低谷时段向外部电网出售的电力。

由于连接 MG 与主网的变压器容量有限, MGO 与主网交互的功率需要满足以下约束:

$$0 \leq P_t^P \leq P^{tr}, 0 \leq P_t^S \leq P^{tr}, \quad (3-12)$$

其中  $P^{tr}$  是变压器的最大传输功率。由于 MG 与主电网之间的电力交互, MGO 的成本和收益如下:

$$C_t^P = \lambda_t^P P_t^P, U_t^S = \lambda_t^S P_t^S, \quad (3-13)$$

其中 $C_t^P$ 和 $U_t^S$ 分别是主电网购买电力的成本和向主电网出售电力的收入。 $\lambda_t^P$ 是外部电力批发市场的 RTP。

MGO 根据 TOU 关税向产消者出售电量。为了鼓励产消者安装分布式光伏，MGO 需要购买产消者所有多余的光伏电量：

$$C_t^I = \begin{cases} \lambda_t^{\text{TOU}} P_{i,t}^{\text{co,L}}, P_{i,t}^{\text{co,L}} \geq 0 \\ \psi \lambda_t^{\text{TOU}} P_{i,t}^{\text{co,L}}, P_{i,t}^{\text{co,L}} < 0 \end{cases}, \quad (3-14)$$

式中 $C_t^I$ 为产消者能源交互产生的成本/收入， $\lambda_t^{\text{TOU}}$ 为分时电价， $\psi$ 为 MGO 购买产消者多余光伏发电的价格系数，为促进光伏发电就地消纳， $\psi \in [0,1]$ 。

时隙  $t$  中 MGO 的总成本如下：

$$C_t^{\text{sum}} = C_t^I + C_t^P + \sum_{i=1}^N C_{i,t}^S + \sum_{i=1}^N \Delta P_{i,t}^L \lambda_t^L - U_t^S, \quad (3-15)$$

微电网运营商的优化目标是最大限度地降低整个周期的总成本：

$$\max \sum_{t=0}^T C_t^{\text{sum}} \text{ s.t. } (3-6) \sim (3-12), \quad (3-16)$$

### 3.1.3 优化问题制定和解决方法

在本节中，我们使用马尔可夫决策过程（MDP）来描述 MGO 的决策动作，然后介绍所提出的两层半耦合优化策略，以及应用马尔可夫采样的方法。采用神经网络的逆向训练来处理不确定性的影响。

#### 3.1.3.1 基于 MDP 的决策问题构建

一个标准的马尔可夫过程由一组 5 个元组组成，即 $M = (S, A, T, R, \gamma)$ ，其中  $S$  是状态空间， $A$  是动作空间， $T$  是动作执行后状态的转移概率， $R$  是动作的奖励， $\gamma$  是折扣系数。

MGO 的行为包括 ES 共享的激励价格 $\lambda_t^{\text{CS}}$ 、ES 和的充放电功率 $P_t^c$ 和 $P_t^d$ 、DR 的激励价格 $\lambda_t^d$ 。由于 ES 容量的共享是在前一天完成的，共享 ES 的 DR 和功率控制是在当天完成的，所以动作空间可以分为两个子空间：

$$a_t^{DA} = (\lambda_t^{CS}), a_t^{RT} = (\lambda_t^L, P_t^c, P_t^d), \quad (3-17)$$

动作 $a_t^{DA}$ 的目的是根据每个时隙所需的 ES 容量 $\hat{O}_t^{target}$ 获得足够的共享容量，因此设置状态为 $S_t^{DA} = (\hat{O}_t^{target})$ 。以收益最大化为目标，其值由 RTP、PV 输出等多种因素决定，受多重不确定性影响，无法直接获取。通过对本文提出的两层半耦合网络进行反向训练和蒙特卡洛采样，可以得到 $\hat{O}_t^{target}$ 的值，下一部分会详细讨论。为了满足功率平衡约束，将交互功率 $P_t^S$ 和 $P_t^P$ 与主电网的交互设置为被动变量，根据功率平衡约束计算它们的值。

在日内优化中，MGO 需要综合考虑 RTP、负荷需求、光伏输出功率、共享 ES 容量等，实现全天累计收益最大化。动作的状态空间是：

$$S_t^{RT} = (P_t^L, P_t^{PV}, \lambda_t^P, \lambda_t^{TOU}, \psi), \quad (3-18)$$

虽然 MGO 的优化目标是全天的收益最大化，但动作 $a_t^{DA}$ 和 $a_t^{RT}$ 的子目标不同。在前一天，MGO 的目标是在确保共享的 ES 容量不小于目标 ES 容量 $\hat{O}_t^{target}$ 的同时，最小化 ES 共享的成本。因此，奖励设置如下：

$$r_t^{DA} = - \sum_{i=1}^N C_{i,t}^S - \varphi \sqrt{\max(\hat{O}_t^{target} - O_t, 0)}, \quad (3-19)$$

其中 $O_t$ 是前一天交易的共享 ES 容量。

奖励由两部分组成。第一个是 ES 共享的成本，后者是共享 ES 容量不足的惩罚。当共享的 ES 容量不足时，MG 的运行效率会降低，甚至可能违反功率平衡约束。因此，惩罚可以确保始终有足够的共享 ES 容量进行实时调整。同时，为了保证惩罚的值不会增长过快，将 $\hat{O}_t^{target}$ 和 $O$ 之间的差值平方，以及一个修正系数 $\varphi$ ，使其值与前一项成本的值相匹配。

在实时调整中，动作的目标是在满足约束条件的前提下最大化收益，因此设置动作的奖励为

$$r_t^{RT} = -C_t^{sum} - D_t^S - D_t^{tr}, \quad (3-20)$$

其中 $D_t^S$ 和 $D_t^{tr}$ 分别是对违反约束 (3-7) 和 (3-12) 的惩罚:

$$D_t^{SOC} = \begin{cases} \max(D_t^S - O_t, 0), & D_t^S \geq 0 \\ -D_t^{SOC}, & D_t^S < 0 \end{cases}, \quad (3-21)$$

$$D_t^{tr} = |\max(P_t^S, P_t^P) - P^{tr}|, \quad (3-22)$$

而后, 累积奖励为:

$$R_t^{DA} = \sum_{k=t}^T \gamma_{DA}^{k-t} r_t^{DA}, R_t^{RT} = \sum_{k=t}^T \gamma_{RT}^{k-t} r_t^{RT}, \quad (3-23)$$

其中 $\gamma_{DA}$ 和 $\gamma_{RT}$ 是折扣因子。

### 3.1.3.2 基于 DDPG 的双层半耦合优化策略

MGO 需要分别在前一天和当天执行操作。但是, 在前一天的操作期间, 目标 ES 容量 $\hat{O}_t^{target}$ 是未知的, 因此无法评估该操作所获得的奖励。受神经网络学习和记忆能力的启发, 提出一种基于 DDPG 的半耦合双层网络来解决这个问题, 其中分别为日前动作和日内动作建立行为者-批评网络, 并在训练过程中引入蒙特卡罗采样, 以应对多重不确定性的影响, 如图 3-3 所示。

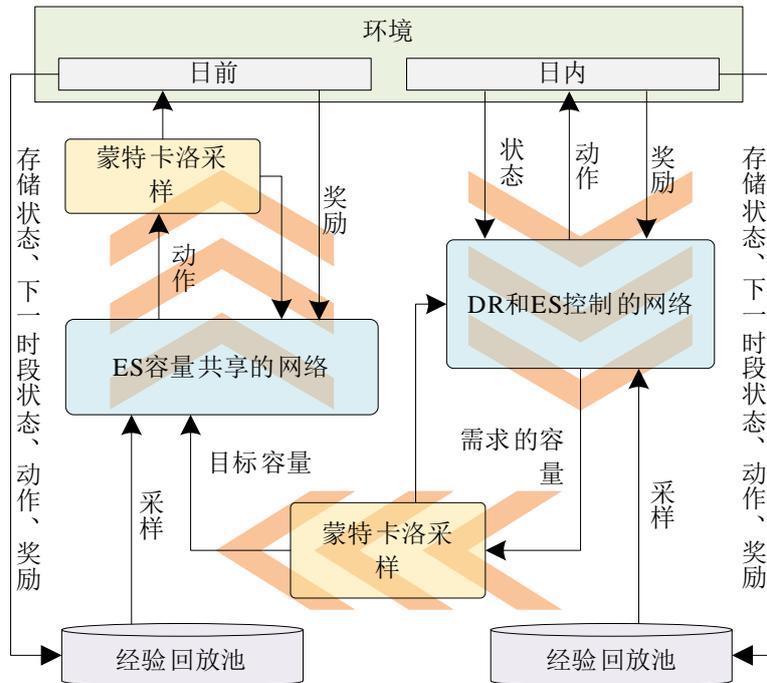


图 3-3 基于 DDPG 的双层半耦合网络

由于 DDPG 是一种离线学习算法，尽管在实践中动作是按时间顺序执行的，但在网络的训练中，可以先训练用于 DR 和 ES 控制的日内网络。共享的 ES 容量是事先未知的，因此我们首先做出以下假设，这些假设保证通过日前的操作来建立。

**假设 1.** 每个时隙所需的 ES 容量始终可以得到满足。

使用假设 1，可以训练 ES 容量共享网络。由于日前行动和日内行动的动作-评价网络都基于 DDPG，因此它们具有类似的训练过程，并且训练过程显示在随后的分析中。为简单起见，省略了一些下标。

为了测量动作  $a_t$  的表现，我们将基于贝尔曼方程的值函数设置为：

$$Q^\mu(S_t, a_t) = E \left[ r_t + \gamma Q^\mu \left( S_{t+1}, \mu \left( S_{t+1} \right) \right) \right], \quad (3-24)$$

其中  $\mu(S_{t+1})$  为动作策略。

使  $\theta^Q$  和  $\theta^\mu$  分别为评价网络和动作网络的参数， $\theta^{Q'}$  和  $\theta^{\mu'}$  分别是目标评价网络和目标动作网络的参数。为了训练评价网络，将损失函数定义为

$$L(\theta^Q) = \frac{1}{H} \sum_i (y_i - Q(S_i, a_i | \theta^Q))^2, \quad (3-25)$$

其中  $H$  是从经验回放时采集的样本数， $y_i$  使用目标网络计算：

$$y_i = r_i + \gamma Q^{\mu'} \left( S_{i+1}, \mu' \left( S_{i+1} | \theta^{\mu'} \right) | \theta^{Q'} \right), \quad (3-26)$$

而后，训练动作网络的策略梯度为：

$$\nabla_{\theta^\mu} J \approx \frac{1}{H} \sum_i \nabla_a Q(S, a | \theta^Q) | S = S_i, a = \mu(S_i) \nabla_{\theta^\mu} \mu(S | \theta^\mu) | S_i, \quad (3-27)$$

其中  $\nabla_{\theta^\mu}$  是  $\theta^\mu$  的梯度，并且  $\nabla_a$  是  $a$  的梯度。

此外，由于 DDPG 是一种确定性策略，因此在探索环境时需要添加随机噪声：

$$\tilde{\mu}(S) = \mu(S | \theta^\mu) + \mathcal{N}_t, \quad (3-28)$$

其中， $\mathcal{N}_t$  是随机噪声。

目标网络的参数通过复制以下内容进行更新：

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1 - \tau)\theta^{Q'}, \quad (3-29)$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1 - \tau)\theta^{\mu'}, \quad (3-30)$$

其中 $\tau$ 是更新速率系数， $\theta^Q$ 和 $\theta^\mu$ 分别是评价网络和动作网络的参数向量， $\theta^{Q'}$ 和 $\theta^{\mu'}$ 分别是目标评价网络和目标动作网络的参数向量。

日前优化的目标是以最低的成本为日内优化提供足够的 ES 容量。优化问题可分为两个子问题，即计算每个时隙的 ES 目标容量 $\hat{O}_t^{\text{taget}}$ ，以及 ES 共享激励价格 $\lambda_t^{\text{CS}}$ 的制定。

PV 输出的不确定性、产消者的负载和响应的不确定性都会影响 $\hat{O}_t^{\text{taget}}$ 的值，因此应用蒙特卡罗采样来获得期望值 $\hat{O}_t^{\text{taget}}$ 。对每次训练中采样 $K_E$ 次，并使用最大值作为目标 ES 容量，如下所示：

$$\hat{O}_t^{\text{taget}} = \max_j O_{t,j}, j = 1, 2, \dots, K_E, \quad (3-31)$$

其中 $O_{t,j}$ 为第 $j$ 个样本中所需的 ES 容量。

特定激励下产消者共享的 ES 容量也具有不确定性。为了使共享的 ES 容量始终满足日内调整的需要，依然采用蒙特卡洛抽样来确定 ES 共享的激励价格。在网络训练中，对于某个动作，即 ES 激励价格 $\lambda_t^{\text{CS}}$ ，所有的产消者共享的 ES 容量均采样 $K_B$ 次，并且用最小值来计算奖励，以确保假设 1 在任何情况下均成立。从而，在网络训练中 $O_t$ 的值如下：

$$O_t(a_t^{\text{DA}}) = \min_j \sum_{i=1}^n O_{i,j,t}(a_t^{\text{DA}}), j = 1, 2, \dots, K_B, \quad (3-32)$$

其中 $O_{i,j,t}(a_t^{\text{DA}})$ 是在动作 $a_t^{\text{DA}}$ 中，第 $i$ 个产消者的第 $j$ 个样本中的共享 ES 容量。然后，可以使用（3-19）计算训练的奖励，并相应地训练网络。

---

#### Algorithm 1

随机初始化评价网络 $Q_{\text{RT}}$ 和DR和ES控制的动作网络 $\mu_{\text{RT}}$ ，设置参数为 $\theta_{\text{RT}}^Q$ 和 $\theta_{\text{RT}}^\mu$ ，

随机初始化评价网络 $Q_{\text{DA}}$ 和储能共享的动作网络 $\mu_{\text{DA}}$ ，设置参数为 $\theta_{\text{DA}}^Q$ 和 $\theta_{\text{DA}}^\mu$

初始化目标网络： $\theta_{\text{RT}}^{Q'} \leftarrow \theta_{\text{RT}}^Q$ ， $\theta_{\text{RT}}^{\mu'} \leftarrow \theta_{\text{RT}}^\mu$ ， $\theta_{\text{DA}}^{Q'} \leftarrow \theta_{\text{DA}}^Q$ ， $\theta_{\text{DA}}^{\mu'} \leftarrow \theta_{\text{DA}}^\mu$

初始化经验回放池 $R_{\text{RT}}, R_{\text{DA}}$

for episode = 1, M do

---

---

```

接收初始观测状态 $S_1^{RT}$ 
for  $t = 1, T$  do
    根据(3-28), (3-29)选择动作 $a_t^{RT}$ 
    实行动作 $a_t^{RT}$ 并观测奖励和下一时段状态
    存储状态转移 $(S_t^{RT}, S_{t+1}^{RT}, a_t^{RT}, r_t^{RT})$ 到经验回放池 $R_{RT}$ 
    从经验回放池 $R_{RT}$ 中随机采样 $H$ 组样本
    根据 (3-25) 更新评价网络
    根据 (3-27) 更新动作网络
    根据 (3-30) (3-31) 更新目标网络
End for
End for
for episode = 1, M do
    根据 (3-32) 接收初始观测状态
    for  $t = 1, T$  do
        根据(3-28), (3-29)选择动作 $a_t^{DA}$ 
        根据 (3-33) 实行动作 $a_t^{DA}$ 并观测奖励和下一时段状态
        存储状态转移 $(S_t^{DA}, S_{t+1}^{DA}, a_t^{DA}, r_t^{DA})$ 到经验回放池 $R_{DA}$ 
        从经验回放池 $R_{DA}$ 中随机采样 $H$ 组样本
        根据 (3-25) 更新评价网络
        根据 (3-27) 更新动作网络
        根据 (3-30) (3-31) 更新目标网络
    End for
End for

```

---

### 3.1.4 仿真实验

#### 3.1.4.1 仿真环境设置

为了验证所提算法的性能，使用 180 天的数据进行模拟，其中 140 天作为训练集，20 天作为验证集，20 天作为测试集。虽然数据集包含不同季节的数据，但由于产消者的负荷数据和 PV 输出数据包含季节性特征，因此没有特别考虑季节性差异，这可以通过算法来学习。RTP 取自宾夕法尼亚州-新泽西州-马里兰州 (PJM) 电力市场。电力需求和光伏输出功率基于 PJM 电力市场的真实数据<sup>[102]</sup>，但按比例缩小。在本文中，MG 作为独立实体参与电力市场。为了鼓励可再生能源的就地消纳，剩余光伏出售给主电网的价格低于 RTP<sup>[103]</sup>，价格系数 $\psi$ 设定为 0.5。表 3-1 中的 TOU 价格是根据[104]中的现有研究结果设定的：

表 3-1 图 TOU 价格

时段	0:00–6:00 (平时段)	6:00–22:00 (峰时段)	22:00–23:00 (平时段)
价格 (美元/千瓦时)	0.02	0.04	0.02

ES 容量共享激励 $f_t$ 和 DR 激励 $g_t$ 的产消者响应函数均被假定为二次函数，即 $O_t = \alpha_{CS} (\lambda_t^{CS})^2 + \beta_{CS} \lambda_t^{CS} + \zeta_{CS} + \varepsilon_{CS}$ ， $\lambda_t^{DR} = \alpha_{DR} (\Delta P_t)^2 + \beta_{DR} \Delta P_t + \zeta_{DR} + \varepsilon_{DR}$ ，

其中所有参数的值如表 3-2 所示。RTP、PV 输出、产消者的负荷需求、产消者对 DR 激励的响应、对 ES 容量共享激励的响应等不确定性均假设为均值为 0、标准差为 0.03 的正态分布。

表 3-2 参数设置

参数	$\alpha_{CS}$	$\beta_{CS}$	$\zeta_{CS}$	$\alpha_{DR}$	$\beta_{DR}$	$\zeta_{DR}$	$e_{CS}$	$\sigma_{CS}$
取值	$3 \times 10^{-9}$	$1 \times 10^{-7}$	$1 \times 10^{-4}$	$5 \times 10^{-7}$	$1 \times 10^{-6}$	$1 \times 10^{-3}$	0	0.03
参数	$\sigma_{PV}$	$P_c^{\max}$	$P_d^{\max}$	$e_{DR}$	$\sigma_{DR}$	$e_L$	$\sigma_L$	$e_{PV}$
取值	0.03	200	200	0	0.03	0	0.03	0

在表 3-2 中， $e$ 和 $\sigma$ 分别是每个不确定变量的期望值和标准差，所有不确定度的计算方法如下：

$$X_U = X_O \times (1 + N(e, \sigma^2)), \quad (3-33)$$

其中， $X_O$ 表示没有不确定性的原始值， $X_U$ 表示具有不确定性的值， $N(e, \sigma^2)$ 是正态分布，期望为 $e$ ，标准差为 $\sigma$ 。

假设负载反弹受到过去 6 个时隙的影响，由于不确定性的影响，负载反弹系数服从标准差为 0.01 的正态分布，每个时隙的负载回弹系数期望值如表 3-3 所示：

表 3-3 负载反弹系数的期望值

参数	$\eta_{t-6,t}$	$\eta_{t-5,t}$	$\eta_{t-4,t}$	$\eta_{t-3,t}$	$\eta_{t-2,t}$	$\eta_{t-1,t}$
取值	0.01	0.02	0.03	0.05	0.1	0.2

为了验证所提方法的优势，在仿真实验中设置了 4 个对比案例。

案例 1：分布式 ES 容量的共享和调整用于提高微电网运营商的收入，而不考虑 DR。

案例 2：DR 用于提高微电网运营商的收入，但产消者的 ES 闲置容量未得到利用。

案例 3：产消者的共享 ES 容量是固定的，即微电网运营商在每个时隙中聚合的 ES 容量是相同的。

案例 4：在共享 ES 容量时忽略了多种不确定性的影响。

DRL 中的所有网络都采用具有三个隐藏层的全连接网络，其中有 256 个神经

元。两个子代理的动作网络和评价网络的学习速率分别设置为 0.0001 和 0.001。该算法是在 Python 3.7.7 中使用 PyTorch 1.8.1 实现的。仿真实验是在配备 Intel (R) Core (TM) i7-9750H 处理器和单个 NVIDIA GeForce GTX 1660 Ti GPU 的笔记本电脑上进行的。

### 3.1.4.2 日内联合优化的效果分析

在 20 天的测试集中选择一天进行效果展示。当天的 RTP, TOU, 负载需求和 PV 输出如下:

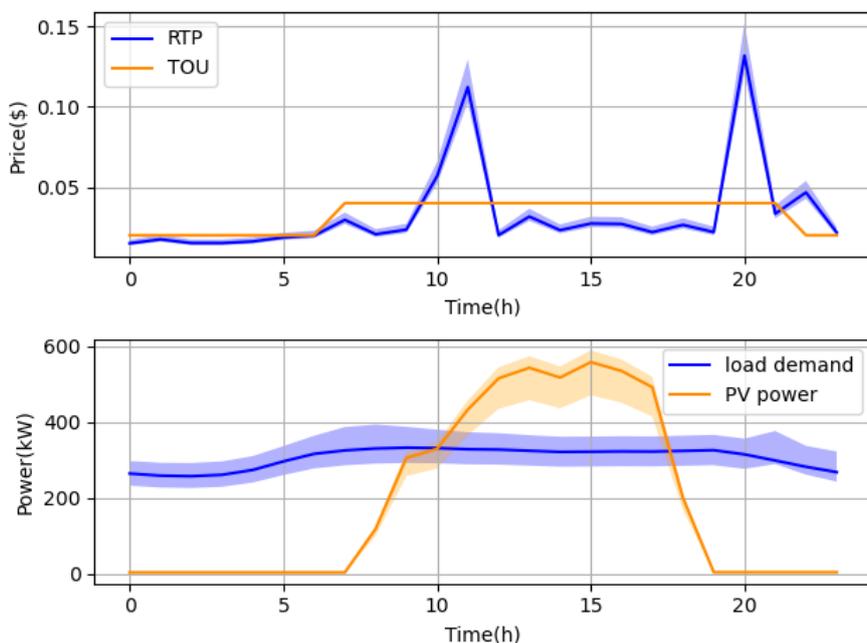


图 3-4 选定日期的价格和功率

由于光伏发电、产消者的负载和 RTP 都是执行操作时的预测值, 因此它们与实际值之间存在误差。图 3-4 中的实线是每个参数的实际值, 阴影是 5000 个蒙特卡罗样本中每个参数的波动范围。需要指出的是, 所选日子的负荷波动不大, 但在数据集中, 有很多天负荷波动较大。

首先分析 MG 的功率需求, 结果如图 3-5 所示。

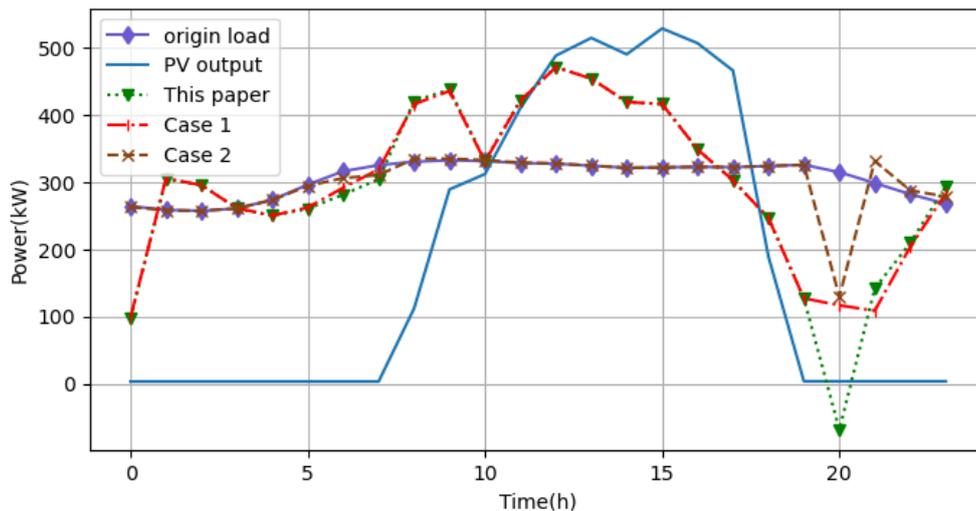


图 3-5 优化前后微电网功率需求的变化

图中蓝线是 MG 中的光伏出力，紫色线是 MG 中的原始负载需求，绿色、红色和棕色线分别是本文案例 1 和案例 2 中调整后的负载需求，包括产消者的负载需求和共享 ES 的充电/放电功率。原始负载需求全天略有波动，但由于光照强度的影响，PV 输出功率变化很大。因此，MG 从 10 点到 17 点钟位置的 PV 输出功率高于负载需求，而 MG 在其他时间段则存在功率短缺。

在本文和案例 1 的优化中，考虑了 ES 容量的共享，微电网运营商可以利用共享的 ES 容量来存储多余的光伏输出，以增加 PV 的本地消耗。全天光伏发电总输出为 4307.4 千瓦时。在优化之前，73.6% 的光伏出力可以在本地消费。本文和案例 1 中利用共享 ES 进行优化后，光伏输出的本地消费比例提高到 86.5%。在不考虑 ES 共享的情况下，PV 输出的本地消耗率与未优化的情况相同，表明 PV 输出的本地消耗率的增加主要是由 ES 共享贡献的。

从图 3-4 和图 3-5 可以看出，DR 主要在 RTP 较高的时间段实施，以减少微电网运营商在特定时间段的亏损，从而提高微电网运营商全天的累计收入。因此，分析每个时间段的收入，如图 3-6 所示：

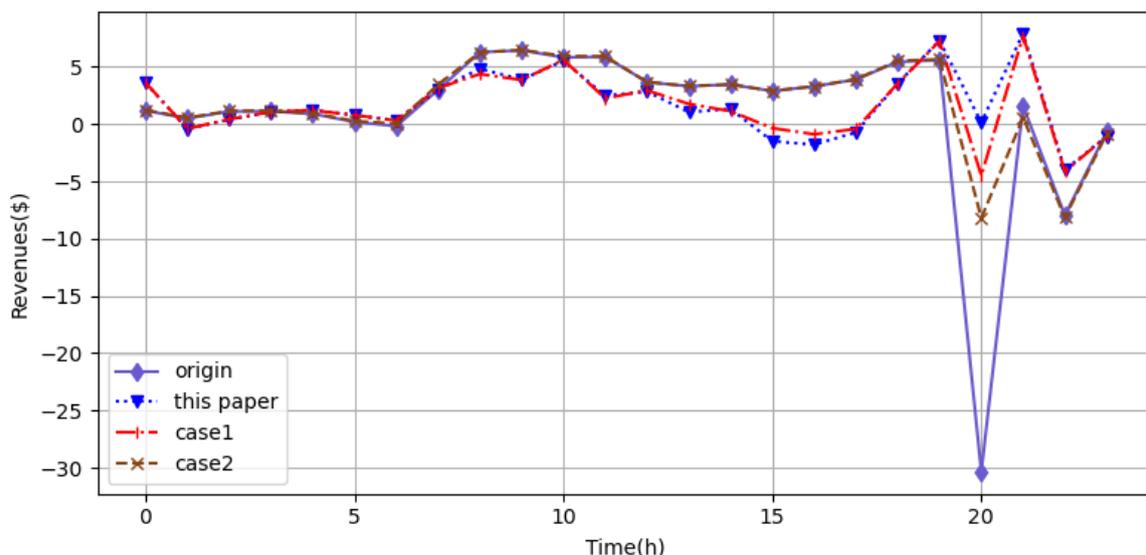


图 3-6 每个时间段的微电网运营商收入

由于 TOU 的价格在大多数时间段都高于 RTP，微电网运营商可以从供电中获得收入，而在其他 RTP 较高的时间段，微电网运营商的收入为负，即它必须承担亏损以满足产消者的能源需求。案例 1 中 ES 的调整和案例 2 中的 DR 都是减少亏损和增加微电网运营商总收入的有效手段。未经优化前微电网运营商全天的总收入为 29.2 美元，而案例 1 的总收入增长了 74.0%，达到 50.8 美元，案例 2 增长了 69.9% 至 49.6 美元。本文提出的算法考虑了 DR 和共享 ES，微电网运营商的收益增加了 113.4%，达到 62.3 美元。结果表明，与单独使用每种方法相比，通过 DR 和 ES 共享的合作，微电网运营商的收入可以进一步提高。由于这两种方法对微电网运营商各时间段全天的收入影响不同，因此与未优化相比，每种方法的收入改善情况分析如下：

由于 RTP 极高，微电网运营商的收入差异主要出现在时间段 20。虽然时隙 11 中的 RTP 也很高，但 PV 的输出可以满足当地的负载需求，因此 RTP 的变化对微电网运营商的营收没有太大的影响。受负载反弹效应的影响，后续时间段的负载需求增加，从而改变微电网运营商的收入。在时隙 20 中的 DR 之后，时隙 21 中的负载需求增加，并且由于 RTP 仍然高于时隙 21 的 TOU，情况 2 中 MGO 的收入减少。相比之下，ES 共享主要影响 MGO 在先前时隙中的收入。在本文和案例 1 的算法中，MGO 的收入在高光伏输出的时隙中减少，因为 MGO 需要支付共享 ES 容量以存储过剩的光伏出力。而后，存储的电量用于满足具有高 RTP 的时隙中的消费者的负载需求，从而提高全天的总收入。

为了验证算法优化效果的稳定性，在测试集中统计连续 10 天的优化效果，这 10 天的总收益如表 3-4 所示。可以看出，微电网运营商在这 10 天的总收入已达到 59.37 美元，分别比案例 1 和案例 2 高出 29.30% 和 9.18%，表明本文提出的算法在连续运行中具有稳定的性能。

表 3-4 微电网运营商连续 10 天的总收入

参数	本文	案例 1	案例 2
收入	59.37 美元	45.87 美元	54.38 美元

### 3.1.4.3 日前的 ES 容量共享的性能分析

MGO 需要支付共享的 ES 容量，所需 ES 容量的变化以及每个时隙中相应的功率如图 3-7 所示。

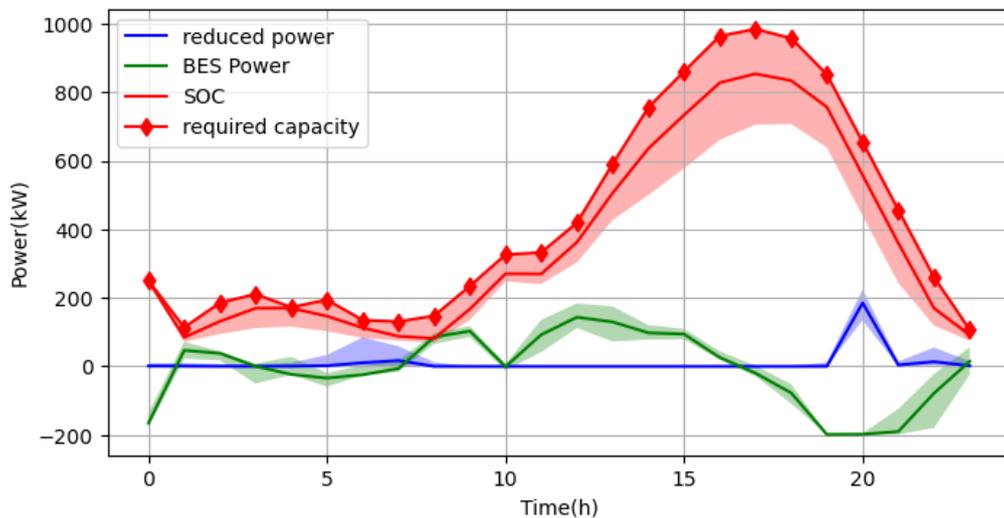


图 3-7 共享 ES 的功率和 DR 中减少的负荷

图 3-7 中未标记的实线是不受不确定性影响的每个变量的值，阴影是 5000 个蒙特卡洛样本中每个变量的波动范围，带菱形标记的红色实线是所需 ES 容量的上限，即日前策略中的目标 ES 容量。

如图 3-8 所示，在具有高 PV 输出的时隙中，多余的 PV 出力存储在共享 ES 中，而后存储的电量，在外部电价较高时，向用户供电。如图 3-5 所示，由于 ES 共享的单位成本是递增，过高的共享 ES 容量会导致总效益下降，因此多余的光伏出力没有被完全存储。除了控制共享 ES 吸收多余的光伏功率外，本文中的算法还可以跟踪 RTP 的变化，在 RTP 较低时（例如时隙 8 和 9）存储电量，并在随后的时隙中向生产消费者供电，从而进一步提高 MGO 的收入。

每个时隙的共享 ES 容量也受到不确定性的影响，ES 共享的结果如下：

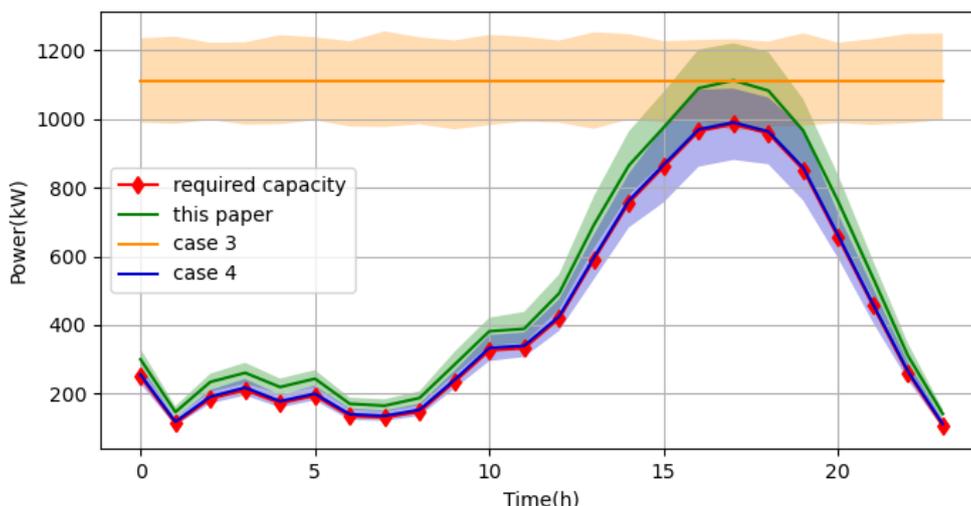


图 3-8 ES 容量共享的结果

图 3-8 中的实线表示微电网运营商获得的共享 ES 容量的预期值，阴影部分是微电网运营商在 5000 个蒙特卡罗样本中获得的 ES 容量的分布区间。在案例 3 中，微电网运营商预测第二天所需的最大 ES 容量，然后根据此值聚合共享 ES。微电网运营商不需要为每个时隙做一个决策，这降低了 ES 共享的难度，但共享的 ES 容量没有得到充分利用，从而减少了微电网运营商的收入。在案例 4 中，ES 共享的不确定性尚未得到充分考虑。虽然共享 ES 容量的成本最低，但用于日内优化的 ES 容量可能不足。本文提出的算法中增加了蒙特卡罗采样过程，保证了共享 ES 容量始终能够满足日内优化的需求，同时最大限度地降低共享 ES 容量的成本。产消者可以通过共享 ES 容量获得收入，图 3-9 显示了其一天中每个时间段的收入。

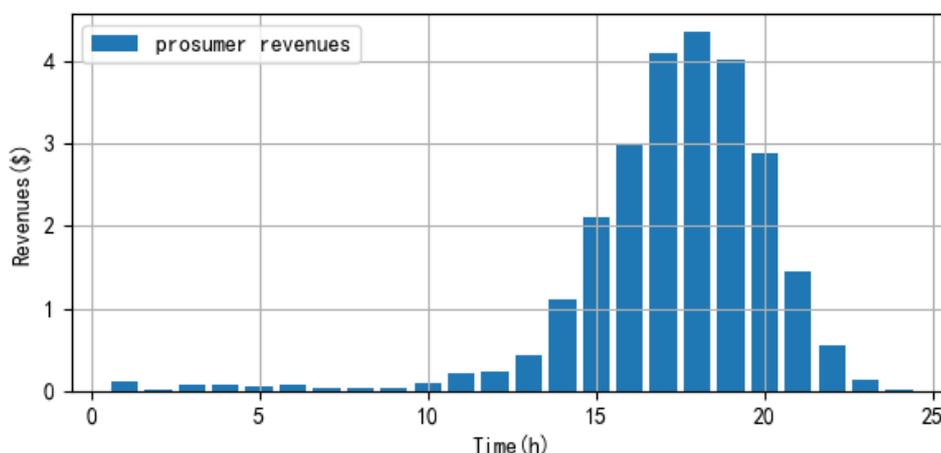


图 3-9 通过共享 ES 容量为产消者提供的收入

从图中可以看出，产消者可以通过 ES 容量共享获得可观的收入，并且收入与共享的 ES 容量呈正相关。

### 3.1.4.4 算法的性能分析

蒙特卡罗采样的性能决定了在最坏的情况下是否可以满足日内优化所需的 ES 容量。所提出的算法中有两个独立的蒙特卡罗样本。由于在确定所需的 ES 容量时存在很多的不确定因素，因此选择此采样过程来分析不同采样次数对确定所需 ES 容量的影响，如图 3-10 所示：

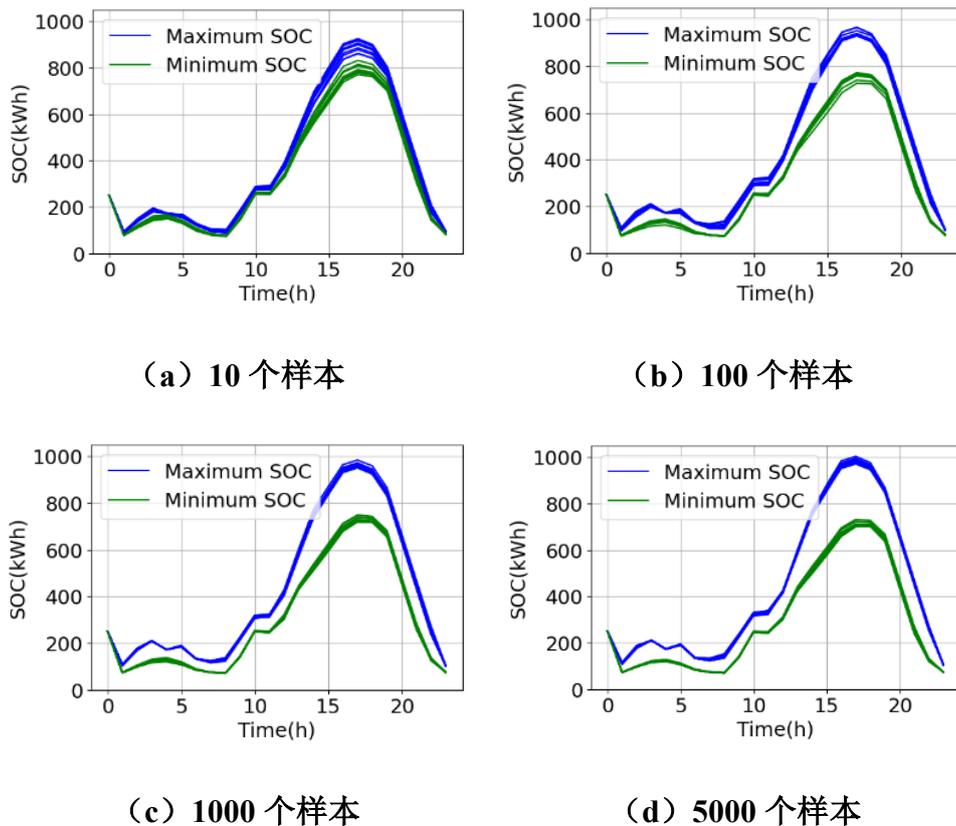


图 3-10 不同蒙特卡罗采样次数的影响

为了应对多重不确定性的影响，使共享 ES 容量满足最坏情况下日内优化的要求，应用蒙特卡罗采样来查找最坏情况下所需的 ES 容量。蒙特卡罗采样次数过多会消耗大量的计算资源，而蒙特卡罗采样次数太少可能难以准确反映不确定性的影响。因此，设置 10、100、1000 和 5000 个蒙特卡罗样本，以验证采样次数对所需 ES 容量的影响。每种类型的采样运行 10 次，结果如图 3-10 所示。

从图 10 (a) 可以看出，由于采样次数太少，每组蒙特卡罗采样所获得的 ES 容量的边界是完全不同的。而且，与其他采样次数组相比，由具有 10 次采样次

数的组确定的所需 ES 容量的上限较低，这并不能反映最坏的情况。随着采样次数的增加，由每组样品确定的所需 ES 容量的上限逐渐稳定。在图 11 (c) (d) 中，每组样品确定的上限相对较近，误差小于 35kWh，最大所需 ES 容量约为 1000kWh，表明在最坏情况下所需的 ES 容量上限可以在足够的采样次数下稳定地找到。

算法的收敛性和稳定性是评估其性能的重要因素。本文显示了网络在 DR 和 ES 控制中的损耗收敛和奖励变化，如图 3-11 所示。

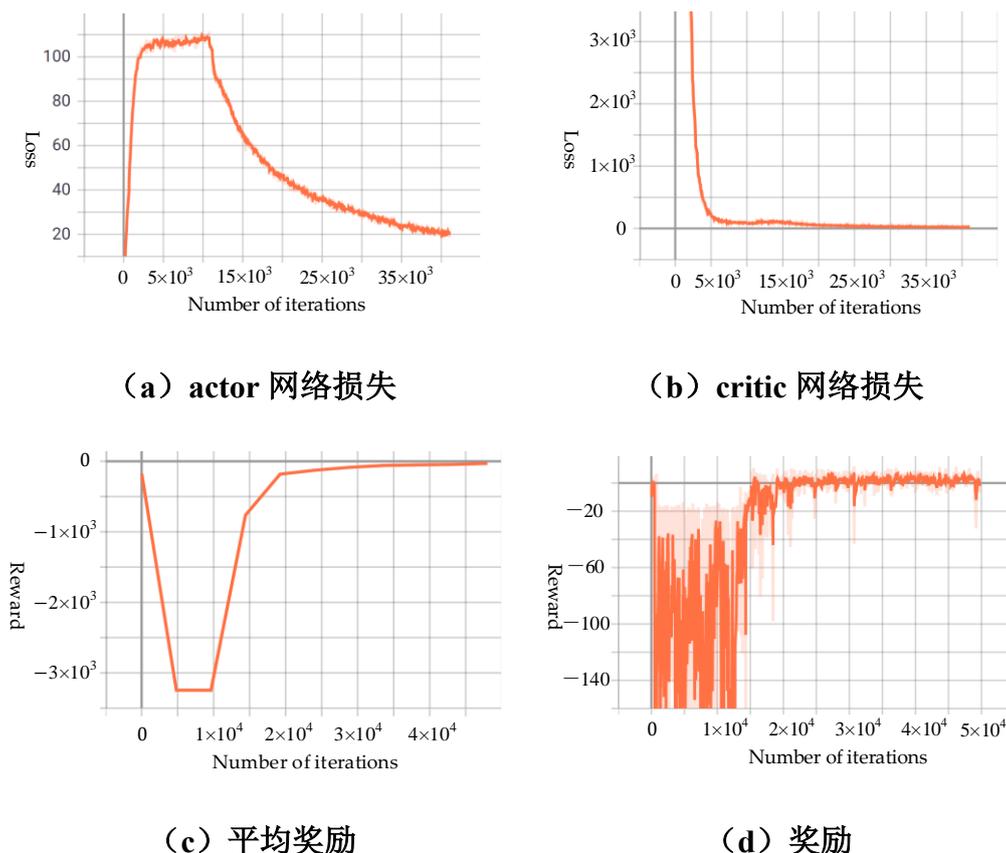


图 3-11 不同蒙特卡罗采样次数的影响

经过大约十分钟的训练，算法的网络损失和奖励趋于稳定，说明该算法具有较高的训练效率。可以看出，actor 网络和 critic 网络的损失变化相对稳定，训练开始时波动后奖励也能顺利收敛。本文尝试了多次训练，大多数算法表现出相似的收敛特性，表明该算法具有较高的稳定性。提高学习速度可以提高收敛速度，但学习速度过高可能导致不收敛。因此，在实际应用中，应在适当的范围内使用高学习率。此外，网络的深度和宽度也会影响算法的性能。网络过大不利于训练，过小的网络不能满足优化的要求。在实践中，有必要根据优化任务的复杂性构建适当的网络。

### 3.2 基于合作博弈的储能共享方式

除了以售电商为核心的储能共享方式外，储能的拥有者还可以通过合作的方式实现储能的共享。本节以微电网（Microgrid, MG）为研究对象，探索微电网组成联盟储能共享及参与市场竞价的优化策略。

#### 3.2.1 微电网联盟架构

本文考虑由  $n$  个 MG 组成的 MG 联盟，其协调架构如图 3-12 所示。多个 MG 的协作通过中央协调器（CentralCoordinator, CC）实现。与中央控制模式不同的是，CC 不能直接控制每个 MG，而是通过机制设计和协商来协调 MG。

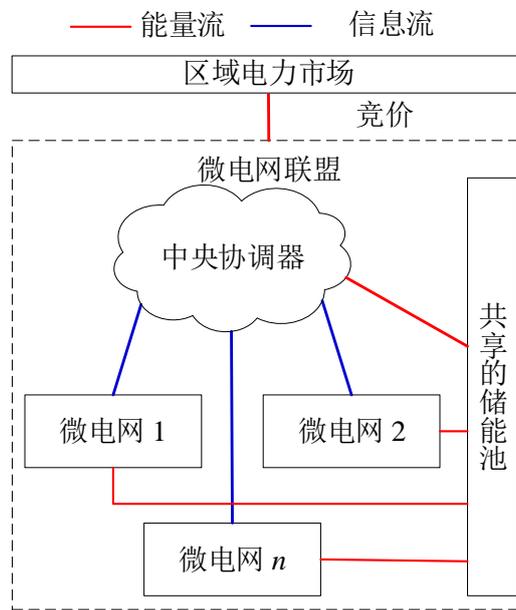


图 3-12 微电网联盟架构

MG 联盟通过 CC 作为独立实体参与区域电力市场。在区域电力市场中，CC 提供投标或报价，以购买需求电量，或出售从分布式能源资源（如光伏（PV）、风力发电等）中获得的过剩的电量。在 MG 联盟中，闲置能量存储由 CC 汇总，MG 联盟的总效益可以通过投标和共享能量存储的控制来提高。

#### 3.2.2 区域电力市场

在区域电力市场中，运营商收集每个参与者的投标/报价曲线，并以社会福利最大化为目标出清市场：

$$\begin{aligned} \max \sum_{i=1}^N \int_0^{P_{i,t}^B} f_{i,t}(x) dx - \sum_{j=1}^M \int_0^{P_{j,t}^S} g_{j,t}(x) dx, \\ \text{s. t. } \sum_{i=1}^N P_{i,t}^B - \sum_{j=1}^M P_{j,t}^S = 0 \leftrightarrow \lambda_t^P, \end{aligned} \quad (3-34)$$

其中  $N$  是买方数量， $M$  是卖方数量。 $f_{i,t}(x)$ 和 $g_{j,t}(x)$ 分别是第  $i$  个买方和第  $j$  个卖方的时隙  $t$  的投标和报价。 $P_{i,t}^B/P_{j,t}^S$ 是购/售电量，通过电力平衡约束，可以获得市场清算价格 $\lambda_t^P$ 。

由于  $MG$  联盟的负荷需求足够大，在区域电力市场中，其投标将影响市场的出清价格。当其电力需求较高时，出清价格增加，否则出清价格降低。对于区域电力市场的特定参与者，其关注剩余需求/供应曲线。剩余需求曲线可根据其他买方的需求和卖方的总供应量获得<sup>[49]</sup>。以竞价为例，在电力市场中，每个参与者分段竞价，结算价格有两种形式，如图 3-13 所示：

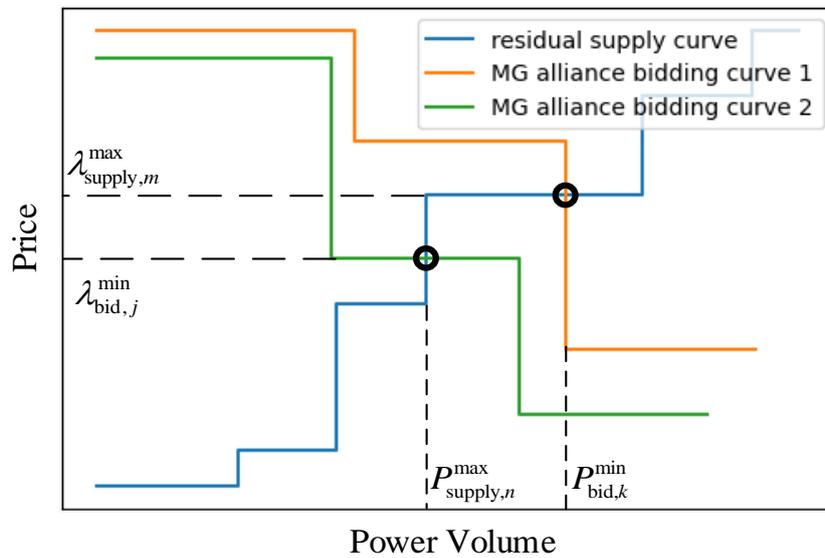


图 3-13 电力市场剩余供给曲线

剩余供给曲线由多个段组成，并且是逐步增加的。每个水平段表示以特定价格供应的电量。类似地， $MG$  联盟的投标曲线也由几个部分组成，但逐步减少，每个水平部分代表特定价格下所需的电量。每个段由两对端点标记，即最低价格和功率 $(\lambda_{bid,j}^{\min}, P_{bid,j}^{\min})$ ，以及最高价格和功率 $(\lambda_{bid,j}^{\max}, P_{bid,j}^{\max})$ 。对于每个水平段， $\lambda_{bid,j}^{\min} = \lambda_{bid,j}^{\max}$ ，对于每个垂直段， $P_{bid,j}^{\min} = P_{bid,j}^{\max}$ 。

剩余供应曲线和竞价曲线的交点是市场出清的电量和电价。如果交叉点位于供应曲线的水平段，即  $MG$  联盟竞价曲线 1， $\lambda_{supply,m}^{\max}$ 和 $P_{bid,k}^{\min}$ 分别用作电力结算价格和电量。在本文中，假设  $MG$  联盟有足够的知识来预测剩余供应曲线，相关工作参见[105][106]。

### 3.2.3 模型构建

#### 3.2.3.1 共享储能模式

在每个时隙中, CC 可以控制共享储能的充放电功率, 储能的状态变化如下:

$$SOC_t = SOC_{t-1} + \eta_t^c P_t^c \Delta t - \frac{1}{\eta_t^d} P_t^d \Delta t, \quad (3-35)$$

$$0 \leq SOC_t \leq \sum_{i=1}^N SOC_i^{shared}, P_t^c \leq P_t^{c,max}, P_t^d \leq P_t^{d,max}, \quad (3-36)$$

其中  $P_t^c$  和  $\eta_t^c$  分别为储能的充电功率和效率,  $P_t^d$  和  $\eta_t^d$  分别是能量存储的放电功率和效率, 并且时隙  $t$  中存储的能量不应超过共享能量存储容量  $\sum_{i=1}^N SOC_i^{shared}$ 。

MG 只有在获得利益时才愿意参与能量存储的共享。由于 MG 联盟的利益受共享储能容量的影响, 因此基于共享储能容量计算给予 MG 的激励  $R_{sum,t}$ , 即,

$$R_{sum,t} = h \left( \sum_{i=1}^N SOC_{i,t}^{shared} \right) \sum_{i=1}^N SOC_{i,t}^{shared}, \quad (3-37)$$

其中  $h$  是一个递增函数, 因为在一定范围内, MG 联盟的收益随着共享能量存储容量的增加而增加, 因此给予 MG 的激励也增加。

由于共享储能池由联盟中的所有 MG 组成, 因此 MG 可以通过共享储能获利。为了反映边际成本的变化, 引入 Shapley 值法计算各 MG 的收入:

$$R_i = \sum_{\phi \in n} \frac{[(|\phi|-1)!(k-|\phi|)!]}{k!} [R_{sum}(\phi) - R_{sum}(\phi \setminus \{i\})], \quad (3-38)$$

其中  $\phi$  是由  $n$  个 MG 组成的联盟的任何子集,  $k$  是从联盟  $\phi$  中移除第  $i$  个 MG 后剩余 MG 的数量。

#### 3.2.3.2 目标函数

CC 的优化目标是 minimized MG 联盟的运营成本, 包括区域电力市场的能源成本、储能共享激励和储能循环寿命损失成本:

$$\min C_a = \sum_{t=0}^T R_{sum,t} + \sum_{t=0}^T \lambda_{clear,t} P_{clear,t} + \sum_{t=0}^T C_{ES,t}, \quad (3-39)$$

其中  $\lambda_{clear,t}$  和  $P_{clear,t}$  分别是区域电力市场时隙  $t$  内的出清价格和出清功率。储

能损失主要由充电/放电引起，并可在一定范围内近似线性化，如下所示：

$$\tau |SOC_t - SOC_{t-1}|, \quad (3-40)$$

其中 $\tau$ 是损耗因子。在优化中，必须满足功率平衡约束：

$$P_{clear,t} + \sum_{i=0}^n P_{t,t}^{MG} + P_t^{ES} - \sum_{i=0}^n P_{t,i}^{PV} = 0, \quad (3-41)$$

其中 $P_t^{ES}$ 是共享储能的充电/放电功率。当储能装置充电时， $P_t^{ES} = P_t^c$ ，而当储能装置放电时， $P_t^{ES} = -P_t^d$ 。

### 3.2.4 基于 soft actor critic 的竞价策略

#### 3.2.4.1 马尔可夫决策过程

投标过程可以用五元组马尔可夫决策过程来描述，即 $M = (S, A, T, r, \gamma)$ 。其中 $S$ 是状态空间，包括剩余供电曲线、负载、充电状态和 PV 出力。 $A$ 是行动空间，即投标策略。 $\gamma$ 是折扣系数。 $r$ 是动作的奖励。

1) 状态：CC 参与市场竞价时，需要了解 MG 联盟的负荷需求和光伏输出，以便了解需要从市场购买多少电力。大量研究致力于预测负荷和光伏输出，因此预测负荷需求 $\hat{P}_t^{MG}$ 和光伏输出 $\hat{P}_t^{PV}$ 被用作状态的元素。其他参与者的出价直接决定了市场的清算结果。因此，市场中的预测竞价对 $(\hat{\lambda}_{supply}, \hat{P}_{supply})$ 也作为状态的元素。最后，储能中存储的电能不能违反约束，因此荷电状态 (SOC) 也是一个状态元素。从而，时隙  $t$  的状态如下：

$$S_t = \{\hat{P}_t^{MG}, \hat{P}_t^{PV}, \hat{\lambda}_{supply}, \hat{P}_{supply}, SOC_t\}. \quad (3-42)$$

2) 动作：CC 参与市场竞价，需要向市场提交价格对 $(\lambda_{bid}, P_{bid})$ 。在 MG 联盟中，CC 还需要控制储能的充放电功率。然而，由于功率平衡的约束，储能功率被视为一个被动变量，即储能功率由(3-40)计算。因此，动作向量是 $(\lambda_{bid}, P_{bid})$ 。

3) 奖励：算法的优化目标是使成本最小化，因此成本是奖励的组成部分之一。此外，由于 BES 的功率是一个被动变量，储能有违反约束的风险，因此应从奖励中减去超过储能极限的处罚。从而奖励如下：

$$r_t = -\vartheta_1 R_{sum,t} - \vartheta_2 \lambda_{clear,t} P_{clear,t} + \vartheta_3 E_t, \quad (3-43)$$

其中 $\vartheta_1$ ， $\vartheta_2$ ， $\vartheta_3$ 分别是每个奖励项的调整参数，使算法获得的奖励保持在合

理的范围內，从而促进算法的收敛性。 $E_t$ 是违反约束的惩罚。在本文中，SOC 的充放电功率受动作值范围的约束，因此 $E_t$ 的值来自违反 SOC 的容量约束：

$$E_t = \begin{cases} \left| SOC_t - \sum_{i=1}^N SOC_i^{shared} \right|, & SOC_t > \sum_{i=1}^N SOC_i^{shared} \\ |SOC_t|, & SOC_t \leq 0. \end{cases}, \quad (3-44)$$

最终所有时段的累积奖励是：

$$R_t = \sum_{\tau=t}^{T-1} \gamma^{\tau-t} r_\tau. \quad (3-45)$$

### 3.2.4.2 基于 SAC 的竞价算法

由于很难为市场上其他参与者的竞价建立一个准确的行为模型，本文引入了一种无模型 DRL 算法来解决优化问题。与其他旨在最大化预期奖励值的算法不同，SAC 训练网络的目的是最大化熵。在竞价过程中，可能有多种动作可以达到最优目标，因为在市场出清规则下，只要分段竞价在特定范围内，就可以获得相同的出清结果。基于 SAC 的算法可以避免竞价策略只选择某个动作，而忽略其他可能的优化动作。因此，SAC 的最优策略是：

$$\pi^* = \arg \max_{E_{(s_t, a_t)} \sim \rho_\pi} \left[ \sum_t R(s_t, a_t) + \alpha H(\pi(\cdot | s_t)) \right], \quad (3-46)$$

其中， $H(\pi(\cdot | s_t))$ 是熵， $\alpha$ 是决定熵重要性的权重系数。

为了获得 $\pi^*$ ，在 SAC 算法中设置了状态值网络 $V_\psi(s_t)$ 、动作值网络 $Q_\theta(s_t, a_t)$ 。优化网络 $V_\psi(s_t)$ 的目标函数为：

$$J_V(\psi) = E_{s_t \sim D} \left[ \frac{1}{2} \left( V_\psi(s_t) - E_{a_t \sim \pi_\phi} [Q_\theta(s_t, a_t) - \log \pi_\phi(a_t | s_t)] \right)^2 \right], \quad (3-47)$$

其中  $D$  是经验回放池，其 $J_V(\psi)$ 的梯度可以被估计如下：

$$\hat{\nabla}_\psi J_V(\psi) = \nabla_\psi V_\psi(s_t) (V_\psi(s_t) - Q_\theta(s_t, a_t) + \log \pi_\phi(a_t | s_t)). \quad (3-48)$$

$Q_\theta(s_t, a_t)$ 网络优化的目标函数为：

$$J_Q(\theta) = E_{(s_t, a_t) \sim D} \left[ \frac{1}{2} \left( Q_\theta(s_t, a_t) - \hat{Q}(s_t, a_t) \right)^2 \right], \quad (3-49)$$

其中 $\hat{Q}(s_t, a_t)$ 使用目标值网络 $V_{\bar{\psi}}(s_{t+1})$ 获得，如下所示：

$$\hat{Q}(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1} \sim p} [V_{\bar{\psi}}(s_{t+1})]. \quad (3-50)$$

$J_Q(\theta)$ 的梯度为：

$$\hat{\nabla}_{\theta} J_Q(\theta) = \nabla_{\theta} Q_{\theta}(a_t, s_t) \left( Q(s_t, a_t) - r(s_t, a_t) - \gamma V_{\bar{\psi}}(s_{t+1}) \right), \quad (3-51)$$

优化 $\pi_{\phi}(a_t|s_t)$ 网络的目标函数为：

$$J_{\pi}(\phi) = E_{s_t \sim D} \left[ D_{KL} \left( \pi_{\phi}(\cdot | s_t) \right) \left\| \frac{\exp(Q_{\theta}(s_t, \cdot))}{Z_{\theta}(s_t)} \right\| \right], \quad (3-52)$$

其中是 $D_{KL}$  Kullback-Leibler 散度， $Z_{\theta}(s_t)$ 是分区函数。为了最小化 $J_{\pi}(\phi)$ ，使用神经网络对策略进行重新参数化：

$$a_t = f_{\phi}(\epsilon_t; s_t), \quad (3-53)$$

其中 $\epsilon_t$ 是 随机变量，并且 (3-52) 可以被重写为：

$$J_{\pi}(\phi) = E_{s_t \sim D, \epsilon_t \sim N} \left[ \log \pi_{\phi}(f_{\phi}(\epsilon_t; s_t) | s_t) - Q_{\theta}(s_t, f_{\phi}(\epsilon_t; s_t)) \right], \quad (3-54)$$

$J_{\pi}(\phi)$ 的梯度可以写为：

$$\begin{aligned} \hat{\nabla}_{\phi} J_{\pi}(\phi) &= \nabla_{\phi} \log \pi_{\phi}(a_t | s_t) \\ &+ \left( \nabla_{a_t} \log \pi_{\phi}(a_t | s_t) - \nabla_{a_t} Q(s_t, a_t) \right) \nabla_{\phi} f_{\phi}(\epsilon_t; s_t), \end{aligned} \quad (3-55)$$

SAC 算法的伪代码如算法 1 所示。

---

#### 算法 1 Soft Actor Critic 算法

---

初始化参数向量  $\psi, \bar{\psi}, \theta, \phi$

for 每次迭代 do

  for 每个环境步 do

$a_t \sim \pi_{\phi}(a_t | s_t)$ ,  $s_{t+1} \sim p(s_{t+1} | s_t, a_t)$ ,  $D \leftarrow DU\{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$

  end for

  for 每个梯度步 do

$\psi \leftarrow \psi - \lambda_V \hat{\nabla}_{\psi} J_V(\psi)$ ,  $\theta_i \leftarrow \theta_i - \lambda_Q \hat{\nabla}_{\theta_i} J_Q(\theta_i)$  for  $i \in \{1, 2\}$

$\phi \leftarrow \phi - \lambda_{\pi} \hat{\nabla}_{\phi} J_{\pi}(\phi)$ ,  $\bar{\psi} \leftarrow \tau \psi + (1 - \tau) \bar{\psi}$

  end for

end for

---

### 3.2.5 仿真分析

#### 3.2.5.1 仿真设置

负载需求和光伏输出来自 PJM 电力市场，时间段为 2021 年 1 月 1 日至 2021 年 12 月 31 日<sup>[102]</sup>。由于 MG 的负载需求和 PV 出力较小，来自 PJM 的所有数据都按比例缩放。需要注意的是，负载和 PV 出力的波动被保留了。剩余供应曲线基于 PJM 电力市场的清算价格生成。假设剩余供给曲线由 20 段组成。基于基本出清价格 $\lambda_t$ ，在每个时隙内随机生成 $[0, 2\lambda_t]$ 范围内的 20 个价格，并在 $[0, 20\text{kW}]$ 范围内随机生成与每个价格对应的电力。储能的充放电效率系数 $\eta_t^c$ 和 $\eta_t^d$ 均设置为 1，损耗系数 $\tau$ 设置为 0.05。

将储能的 SOC、光伏输出、负荷需求和剩余供应曲线设置为算法的输入，因此输入维度设置为 43。假设 MG 联盟在市场上上传 3 个价格对，则动作维度设置为 6。价值网络、soft Q 网络和策略网络都有 3 层，每个隐藏层有 256 个节点。学习率被设置为 0.00001，折扣系数被设置为 1.99。仿真使用 Python 中的 PyTorch 在配备 Intel (R) Core (TM) i7-9750H 处理器和单个 NVIDIA GeForce GTX 1660 Ti GPU 的笔记本电脑上实现。该算法使用了 365 天的历史数据，其中训练集、验证集和测试集的比例分别为 80%、10% 和 10%。随机选取测试集中的一天用于演示算法的效果，其负荷需求、光伏输出和基本出清价格如图 3-14 所示：

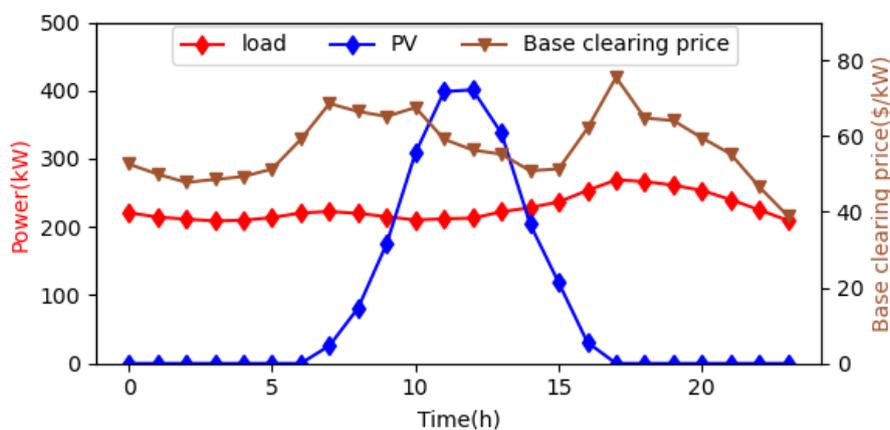


图 3-14 所选日期的负载需求和光伏输出

#### 3.2.5.2 仿真结果分析

每个时间段的出清功率和相应的价格如图 3-15 所示：

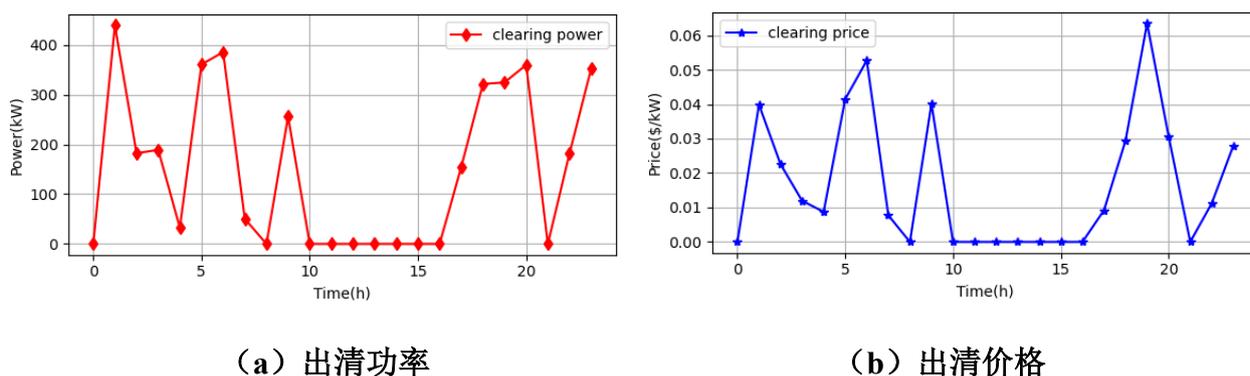
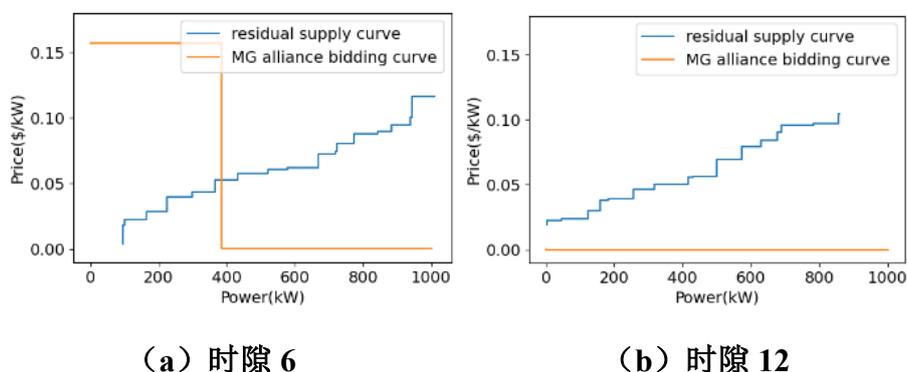


图 3-15 全天每个时间段的清算结果

由于市场的出清结果受 MG 联盟基于其负荷需求的竞价策略的影响，每个时段从市场购买的电力不同，出清价格也波动。中午，由于光伏的高输出，购买的电力为零。在这些时隙中，光伏输出与负载需求相比是过剩的，因此 MG 联盟不需要任何额外的电力。在其他时隙中，例如时隙 9，购买的电力也为零而不超过 PV 输出，因为负载需求由储能来满足。

此外，作为价格制定者，出清价格与 MG 联盟的电力需求正相关，但也受供应商的报价影响。例如，在时隙 20 中，负载需求约为 350kW，出清价格约为 0.03\$/kW。相比之下，在时隙 19 中，出清价格远高于 0.06\$/kW，而负载需求仅约为 320kW。这是因为整个系统的负荷需求在时隙 19 中较高，因此剩余供应曲线上移，显示出较高的价格。

为了确保 MG 联盟始终能够获得所需电力，投标策略应能够适应剩余供应曲线的变化。典型时隙的投标和市场清算如图 3-16 所示：



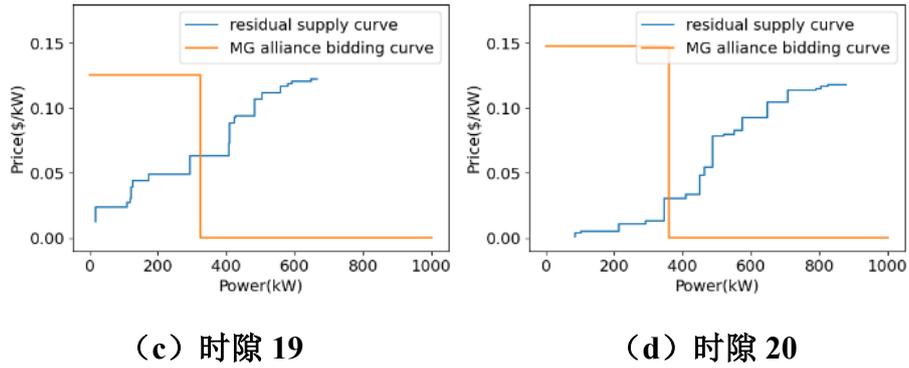


图 3-16 不同时间段的竞价曲线

在竞价中，MG 联盟在每个时隙提交三个价格-功率对，以形成投标曲线。在时隙 12 中，MG 联盟不需要从市场购买电力，因此竞价为 0，在其他时隙中，算法可以根据报价制定适当的竞价策略。为了确保满足电力需求，该算法倾向于在第一个价格电力对中提交较高的价格，剩余供应曲线和投标曲线的交点总是出现在投标曲线的垂直部分。

储能的利用可以提高光伏的本地消耗，储能的 SOC 变化如图 3-17 所示：

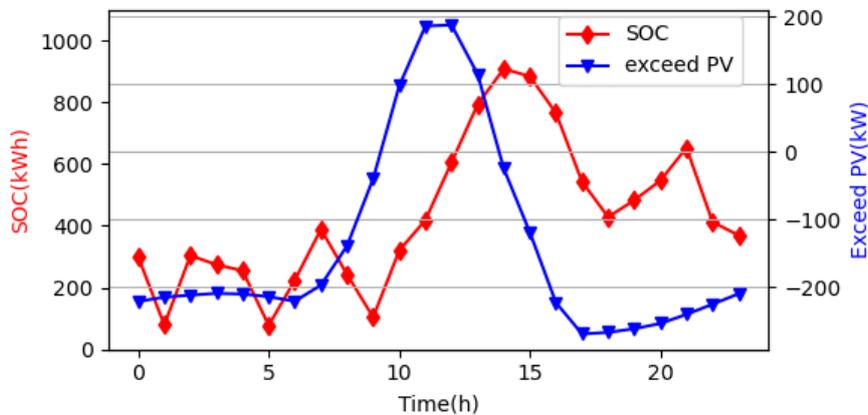


图 3-17 每个时隙中储能的 SOC

在图 3-17 中，蓝线表示过剩的 PV，即满足负载需求后的剩余 PV 输出，红线表示每个时隙中的 SOC。可以看出，储能可以存储中午的过剩光伏输出，并满足后续时隙的负载需求。由于光伏输出的边际成本非常低，提高光伏的本地消耗也可以降低 MG 联盟的电力成本。此外，储能还可以以较低的价格在时隙中存储电力，同时以较高的价格在后续时段中放电，以降低总成本。

### 3.3 本章小结

本节针对分布式储能的共享，探讨了两种模式，一种是以负荷服务商为核心，

采用主从博弈的模式聚合闲置储能资源，另一种是以合作的模式，采用合作博弈的方式实现储能的共享。两种模式均能够实现储能容量的共享，同时基于本文提出的优化算法，能够有效降低系统的运行成本。两种模式有着各自不同的适应场景，在未来的研究中，一方面需要更加深入地探讨这两种模式的实施方式，另一方面需要开发更多的储能共享模式，以满足更多用户的需求。

## 4 分布式数据中心信息能量协同优化策略

随着数字化社会的发展，数据中心的规模在不断扩大，其能耗问题也日益凸显。在最近的几年间，中国数据中心的用电量每年的增幅超过 10%，2020 年突破 2000 亿 kWh，占全社会总用电量的 2.7%。据预测，到 2030 年，全国数据中心的总用电量会在 2020 年的基础上翻倍，占全社会用电量的比重也将上升至 3.7%<sup>[107]</sup>。

与此同时，中国提出“碳中和、碳达峰”的目标，促使电力系统向着以清洁能源为主的结构转变。但可再生能源出力的波动性和不可调节<sup>[108]</sup>的特性使高比例可再生能源电网的供给侧灵活性大大降低，需要充分利用储能<sup>[109]</sup>、可调节负荷等需求侧可调节资源提升电网灵活性，消纳可再生能源。数据中心作为高能耗负荷之一，目前的能量来源还是以传统化石能源为主，2018 年，可再生能源仅仅供应了 23% 的数据中心能耗，低于可再生能源在中国全社会电力消费量的占比 26.5%<sup>[110]</sup>。因此，高能耗和能源供给中可再生能源占比低成为数据中心运营中面临的两大问题。

通过不同数据中心之间的计算任务调度，能实现数据中心功耗的再分配，提升系统的总体效益。但其前提假设是在特定的计算任务下数据中心的功耗是确定的。事实上，根据计算任务时延的需求，可以通过服务器的开关以及 CPU 频率的调整动态调整数据中心的使用率，进一步提升系统总体效率。数据中心的能耗优化问题，就转化为 2 个子问题：1) 考虑能源层经济成本的信息层计算任务分配；2) 考虑信息层计算时延的能源层数据中心能耗的确定。

上述 2 个子问题对于系统总体效益的影响是相互耦合的<sup>[54]</sup>。针对上述优化问题，文[111]构建了双层优化模型，在日前根据预测的计算任务量及时延需求，计算出数据中心可调节功率范围并参与电力市场出清，日内根据出清结果调节数据中心的任务量及功率。文[112]也构建了电网和信息双层优化架构，通过两层间的交互优化，保证计算效率的同时，降低电网阻塞率，同样需要在日前向电网提供数据中心可调度的功率范围。文[113]则是在计算任务预测的基础上，通过两层之间的多次迭代，逐渐逼近最优决策。文[114]综合考虑数据中心和电动汽车的调度，将求解问题转化为随机规划问题并基于交替方向乘子法 (alternating direction method of multiplies, ADMM) 求解该问题。文[115]提出时空双维度任务迁移机制，通过延时容忍型任务时间上的优化和多数据中心间计算任务空间上的迁移，实现

计算负荷与清洁能源出力的匹配。

上述研究中，均构建了能量和信息的双层优化调节架构，通过数据预测或信息交互迭代实现数据中心和电网的联合优化，实现了小时级或分钟级的优化调节。但数据中心计算任务存在较大的波动性<sup>[116]</sup>，且常常会面临难以预测的流量突发现象<sup>[57]</sup>。同时，可再生能源的出力也具有很大的波动性，信息层和能量层的双重不确定性对于控制的反馈速度提出了更高的要求。相比迭代的求解算法，最优控制能够实现毫秒级的反馈，并且已经在火箭运行姿态调整<sup>[117]</sup>、电力系统稳定控制<sup>[118]</sup>、燃料电池的能量管理<sup>[119]</sup>等领域取得了较好的应用效果。

因此，为了提升分布式数据中心控制的速度以更好地应对突发计算任务以及可再生能源出力带来的波动，本文引入最优控制理论，针对目前研究中多数据中心计算任务的分配与单数据中心功率控制协同优化的问题，基于动态微分方程，构建多数据中心计算任务分配及数据中心功率优化动态控制统一模型，实现任务分配与功率控制的同步优化。进而，采用最优控制的极小值原理求解对控制问题求解，有效提升了系统控制的速度，保证系统能够及时随外界环境变化做出调整，使系统能够根据外界环境的波动修正最优控制决策，并对系统控制速度与效益的关系进行了分析。此外，考虑到数据处理的服务质量（quality of service, QoS）与经济成本是相互矛盾的，本文在调节目标中兼顾了延迟成本与经济效益，并探讨了 QoS 与经济效益之间的关系。

#### 4.1 分布式数据中心架构

分布式数据中心分布在电网的不同节点，可以依托电网供电，也可以依托附近可再生能源供电<sup>[120]</sup>。在信息层，各个数据中心互联互通，来自于前端的计算任务可以自由分配，分布式数据中心的信息和能源架构如图 1 所示。

在每个时刻，有  $K$  个前端申请计算任务，每个前端申请的计算任务量为  $l_k(t)$ ，所有计算任务可以分配给  $n$  个分布式数据中心协同完成，第  $i$  个数据中心承担的计算任务为  $L_{dc,i}(t)$ 。各个数据中心之间在信息层相互连通，但由于地理上的分布特性，其通过不同的主网节点接入电网，而每一个数据中心的实时节点电价各不相同<sup>[121]</sup>。数据中心除了能从电网购电外，还可以依托本地可再生能源功能，充分利用可再生能源“剩余”电量，以减少“弃风弃光”现象的发生。由于各地资源禀赋及负荷特性不同，能够给数据中心供能的可再生能源的容量及时域特性也各

不相同。

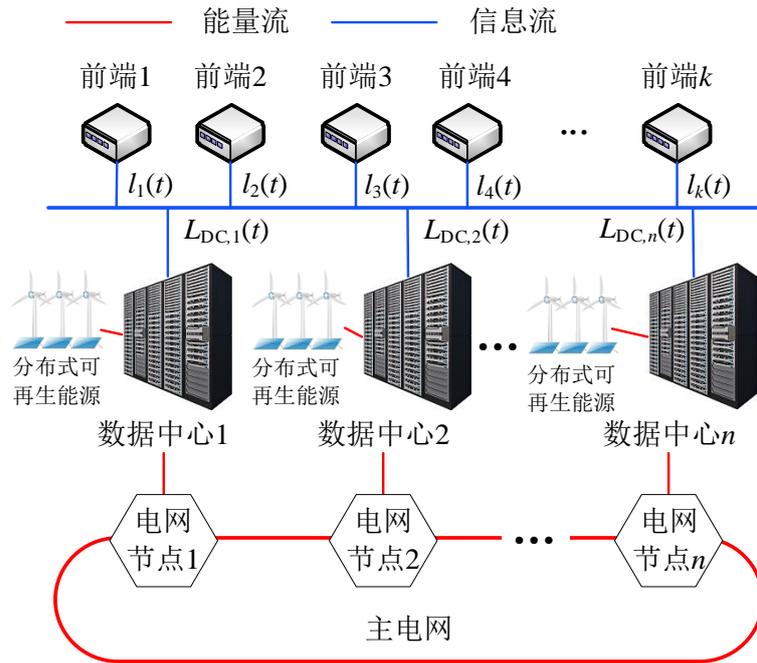


图 4-1 分布式数据中心信息能量交互架构

## 4.2 分布式数据中心信息能量模型

### 4.2.1 数据中心能耗控制模型

数据中心的能耗与承担的计算任务量相关，通过数据中心服务器的启停<sup>[111]</sup>以及 CPU 频率的调节<sup>[113]</sup>，可以控制数据中心的服务器率，使数据中心不总运行在满功率的状态下，以提升数据中心的能效。

数据中心的能耗主要包括服务器能耗以及支持服务器正常运行的散热设备能耗<sup>[122]</sup>。其中，服务器功率主要为 CPU 的功率，可以用如下模型表示：

$$P_{server,i,j}(t) = c_{i,j,0} + A_{i,j}E_{i,j}V_{i,j}^2(t)f_{i,j}(t). \quad (4-1)$$

式(4-1)中数据中心的服务器功率主要包括静态功率和动态功率。 $c_{i,j,0}$ 为第  $i$  个数据中心第  $j$  个服务器 CPU 的静态功率，如果硬件环境不变，则其为固定常量。服务器的动态功率则由电容 $E_{i,j}$ 、时钟频率 $f_{i,j}(t)$ 、CPU 电压 $V_{i,j}(t)$ 共同构成，其能耗与每个时钟周期内的开关次数 $A_{i,j}$ 正相关。其中 $A_{i,j}$ 和 $E_{i,j}$ 在相同的硬件环境下也为固定常量。上述固定常量均可以查阅设备参数信息或通过简单的测量获得。

根据数据中心的典型电能利用效率（power usage effectiveness, PUE）<sup>[124]</sup>，可以得到数据中心服务器功率与总功率的关系：

$$P_{DC,i}(t) = \sum_{j=1}^{M_i} P_{server,i,j}(t)U_i. \quad (4-2)$$

其中： $U_i$ 为第*i*个数据中心的PUE值。如前文所述，可以通过控制开启服务器的数量<sup>[11]</sup>或调节CPU频率<sup>[113]</sup>实现数据中心的功率控制，同时，相对于整个数据中心而言，功率可调节的颗粒度很小，因此认为数据中心的功率是连续可调的，可以定义数据中心使用率控制信号为 $u_{P,i}(t) \in [0,1]$ ，以控制数据中心功率的变化。

对于一个数据中心而言，如果没有新到计算任务，随着数据中心存量任务的处理，数据中心所需处理的计算负载在逐渐下降，则维持同样计算时延所需的功率也在逐渐下降。假设数据中心的功率能够追踪计算负载的变化，则在固定计算时延下，数据中心功率与负载之间存在正相关的关系：

$$P_{DC,i}(t) = k_i L_{DC,i}(t) + P_{base,i}(t). \quad (4-3)$$

其中： $\gamma_i$ 是默认情况下，数据中心功率 $P_{DC,i}(t)$ 与计算任务 $L_{DC,i}(t)$ 间的比例系数，使数据中心的功率能随计算任务量的下降而降低； $P_{base,i}(t)$ 是无计算任务时数据中心的基础功率。

同时，数据中心对任务的处理速度也与功率相关，则剩余计算任务的变化量为

$$dL_{DC,i}(t) = -v_i P_{DC,i}(t). \quad (4-4)$$

其中 $v_i$ 是计算任务变化量与数据中心功率的比例系数。 $P_{DC,i}(t)$ 越大，数据处理速度越快，则单位时间被处理完成的计算任务 $dL_{DC,i}(t)$ 越多。

结合式(4-3)和(4-4)，在特定控制信号下，每个数据中心的功率随时间的动态变化可以表示为

$$dP_{DC,i}(t) = -k_i v_i P_{DC,i}(t) + \beta_{P,i} u_{P,i}(t). \quad (4-5)$$

其中： $\alpha_{P,i}$ 为时间常数， $\beta_{P,i}$ 为最大可到达计算任务数量， $\beta_{P,i} u_{P,i}(t)$ 表示*t*时隙中第*i*个数据中心新增的计算任务数量。

数据中心功率的变化会影响数据中心能量层的能源成本。每个数据中心的能量来源一方面来自本地的分布式可再生能源，一方面来自于电网，且当本地可再生能源过多时，可以出售给电网，获得一定的售电收益，因此，由*n*个数据中心组成的数据中心系统的能源成本在*t*时刻的能源成本为

$$C(t) = \sum_{i=1}^n (P_{B,i}(t)R_{B,i}(t)\Delta t - P_{S,i}(t)R_{S,i}(t)\Delta t), \quad (4-6)$$

$$s.t. P_{PV,i}(t) + P_{B,i}(t) - P_{S,i}(t) = P_{DC,i}(t).$$

当  $t$  时隙数据中心  $i$  的本地可再生能源发电量不能满足数据中心能耗时，其会向电网购买电量  $P_{B,t}(t)$ ，其接入的主网处的购电电价为  $R_{B,i}(t)$ 。当  $t$  时隙数据中心  $i$  的本地可再生能源发电量超过数据中心能耗需求时，剩余部分电量会出售给电网，出售电量用  $P_{S,i}(t)$  表示，上网价格为  $R_{S,i}(t)$ 。同时，每个数据中心可再生能源出力、数据中心功率以及与电网之间的交互功率应实时平衡，即满足功率平衡约束。

#### 4.2.2 多数据中心间计算任务分配模型

图 4-1 中，在每个时隙  $t$ ，数据中心系统根据  $k$  个前端的计算请求分配每个数据中心计算的任务量，因此定义数据中心计算任务的分配控制信号为  $u_{L,i}(t) \in [0,1]$ ，每个数据中心每个时隙承担的计算任务量的动态变化为

$$dL_{DC,i}(t) = -k_i v_i L_{DC,i}(t) + \beta_{L,i} u_{L,i}(t) \quad (4-7)$$

其中： $u_{L,i}(t)$  为分配给第  $i$  个数据中心的任务的控制量， $\beta_{L,i}$  为可以分配给第  $i$  个数据中心的最大的任务量， $\beta_{L,i} u_{L,i}(t)$  为  $t$  时隙中分配给第  $i$  个数据中的计算任务量。为了保证在前端的计算请求均能够被数据中心处理，设置约束如下：

$$\int_0^T \sum_{i=1}^n L_{DC,i}(t) dt \geq \int_0^T \sum_{j=1}^k l_j(t) dt \quad (4-8)$$

式(4-8)保证了所考虑的时间段  $T$  内，每个数据中心承担的计算任务量  $L_{DC,i}(t)$  的总和始终大于等于来自于前端计算请求  $l_j(t)$  的总和。

#### 4.2.3 考虑延迟的信息能量耦合模型

就信息层而言，数据处理的延迟是数据中心服务质量的重要指标，数据中心计算任务的平均延迟又与数据中心承担的计算任务量和功率有关。在目前的研究中，常用 M/M/n 排队模型计算数据中心的平均时延情况<sup>[111][112]</sup>：

$$D_i^P = \frac{1}{\sum_{j=1}^{M_i} \mu_{i,j}(t) - \sum_{j=1}^{M_i} L_{DC,i,j}(t)} \quad (4-9)$$

其中： $D_i^P$ 为第*i*个数据中心的平均时延； $\mu_{i,j}(t)$ 为数据中心服务器的服务率，服务率越高，数据中心的延迟越低； $M_i$ 为第*i*个数据中心服务器的数量。数据中心功率的调节手段有开关服务器<sup>[111]</sup>以及调节服务器的CPU频率<sup>[113]</sup>。关闭服务器会增加每个服务器分担的网络负载，在计算延迟时，可以认为开启的服务器数量不变，每台服务器的服务率 $\mu_{i,j}(t)$ 下降，从而在计算时总有 $M_i$ 台虚拟服务器开启，以方便延迟的计算以及后续优化。

此外，由于网络阻塞的原因，计算任务从前端分配至每个数据中心也会带来时延，同样可以根据M/M/1排队模型，假设计算任务的数据大小呈指数分布，考虑了网络阻塞后的数据中心平均时延为<sup>[126]</sup>

$$D_i = \frac{\varpi}{\sum_{j=1}^{M_i} \mu_{i,j}(t) - \varpi \sum_{j=1}^{M_i} L_{DC,i,j}(t)} \quad (4-10)$$

其中 $\varpi$ 是网络时延参数，取决于网络没有阻塞时的网络传输时延。

在能量层面， $\mu_{i,j}(t)$ 与CPU频率正相关<sup>[127]</sup>，即 $\mu_{i,j}(t) \propto f_{i,j}(t)$ ，又由式(4-1)可知，CPU频率与功率是近似线性关系，因此数据中心服务率与功率之间的关系可近似表示为

$$\sum_{j=1}^{M_i} \mu_{i,j}(t) = \varphi_i \sum_{j=1}^{M_i} P_{server,i,j}(t) \quad (4-11)$$

其中 $\varphi_i$ 为数据中心服务器功率与服务率的转换系数。联立式(4-2)、(4-10)、(4-11)，可得数据中心功耗与计算延迟之间的关系：

$$D_i = \frac{\varpi}{\frac{\varphi_i}{U_i} P_{DC,i}(t) - \varpi \sum_{j=1}^{M_i} L_{DC,i,j}(t)} \quad (4-12)$$

综上，本文构建了能量层的数据中心的能耗模型、信息层的计算任务分配模型以及数据中心功耗与平均计算延迟的能量信息耦合模型。

## 4.3 问题构建及求解

### 4.3.1 问题构建

从前述模型中可以看出，在每个时隙，数据中心需要决策的变量为分配的计算任务量以及数据中心的功率。对于 $t \in [0, T]$ ，数据中心任务量分配以及功率调

节的控制系统可以定义为

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ x(0) = x_0. \end{cases} \quad (4-13)$$

其中,  $A$  和  $B$  均为对称系数矩阵,  $x_0$  为 0 时隙的系统初始状态,  $x(t)$  是  $n$  个数据中心的状态变量。

$$x(t) = [L_{DC,1}(t), P_{DC,1}(t), L_{DC,2}(t), P_{DC,2}(t), \dots, L_{DC,n}(t), P_{DC,n}(t)]^T \quad (4-14)$$

$u(t)$  是  $n$  个数据中心的控制输入:

$$u(t) = [u_{L,1}(t), u_{P,1}(t), u_{L,2}(t), u_{P,2}(t), \dots, u_{L,n}(t), u_{P,n}(t)]^T \quad (4-15)$$

根据式(4-5)和(4-7)可得对角系数矩阵  $A$  和  $B$  分别为:

$$A = \text{diag}[-k_1 v_1, -k_1 v_1, -k_2 v_2, -k_2 v_2, \dots, -k_n v_n, -k_n v_n] \quad (4-16)$$

$$B = \text{diag}[\beta_{L,1}, \beta_{P,1}, \beta_{L,2}, \beta_{P,2}, \dots, \beta_{L,n}, \beta_{P,n}] \quad (4-17)$$

数据中心的能耗成本及功率平衡约束如式(4-6)所示, 除此之外, 数据中心  $t$  时隙的购电电量应大于等于 0, 即  $P_{B,i}(t) \geq 0$ , 数据中心  $t$  时隙出售给电网的电量应介于可再生能源出力与 0 之间, 即  $P_{PV,i}(t) \geq P_{S,i}(t) \geq 0$ 。为了促进可再生能源的就地消纳, 可再生能源的上网价格会低于购电价格<sup>[128]</sup>, 因此, 数据中心会优先使用可再生能源发电为数据中心供电, 则:

$$P_{S,i}(t) = \begin{cases} P_{PV,i}(t) - P_{DC,i}(t), & P_{PV,i}(t) \geq P_{DC,i}(t); \\ 0, & P_{PV,i}(t) < P_{DC,i}(t). \end{cases} \quad (4-18)$$

$$P_{B,i}(t) = \begin{cases} P_{DC,i}(t) - P_{PV,i}(t), & P_{PV,i}(t) < P_{DC,i}(t); \\ 0, & P_{PV,i}(t) \geq P_{DC,i}(t). \end{cases} \quad (4-19)$$

从而, 式(4-6)中的数据中心在  $t$  时隙的单位成本可以改写为

$$C(t) = \sum_{i=1}^n (P_{DC,i}(t) - P_{PV,i}(t)) R_i(t) \Delta t \quad (4-20)$$

其中:

$$R_i(t) \begin{cases} R_{B,i}(t), & P_{PV,i}(t) < P_{DC,i}(t); \\ R_{S,i}(t), & P_{PV,i}(t) \geq P_{DC,i}(t). \end{cases} \quad (4-21)$$

除了经济成本之外, 服务质量也是数据中心能耗优化需要考虑的重要因素之

一。因此，结合式(4-12)，定义目标函数如下：

$$L[x(t)] = \sum_{i=1}^n \left( \gamma D_i + \left( P_{DC,i}(t) - P_{PV,i}(t) \right) R_i(t) \right) \quad (4-22)$$

其中 $\gamma \geq 0$ ，为权重因子， $\gamma$ 越大表示目标对延迟的要求越高。对于延迟而言，一方面希望实现其与经济成本之间的均衡，另一方面其也应满足最大延迟约束，结合式(4-12)可定义延迟约束：

$$D_i \leq t_{delay,i}, i = 1, 2, \dots, n \quad (4-23)$$

其中 $t_{delay,i}$ 即为第 $i$ 个数据中心的延迟上限。将式(4-12)代入式(4-23)可得：

$$\frac{\varphi_i}{U_i} P_{DC,i}(t) - \varpi L_{DC,i}(t) \geq \frac{\varpi}{t_{delay,i}}, i = 1, \dots, n \quad (4-24)$$

为了在控制中满足延迟约束，结合式(4-24)定义： $L_1(x(t)) = Sx(t) + D$ 。

$$S = \text{diag} \left[ \varpi, -\frac{\varphi_1}{U_1}, \varpi, -\frac{\varphi_2}{U_2}, \dots, \varpi, -\frac{\varphi_n}{U_n} \right] \quad (4-25)$$

$$D = \left[ \frac{\varpi}{t_{delay,1}}, \frac{\varpi}{t_{delay,2}}, \dots, \frac{\varpi}{t_{delay,n}} \right]^T \quad (4-26)$$

则不等式约束可以表示为

$$L_1[x(t)] < 0 \quad (4-27)$$

其中符号“ $<$ ”表示矩阵 $L_1[x(t)]$ 中的每一个元素均小于等于0。此外，式(4-13)所表示的系统应在考虑的时间内，处理完所有用户的任务请求，即满足式(4-8)所表示的约束，对其移项变形可得：

$$\int_0^T \left( \sum_{i=1}^n L_{DC,i}(t) - \frac{1}{T} \sum_{j=1}^k l_j(t) \right) dt \geq 0 \quad (4-28)$$

为了在控制中满足任务处理量的约束，根据式(4-28)定义 $L_2$ ：

$$L_2(x(t)) = \frac{1}{T} \sum_{j=1}^k l_j(t) - a^T x(t) \quad (4-29)$$

其中 $a = [1, 0, 1, 0, \dots, 1, 0]^T$ ，为一个 $2n$ 维向量，则式(4-28)可以表示为

$$\int_0^T L_2(x(t)) \leq 0 \quad (4-30)$$

每个数据中心的任务分配量以及功率可以在 0 到最大值之间被调节。因此，定义每个数据中心任务分配量和功率的控制集  $\mathbf{u}(t)$  应满足如下约束：

$$0 < \mathbf{u}(t) < 1 \quad (4-31)$$

最终构建数据中心任务分配及功率确定的最优控制问题如下：

$$\begin{aligned} \min J(0, \mathbf{x}(0), \mathbf{u}(\cdot)) &= \int_0^T L[\mathbf{x}(t)] dt \\ \text{s.t. } L_1[\mathbf{x}(t)] &< 0, t \in [0, T] \\ \int_0^T L_2[\mathbf{x}(t)] dt &\leq 0 \end{aligned} \quad (4-32)$$

经过上述推导与转换后，数据中心计算任务分配以及最佳功率的选择就被转化为一个最优控制问题，而后就需要对该系统求解，以获得数据中心每一时隙最优的控制策略。

### 4.3.2 问题求解

针对式(4-13)的系统，式(4-32)的最优控制问题既有不等式过程约束，也有不等式积分约束。为了在最优控制的过程中满足约束，引入 Lagrange 乘子： $\lambda(t)$ 、 $\eta(t)$ 、 $\zeta$ 。首先考虑不等式积分约束，定义 Hamilton 函数  $H$ ：

$$H[x(t), u(t), \lambda(t), \zeta, t] = L[x(t)] + \lambda^T(t)[Ax(t) + Bu(t)] + \zeta L_2[x(t)] \quad (4-33)$$

其中： $\lambda(t)$ 为系统约束的参数， $\zeta$ 为不等式积分约束项的参数。进而考虑不等式过程约束，定义 Lagrange 函数  $\hat{L}$ ：

$$\hat{L}[x(t), u(t), \lambda(t), \eta(t), \zeta, t] = H[x(t), u(t), \lambda(t), t] + \eta^T(t) L_1[x(t)] \quad (4-34)$$

其中  $\eta(t)$  为不等式过程约束项的参数。

由最优控制理论的极小值原理<sup>[129]</sup>，式(4-32)中问题的最优控制信号  $\mathbf{u}^*(t)$  和最优轨线，即每个数据中心的负荷分配量与功率随时间的最佳状态变化曲线  $\mathbf{x}^*(t)$  应满足最优控制原理中所需的方程、等式和不等式。将数据中心计算任务分配以及功率控制的相关变量带入，并转化为最优控制中可以求解的形式，可以获得系统求解时应满足的约束条件：

$$\left\{ \begin{array}{l} \zeta \geq 0 \\ \int_0^T L_2[x(t)]dt \leq 0 \\ \zeta \int_0^T L_2[x(t)]dt = 0 \\ \eta(t) > 0 \\ L_1[x(t)] < 0 \\ \eta^T(t) L_1[x(t)] = 0 \end{array} \right. \quad (4-35)$$

通过 GEKKO 工具包<sup>[130]</sup>求解该方程组，可得到 $u^*(t)$ 和对应的负荷分配量以及 $x^*(t)$ 。

## 4.4 仿真实验

### 4.4.1 仿真环境设置

本文以 3 个数据中心为例进行仿真，数据中心的参数如表 4-1 所示：

表 4-1 数据中心仿真参数

数据中心	DC1	DC2	DC3
服务器台数	$2 \times 105$	$2 \times 105$	$2 \times 105$
服务器最大功率/W	300	300	300
服务器每秒最大处理任务个数	30	30	30
$\varphi$	0.1	0.1	0.1
PUE	1.2	1.3	1.4
kivi	1	1	1
$\beta$	1	1	1
$\varpi$	1	1	1

为了验证本文方法的适用性，针对可再生能源过剩与可再生能源不足 2 种情况分别设置场景 1 和 2。

每个数据中心处于不同的电网节点，即每个数据中心有不同的节点边际电价。目前美国 PJM 电力市场和 ERCOT 电力市场的结算周期为 5min<sup>[131]</sup>，而随着可再生能源的渗透，未来的结算周期可能会进一步降低。因此，本文模拟 5min 内价格信号波动 2 次。

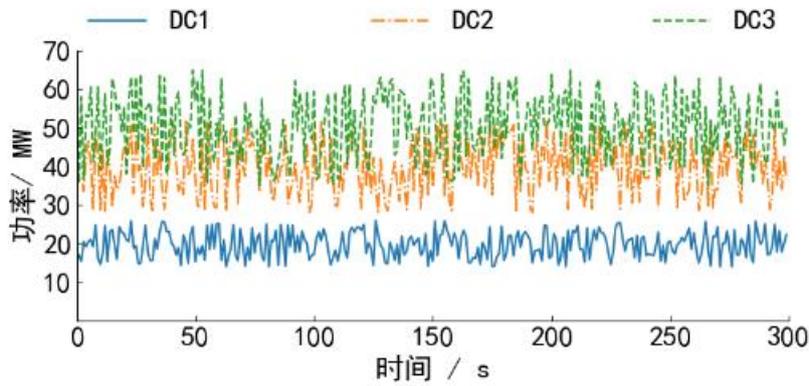
同时，每个数据中心的所获得的可再生能源电量也不同，由于受外界环境影响，可再生能源的出力具有波动性，从前端传输的数据中心的计算任务同样存在较大的波动。基于文[115]中真实的负载数据，结合文[132]中随机场景的生成方法，本文设置了 3 个数据中心的可再生能源出力及计算任务数据，并在 $\pm 30\%$ 的范围

内按照均匀分布，增加随机扰动  $\varepsilon$ ，以模拟可再生能源出力及计算任务的不确定性。为了提升可再生能源的就地消纳率，可再生能源的出售给电网的价格常常低于购电价格<sup>[128]</sup>，各项参数具体取值如表 4-2 所示。

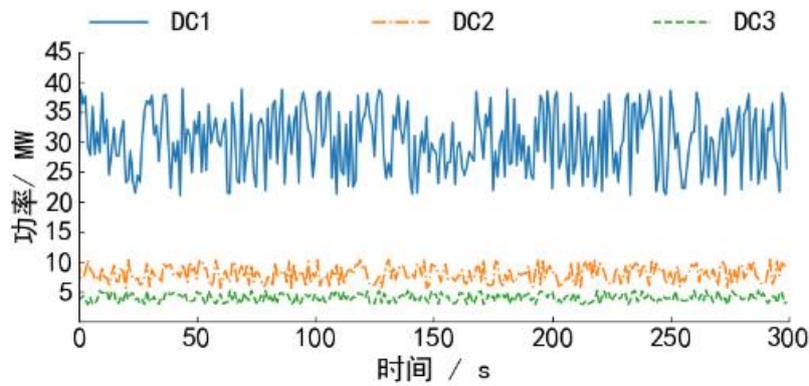
表 4-2 仿真环境参数

数据中心		DC1	DC2	DC3
可再生能源/MW	场景 1	$20 + \varepsilon_{1,1}$	$40 + \varepsilon_{1,2}$	$50 + \varepsilon_{1,3}$
	场景 2	$30 + \varepsilon_{2,1}$	$8 + \varepsilon_{2,2}$	$4 + \varepsilon_{2,3}$
购电价格/(元·kW <sup>-1</sup> )	0~100 s	0.8	0.6	0.7
	101~200 s	0.5	0.7	0.6
	201~300 s	0.8	0.9	0.6
售电价格/(元·kW <sup>-1</sup> )	0~100 s	0.3	0.4	0.2
	101~200 s	0.4	0.2	0.2
	201~300 s	0.2	0.3	0.5
每秒需处理总任务个数			$3 \times 10^6 + \varepsilon_L$	
延迟性能的权重 $\gamma$			$1 \times 10^4$	

增加扰动后可再生能源出力如图 2 所示。



(a) 场景 1



(b) 场景 2

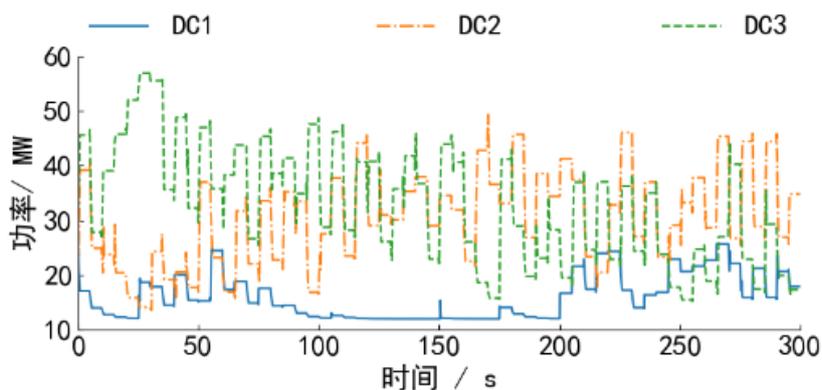
图 4-2 两种场景下各个数据中心可再生能源出力 W

现有研究已经证明，相比于仅考虑数据中心间的计算负载分配<sup>[60]</sup>，或仅考虑单个数据中心的功率优化<sup>[61]-[63]</sup>，综合考虑二者的协同优化能够降低数据中心运行成本<sup>[111]-[114]</sup>。但目前的研究仅实现了最小分钟级的优化，如文[63][111][112][114]为小时级优化，文[62][113][115]为分钟级优化，未发现有研究实现秒级优化。因此，本文基于文[62][113][115]的研究，设置每次控制间隔 60s 为对比算法，进而进一步分析不同控制频率对系统效益带来的影响。

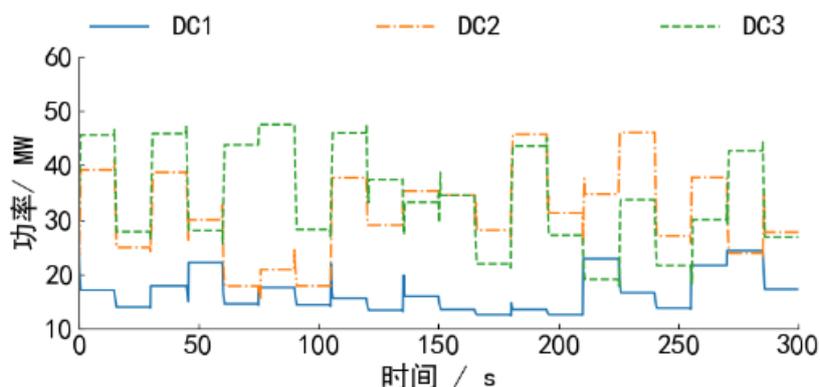
#### 4.4.2 控制频率影响分析

由于可再生能源出力以及计算任务均存在较大的波动性，且难以被十分准确地预测，提升控制频率有助于降低不确定性带来的影响，本文首先对比了控制频率对经济性的影响。基于最优控制，每次优化可以在 100ms 之内完成求解，但考虑到调节过程中的还存在任务分解、通信等会耗时，本文取最小每次控制间隔为 5s。

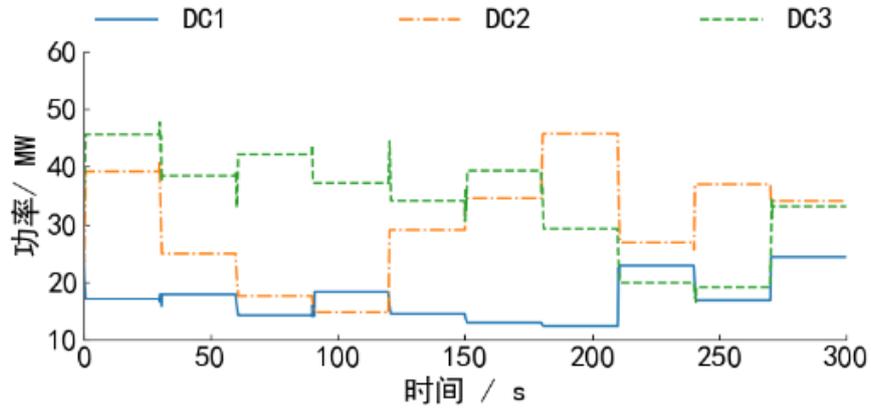
2 种场景下 3 个数据中心任务分配和功率情况分别如图 4-3 和 4-4 所示。



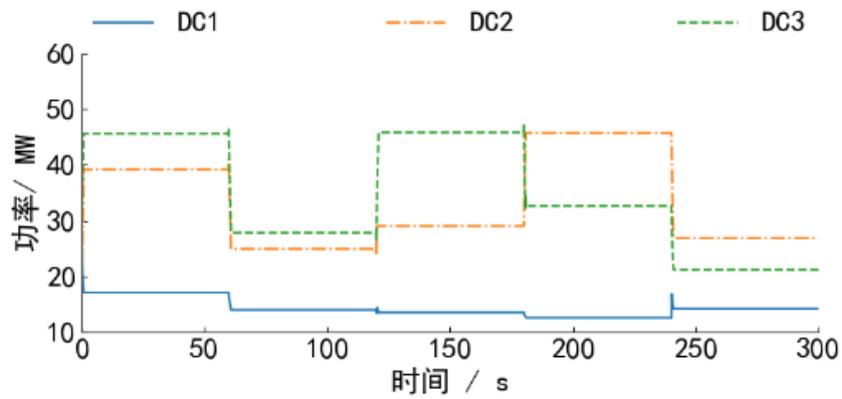
(a) 控制间隔 5 s



(b) 控制间隔 15 s

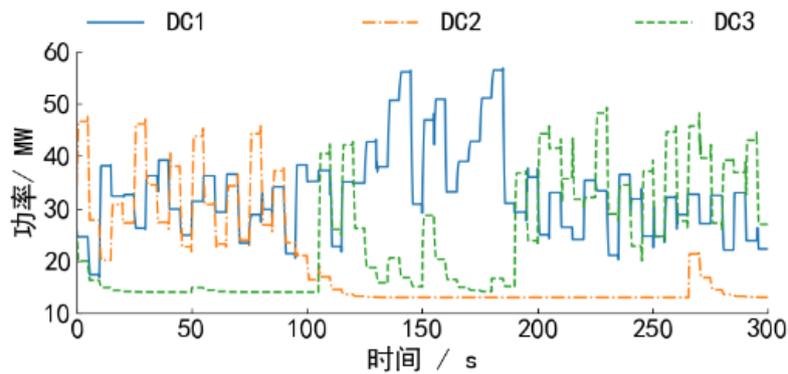


(c) 控制间隔 30 s

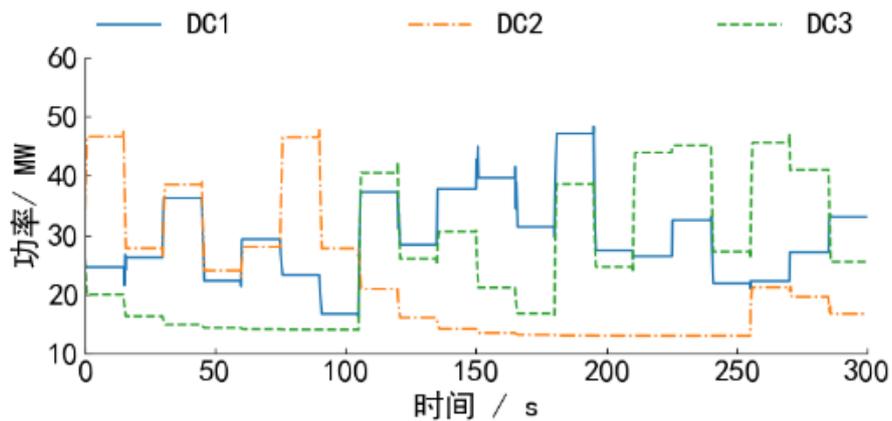


(d) 控制间隔 60 s

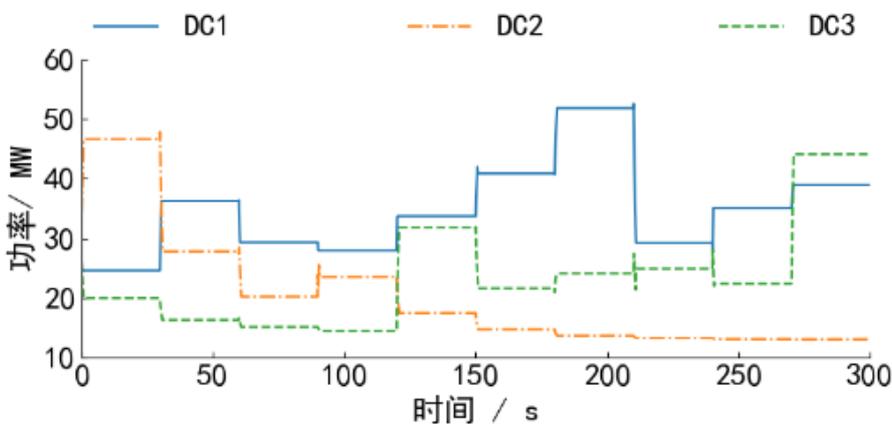
图 4-3 场景 1 数据中心的功率变化



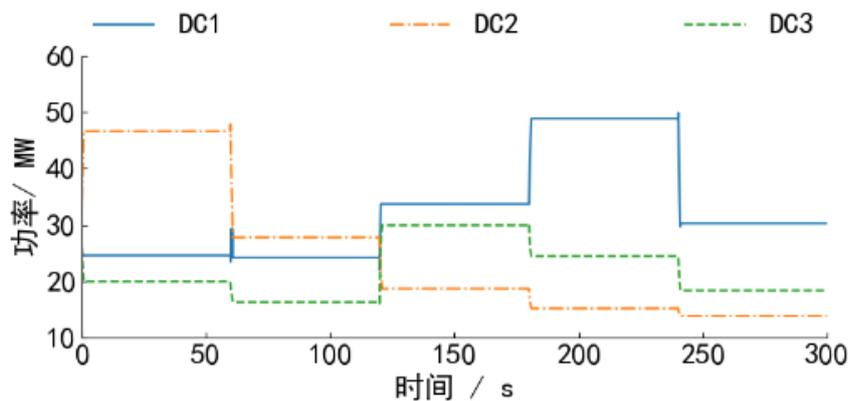
(a) 控制间隔 5s



(b) 控制间隔 15s



(c) 控制间隔 30s



(d) 控制间隔 60s

图 4-4 场景 2 数据中心的功率变化

2 种场景下，不同控制间隔在 5 min 中的总效益如表 4-3 所示。

表 4-3 不同控制频率下的效益

场景	5 s	15 s	30 s	60 s
1	10659.16	9996.86	7254.43	6689.15
2	-18897.58	-20310.85	-20002.74	-20621.53

从图 4-3 可以看出, 在场景 1 下, 3 个数据中心周边可再生能源的发电总量大于数据中心的能源需求, 因此除数据中心自身的能量消耗外, 还可以将多余光伏出售给电网。在 0~100 s 时, DC3 的售电价格最低, 因此尽可能将计算任务调度至 DC3 处, 以提升可再生能源的售电效益。在 101~200 s 时, DC1 处的售电价格最高, 但其可再生能源的容量较小, 因此其功率依然维持在较小的水平。在 201~300 s 时, DC3 处的售电价格变为最高, 从而分配给其的任务被相应减少, 以提升 3 个数据中心总的售电收益。由于外部价格有变化, 同时可再生能源出力以及计算任务有波动, 因此高控制频率能够更快的追踪不确定性带来的波动, 从而提升系统总体效益。从表 3 也可以看出, 在场景 1 中, 控制间隔为 5s 时, 收益为 10 659.16 元, 比控制间隔为 60 s 时提升了 59.35%。

图 4-4 是场景 2 可再生能源出力不足场景下 3 个数据中心的功率分配。由于可再生能源出力较小, 数据中心所需电量主要来自于外电网购电。由于外部电网的购电价格总是高于对外售电的价格, 因此数据中心的任务分配优先使用本地的可再生能源, 不足的部分再从外电网购买。在 0~100 s 时, 系统尽可能的从 DC2 处购电以满足功率需求, 在 101~200 s 时, DC2 承担了较多的计算任务, 同时由于其本身的可再生能源出力就较高, 其接近于满载运行。在 201~300 s 时, 则尽可能将功率分配至购电价格较低的 DC3 处。由于高频率控制的灵活性, 场景 2 下的总成本也能被有效地降低。

本文统计了 2 种场景下可再生能源就地消纳的情况, 如表 4-4 所示。

表 4-4 不同控制间隔下的可再生能源就地消纳率

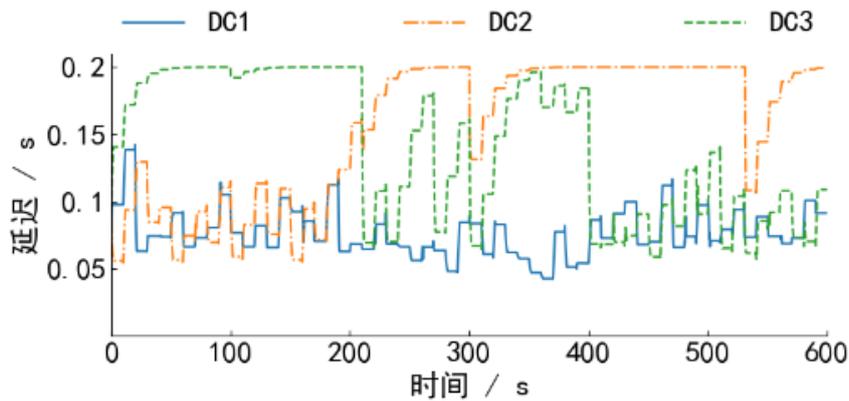
场景	5 s	15 s	30 s	60 s	%
场景 1	12.17	4.34	9.19	5.93	
场景 2	97.84	97.48	97.91	97.66	

从表 4-4 可以看出, 场景由于高频控制能够更好地追踪可再生能源出力的波动, 有效地提升了可再生能源就地消纳率, 相比 60 s 的控制间隔, 5 s 的控制间隔下可再生能源的就地消纳率提升了 6.24%。而在场景 2 中, 由于可再生能源出力较小, 因此不同控制频率下的就地消纳率差别不大, 均在 97% 以上。

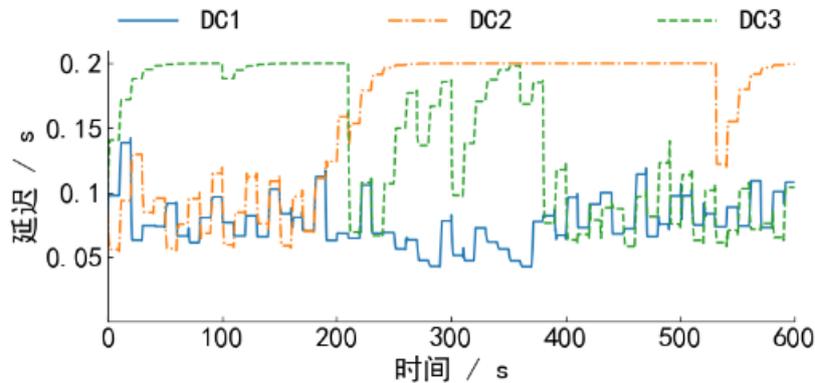
场景 2 中，控制间隔为 30 s 时可再生能源的就地消纳率最高，这是由于场景 2 下各个控制频率对可再生能源消纳的影响很小，而对于实时电价和计算任务不确定性的追踪对总体效益的影响更大，因此高频控制下，虽然可再生能源就地消纳率稍有降低，但成本、计算时延等指标结果更优，从而实现了更好的总体效益。因此，本文继续就不同控制频率下，经济性和延迟指标的优化结果进行分析。

#### 4.4.3 经济性与延迟优化分析

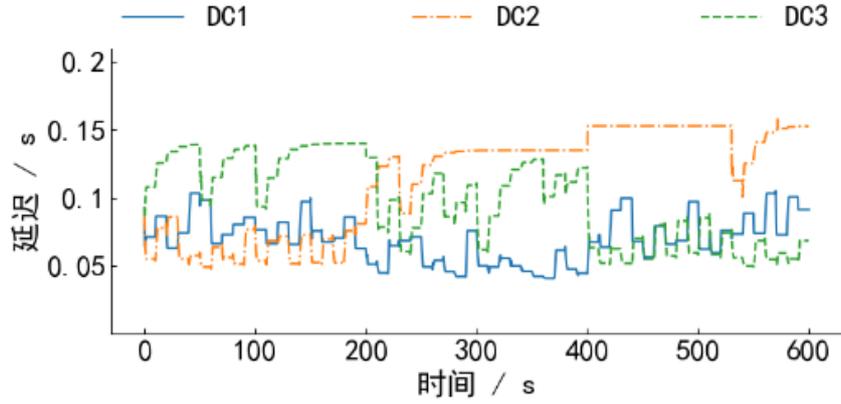
前述仿真实验验证了本文算法中计算任务调度所带来的效益提升，在前文的理论分析中，在相同的计算任务下，数据中心的功率与延迟存在一定的关系，从而适当提升数据中心的计算时延，能够提升系统总体的运营效益。本文基于场景 2，又对数据中心计算时延与系统的运行成本之间的关系进行了仿真实验，验证不同  $\gamma$  下算法控制的效果，实验结果如图 4-5 和 4-6 所示。



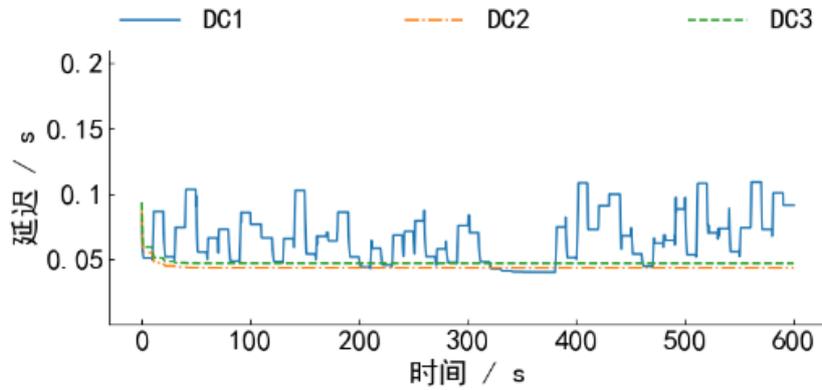
(a)  $\gamma=1 \times 10^3$



(b)  $\gamma=1 \times 10^4$



(c)  $\gamma=1\times 10^5$



(d)  $\gamma=1\times 10^6$

图 4-5 不同  $\gamma$  下的延迟

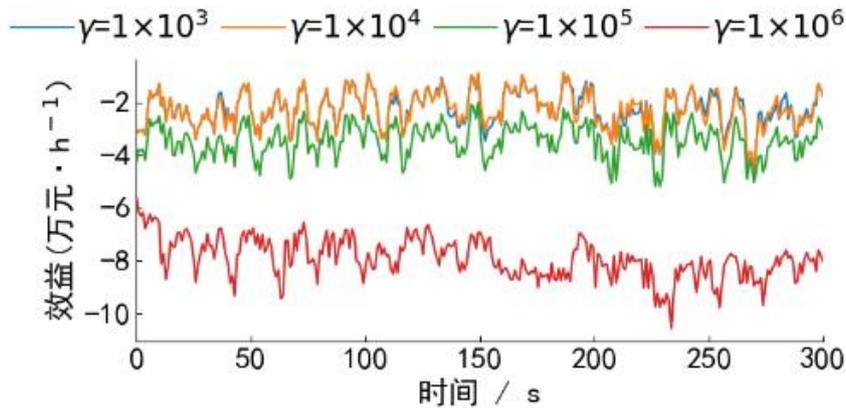


图 4-6 不同  $\gamma$  下的总效益

本文算法在设计目标函数时，考虑了信息层面服务质量这一因素，通过对目标函数中  $\gamma$  的调节能够很方便地变动经济成本和延时成本之间的权重关系。从图 5 和 6 可以看出，延时的约束上限为 200 ms，在  $\gamma$  较小时，系统优化时主要考虑

经济成本，因此 3 个数据中心的计算时延普遍较高，此时经济成本较低。而随着  $\gamma$  的增加，系统在进行优化调节时，更多的考虑了信息层面的服务质量，系统的计算时延有了明显的下降，但随着系统的成本也有了明显的上升。同时，对比  $\gamma=1\times 10^5$  和  $\gamma=1\times 10^6$  的仿真结果可以看出，当计算时延较低时，继续降低计算时延，需要牺牲较多的经济成本。因此，在实际环境中，可以根据需求动态的调整  $\gamma$ ，以实现经济成本与时延成本之间的均衡。

#### 4.5 本章小结

本文基于最优控制理论，实现多个数据中心的任务分配以及每个数据中心功率与计算时延的统一优化。同时，采用动态微分方程对所求问题建模，实现数据中心任务的快速控制。理论分析与仿真结果表明：

1) 通过多个数据中心计算任务的调度，能够有效降低系统总体运行成本，且能够有效提升可再生能源的就地消纳率；

2) 采用动态微分方程对所求问题建模，能够实现数据中心的秒级的优化控制，从而更好地追踪由于可再生能源出力以及计算任务不确定性带来的波动，以提升系统总体的经济效益；

3) 在满足计算延迟约束的前提下，提高目标函数中延迟项的权重，能够有效降低计算延迟，但也会大幅提升系统运行总成本，在实际运行中可以根据需求灵活调节权重。

综上，本文所提基于最优控制的能量信息协同优化策略能够综合考虑能源层的能量成本以及信息层的计算时延，实现计算任务分配及数据中心功率控制的快速联合优化控制，降低系统运营成本，提升可再生能源就地消纳率，并能够根据需求调整计算时延。未来可以进一步细化信息层和能量层的研究。在能量层面，可以细化数据中心的能耗组成，考虑冷热电等多种形式综合能源供给系统的调节；在信息层面，可以细化计算任务，包括不同任务之间不同的复杂程度及不同的时延需求等。

## 5 需求侧电-碳耦合交易机制

大量温室气体排放造成的气候变化是人类迫切需要解决的最大挑战之一。根据研究，二氧化碳排放量占温室气体排放量的 70%以上<sup>[133]</sup>，超过 40%的二氧化碳排放量来自发电中化石燃料的燃烧<sup>[134]</sup>。因此，电力部门有必要通过整合更多零碳或低碳能源资源来优化能源结构，实现碳减排。然而，应注意的是，尽管碳排放是由发电厂直接产生的，但消费者应该对碳排放负责，因为是需求方推动了电力的生产<sup>[135][136]</sup>。

尽管现有的低碳政策，如碳税，可以在需求侧征收部分碳排放责任（carbon emission liability, CEL），但仍然缺乏可操作的机制来改变用户的能源消费行为。在现行政策下，用户的碳附加费与用电量成比例，尚未考虑到不同的碳排放量。另一方面，需求侧管理（demand-side management, DSM）是一种鼓励消费者在高峰时段使用较少能源的措施，或通过经济激励/惩罚措施将能源使用时间从高峰时段转移到非高峰时段。将需求侧管理与低碳政策相结合，可以设计包括发电、输电和碳排放成本在内的价格信号，以引导用户更绿色地使用电力，从而提高可再生能源的消费比例。

本工作旨在研究考虑时变碳排放的需求侧定价机制和最优交易策略，从而通过负荷调节减少碳排放。主要的挑战是将价格信号与需求侧的碳排放结合起来，同时设计电力-碳耦合市场，以获得期望的负荷消耗行为。此外，从用户的角度来看，需要制定市场中的负荷调节策略和竞价策略，以使电力消费成本最小化。

尽管需求侧监管可以实现碳减排，但该领域的研究很少。其中一个主要原因是，在当前的碳测量机制中，单位电力消耗的 CEL 是固定的，因此缺乏明确的价格信号，用于负荷调节以减少碳排放。比例分摊原则（proportional sharing theorem, PST）可以实现碳排放的实时测量，但 CEL 仍然难以在用户之间灵活传输。这是由于在 PST 中，CEL 的计算严格与功率流成比例。由于电力需要实时平衡，因此对 CEL 施加了不必要的实时平衡约束，这限制了用户之间 CEL 的灵活传输。

受能量存储灵活性的启发，本文提出了“虚拟碳存储”的概念，以放松碳和电之间的紧密耦合。通过“虚拟碳存储”，CEL 可以临时存储，无需立即支付，因此可以通过交易在用户之间灵活转移 CEL。然后，本文提出了一种联合电力碳交易机制，以实现用户之间电力和 CEL 的综合交易。最后，提出了一种基于蒙特

卡罗的 DRL 算法，用于在多源不确定性环境中获得无显式模型用户的最优报价和控制策略。本工作的主要贡献可概括如下：

1) 提出了“虚拟碳存储”的概念，松弛松 PST 对碳排放施加的不必要的实时平衡约束，其中每个用户承担的 CEL 严格按照潮流计算。通过虚拟碳存储，CEL 可以临时存储，从而可以在用户之间灵活转移，同时确保供应侧产生的所有 CEL 可以由需求侧支付。

2) 为了在考虑经济成本的同时提高需求侧的碳减排，提出了一种需求侧电-碳联合交易机制，以实现用户之间的 CEL 灵活分配。在该机制下，高灵活性的用户可以通过负荷调节进一步减少碳排放，同时通过交易获得利益。

3) 面对投标和调节过程中的多源不确定性带来的挑战，并且由于在实践中缺乏精确的显式模型，例如舒适度损失函数，提出了一种基于蒙特卡罗的近端策略优化 (PPO) 算法，用于用户的投标和控制策略。这些策略是通过数据驱动方法获得的，并且通过蒙特卡洛采样，优化算法的收敛速度显著提高。

## 5.1 电-碳耦合市场模型

### 5.1.1 系统结构

假设存在  $n$  个互联用户和  $m$  个发电厂以形成区域电力系统。每个用户配备电池储能 (BES)、光伏 (PV) 和负载。用户侧的所有设备由运营商通过能量路由器控制，用户之间的能量交互也可以通过能量路由器实现<sup>[137][138]</sup>。

在区域电力系统中，碳排放直接由化石燃料发电厂产生。由于可再生能源的碳排放非常低，增加可再生能源在电力供应中的比例可以有效减少碳排放。然而，电力供应主要由电力需求决定，因此需要通过需求侧的负荷调节来实现碳减排。

将碳排放成本从供给侧转移到需求侧，是引导用户改变用电量以减少碳排放的有效措施。碳排放成本由两个因素决定，即单位电力消耗的碳排放量和碳排放的单价。为简单起见，CEL 被定义为用户应承担的碳排放量，并且与电力消耗相关。由于 PST 本质上是一种碳排放成本的分配方法，它根据电力流将供应侧的碳排放分配给不同的用户，因此需求侧的 CEL 总量等于供应侧实际产生的碳排放。

PST 用于计算电力消耗的碳排放，如图 5-1 所示。

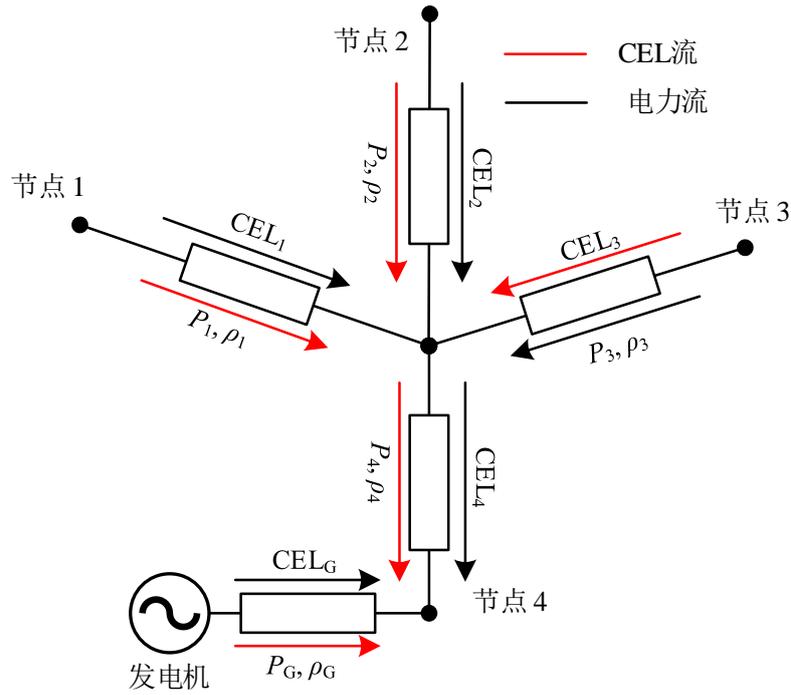


图 5-1 比例共享定理示意图

假设节点 4 消耗的电力来自节点 1、节点 2、节点 3 和本地发电机，来自每个节点和本地发电机的电力分别为 $P_1$ 、 $P_2$ 、 $P_3$ 和 $P_G$ 。由于每个节点的发电机配置不同，来自不同节点的电具有不同的碳强度，即 $\rho_1$ 、 $\rho_2$ 、 $\rho_3$ 和 $\rho_G$ 。因此，节点 4 中注入的 CEL 可以通过以下公式计算：

$$CEL_i = P_i \rho_i. \quad (5-1)$$

从而，节点 4 中的总消耗电量和相应的 CEL 如下：

$$P_{c,4} = \sum_{i=1}^3 P_i + P_G, \quad (5-2)$$

$$CEL_{c,4} = \sum_{i=1}^3 P_i \rho_i + P_G \rho_G, \quad (5-3)$$

其中 $P_{c,4}$ 是节点 4 消耗的电量， $CEL_{c,4}$ 为节点 4 中用户产生的相应碳排放量。

为了实现用户之间电力和 CEL 的灵活交易，本文提出了一个两阶段市场框架，如图 5-1 所示。

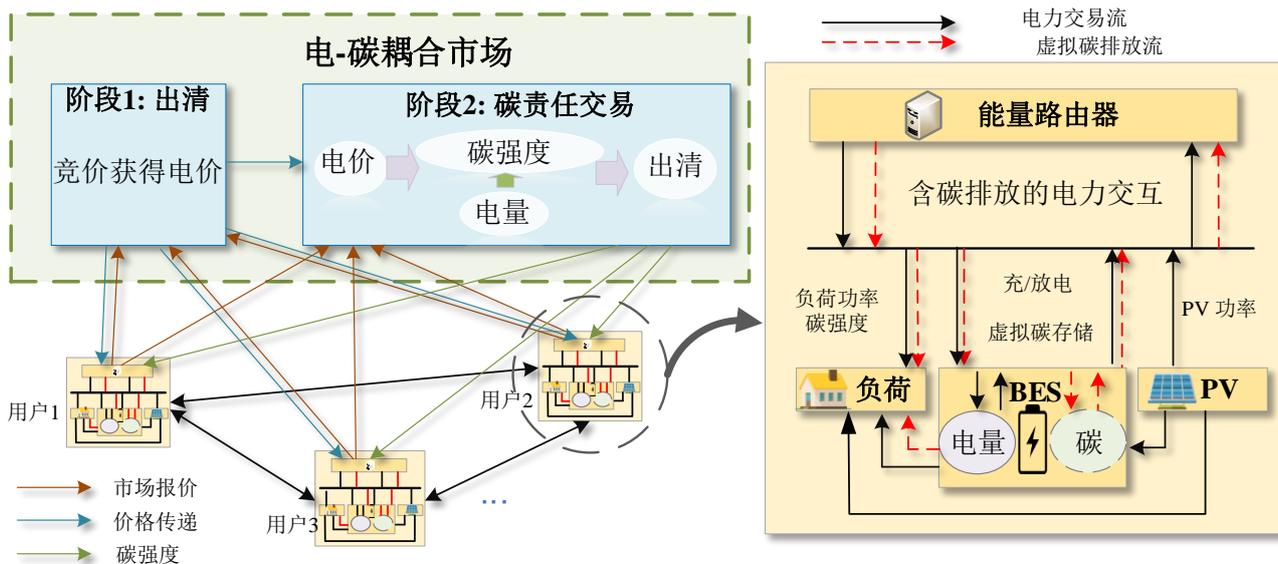


图 5-2 电-碳耦合市场架构

需要注意的是，所提的电力碳耦合市场是零售市场，所有交易都在一天内完成。动作顺序如图 5-3 所示。

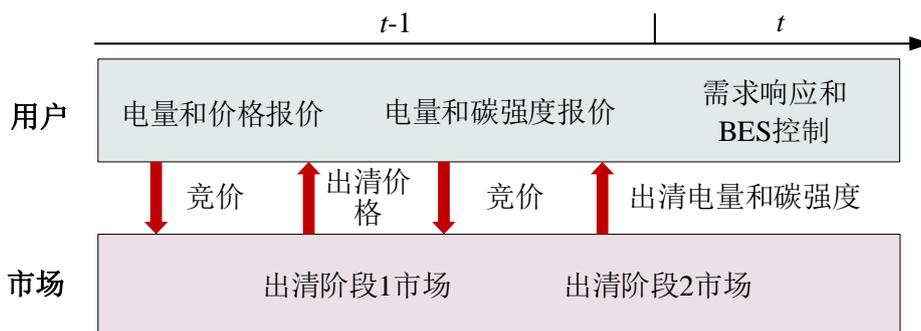


图 5-3 电碳耦合市场的行动顺序。

在时隙  $t-1$  中，每个用户首先在阶段 1 中为时隙  $t$  提供虚拟投标/报价，并且相应地市场被出清。然后，将结算价格发送回用户，用户在第 2 阶段对交易电力和相应的碳强度进行竞价。碳强度是每单位电力的 CEL。市场清零后，确定时间段  $t$  的价格、交易电量和相应的碳强度。在时隙  $t$  开始时，每个用户根据时隙  $t$  中的清除结果调整 BES 的充电/放电功率和负载需求，以保持功率平衡。

### 5.1.2 虚拟碳存储

为了提高碳排放交易的灵活性，本文提出了“虚拟碳存储”的概念。虚拟碳存储可以被视为一个容器，用于临时存储用户的 CEL，而无需立即支付。CEL 仅在用户使用电力时支付，否则可随电力交易转移给其他用户。

随着电力碳耦合市场和虚拟碳存储，由 PST 引起的 CEL 和电力消耗之间的紧密耦合被放松，并且 CEL 可以在用户之间灵活转移。在此基础上，用户的 CEL 可以在不改变特定电量的情况下降低。例如，当用户需要消耗电力，但由于碳帽的限制而无法承担相应的 CEL 时，用户可以以更高的电价购买低碳强度的电力，以满足其电力消耗需求。后续会对虚拟碳存储的模型详细描述。

### 5.1.3 结算价格的确定

每个时隙的清算价格在第 1 阶段通过参与者的虚拟投标/报价获得。通常，投标/报价曲线是阶梯函数[139]，[140]，但在一些电力市场，如宾夕法尼亚-新泽西-马里兰州互联（PJM），投标曲线可以设置为分段线性函数[141]，这有利于更准确地反映边际成本的变化。因此，本研究在投标/报价中使用分段线性函数，如图 5-4 所示。

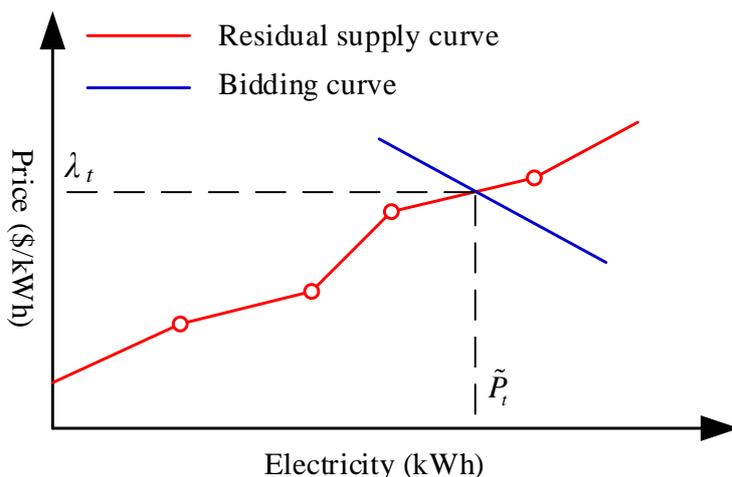


图 5-4 在第一阶段竞标价格

对于每个买方，单调递减的线性函数作为投标提交给市场。类似地，每个卖方向市场提交单调递增的线性函数作为报价。由于需求侧区域市场参与者数量有限，用户竞价可能会影响市场出清价格。因此，图 5-4 中剩余供应曲线和用户竞价的交点是市场的清算结果，即价格  $\lambda_t$  和相应的电力  $\tilde{P}_t$ 。

### 5.1.4 电量和碳强度的确定

根据第一阶段的结算价格，通过第二阶段的投标确定交易电量和相应的碳排放强度。与电力交易类似，参与者使用分段线性函数对电量和碳排放强度进行投标，如图 5-5 所示。

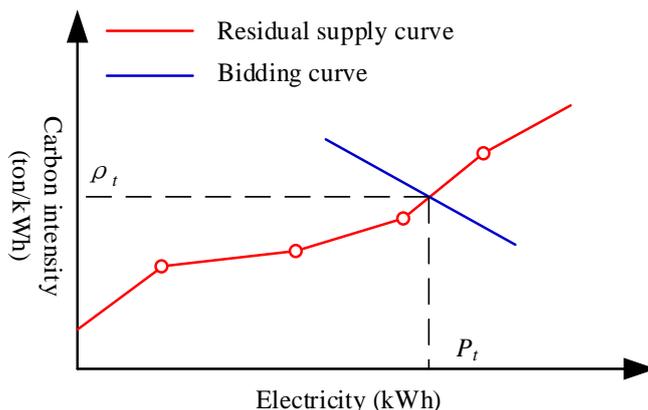


图 5-5 在第二阶段竞价碳强度

从买方的角度来看，剩余供应曲线由多个递增段组成。买方提交一条递减的投标曲线，其与剩余供应曲线的交点是买方可从市场获得的电量 $P_t$ 和相应的碳强度 $\rho_t$ 。

在竞价开始时，卖方期望出售其多余的电力，因此他们倾向于为每单位电力附加低碳强度，而买方需要满足其基本电力需求，从而能够承受更多的碳排放。随着电力需求的增加，买家的边际效用逐渐降低，他们能够承受的碳强度逐渐降低，因此第 $i$ 个买家的竞价曲线单调下降。相反，随着售电量的增加，售电意愿也降低，因此单位电量的额外碳强度逐渐增加。

## 5.2 问题构建

### 5.2.1 通过虚拟碳储存调整碳强度

CEL 是根据发电厂的实际碳排放量产生的，所有 CEL 都应支付。因此，对于用户而言，其输入、输出和支付的 CEL 应当平衡，即

$$\rho_{in,t}P_{in,t} - \rho_{BES,t}P_{BES,t} - \rho_{out,t}P_{out,t} - \rho_{L,t}P_{L,t} = 0 \quad (5-4)$$

其中， $P_{in,t}$ 是在时隙 $t$ 内购买的电力， $P_{out,t}$ 是时隙 $t$ 中售出的电， $P_{BES,t}$ 是时隙 $t$ 中的充放电功率。 $P_{L,t}$ 是时隙 $t$ 中负载的功率。

在(5-4)中，CEL 仍然与功耗紧密耦合。通过虚拟碳储存，可以改变 BES 的碳强度，从而可以松弛 CEL 与电力之间的耦合，如下所述。

负荷需求由购买的电力和 BES 满足。根据(5-4)，所消耗电力的碳强度 $\rho_{L,t}$ 计算如下：

$$\rho_{L,t} = \frac{\rho_{in,t}P_{in,t} + \rho_{BES,t}P_{BES,t}}{P_{in,t} + P_{BES,t}} \quad (5-5)$$

其中,  $\rho_{in,t}P_{in,t}$  是指所购电量附带的 CEL,  $\rho_{BES,t}P_{BES,t}$  是指 BES 中电量的 CEL。

为简单起见, 让售出的电力的碳强度等于所消耗电力的碳强度, 即  $\rho_{out,t} = \rho_{L,t}$ 。将时隙  $t$  中的负载功率表示为  $P_{PV,t}$ 。假设可再生能源的碳排放量为 0, 根据 PST<sup>[70]-[72]</sup>, 每个时隙出售的电力的碳强度为

$$\rho_{out,t} = \frac{\rho_{in,t}P_{in,t} - \rho_{BES,t}P_{BES,t} - \rho_{L,t}P_{L,t}}{P_{in,t} - P_{BES,t} - P_{L,t} + P_{PV,t}} \quad (5-6)$$

当一个实体在时间段  $t$  内出售电力时, 有  $P_{in,t} = 0$ 。将  $P_{in,t} = 0$  和 (5-5) 带入 (5-6), 有

$$\rho_{out,t} = \rho_{BES,t} \frac{-P_{BES,t}}{-P_{BES,t} + P_{PV,t}} \quad (5-7)$$

其中  $\rho_{BES,t}$  是 BES 输出功率的碳强度。在 (5-7) 中,  $P_{PV,t}$  是不可控变量,  $P_{BES,t}$  是可控变量。为了提高  $\rho_{out,t}$  的灵活性, 本文提出的虚拟碳存储可以成功地将  $\rho_{BES,t}$  从不可控变量转换为可控变量。

由于与电力相关的 CEL 在电力消费时被支付, 当电力存储在 BES 中时, 相关 CEL 也临时放置在 BES。通过引入“虚拟碳存储”, 在不违反 PST 的情况下, 松弛了 PST 引起的不必要的碳排放实时平衡约束, 从而可以更灵活地调整用户的电力需求以跟踪外部市场的变化, 并且可以满足约束 (5-4)。

由于 BES 在每个时隙中充电或放电的电的碳强度不同, 因此存储在 BES 中的电的碳强度也在变化。因此, 有

$$\rho_{BES,t+1}^{stored} = \frac{C_{BES}SOC_{BES,t}^{stored} \rho_{BES,t}^{stored} + P_{BES,t}\rho_{BES,t}\Delta t}{(P_{BES,t}^{stored} + P_{BES,t})\Delta t} \quad (5-8)$$

其中  $C_{BES}$  是 BES 的容量。在时隙  $t$  中,  $SOC_{BES,t}^{stored}$  是 BES 的荷电状态 (SOC),  $\rho_{BES,t}^{stored}$  是相对于的碳强度。为了将 BES 中的碳强度保持在合理范围内, 需要满足以下约束条件  $0 \leq \rho_{BES,t}^{stored} \leq \rho_{BES,max}^{stored}$ 。其中,  $\rho_{BES,max}^{stored}$  是  $\rho_{BES,t}^{stored}$  的最大值。

## 5.2.2 用户模型

在本文中, 可控设备主要包括 BES 和柔性负载。BES 的 SOC 动态变化为:

$$SOC_t = SOC_{t-1} + \eta_c P_{BES,t}^c \Delta t + \frac{1}{\eta_d} P_{BES,t}^d \Delta t \quad (5-9)$$

需要满足约束：

$$SOC_{min} \leq SOC_t \leq SOC_{max} \quad (5-10)$$

$$0 \leq P_{BES,t}^c \leq P_{BES,max} \quad (5-11)$$

$$P_{BES,min} \leq P_{BES,t}^d \leq 0 \quad (5-12)$$

其中 $\eta_c$ 和 $\eta_d$ 分别是 BES 的充放电效率。 $P_{BES,t}^c$ 和 $P_{BES,t}^d$ 分别是 BES 的充放电功率，为控制变量。 $P_{BES,max}$ 和 $P_{BES,min}$ 分别是最大充电功率和放电功率。 $SOC_{max}$ 和 $SOC_{min}$ 分别是最大和最小 SOC。

BES 的损失 $C_{BES,t}$ 与其放电深度（depth of discharge, DOD）呈正相关，BES 的成本可以表示为<sup>[142]</sup>：

$$C_{BES,t} = \tau |SOC_t - SOC_{t-1}| \quad (5-13)$$

在一定范围内，BES 的成本与 SOC 的变化近似成线性关系， $\tau$ 是成本系数。

在高峰期间可以减少柔性负载，但这将导致用户舒适度的一定损失。令 $C_{L,t} = B(\Delta P_{L,t})$ 表示舒适度损失<sup>[4]</sup>，其中  $B$  是单调递增函数， $\Delta P_{L,t}$ 是负载减少量。由于用户的不同， $B$  可能有不同的形式。用户负载的灵活性有限。为了保证用户的基本负荷需求，负荷减少量需要满足 $0 \leq \Delta P_{L,t} \leq \Delta P_{t,max}$ ，其中 $\Delta P_{t,max}$ 是在时隙  $t$  中可以减少的最大负载功率。

此外，需要满足功率平衡约束 $P_{PV,t} + P_{in,t} + P_{BES,t} = P_{out,t} + P_{L,t} - \Delta P_{L,t}$ ， $P_{PV,t}$ 是 PV 在时隙  $t$  中的功率。用户购买和出售的电力功率不应大于联络线的最大功率，即 $0 \leq P_{in,t} \leq P_{max}$ ， $0 \leq P_{out,t} \leq P_{max}$ 。

### 5.2.3 目标函数

由于无法事先知道市场出清结果，因此在计算投标成本时只能使用预测数据。目前，有许多文献关注市场出清结果的预测，取得了令人满意的预测效果。用户从市场购买电力所产生的成本，用 $C_{in,t}^E$ 表示，根据第 1 阶段的出清价格结算：

$$C_{in,t}^E = (\hat{\lambda}_t + \varepsilon_{\lambda,t})(\hat{P}_{in,t} + \varepsilon_{P,in}) \quad (5-14)$$

其中， $\hat{\lambda}_t$ 是时隙  $t$  中的预测出清价格， $\hat{P}_{in,t}$ 是相应的出清功率。在 (5-14) 中，

$\varepsilon_{\lambda,t}$ 和 $\varepsilon_{P,in}$ 分别是价格和功率的预测值和真实值之间的误差。用户的剩余电力也可以在市场上出售以获取利润。将 $U_{out,t}^E$ 表示为电力销售收入， $\hat{P}_{out,t}$ 表示为向市场出售的预测出清电量。从而

$$U_{out,t}^E = (\hat{\lambda}_t + \varepsilon_{\lambda,t})(\hat{P}_{out,t} + \varepsilon_{P,in}) \quad (5-15)$$

在每个时隙中，与消耗功率相对应的碳排放应由用户承担。碳排放引起的成本，表示为 $C_{in,t}^C$ ，可根据年度碳市场的碳价格确定，即，

$$C_{in,t}^C = \lambda_c(\hat{\rho}_{L,t} + \varepsilon_\rho)P_{L,t} \quad (5-16)$$

其中 $\lambda_c$ 是年度碳市场中的碳价格， $\hat{\rho}_{L,t}$ 是市场中出清的预测碳强度， $\varepsilon_\rho$ 是实际值和预测值之间的对应误差。将 $C_{MG,t}$ 表示为用户在时隙 $t$ 内的总运营成本，从而：

$$C_{MG,t} = C_{BES,t} + C_{L,t} + C_{in,t}^E + C_{in,t}^C - U_{out,t}^E \quad (5-17)$$

为了最大限度地降低用户全天的成本，用户的优化目标是

$$\min \sum_{t=0}^T C_{MG,t} \quad (5-18)$$

本节阐述了用户在电碳耦合市场中的投标以及 BES 和可调负载的优化目标，通过提出的“虚拟储能”概念，增强了用户的灵活性，通过下一节详细研究的优化方法，可以进一步降低运营成本。

### 5.3 优化问题的求解

由于所考虑情景的复杂性，在实践中很难获得精确的数学模型和相应的系统参数。本文采用无模型 DRL 方法来获得优化策略。本文将所研究的用户环境视为马尔可夫决策过程 (MDP)，其中考虑了光伏输出、可控负荷、投标和报价。在本节中，我们首先将决策过程表述为 MDP，然后应用 DRL 算法来解决优化问题。

#### 5.3.1 基于 MDP 的强化学习

典型的 MDP 可以描述为 4 元组 $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$ ，其中 $\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}$ 分别是状态空间，动作空间，状态迁移和奖励函数。在时隙 $t$ 中，代理从状态空间 $\mathcal{S}$ 中观察状态 $s_t$ ，并从动作空间 $\mathcal{A}$ 中选择动作 $a_t$ 。在动作 $a_t$ 被执行后，状态 $s_t$ 将以概率 $\mathcal{P}(s_t, a_t)$ 转移至状态 $s_{t+1}$ 。此外，代理接收标量奖励 $r_t = \mathcal{R}(s_t, a_t)$ 。在本文中，代理从环境中观察

状态，并在电力碳耦合市场中对 BES 控制、负荷调整和投标策略做出决策。MDP 模型的细节如下所示。

1) 状态：状态  $s_t$  包含时隙  $t$  中所考虑用户的完整信息，其表示如下：

$$s_t = [P_t^{PV}, P_{L,t}, p_t, \alpha_t, \beta_{in,t}, \beta_{out,t}, SOC_t, \rho_{BES,t}^{stored}] \quad (5-19)$$

其中， $\alpha_t$  表示投标和报价中电力的基本碳强度， $\beta_{in,t}$  表示购电中碳强度的变化系数， $\beta_{out,t}$  代表售电中碳强度的变化系数。这三个变量对应于图 5-5 中的投标曲线。当用户是市场中的买方时， $\alpha_t$  表示其愿意购买电力的最大碳强度， $\beta_{in,t} \leq 0$  确定投标曲线的斜率。当用户是市场中的卖方时， $\alpha_t$  表示其愿意出售电力的最小碳强度， $\beta_{out,t} \geq 0$  确定提供曲线的斜率。

2) 动作：对于每个用户，可控设备包括 BES 和负载设备。假设用户在中购买电力，连续动作空间  $\mathcal{A}$  中的动作定义为

$$a_t = [P_{in,t}, \Delta P_{L,t}, P_{BES,t}, \rho_{BES,t}] \quad (5-20)$$

其中， $P_{in,t}$  是所需的购买电力， $\Delta P_{L,t}$  和  $P_{BES,t}$  分别是需求响应中的功率降低和时隙  $t$  中的 BES 功率， $\rho_{BES,t}$  是与 BES 功率相关的碳强度。为了稳定训练策略， $a_t$  和  $s_t$  通过除以其可能的最大值进行预处理，使得它们被限制在范围  $[0,1]$ 。

3) 状态迁移：MDP 将时隙  $t$  中动作  $a_t$  的状态  $s_t$  映射到时隙  $t+1$  中用户的状态  $s_t$ 。通常，在状态转换过程中，光伏输出、电力需求等的值无法提前知道，因此这些参数的预测值用于计算下一个状态。根据 [143][144]，预测结果是可信的。

4) 奖励：用户控制方案的目标是最小化总体运营成本。由于系统中存在多源不确定性，在同一状态下采取行动的益处可能不同。这将干扰算法的训练。因此，创新性地使用蒙特卡罗抽样来计算奖励，以提高奖励的稳定性。时隙  $t$  中的奖励，表示为  $r_t$ ，由  $\omega$  次蒙特卡洛样本确定：

$$r_t = \left( \sum_{k=1}^{\omega} C_{MG,t,k} + g_k(s_{t+1}) + P_{ub,k} \right) / \omega \quad (5-21)$$

其中

$$g(s_{t+1}) = \begin{cases} 0, & \text{if constraints are satisfied,} \\ \sigma, & \text{otherwise,} \end{cases} \quad (5-22)$$

$$P_{ub,t} = P_{PV,t} + P_{in,t} + P_{BES,t} - P_{out,t} - P_{L,t} + \Delta P_{L,t} \quad (5-23)$$

其中 $\sigma$ 表示惩罚参数， $P_{ub,t}$ 表示用户中的功率不平衡。当 $s_t$ 根据 $a_t$ 被更新为 $s_{t+1}$ ，并且状态量超过设定值时，返回一个大的惩罚项以使代理知道当前策略不合适。总之，我们将联合电力碳交易的优化问题制定为 MDP，从而相应的强化学习（reinforcement learning, RL）设置如下。

RL 本质上是一个在一系列离散时隙中的代理和环境之间的交互过程，旨在找到最佳策略 $\pi^*$ ，以最大化预期累积回报<sup>[145]</sup>。在每个时隙 $t$ 中，代理获得用户的当前状态 $s_t$ ，然后根据其策略 $\pi$ 从动作空间 $\mathcal{A}$ 中选择动作 $a_t$ 。在获得 $r_t$ 和 $s_{t+1}$ 之前，建立蒙特卡罗采样，以减少状态中不确定因素的影响。在 $\omega$ 次蒙特卡罗抽样之后，将 $\omega$ 次的 $r_t$ 的平均值输入网络。

将时间段 $t$ 内的折扣累积奖励定义为

$$R_t = \sum_{\tau=t}^{T-1} \gamma^{\tau-t} r_{\tau} \quad (5-24)$$

其中 $\gamma \in (0,1]$ 是折扣因子，通过调整 $\gamma$ 可以获得不同的近似估计值。RL 的目标是训练一个最优策略 $\pi^*$ 来最大化 $R_t$ 。

### 5.3.2 近端策略优化

根据上述分析，连续行动空间和多源不确定性使得难以获得期望的优化策略。为了解决这个问题，深度神经网络由于其通用拟合能力而被用于 RL。通常，用于连续动作空间的几种 DRL 算法已被应用于解决优化问题，如异步优势演员-评论家算法（A3C）、深度确定性策略梯度算法（DDPG）、PPO 等。考虑到 PPO 可以实时评估策略，本文采用 PPO 作为基准算法。

PPO 是一种演员-评论家算法，它有演员和评论家两个网络。演员网络用于学习控制策略。不同于（5-24），优势 $A_t$ 作为强化信号来衡量用户优化策略的质量。为了减少估计值的方差，使用广义优势估计（GAE）方法来估计优势 $A_t$ <sup>[146]</sup>，其表示如下：

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \dots + (\gamma\lambda)^{T-t+1}\delta_{T-1} \quad (5-25)$$

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \quad (5-26)$$

其中 $\lambda$ 是 $[0,1]$ 中的超参数。 $V(\cdot)$ 是学习的状态值函数，用于计算批评家网络学习的当前值。 $\delta_t$ 为时间差分误差，用于度量 $V(s_t)$ 的表现。

PPO 算法将更新参数的值限制在某个信任区域。代理的目标函数如下<sup>[147]</sup>:

$$J_{PPO}(\theta) = \mathbb{E}_t \min[r_{im}(\theta)\hat{A}_t, \text{clip}(r_{im}(\theta), 1 - \varepsilon, 1 + \varepsilon)\hat{A}_t] \quad (5-27)$$

其中， $\mathbb{E}_t$ 表示采样和优化计算中样本的经验平均值;在（5-27）中， $r_{im} = \pi_{\theta}(a(t)|s(t))/\pi_{\theta_{old}}(a(t)|s(t))$ 表示用于测量概率分布与目标分布之差的重要性抽样比； $\pi_{\theta}(\cdot)$ 和 $\pi_{\theta_{old}}(\cdot)$ 表示使用参数 $\theta$ 前后的策略， $\hat{A}_t$ 表示优势函数的估计值。削波函数表示目标的削波概率，其去除了范围 $(1 - \varepsilon, 1 + \varepsilon)$ 之外的 $r_{im}$ 值； $\varepsilon$ 是一个超参数。通过使用（5-27）中的削波函数，采样模型和实际模型之间的差距将缩小。图 5-6 展示了联合电力碳竞价和报价机制下用于优化的 PPO 结构。

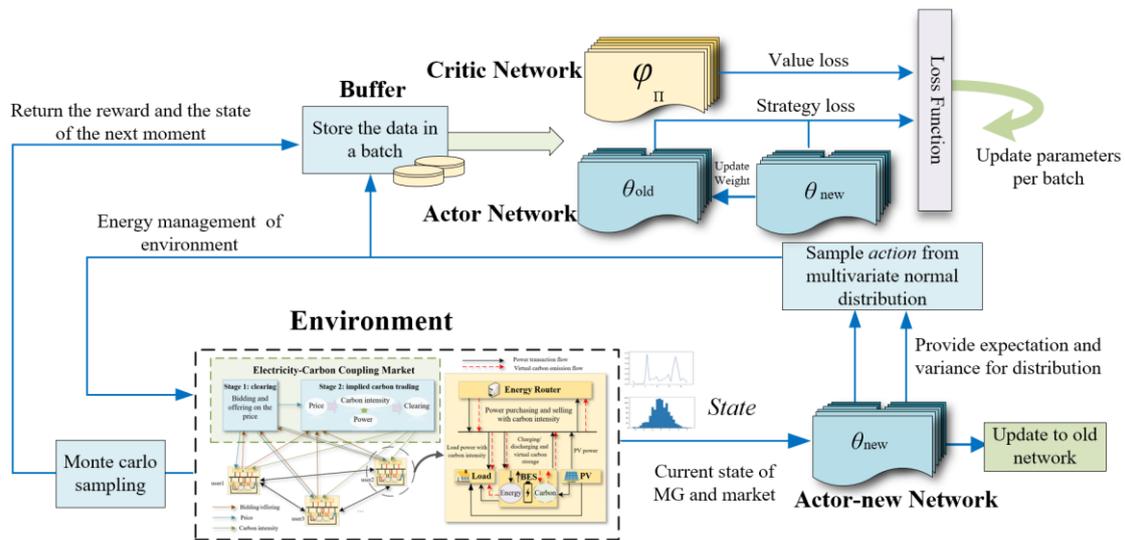


图 5-6 PPO 的网络结构

将（5-19）中的八个元素作为网络的输入，并将采样策略作为输出。而后，缓冲区记录每个批次中的信息，并将其导出到演员和评论家网络，以更新深度神经网络。PPO 训练过程的伪代码如算法 1 所示。

算法 1 控制策略的训练过程

为演员-评论家网络初始化参数 $\theta_0$ 和 $\varphi_0$

for  $k \in \{1, 2 \dots, N\}$  do

    在一天内运行  $T$  个时间步长的策略 $\pi_{\theta}$ ,

    并在用户环境中收集轨迹集 $\mathcal{D}_k = \{s_t, a_t\}$

    计算奖励 $\hat{R}_t$

    根据（5-25）计算优势估计值 $\hat{A}_t$

    根据（5-27）使用梯度上升更新策略参数 $\theta$

    通过均方误差回归更新策略 $\varphi$ :

$$L(\varphi) = \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T (V(s_t) - \hat{R}_t)^2$$

end for

## 5.4 仿真分析

在本节中，首先介绍仿真参数。然后，通过 3 组案例验证了所提机制的优越性。用于算法训练的计算机配置是：i7-11800H CPU，NVIDIA GeForce RTX 3070 GPU 和 8GB 内存。

### 5.4.1 仿真设置

由于 PPO 是一种无模型算法，因此优化不需要显式模型。为了便于说明，损失函数在仿真实验中采用如下形式：

$$C_{L,t} = \alpha_L (\Delta P_{L,t})^2 + \beta_L \Delta P_{L,t} \quad (5-28)$$

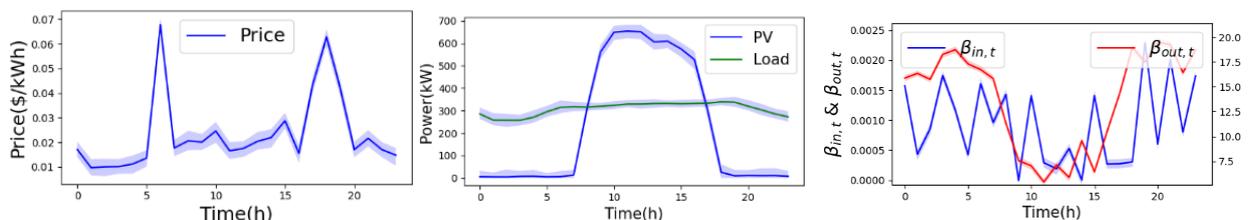
其中  $\alpha_L$  和  $\beta_L$  是舒适性损失函数的系数。上述函数只是特殊情况之一，其他形式的函数也可以通过我们提出的算法进行求解。

负荷设备、BES 和 PV 的相关参数如表 5-1 所示。

表 5-1 系统参数

参数	价值	参数	价值
$SOC_{min}$	0.2	$C_{BES}$ (千瓦时)	2000
$SOC_{max}$	0.8	$P_{max}$ (千瓦)	500
$\lambda_c$ (\$/kg)	0.02	$P_{t,max}^{PV}$ (千瓦)	1500
$\rho_{BES,max}^{stored}$	2	$\Delta P_{t,max}$ (千瓦)	500
$\tau$	0.01	$P_{L,max}$ (千瓦)	1500
$\eta_c$	0.95	$P_{BES,max}$ (千瓦)	1000
$\eta_d$	0.95	$P_{BES,min}$ (千瓦)	-1000
$\alpha_L$	0.001	$p_{max}$ (\$/千瓦时)	0.08
$\beta_L$	0.003	$\sigma$	100

日前电价、光伏输出功率和负载需求取自 PJM 电力市场<sup>[148]</sup>，其中光伏输出功率和负载需求按比例缩小。在仿真中，使用 170 天的数据，其中训练集，验证集和测试集分别占 80%，10% 和 10%。随机选择测试集中的一天用于显示仿真结果，各项参数如图 5-7 所示。



(a) 日前电价

(b) 光伏和负荷

(c) 竞价曲线参数

图 5-7 所选日期的预测数据。

由于预测值和实际值之间的不确定性，预测误差采用正态分布模拟。每个时隙的期望值如图 5-7 中实线上的值表示，负荷和 PV 输出的方差分别为 10 和 20。图 5-7 中的阴影区域表示具有 90%置信率的 500 个蒙特卡洛样本中样本分布的区间。此外，BES 中一天的初始 SOC 被设置为 0.2 至 0.8 的随机值。

演员网络和评论家网络都包含两个具有 64 个节点的隐藏层。演员网络的第一层是具有 64 个节点的隐藏层，随后是具有 128 个节点的隐层。PPO 算法中的折扣因子 $\gamma$ 、超参数 $\epsilon$ 和 $\lambda$ 、更新周期 $k$ 、行动者网络和评论家网络的学习率分别设置为 0.99、0.2、0.95、50、0.003 和 0.001。

设置了三个比较集，分别在经济成本、碳减排效果和收敛性能方面验证了算法的有效性，如下所示：

**基线 1 和基线 2：**这组案例旨在说明所提出算法在经济成本方面的优势。由于经济成本降低主要由 BES 和需求响应贡献，基线 1 设置为无需求响应，而基线 2 设置为无 BES。

**基线 3 和基线 4：**这组案例旨在说明所提的电力碳耦合市场和虚拟碳存储在碳减排方面的优势。在基线 3 中，用户参与电力碳耦合市场，而虚拟碳存储未被考虑，而在基线 4 中，用户仅参与传统电力市场。

**基线 5 和基线 6：**这组案例用于说明算法的收敛性能。蒙特卡洛抽样的数量影响奖励的稳定性，进而又影响收敛的稳定性。在基线 6 中，蒙特卡洛抽样的数量减少到 25，而在基线 5 中，不应用蒙特卡洛抽样。

#### 5.4.2 成本分析

通过需求响应和 BES，一个用户可以根据市场电价的变化改变不同时段的购电量，以降低经济成本。在不同时隙从市场购买的电力如图 5-8 所示：

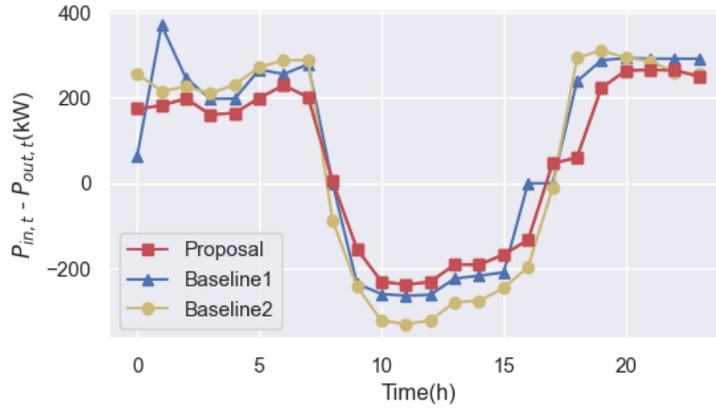


图 5-8 用户与外部市场之间的电力交易

正值表示购买电力，而负值表示出售电力。从图 5-8 中可以看出，所提出的策略购买和出售的最大电量减少，这主要是由于调整能量存储以提高光伏输出的本地消耗。所提算法全天的总成本分别比基线 1 和基线 2 低 8.7% 和 8.0%，表明需求响应和 BES 利用率对成本降低有显著影响。

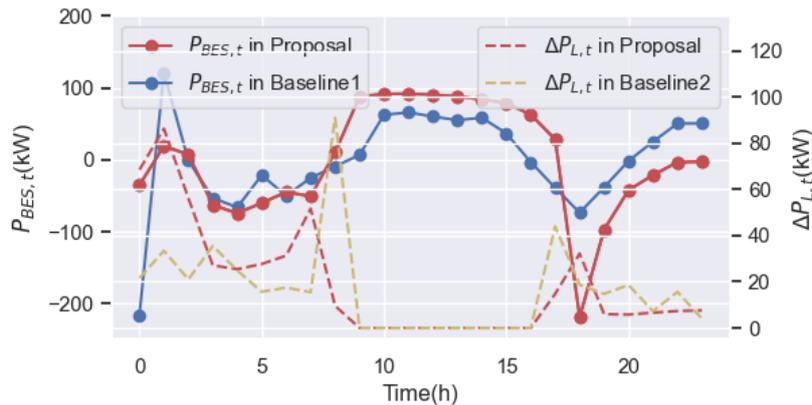


图 5-9 BES 的充电/放电和负荷削减

BES 的充放电功率和需求响应的实现如图 5-9 所示。市场电价和光伏输出功率都会影响 BES 的充电/放电功率以及需求响应的深度。当本地可再生能源过剩时，BES 将被充电以消耗光伏出力，而需求响应不起作用。相反，BES 放电，并执行需求响应以降低购电成本。当电价相对较高时，所提出的算法还可以通过在 BES 调节的基础上深化需求响应，进一步减少购买电量。

由于考虑了碳排放成本，BES 的行动和需求响应不仅受到电价的影响，还受到不同时间段与电力相关的碳排放强度的影响。因此，下一小节将进一步分析所提出算法的碳减排。

### 5.4.3 碳减排分析

基线 3、基线 4 及所提出的机理和策略的碳排放如图 5-10 所示。

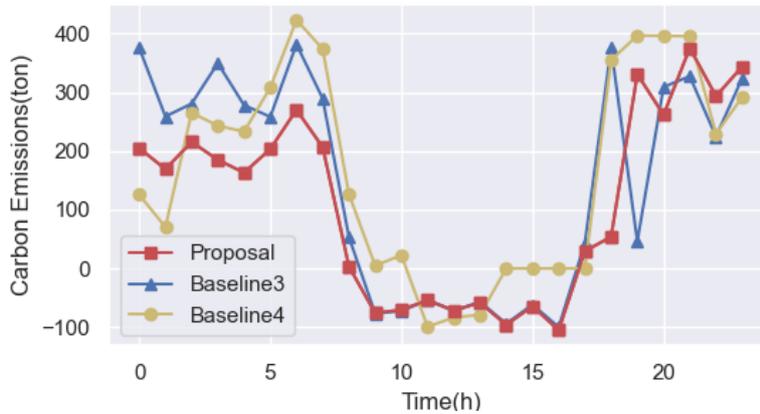
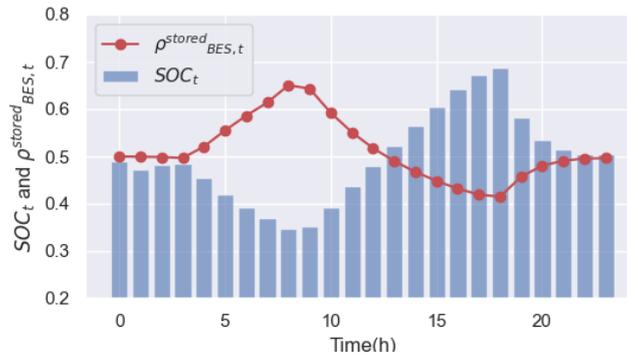
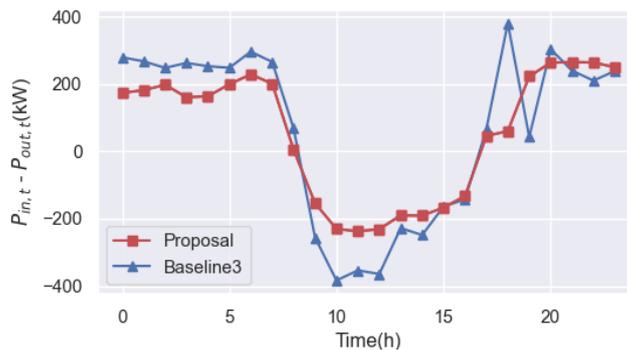


图 5-10 所提策略和对比情况下的碳排放

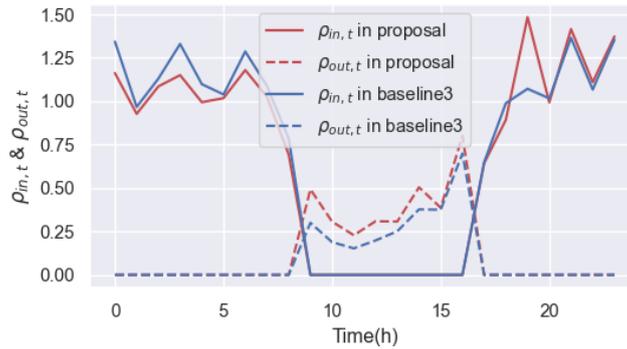
图 5-10 所示的碳排放包括净流入和净流出，当碳排放为净流入时，其值为正值，否则为负值。可以看出，早晚的碳排放量相对较大，而中午的碳排放因光伏输出量较高而明显较低。所提策略的平均每日碳排放量分别比基线 3 和基线 4 低 27.0% 和 31.6%，表明所提机制和优化策略能够有效减少用户的碳排放。虚拟碳存储可以在电力碳耦合市场的基础上进一步减少碳排放。接下来，进一步分析虚拟碳存储对碳减排的影响，结果如图 5-11 所示。



(a) 所提策略中的 SOC 和平均储存碳强度



## (二) 用户与外部市场之间的电力交易



(c) 购售电中的碳强度

图 5-11 虚拟碳储存相关的仿真结果

由于基线 4 不参与碳排放市场，因此将所提的机制和策略与基线 3 进行比较，以证明所提的虚拟碳储存对碳减排的有效性。在图 11 (a) 中，由于 CEL 可以通过虚拟碳存储临时存储，用户可以以低碳强度消耗 BES 中存储的电，但 BES 中剩余电的碳强度显著增加。由于光伏输出的碳强度为 0，BES 在中午充电，BES 中的电的碳强度逐渐降低。从图 11 (b) 可以看出，在提出的策略中，中午出售的电力减少，更多的光伏输出存储在 BES 中，因为 BES 中碳强度的变化提高了用户通过存储低碳电力可以获得的效益。从图 11 (c) 中碳强度的交易结果可以看出，虚拟碳存储也有助于用户在特定时间段购买低碳电力并出售更多的 CEL。

## 5.4.4 不确定性分析

图 5-12 显示了不同蒙特卡洛采样时间的算法收敛性。训练期间的蒙特卡洛采样用于处理多重不确定性的影响。采样时间 $\omega=0$ 、25 和 50 分别作为基线 5、6 和所提的策略。在仿真实验中，我们提的策略采用 $\omega=50$ 。经过  $6 \times 10^6$  次训练，获得了算法的优化结果，耗时 5 小时 15 分钟。经过  $5.5 \times 10^5$  步的训练，算法就已经收敛了。

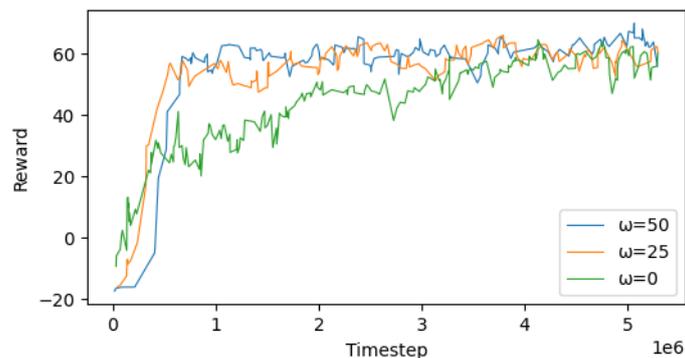


图 5-12 算法训练过程中的奖励变化

基线 5 的结果表明，蒙特卡罗抽样可以提高优化算法的收敛速度。然而，25 和 50 个样本的收敛速度相似，表明当蒙特卡洛样本数量达到一定程度时，算法的收敛性能很难通过增加蒙特卡洛样本的数量来提高。虽然蒙特卡洛采样可以加快算法收敛速度，但它确实消耗了更多的计算能力和时间。在实践中，有必要根据实际应用中的不确定性选择适当的采样数。

## 5.5 本章小结

本文提出了面向需求侧的电-碳联合交易机制，以实现不同终端用户之间碳排放责任的有效分配。引入了“虚拟碳存储”的概念，以松弛对 PST 造成的碳排放施加的不必要的实时平衡约束。提出了一种基于蒙特卡罗的奖励方法，以加快优化算法的收敛速度。结果表明，通过提出的机制和算法，由于 BES 和需求响应的贡献，用户的总经济成本可减少 8.7%，而碳排放可减少 31.6%，这归因于所提出的碳-电耦合市场和虚拟碳存储。此外，基于蒙特卡罗抽样，算法的收敛速度显著提高。

在本文中，仅考虑了电力消耗和碳排放之间的耦合关系，而其他形式的能源，如冷、热和气体也会导致碳排放。在未来的工作中，需要进一步研究考虑不同能源载体碳排放的需求方综合能源交易机制。

## 6 总结与展望

报告面向电力需求侧分布式灵活资源的聚合与优化调节方法开展研究，首先，针对分布式用户差异化的特点提出需求响应增量激励机制，更加充分地释放了用户的可调节潜力；其次，针对分布式储能归属于多主体的现状，提出了两种储能共享模式，充分利用储能的闲置容量，提升用户效益的同时更多地消纳可再生能源出力；再次，针对分布式数据中心，提出了信息-能量融合的调节模型，在保障计算任务被满足的前提下，利用数据中心计算任务卸载的能力，实现电力负荷在地理上的转移，以增加可再生能源的消纳能力；最后，针对需求侧调节的碳减排问题，提出了电-碳耦合市场架构，实现了碳排放责任在需求侧的灵活交易，进而通过负荷的调节达到碳减排的目的。

本文主要针对电力需求侧的一些典型负荷展开研究，在未来的工作中，一方面需要拓展更多的灵活性负荷，如空调、电动汽车等等，另一方面，需要研究各种类型负荷的聚合方法，以形成更大的调节能力。此外，负荷的聚合调节还可以拓展至总能能源领域，利用多能互补提升综合能源系统的运行效率。

## 参考文献

- [1] M. R. Sarker, M. A. Ortega-Vazquez and D. S. Kirschen, “Optimal Coordination and Scheduling of Demand Response via Monetary Incentives,” *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1341-1352, Dec. 2015.
- [2] Yang, J.; Zhao, J.; Wen, F.; Dong, Z.Y. A framework of customizing electricity retail prices. *IEEE Trans. Power Syst.* 2018, 33, 2415–2428.
- [3] Jacquot, P.; Beaudé, O.; Gaubert, S.; Oudjane, N. Analysis and implementation of an hourly billing mechanism for demand response management. *IEEE Trans. Smart Grid* 2019, 10, 4265–4278.
- [4] Li, P.; Wang, H.; Zhang, B. A distributed online pricing strategy for demand response programs. *IEEE Trans. Smart Grid* 2017, 10, 350–360.
- [5] Kim, S.; Giannakis, G.B. An online convex optimization approach to real-time energy pricing for demand response. *IEEE Trans. Smart Grid* 2017, 8, 2784–2793.
- [6] Rana, M.; Rahi, K.; Ray, T.; Sarker, R. An efficient optimization approach for flexibility provisioning in community microgrids with an incentive-based demand response scheme. *Sustain. Cities Soc.* 2021, 74, 103218.
- [7] Ellman, D.; Xiao, Y. Incentives to manipulate demand response baselines with uncertain event schedules. *IEEE Trans. Smart Grid* 2021, 12, 1358–1369.
- [8] Chai, Y.; Xiang, Y.; Liu, J.; Gu, C.; Zhang, W.; Xu, W. Incentive-based demand response model for maximizing revenue of electricity retailers. *J. Mod. Power Syst. Clean Energy* 2019, 7, 1644–1650.
- [9] Wei, C.; Xu, J.; Liao, S.; Sun, Y. Aggregation and scheduling models for electric vehicles in distribution networks considering power fluctuations and load rebound. *IEEE Trans. Sustain. Energy* 2020, 11, 2755–2764.
- [10] Muthirayan, D.; Kalathil, D.; Poolla, K.; Varaiya, P. Mechanism design for demand response programs. *IEEE Trans. Smart Grid* 2020, 11, 61–73.
- [11] Li, G.; Huang, Y.; Bie, Z. Reliability evaluation of smart distribution systems considering load rebound characteristics. *IEEE Trans. Sustain. Energy* 2018, 9, 1713–1721.
- [12] Georges, E.; Cornélusse, B.; Ernst, D.; Lemort, V.; Mathieu, S. Residential heat pump as flexible load for direct control service with parametrized duration and rebound effect. *Appl. Energy* 2017, 187, 140–153.
- [13] Wei, C.; Wu, Q.; Xu, J.; Sun, Y.; Jin, X.; Liao, S.; Yuan, Z.; Yu, L. Distributed scheduling of smart buildings to smooth power fluctuations considering load rebound. *Appl. Energy* 2020, 276, 115396.
- [14] L. Wen, K. Zhou, J. Li, S. Wang, “Modified deep learning and reinforcement learning for an incentive-based demand response model,” *Energy*, vol. 205, pp. 118019, Jun. 2020.
- [15] M. Rana, K. Rahi, T. Ray, R. Sarker, “An efficient optimization approach for flexibility provisioning in community microgrids with an incentive-based demand response scheme,” *Sustainable Cities and Society*, vol.74, pp. 103218, Aug. 2021.

- [16] D. Muthirayan, D. Kalathil, K. Poolla and P. Varaiya, "Mechanism design for demand response programs," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 61-73, Jan. 2020.
- [17] Y. Chai, Y. Xiang, J. Liu, C. Gu, W. Zhang and W. Xu, "Incentive-based demand response model for maximizing benefit of electricity retailers," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 6, pp. 1644-1650, Nov. 2019.
- [18] H. Aalami, H. Pashaei-Didani, S. Nojavan, "Deriving nonlinear models for incentive-based demand response programs," *Int. J. Electr. Power Energy Syst.*, vol. 106, pp. 223-231, Oct. 2019.
- [19] J. Lin, J. Dong, X. Dou, Y. Liu, P. Yang, T. Ma, "Psychological insights for incentive-based demand response incorporating battery energy storage systems: A two-loop Stackelberg game approach," *Energy*, vol. 239, pp. 122192, Sep. 2022.
- [20] M. Yu, S. H. Hong, Y. Ding and X. Ye, "An incentive-based demand response (DR) model considering composited DR resources," *IEEE Trans. Ind. Elec.*, vol. 66, no. 2, pp. 1488-1498, Feb. 2019.
- [21] Jindal, M. Singh and N. Kumar, "Consumption-aware data analytical demand response scheme for peak load reduction in smart grid," *IEEE Trans. Ind. Elec.*, vol. 65, no. 11, pp. 8993-9004, Nov. 2018.
- [22] H. Yang, J. Zhang, J. Qiu, S. Zhang, M. Lai and Z. Y. Dong, "A practical pricing approach to smart grid demand response based on load classification," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 179-190, Jan. 2018.
- [23] Y. Astriani, G. Shafiullah, F. Shahnia, "Incentive determination of a demand response program for microgrids," *Applied Energy*, vol. 292, pp. 116624, Mar. 2021.
- [24] Q. Hu, F. Li, X. Fang and L. Bai, "A framework of residential demand aggregation with financial incentives," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 497-505, Jan. 2018.
- [25] M. Vanouni and N. Lu, "A reward allocation mechanism for thermostatically controlled loads participating in intra-hour ancillary services," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4209-4219, Sept. 2018.
- [26] H. Zhong, L. Xie and Q. Xia, "Coupon incentive-based demand response: theory and case study," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1266-1276, May 2013.
- [27] Yan, X.; Gu, C.; Wyman-Pain, H.; Li, F. Capacity share optimization for multiservice energy storage management under portfolio theory. *IEEE Trans. Ind. Electron.* 2019, 66, 1598–1607.
- [28] Mediwaththe, C.P.; Shaw, M.; Halgamuge, S.; Smith, D.B.; Scott, P. An incentive-compatible energy trading framework for neighborhood area networks with shared energy storage. *IEEE Trans. Sustain. Energy* 2020, 11, 467–476.
- [29] Szabó, D.Z.; Duck, P.; Johnson, P. Optimal trading of imbalance options for power systems using an energy storage device. *Eur. J. Oper. Res.* 2020, 285, 3–22.

- [30] Rahbar, K.; Xu, J.; Zhang, R. Real-Time Energy Storage Management for Renewable Integration in Microgrid: An Off-Line Optimization Approach. *IEEE Trans. Smart Grid* 2015, 6, 124–134.
- [31] Ju, C.; Wang, P.; Goel, L.; Xu, Y. A two-layer energy management system for microgrids with hybrid energy storage considering degradation costs. *IEEE Trans. Smart Grid* 2018, 9, 6047–6057.
- [32] Wang, B.; Zhang, C.; Dong, Z.Y. Interval optimization based coordination of demand response and battery energy storage system considering SOC management in a microgrid. *IEEE Trans. Sustain. Energy* 2020, 11, 2922–2931.
- [33] Du, Y.; Wu, J.; Li, S.; Long, C.; Onori, S. Hierarchical coordination of two-time scale microgrids with supply-demand imbalance. *IEEE Trans. Smart Grid* 2020, 11, 3726–3736.
- [34] Walker, A.; Kwon, S. Analysis on impact of shared energy storage in residential community: Individual versus shared energy storage. *Appl. Energy* 2021, 282, 116172.
- [35] Zhu, H.; Ouahada, K. A distributed real-time control algorithm for energy storage sharing. *Energy Build.* 2021, 230, 110478.
- [36] Lai, S.; Qiu, J.; Tao, Y. Credit-based pricing and planning strategies for hydrogen and electricity energy storage sharing. *IEEE Trans. Sustain. Energy* 2022, 13, 67–80.
- [37] Zhong, S.; Qiu, J.; Sun, L.; Liu, Y.; Zhang, C.; Wang, G. Coordinated planning of distributed WT, shared BESS and individual VESS using a two stage approach. *Int. J. Electr. Power Energy Syst.* 2020, 114, 105380.
- [38] Hafiz, F.; de Queiroz, A.R.; Fajri, P.; Husain, I. Energy management and optimal storage sizing for a shared community: A multi-stage stochastic programming approach. *Appl. Energy* 2019, 236, 42–54.
- [39] Haochen Hua, Zhiqian Wei, Yuchao Qin, Tonghe Wang, Liuying Li, and Junwei Cao, "A review of distributed control and optimization in energy Internet: From traditional methods to artificial intelligence-based methods," *IET Cyber-Physical Systems: Theory & Applications*, vol. 6, no. 2, pp. 63-79, 2021.
- [40] C. Ju, P. Wang, L. Goel and Y. Xu, "A Two-Layer Energy Management System for Microgrids With Hybrid Energy Storage Considering Degradation Costs," in *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6047-6057, Nov. 2018, doi: 10.1109/TSG.2017.2703126.
- [41] Y. Xu and X. Shen, "Optimal Control Based Energy Management of Multiple Energy Storage Systems in a Microgrid," in *IEEE Access*, vol. 6, pp. 32925-32934, 2018, doi: 10.1109/ACCESS.2018.2845408.
- [42] H. -C. Jo, G. Byeon, J. -Y. Kim and S. -K. Kim, "Optimal Scheduling for a Zero Net Energy Community Microgrid With Customer-Owned Energy Storage Systems," in *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2273-2280, May 2021, doi: 10.1109/TPWRS.2020.3036077.
- [43] J. S. Giraldo, J. A. Castrillon, J. C. López, M. J. Rider and C. A. Castro, "Microgrids Energy Management Using Robust Convex Programming," in *IEEE*

- Transactions on Smart Grid, vol. 10, no. 4, pp. 4520-4530, July 2019, doi: 10.1109/TSG.2018.2863049.
- [44] Zhaoming Qin, Di Liu, Haochen Hua, and Junwei Cao, "Privacy preserving load control of residential microgrid via deep reinforcement learning," IEEE Transactions on Smart Grid, vol. 12, no. 5, pp. 4079-4089, Sept. 2021.
- [45] S. Cui, Y. -W. Wang, X. -K. Liu, Z. Wang and J. -W. Xiao, "Economic Storage Sharing Framework: Asymmetric Bargaining-Based Energy Cooperation," in IEEE Transactions on Industrial Informatics, vol. 17, no. 11, pp. 7489-7500, Nov. 2021, doi: 10.1109/TII.2021.3053296.
- [46] J. Jo and J. Park, "Demand-Side Management With Shared Energy Storage System in Smart Grid," in IEEE Transactions on Smart Grid, vol. 11, no. 5, pp. 4466-4476, Sept. 2020, doi: 10.1109/TSG.2020.2980318.
- [47] S. Lai, J. Qiu and Y. Tao, "Credit-Based Pricing and Planning Strategies for Hydrogen and Electricity Energy Storage Sharing," in IEEE Transactions on Sustainable Energy, vol. 13, no. 1, pp. 67-80, Jan. 2022, doi: 10.1109/TSTE.2021.3103886.
- [48] M. Song and M. Amelin, "Price-Maker Bidding in Day-Ahead Electricity Market for a Retailer With Flexible Demands," in IEEE Transactions on Power Systems, vol. 33, no. 2, pp. 1948-1958, March 2018, doi: 10.1109/TPWRS.2017.2741000.
- [49] M. Kohansal and H. Mohsenian-Rad, "Price-Maker Economic Bidding in Two-Settlement Pool-Based Markets: The Case of Time-Shiftable Loads," in IEEE Transactions on Power Systems, vol. 31, no. 1, pp. 695-705, Jan. 2016, doi: 10.1109/TPWRS.2015.2405084.
- [50] S. Wang, S. Bi and Y. -J. A. Zhang, "Demand Response Management for Profit Maximizing Energy Loads in Real-Time Electricity Market," in IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6387-6396, Nov. 2018, doi: 10.1109/TPWRS.2018.2827401.
- [51] Y. Wang, D. Wang, H. Zhang and K. Zhang, "Optimal Bidding of Price-Maker Retailers With Demand Price Quota Curves Under Price Uncertainty," in IEEE Access, vol. 8, pp. 120746-120756, 2020, doi: 10.1109/ACCESS.2020.3005932.
- [52] S. Shafiee, H. Zareipour and A. M. Knight, "Developing Bidding and Offering Curves of a Price-Maker Energy Storage Facility Based on Robust Optimization," in IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 650-660, Jan. 2019, doi: 10.1109/TSG.2017.2749437.
- [53] 王继业, 周碧玉, 刘万涛, 等. 数据中心跨层能效优化研究进展和发展趋势[J]. 中国科学:信息科学, 2020, 50(1): 1-24.
- [54] 冯成, 王毅, 陈启鑫, 等. 能源互联网下的数据中心能量管理综述[J]. 电力自动化设备, 2020, 40(7): 1-9.
- [55] Yao F, Demers A, Shenker S. A scheduling model for reduced CPU energy[C]//Proc of the 36th Annual Symp on Foundations of Computer Science. Piscataway NJ: IEEE, 1995: 374-382.
- [56] Albers S. Energy-efficient algorithms[J]. Communications of the ACM, 2010, 53(5): 86-96.

- [57] 李文信, 齐恒, 徐仁海, 等. 数据中心网络流量调度的研究进展与趋势[J]. 计算机学报, 2020, 43(4): 600-617.
- [58] 吴刚, 高赐威, 陈宋宋, 等. 考虑需求响应的数据中心用电负荷优化研究综述[J]. 电网技术, 2018, 42(11): 3782-3788.
- [59] 高赐威, 曹晓峻, 闫华光, 等. 数据中心电能管理及参与需求侧资源调度的展望[J]. 电力系统自动化, 2017, 41(23): 1-7.
- [60] 宋杰, 孙宗哲, 刘慧, 等. 混合供电数据中心能耗优化研究进展[J]. 计算机学报, 2018, 41(12): 2670-2688.
- [61] YU L, JIANG T, ZOU Y. Real-time energy management for cloud data centers in smart microgrids[J]. IEEE Access, 2016, 4: 941-950.
- [62] YU L, JIANG T, ZOU Y. Distributed real-time energy management in data center microgrids[J]. IEEE Transactions on Smart Grid, 2018, 9(4): 3748-3762.
- [63] 曹晓峻, 高赐威, 李德智, 等. 数据网络与电力网络混合运行建模及其参与系统经济运行[J]. 中国电机工程学报, 2018, 38(5): 1448-1456.
- [64] J. Knudsen, J. Hansen, and A. M. Annaswamy, "A dynamic market mechanism for the integration of renewables and demand response," IEEE Trans. Cont. Syst. Tech., vol. 24, no. 3, pp. 940-955, May 2016.
- [65] M. S. H. Nizami, M. J. Hossain, and K. Mahmud, "A Nested Transactive Energy Market Model to Trade Demand-Side Flexibility of Residential Consumers," IEEE Trans. Smart Grid, vol. 12, no. 1, pp. 479-490, Jan. 2021.
- [66] R. Chandra, S. Banerjee, K. K. Radhakrishnan, and S. K. Panda, "Transactive energy market framework for decentralized coordination of demand side management within a cluster of buildings," IEEE Trans. Ind. Appl., vol. 57, no. 4, pp. 3385-3395, July-Aug. 2021.
- [67] M. Pourakbari-Kasmaei, M. Lehtonen, J. Contreras, and J. Mantovani, "Carbon footprint management: A pathway toward smart emission abatement," IEEE Trans. Ind. Inform., vol. 16, no. 2, pp. 935-948, Feb. 2020.
- [68] Y. Cheng, N. Zhang, B. Zhang, C. Kang, W. Xi, and M. Feng, "Low-carbon operation of multiple energy systems based on energy-carbon integrated prices," IEEE Trans. Smart Grid, vol. 11, no. 2, pp. 1307-1318, Mar. 2020.
- [69] M. Fang, Y. Xiang, and J. Li, "Locational electricity-carbon price model: Design and analysis," Energy Rep., vol. 8, no. 10, pp. 721-728, Jun. 2022.
- [70] C. Kang, T. Zhou, Q. Chen, Q. Xu, Q. Xia, and Z. Ji, "Carbon emissions flow in networks," Sci. Rep., vol. 2, pp. 479, Jun. 2012.
- [71] Y. Cheng, N. Zhang, Z. Lu, and C. Kang, "Planning multiple energy systems towards low-carbon society: a decentralized approach," IEEE Trans. Smart Grid, vol. 10, no. 5, pp. 4859-4869, Sep. 2019.
- [72] C. Kang et al., "Carbon emissions flow from generation to demand: a network-based model," IEEE Trans. Smart Grid, vol. 6, no. 5, pp. 2386-2394, Sep. 2015.
- [73] D. Qiu, D. Papadaskalopoulos, Y. Ye and G. Strbac, "Investigating the effects of demand flexibility on electricity retailers' business through a tri-level

- optimisation model,” *IET Gener. Transm. Distrib.*, vol. 14, no. 9, pp. 1739-1750, May, 2020.
- [74] R. Henríquez, G. Wenzel, D. E. Olivares and M. Negrete-Pincetic, “Participation of demand response aggregators in electricity markets: optimal portfolio management,” *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4861-4871, Sept. 2018.
- [75] E. Mahboubi-Moghaddam, M. Nayeripour, J. Aghaei, A. Khodaei and E. Waffenschmidt, “Interactive robust model for energy service providers integrating demand response programs in wholesale markets,” *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2681-2690, Jul. 2018.
- [76] S. R. Konda, A. S. Al-Sumaiti, L. K. Panwar, B. K. Panigrahi and R. Kumar, “Impact of load profile on dynamic interactions between energy markets: a case study of power exchange and demand response exchange,” *IEEE Trans. Ind. Inform.*, vol. 15, no. 11, pp. 5855-5866, Nov. 2019.
- [77] Wei, Z.; Quan, Z.; Wu, J.; Li, Y.; Pou, J.; Zhong, H. Deep deterministic policy gradient-drl enabled multiphysics-constrained fast charging of lithium-ion battery. *IEEE Trans. Ind. Electron.* 2022, 69, 2588–2598.
- [78] Nicola, M.; Nicola, C.-I.; Selişteanu, D. Improvement of PMSM sensorless control based on synergetic and sliding mode controllers using a reinforcement learning deep deterministic policy gradient agent. *Energies* 2022, 15, 2208.
- [79] Zhang, J.; Li, Y.; Wu, Z.; Rong, C.; Wang, T.; Zhang, Z.; Zhou, S. Deep-reinforcement-learning-based two-timescale voltage control for distribution systems. *Energies* 2021, 14, 3540.
- [80] H. Xu, H. Sun, D. Nikovski, S. Kitamura, K. Mori and H. Hashimoto, “Deep reinforcement learning for joint bidding and pricing of load serving entity,” *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6366-6375, Nov. 2019.
- [81] Hua, H.; Wei, Z.; Qin, Y.; Wang, T.; Li, L.; Cao, J. A review of distributed control and optimization in energy Internet: From traditional methods to artificial intelligence-based methods. *IET Cyber-Phys. Syst. Theory Appl.* 2021, 6, 63–79.
- [82] J. Shu, R. Guan, L. Wu and B. Han, “A bi-level approach for determining optimal dynamic retail electricity pricing of large industrial customers,” *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2267-2277, Mar. 2019.
- [83] Z. Wang, H. Li, N. Deng, et al., “How to effectively implement an incentive-based residential electricity demand response policy? Experience from large-scale trials and matching questionnaires,” *Energy Policy*, vol. 141, pp. 111450, Apr. 2020.
- [84] I. Antonopoulos, V. Robu, B. Couraud, D. Flynn, “Data-driven modelling of energy demand response behaviour based on a large-scale residential trial,” *Energy and AI*, vol. 4, pp. 100071, Apr. 2021.
- [85] P. R. Thimmapuram and J. Kim, “Consumers’ Price Elasticity of Demand Modeling with Economic Effects on Electricity Markets Using an Agent-Based Model,” *IEEE Trans. on Smart Grid*, vol. 4, no. 1, pp. 390-397, March 2013.

- [86] M. Muratori and G. Rizzoni, "Residential demand response: dynamic energy management and time-varying electricity pricing," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1108-1117, Mar. 2016.
- [87] N. Mazzi, J. Kazempour and P. Pinson, "Price-taker offering strategy in electricity pay-as-bid markets," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2175-2183, Mar. 2018.
- [88] Z. Baharlouei and M. Hashemi, "Efficiency-fairness trade-off in privacy-preserving autonomous demand side management," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 799-808, March 2014.
- [89] Y. Li, C. Wang, G. Li, C. Chen, "Optimal scheduling of integrated demand response-enabled integrated energy systems with uncertain renewable generations: a stackelberg game approach," *Energy Convers. Manag.*, Vol. 235, 2021.
- [90] O. Jogunola et al., "Consensus algorithms and deep reinforcement learning in energy market: a review," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4211-4227, 15 March 2021.
- [91] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, D. d. L. Casas, D. Budden, A. Abdolmaleki, J. Merel, A. Lefrancq, et al., "Deepmind control suite," arXiv preprint arXiv:1801.00690, 2018.
- [92] S. Devlin, L. Yliniemi, D. Kudenko, and K. Tumer, "Potential-based difference rewards for multiagent reinforcement learning," *International Conference on Autonomous Agents and Multi-agent Systems*, 2014.
- [93] Q. Zang and L. Zhang, "Asymptotic behaviour of the trajectory fitting estimator for reflected Ornstein–Uhlenbeck processes," *J. Theor. Probability*, vol. 3, pp. 1–19, 2017.
- [94] Pecan Street Database. [Online]. Available: <http://www.pecanstreet.org/>.
- [95] Morstyn, T.; Hredzak, B.; Agelidis, V.G. Distributed cooperative control of microgrid storage. *IEEE Trans. Power Syst.* 2015, 30, 2780–2789.
- [96] Trivedi, R.; Khadem, S. Implementation of artificial intelligence techniques in microgrid control environment: Current progress and future scopes. *Energy AI* 2022, 8, 100147.
- [97] Shcherbakova, A.A.; Shvedov, G.V.; Morsin, I.A. Power consumption of typical apartments of multi-storey residential buildings. In *Proceedings of the 2020 International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, Moscow, Russia, 12–14 March 2020.
- [98] Satpathy, P.R.; Babu, T.S.; Shanmugam, S.K.; Popavath, L.N.; Alhelou, H.H. Impact of uneven shading by neighboring buildings and clouds on the conventional and hybrid configurations of roof-top PV arrays. *IEEE Access* 2021, 9, 139059–139073.
- [99] Power Wall. Available online: <https://www.tesla.cn/powerwall> (accessed on 12 April 2022).
- [100] Shuai, H.; He, H. Online scheduling of a residential microgrid via Monte-Carlo tree search and a learned model. *IEEE Trans. Smart Grid* 2021, 12, 1073–1087.

- [101] Hua, H.; Qin, Y.; Hao, C.; Cao, J. Optimal energy management strategies for energy Internet via deep reinforcement learning approach. *Appl. Energy* 2019, 239, 598–609.
- [102] PJM. Real-Time Hourly LMPs. Available online: [https://dataminer2.pjm.com/feed/rt\\_hrl\\_lmpps](https://dataminer2.pjm.com/feed/rt_hrl_lmpps) (accessed on 21 March 2022).
- [103] Liu, N.; Yu, X.; Wang, C.; Li, C.; Ma, L.; Lei, J. Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers. *IEEE Trans. Power Syst.* 2017, 32, 3569–3583.
- [104] Zhao, X.; Gao, W.; Qian, F.; Ge, J. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system. *Energy* 2021, 229, 120538.
- [105] A. Mu, S. Roque, F. S. Eugenio, and J. Garc, “Residual demand curves for modeling the effect of complex offering conditions on day-ahead electricity markets,” *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 50–61, Jan. 2017.
- [106] G. Aneiros, J. M. Vilar, R. Cao, and A. M. San Roque, “Functional prediction for the residual demand in electricity spot markets,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4201–4208, Nov. 2013.
- [107] 张素芳, 王鹏. 数据中心行业在能源转型中将发挥重要作用[N/OL]. (2020-07-24) [2022-03-10]. <https://baijiahao.baidu.com/s?id=1673076860372734884&wfr=spider&for=pc>.
- [108] 李洋, 肖泽青, 聂松松, 等. 生成对抗网络及其在新能源数据质量中的应用研究综述[J]. *南方电网技术*, 2020, 14(2): 25-33.
- [109] HUA H C, QIN Y C, CAO J W. Stochastic optimal control for energy internet: a bottom-up energy management approach[J]. *IEEE Transactions on Industrial Informatics*, 2019, 5(3): 1788-1797.
- [110] 绿色和平组织, 华北电力大学. 点亮绿色云端:中国数据中心能耗与可再生能源使用潜力研究[R]. 北京: 绿色和平组织, 2019.
- [111] 陈敏, 高赐威, 陈宋宋, 等. 考虑数据中心用电负荷调节潜力的双层经济调度模型[J]. *中国电机工程学报*, 2019, 39(5): 1301-1314.
- [112] 王晴, 刘友波, 黄杨, 等. 考虑数据中心需求响应的城市电网阻塞管理[J]. *电网技术*, 2020, 44(8): 3129-3138.
- [113] 高赐威, 吴刚, 陈宋宋. 考虑地理分散的数据中心服务器频率调节的电网降损模型[J]. *中国电机工程学报*, 2019, 39(6): 1673-1681, 1863.
- [114] YU L, JIANG T, ZOU Y, Distributed online energy management for data centers and electric vehicles in smart grid[J]. *IEEE Internet Things J.*, 2016, 3(6): 1373–1384.
- [115] 杨挺, 姜含, 侯昱丞, 等. 基于计算负荷时-空双维迁移的互联多数据中心碳中和调控方法研究[J]. *中国电机工程学报*, 2022, 42(1): 164-177.
- [116] GOOGLE. Cluster workload traces[Z/OL]. (2021-10-22) [2022-03-11]. <https://github.com/google/cluster-data>.

- [117] 赵剑, 黄悦琛, 李海阳, 等.垂直起降运载火箭返回轨迹不确定性优化[J]. 航空学报, 2021, 42(11): 262-276.
- [118] 黄少锋, 李慧, 李轶凡, 等.换相序技术应用于稳定控制的条件及最优控制策略[J]. 电工技术学报, 2021, 36(11): 2245-2254.
- [119] 李奇, 王晓锋, 孟翔, 等.基于在线辨识和极小值原理的 PEMFC 混合动力系统综合能量管理方法[J].中国电机工程学报, 2020, 40(21): 6991-7002.
- [120] PETER A. Data centers and advanced microgrids[R]. Boulder: Navigant Consulting, 2017.
- [121] LU X, KONG F, LIU X, et al. Bulk savings for bulk transfers: minimizing the energy-cost for geo-distributed data centers[J]. IEEE Transactions on Cloud Computing, 2020, 8(1): 73-85.
- [122] CHEUNG H, WANG S W, ZHUANG C Q, et al. A simplified power consumption model of information technology (IT) equipment in data centers for energy system real-time dynamic simulation[J]. Applied Energy, 2018, 222: 329-342.
- [123] DAYARATHNA M, WEN Y, FAN R. Data center energy consumption modeling: a survey[J]. IEEE Communications Surveys & Tutorials, 2016, 118(1): 732-794.
- [124] KHOSRAVI A, ANDREW L L H, BUYYA R. Dynamic VM placement method for minimizing energy and carbon cost in geographically distributed cloud data centers[J]. IEEE Transactions on Sustainable Computing, 2017, 2(2): 183-196.
- [125] LIU Z, LIN M, WIERMAN A, et al. Greening geographical load balancing[J]. IEEE/ACM Transactions on Networking, 2015, 23(2): 657-671.
- [126] CHEN L, ZHOU S, XU J. Computation peer offloading for energy-constrained mobile edge computing in small-cell networks[J]. IEEE/ACM Transactions on Networking, 2018, 26(4): 1619-1632.
- [127] AHMAD F, VIJAYKUMAR T N. Joint optimization of idle and cooling power in data centers while maintaining response time[C]. Fifteenth Edition of ASPLOS on Architectural Support for Programming Languages and Operating Systems. ACM, 2010: 243-256.
- [128] LIU N, YU X, WANG C, et al. Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers[J]. IEEE Transactions on Power Systems, 2017, 32(5): 3569-3583.
- [129] FRANK L, LEWIS, DRAGUNA L, et al. Optimal control[M]. New Jersey: John Wiley & Sons, 2012.
- [130] BEAL L D R, HILL D, MARTIN R A, et al. GEKKO optimization suite [J]. Processes, 2018, 6(8): 106.
- [131] 武昭原, 周明, 王剑晓, 等.双碳目标下提升电力系统灵活性的市场机制综述 [J/OL]. 中国电机工程学报 . [2022-03-11]. DOI:10.13334/j.0258-8013.pcsee.212117.

- [132] 吴云芸, 方家琨, 艾小猛, 等. 计及多种储能协调运行的数据中心实时能量管理[J/OL]. 电力自动化设备, 2021, 41(10): 82-89.
- [133] EPA, Global greenhouse gas emissions data [Online]. Available: <https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data>.
- [134] Y. Wang, J. Qiu, Y. Tao, and J. Zhao, "Carbon-oriented operational planning in coupled electricity and emission trading markets," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3145-3157, Jul. 2020.
- [135] Y. Sun, C. Kang, Q. Xia, Q. Chen, N. Zhang, and Y. Cheng, "Analysis of transmission expansion planning considering consumption-based carbon emissions accounting." *Appl. Energy*, vol. 193, pp. 232-242, May. 2017.
- [136] C. Shao, Y. Ding, and J. Wang. "A low-carbon economic dispatch model incorporated with consumption-side emission penalty scheme." *Appl. Energy*, vol. 238, pp. 1084-1092, Mar. 2019.
- [137] H. Liang, et al., "Stochastic optimal energy storage management for energy routers via compressive sensing," *IEEE Trans. Ind. Inform.*, vol. 18, no. 4, pp. 2192-2202, Apr. 2022.
- [138] H. Hua, Y. Qin, Z. He, L. Li, and J. Cao, "Energy sharing and frequency regulation in energy Internet via mixed  $H_2/H_\infty$  control with Markovian jump," *CSEE J. Power & Energy Syst.*, vol. 7, no. 6, pp. 1302-1311, Nov. 2021.
- [139] M. Shamsi and P. Cuffe, "A prediction market trading strategy to hedge financial risks of wind power producers in electricity markets," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4513-4523, Sep. 2021.
- [140] R. Zhang, T. Jiang, F. Li, G. Li, H. Chen, and X. Li, "Coordinated bidding strategy of wind farms and power-to-gas facilities using a cooperative game approach," *IEEE Trans. Sustain Energy*, vol. 11, no. 4, pp. 2545-2555, Oct. 2020.
- [141] PJM, Hourly Market Bid Data for Demand Response [Online]. Available: <https://www.pjm.com/-/media/markets-ops/energy/real-time/data-description-ido-for-demand-response-data.ashx>
- [142] Y. Qin, H. Hua, and J. Cao, "Stochastic optimal control scheme for battery lifetime extension in islanded microgrid via a novel modeling approach," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4467-4475, Jul. 2019.
- [143] D. Cao, et al., "Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices." *J. Mod. Power Syst. Clean Energy*, vol. 9, no. 5, pp. 1101-1110, Sep. 2021.
- [144] Y. Li, R. Wang, and Y. Zhen. "Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting." *IEEE Trans. Sust. Energy*, vol. 13, no. 1, pp. 159-169, Jan. 2022.
- [145] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," arXiv preprint arXiv:1811.12560, 2018.

- [146] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, “High-dimensional continuous control using generalized advantage estimation,” arXiv preprint arXiv:1506.02438, 2015.
- [147] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” arXiv preprint arXiv: 1707.06347, 2017.
- [148] PJM, Data directory [Online]. Available: <https://www.pjm.com/markets-and-operations/data-dictionary>.

## 项目及研究成果

### 参与项目：

#### (1) 南方电网公司科技项目：

基于新一代人工智能和 5G 通信技术的电能质量数据监测管理与分析技术

#### (2) 清华-丰田研究院跨学科基金项目：

信息物理融合的城市能源系统综合状态感知、安全评估以及能量管理

#### (3) 国家高铁创新中心项目：

新型高效能城市轨道交通系统建模与综合智能优化

#### (4) 深圳市腾讯计算机系统有限公司项目：

基于能量路由器的数据中心与储能系统融合架构研究

### 期刊论文：

- [1] **D. Liu**, J. Cao, M. Liu. Joint Optimization of Energy Storage Sharing and Demand Response in Microgrid Considering Multiple Uncertainties. *Energies*[J]. 2022, 15: 3067.
- [2] Z. Qin, **D. Liu**, H. Hua and J. Cao. Privacy Preserving Load Control of Residential Microgrid via Deep Reinforcement Learning[J]. *IEEE Transactions on Smart Grid*, 2021, 12(5): 4079-4089.
- [3] 刘迪,曹军威,刘明爽.分布式数据中心信息能量协同优化策略[J/OL].清华大学学报(自然科学版):1-11[2022-09-15].DOI:10.16511/j.cnki.qhdxxb.2022.21.016.
- [4] **D. Liu**, Z. Qin, H. Hua, Y. Ding, and J. Cao. Incremental Incentive Mechanism Design for Diversified Consumers in Demand Response, *Applied Energy* (在审, 第一作者)
- [5] H. Hua, **D. Liu**, Z. Qin, J. Sun, L. Gan, M. Yin, K. Li, X. Chen, and J. Cao. A Demand Side Joint Electricity and Carbon Trading Mechanism, *IEEE Transactions on Smart Grid* (在审, 通信作者)

### 申请发明专利：

- [1] 基于分布式储能的电力信息处理方法及系统 (第一发明人, 申请号: 202110789445.2)

## 个人简介

刘迪，男，汉族，1990年3月生，河南开封人。

2009-2013年，华北电力大学电气与电子工程学院电力工程与管理专业攻读学士学位；

2015-2017年，华北电力大学电气与电子工程学院电子与通信工程专业攻读硕士学位；

2017-2020年，华北电力大学电气与电子工程学院电气工程专业攻读博士学位；

2020-2022年，清华大学自动化系控制科学与工程流动站从事博士后工作。