arXiv:2602.16063v1 [eess.SY] 17 Feb 2026

MARLEM：用于去中心化本地能源市场隐式合作的多智能体强化学习仿真框架

Nelson Salazar-Peña a，Alejandra Tabares b，以及 Andrés González-Mancera a,*

a 安第斯大学机械工程系，111711，哥伦比亚波哥大特区　b 安第斯大学工业工程系，111711，哥伦比亚波哥大特区

摘要

分布式能源资源的普及要求本地能源市场（LEM）采用先进的协调机制，以维持电网稳定性和经济效率；而传统的集中式控制难以应对这一挑战。由于缺乏能够有机整合真实的去中心化市场动态、物理电网约束和先进多智能体强化学习（MARL）能力的仿真框架，研究领域存在显著空白，尤其难以研究部分可观测条件下涌现的协调行为。本文介绍一种新颖的开源多智能体强化学习仿真框架，用于研究本地能源市场中的隐式合作。该框架将此问题建模为去中心化部分可观测马尔可夫决策过程，并实现为面向多智能体强化学习的 Gymnasium 环境。框架包含一个模块化市场平台，支持即插即用的出清机制；具有物理约束的智能体模型（包括电池储能）；真实的电网；以及用于评估涌现协调行为的综合分析工具套件。本文的主要贡献是一种促进隐式合作的新方法：通过在智能体的观测和奖励中加入系统级关键绩效指标，使智能体能够独立学习有益于整个系统的策略，并致力于实现集体有利的结果，而无需显式通信。通过具有代表性的案例研究（可在专门的 GitHub代码库中获取：https://github.com/salazarna/marlem），我们展示了该框架分析不同市场配置（例如储能部署变化）如何影响系统性能的能力。这体现了该框架在促进涌现协调、提高市场效率和增强电网稳定性方面的潜力。所提出的仿真框架是一种灵活、可扩展且可复现的工具，可供研究人员和从业者设计、测试并验证未来智能化去中心化能源系统的策略。

关键词 本地能源市场、多智能体强化学习、去中心化部分可观测马尔可夫决策过程、隐式合作、去中心化能源系统、仿真框架

*通讯作者**电子邮件地址：**na.salazar10@uniandes.edu.co (Nelson Salazar-Peña)，a.tabaresp@uniandes.edu.co (AlejandraTabares)，angonzal@uniandes.edu.co (Andrés González-Mancera)

1 引言

1.1 研究动机

全球能源行业正经历根本性的范式转变，从历来依赖少数大型、可调度化石燃料电厂的集中式发电模式，转向高度去中心化的系统，其特点是分布式能源资源（DER）的广泛普及 [1, 2]。这类资产包括住宅和商业屋顶太阳能光伏、电池储能系统、电动汽车以及可控负荷 [3]。这一转变由多种因素共同推动：国际气候政策提出的脱碳要求、可再生能源和储能技术成本下降使其对消费者而言具备经济可行性，以及人们对能源自主性和抵御大规模电网中断能力日益增长的需求 [2]。

这一演变标志着产消者的兴起：他们是积极的市场参与者，不仅消费能源，也生产、储存并可能管理能源 [4]。这种范式转变为实现更高效、更具韧性且更可持续的能源未来带来了重大机遇，但同时也给配电网带来了运行方面的挑战 [5]。传统上，电力从中央发电机组单向流向被动消费者；如今，这一模式正被涉及众多异质参与者的复杂双向潮流所取代。因此，需要采用更复杂的方法进行协调与控制，以应对电网拥塞、电压不稳定以及可再生能源发电固有的波动性等潜在问题 [5]。

在此背景下，本地能源市场（LEM）作为一种有前景的框架应运而生，旨在管理配电层级日益复杂的情况 [1, 2]。通过为特定地理社区内的点对点（P2P）能源交易和灵活性服务提供平台，本地能源市场致力于增强本地电网稳定性，促进本地可再生能源资源的高效利用，减少对大容量输电系统的依赖（从而最大限度地降低损耗），并让产消者能够直接将其灵活性变现，以释放分布式资产的全部经济潜力 [1, 2]。

1.2 问题陈述

尽管本地能源市场潜力巨大，但要有效设计并运行高效、稳定且可扩展的本地能源市场，仍是一项复杂的科学与工程挑战，通常被概括为三难困境 [1, 2]。首先，要在数量可能极其庞大的自主自利型智能体之间实现高效且可扩展的协调，是一个极其复杂的问题 [6]。每个智能体（代表一户家庭、一栋建筑、一辆电动汽车或一块社区电池）都会根据其私人目标（例如降低成本、提高收入）和有限的局部信息作出决策。这种固有的去中心化会形成非平稳环境，在这种环境下，任何单个智能体的最优策略都取决于所有其他参与者同时采取的、往往无法观测到的动作 [2, 7]。

其次，构成市场的经济交易不得损害底层配电网的物理完整性 [1]。能源交易对应着物理功率的注入与取出。即使对个别智能体而言在经济上合理的行动（例如在价格高峰时同时让电池放电），若缺乏协调，也可能导致局部电压越限、线路热过载或电能质量下降至不可接受的程度，从而威胁电网安全 [1, 2]。因此，任何可行的本地能源市场设计都必须内在地遵循这些物理约束，或与这些约束协同管理。本地能源市场仿真软件因而必须将经济决策与物理电网仿真紧密耦合，以捕捉这些技术经济交互。

第三，切实可行的本地能源市场架构必须保障参与者的隐私和自主性 [6]。集中式控制方案要求智能体披露敏感用电数据，或将控制权交给中心实体，因此往往会因隐私问题、网络安全风险以及可能造成单点故障而面临重大阻碍 [2, 6]。可扩展性也会成为一个主要问题，因为负责管理大量（数百或数千个）分布式能源资源的中央控制器将承受巨大的计算负担。因此，尊重智能体自主性和数据隐私的真正去中心化方法备受期待，甚至不可或缺；这也需要能够对这类范式进行建模和严格评估的软件。

因此，核心问题在于如何发现相应的机制和系统设计，并开发和测试这些机制与设计所需的仿真软件，使系统层面的理想目标（供需平衡、整体经济效率、物理电网安全）能够通过在部分可观测条件下运行的独立实体之间未经协调、各自为利的动作自然实现，而不必诉诸难以为继的集中式控制。

1.3 研究空白

要应对本地能源市场三难困境，需要复杂的建模与仿真软件，以捕捉自适应智能体学习（尤其是多智能体强化学习（MARL））、复杂的市场动态和现实物理电网约束之间错综复杂的相互作用。然而，对现有文献的全面综述（详见第 2 节）表明

目前存在一个重大且关键的缺口：尚无现有软件能够在单一、标准化且易于使用的框架中，有效整合现实的电力网络仿真、灵活能源市场建模和基于智能体的多智能体强化学习 [8]。

现有仿真工具往往因碎片化和专业化而存在不足，阻碍了对整体集成本地能源市场的全面理解 [2, 7, 8]：

- **碎片化：** 许多人称电力系统仿真器（例如 GridLAB-D MASPO）擅长物理建模，但缺乏对基于智能体的市场交互和自适应学习能力的原生支持 [9, 10, 11]。相反，面向能源市场的基于智能体的建模平台（如 Lemlab/Hamlet [12]）通常侧重于基于优化的智能体或简化的启发式方法，缺乏研究涌现自适应策略所需的、与最新多智能体强化学习算法的稳健集成 [5, 13]。这迫使研究人员采用复杂的联合仿真方案，或对系统的关键方面进行抽象。
- **抽象：** 许多平台，尤其是专注于特定强化学习应用（如需求响应）的平台（例如 CityLearn [5]），采用“铜板式”电网模型，忽略了拥塞和损耗等关键物理约束。这严重限制了对至关重要的技术经济交互的研究，也限制了市场动作影响电网稳定性的潜力[8]。其他平台则对市场机制本身进行抽象，无法研究不同拍卖类型或定价规则如何影响主体行为 [7]。
- **集中化偏差：** 整合强化学习的框架通常面向批发市场（例如 ASSUME [5]），或从根本上依赖集中式训练与分散式执行（CTDE）范式 [5, 14]。虽然 CTDE 具有实用性，但在学习阶段仍需要大量集中协调，这从根本上违背了真正去中心化、保护隐私且具备韧性的本地能源市场的原则，也未能捕捉完全去中心化学习所面临的独特挑战 [8]。
- **对隐式协调的关注有限：** 关键的是，目前尚无标准化、可扩展且原生支持多智能体强化学习的仿真框架，专门用于研究具有物理约束的去中心化市场中，在极具挑战性的分散式训练与分散式执行（DTDE）范式下的隐式协调——智能体通过响应共享环境信号，在无需显式通信的情况下学会协调 [2, 8]。这种能力对于探索可扩展且保护隐私的协调机制至关重要。

缺乏集成化、标准化且重点恰当的工具，严重阻碍了对真正去中心化、智能化且具备物理感知能力的本地能源市场的理解、设计与部署。

1.4 新颖性与贡献

本研究提出了一种新颖的开源仿真框架，用于研究去中心化本地能源市场中的多智能体强化学习。该软件旨在填补已识别的空白，提供首个将这一研究领域所需组件有机整合的平台。具体而言，本工作通过支持以下研究来应对本地能源市场设计与运行的核心挑战：

- 如何在我们的统一软件中设计多智能体仿真环境，尤其是其信息与激励结构（观测与奖励设计），以促进自利的多智能体强化学习智能体之间涌现隐式合作，从而提升电网稳定性与经济效率。
- 在软件中以模块化方式实现的不同去中心化市场机制（例如拍卖类型、定价规则）如何影响学习动态、智能体的策略行为，以及在统一仿真环境中由此产生的市场层面的协调与经济结果。
- 显式建模的物理电网约束（例如输电损耗与拥塞，通过软件集成的电网组件进行仿真）对多智能体强化学习智能体所采用的交易策略以及整体市场表现与稳定性的影响。

所提出的本地能源市场软件框架具有多项贡献与创新，旨在弥补现有工具中已识别的缺口：

1. **统一的技术经济多智能体强化学习环境。** 它独特地将模块化、兼容多智能体强化学习的市场（包括新颖的基于偏好的匹配）与物理上真实的配电网模型（包括损耗和阻塞）结合在单一、集成的 Gymnasium 环境中，明确克服了先前工具中普遍存在的市场与电网二分法及碎片化问题 [8]。

2. **聚焦完全去中心化**。 与许多偏向集中训练、分散执行的现有工具不同， [5, 15], 我们的软件在架构上原生支持并促进对分散训练、分散执行范式的研究， [16, 17], 从而契合现实世界去中心化系统的原则与要求 [8]。 3. **以隐式合作为核心构念**。 其观测与奖励结构明确采用系统级关键绩效指标（关键绩效指标）进行设计，以便研究由共享环境信号驱动的涌现合作， [18, 19], 从而超越纯粹的利润最大化目标，在完全去中心化情境中走向具有系统意识的智能体学习 [2, 8]。 4. **标准化与模块化**。 遵循广泛采用的 Gymnasium 标准， [20] 确保与庞大的强化学习算法生态系统互操作，同时其模块化设计（尤其针对市场与电网组件）促进可复现性、可扩展性，并便于更广泛的研究界使用 [2, 7, 8]。 5. **全面的分析工具套件**。 它包含一套集成工具，用于量化涌现行为、市场表现、电网稳定性与协调效果，从而提供在软件内直接验证科学主张所需的实证工具 [7]。

通过在一个易于使用的软件包中提供这种独特的功能组合，我们的框架成为一项重要工具，有助于加深对真正去中心化、智能、具备物理感知能力且高效的本地能源市场的理解并推动其发展，使其能够支撑未来的能源电网 [8]。

1.5 论文结构

本文详细阐述了该框架的设计、实现及功能。本文其余部分安排如下：第2 节全面回顾相关研究，并进一步阐明本框架旨在解决的研究空白。第3 节详细介绍框架的方法论与架构，涵盖去中心化部分可观测马尔可夫决策过程形式化、智能体模型、市场设计、电网集成、隐式协作机制以及多智能体强化学习范式。第4 节概述用于展示框架功能的实验设置。第5 节呈现并讨论本次实验的结果。最后，第6 节总结全文，概括主要贡献，并指出未来研究方向。

2 文献综述

我们的仿真框架开发立足于多个不断发展的研究领域的交汇处：去中心化能源系统、多智能体系统、用于电力系统管理的多智能体强化学习，以及专用仿真环境的设计。全面回顾这些领域的最新研究进展，对于阐明我们的贡献背景并指出本研究旨在解决的研究空白至关重要。本节将对此展开综述，分析与本地能源市场相关的现有方法论、算法和仿真工具。

2.1 集中式与去中心化能源市场

传统的电网管理范式本质上是集中式的，依赖单一实体（例如独立系统运营商）来解决大规模的机组组合和经济调度优化问题。这种自上而下的方法历来能够确保电网可靠性。然而，地理分布广泛且具有间歇性的分布式能源资源不断增加，给集中式控制的可扩展性和有效性带来了根本性挑战。协调数百乃至数千个分布式能源资源所需的大量数据和高频决策，使传统方法在计算上难以处理且效率低下。

作为回应，一个重要的研究趋势已转向去中心化能源系统，尤其是以本地能源市场的形式。本地能源市场提出了一种自下而上的方法，使产消者能够以点对点方式直接相互交易能源。其理论优势众多：通过减少对远距离输电的依赖来增强电网韧性，通过最小化交易成本来提高经济效率，以及更大程度地赋权产消者。然而，这些收益取决于能否在本地层级解决复杂的协调问题。主要权衡在于拥有完全信息的集中式控制器的理论最优性，与去中心化系统所提供的可扩展性、鲁棒性和隐私之间。我们的工作定位于后者，研究能够在没有中央权威机构的情况下实现有效协调、从而缩小这一性能差距的机制。

2.2 能源电网管理中的多智能体系统

将电力系统概念化为多智能体系统是一个成熟的研究领域，为建模众多分布式实体之间的交互提供了自然范式 [21, 22]。早期应用主要采用基于规则或基于优化的智能体，以应对负荷削减、故障诊断、电力恢复和微电网控制等挑战 [23]。这些奠基性工作证明了利用预定义通信协议和启发式策略实现去中心化协调的可行性，其重点通常在于实现系统稳定性或特定运行目标 [9, 21]。例如，多智能体系统已被用于资源分配和调度 [23] 以及协调输电系统运营商和配电系统运营商（配电系统运营商） [24, 25]。多智能体系统的优势在于能够对电网这类复杂社会技术系统所固有的异质性、有限理性和涌现现象进行建模 [26, 27]。

然而，这些传统的多智能体系统方法通常假设智能体遵循固定规则，或具备解决复杂局部优化问题的计算能力；对于异质性产消者而言，这种假设可能并不现实 [5, 26]。更重要的是，这些方法通常缺乏在动态且不确定的环境中进行自适应学习的能力；智能体行为是预先编程的，无法针对不断变化的市场条件或其他参与者的策略调整而演变 [26]。尽管这些静态方法有助于展示分布式控制的潜力，但在现代竞争激烈的本地能源市场中，涌现行为和复杂交互占据主导地位，而这些方法无法充分捕捉所需的自适应策略学习 [14, 28]。

2.3 多智能体强化学习在分布式能源资源协调中的应用与局限

多智能体强化学习已成为一种极具前景的范式，有望克服传统多智能体系统的局限，使自主智能体能够通过在复杂的非平稳环境中直接交互，学习复杂且自适应的策略 [29, 30]。其无模型特性使智能体即使在缺乏明确系统模型的情况下也能有效运行，因此尤其适合应对本地能源市场固有的不确定性 [30]。多智能体强化学习在能源系统中的应用涵盖几个关键领域：

- **点对点能源交易：** 这是核心应用之一，多智能体强化学习智能体通过学习竞价和调度策略，直接交易能源盈余与短缺 [4, 6, 31, 32]。不同研究采用的市场机制复杂度差异显著，范围从详细的双边拍卖 [4, 6] 或连续双向拍卖 [6]，到基于供需比 [1] 或抽象出清价格的高度简化机制。双向拍卖具有较高的经济真实性，但在纳入复杂物理约束时，通常会采用简化模型 [1]。有些方法还会将交替方向乘子法等传统分布式优化方法用作组成部分或基线 [31],，但这些方法往往难以应对实际问题，且对智能体的假设并不现实 [1]。

- **辅助服务：** 研究人员探索了运用多智能体强化学习协调分布式能源资源，以提供电网所需的基本支持服务，例如电压控制 [33] 和频率调节 [1, 33]。这些应用通常侧重于物理控制方面，有时会简化其中涉及的经济激励或市场结构，将问题更多地视为协同控制任务，而非市场交互。

- **需求响应：** 多智能体强化学习智能体学习根据价格信号或聚合商、公用事业公司的直接请求来调整能耗模式 [14, 34]。MARL-iDR 等研究 [14] 表明，基于激励的需求响应能够实现有效协调，通常采用集中训练、分散执行范式 [14]。

尽管取得了令人鼓舞的成果，但对现有文献的批判性分析揭示了将多智能体强化学习问题应用于本地能源市场时通常存在的共同局限：

- **状态空间：** 许多研究主要向智能体提供局部状态信息（例如荷电状态（SoC）、利润）和基本市场信号（例如之前的出清价格） [4, 31]。尽管一些研究纳入了局部电网信息（例如电压水平 [1],，但对系统级关键绩效指标（如整体电网拥塞、社会福利或市场失衡）进行战略性利用，将其作为共享观测信号，以便在没有直接通信的情况下促进隐式协调，仍明显不足。智能体通常缺乏理解其动作对更广泛系统影响所必需的上下文信息。

- **动作空间：** 动作空间通常会被简化。许多研究采用离散动作（例如充放电水平 [22]），或将连续动作限制为报价数量，并假设价格固定或由中央统一确定 [35]。对复杂点对点市场进行建模的框架，往往缺少用于策略性选择交易伙伴或进行基于声誉的交互的明确动作组成部分，从而限制了涌现社会动态的潜力。我们的框架采用连续多维动作空间，包含价格、数量、买卖方向和首选合作伙伴标识，从而支持更丰富的策略学习，兼顾经济与关系层面。

• **奖励函数：**主流方法主要依赖于最大化个体智能体利润 [4, 6, 31]。尽管从经济角度看合乎直觉，但纯粹以利润为导向的奖励往往会导致集体层面不理想的结果，例如市场不稳定、策略性持留，或忽视物理电网约束 [1, 35]。虽然有些研究会针对违反约束的情况设置惩罚（例如电压惩罚 [1]），但很少有研究采用结构化的多目标奖励，明确通过将个体奖励与对系统级关键绩效指标的积极贡献直接挂钩，来激励涌现合作。我们的奖励函数通过市场范围的合作因子和个体贡献因子对基础奖励进行调节，是系统性弥补这一不足的新尝试，并通过使集体成功直接有利于个体来促进亲社会学习。

这些应用中普遍存在的一项挑战是市场与电网二分：在市场保真度方面表现出色的研究，往往将电网抽象为不受约束的"铜板" [4, 6, 36], ；而采用详细电网模型的研究，则往往大幅简化市场 [1, 35]。这种权衡阻碍了全面逼真的本地能源市场仿真的开发，使其难以捕捉关键的技术经济相互依赖关系。

2.4 本地能源市场与多智能体强化学习的仿真框架

多智能体强化学习策略的评估与基准测试依赖于合适的仿真环境。尽管现有工具众多，但一项比较综述指出，目前缺少一个能够充分满足去中心化、受物理约束的本地能源市场研究特定需求的平台，尤其是那些聚焦于隐式协调和完全去中心化学习的研究。

• **基于智能体的建模工具：**诸如 Lemlab/Hamlet [12]之类的框架提供了专为本地能源市场研究设计的模块化基于智能体的建模架构 [5]。它们提供了用于测试各种市场设计和出清机制的工具，但主要侧重于传统经济建模和基于优化的智能体。它们通常缺乏对多智能体强化学习算法和训练范式的原生集成支持，因此难以研究自适应学习动态[13]。

• **基于优化的工具：**OPLEM [37] 和 SIMTES [38] 等框架旨在将市场设计与网络运行评估相结合 [5]。OPLEM 包含适用于多种市场类型（点对点、拍卖、分时电价）的模块，而 SIMTES 则采用联合仿真（整合多智能体系统、Pandapower [39], ns-3 [40]）。两者都纳入了详细的电网模型。然而，它们的核心逻辑依赖数学优化求解器（线性规划、混合整数线性规划）来进行智能体决策和市场出清，而非端到端多智能体强化学习 [5]。在简化假设下，它们很适合用于对最优解进行基准测试，但无法体现多智能体强化学习固有的学习动态或潜在次优性。SINTES 似乎也属于这一类，侧重于交易型能源系统中的优化。

• **面向能源领域的专用强化学习环境：**

– **CityLearn：**这是一个广泛采用且兼容 Gymnasium 的环境，专注于多智能体需求响应和建筑能源协调 [5, 15, 20]。其主要优势在于推动需求响应研究标准化。然而，它缺乏明确的点对点市场机制，并对物理电网进行了抽象，因此不适合研究市场与电网的交互、点对点交易动态，或拥塞等物理约束 [5, 20]。– **ASSUME：**这是一个基于智能体的框架，用于对集成强化学习的批发电力市场进行建模（具体采用集中训练、分散执行范式下的多智能体 TD3 算法） [5, 41]。它具有高度模块化的市场设计（包括类似 EUPHEMIA 的拍卖）和详细的电网建模功能（集成 PyPSA） [5]。然而，由于其聚焦于批发市场，并且本质上依赖集中训练、分散执行，因此不适合研究配电层级上的完全去中心化本地能源市场 [5]。– **RL-ADN：**这是一个高性能强化学习环境，专注于主动配电网络中的最优电池储能调度，并配有"张量潮流"求解器 [42]。它在电网控制研究方面表现出色，但完全缺乏市场机制 [42]。– energy-py [43] 等其他工具侧重于单智能体优化（例如电池储能），而 marl-local-electricit

y

应用了多智能体强化学习，但并未被描述为通用的模块化框架 [13]。

• **电力系统仿真器：** GridLAB-D（配电网）和 MATPOWER（输电网）等工具可实现高保真物理电网建模，但它们本身并非基于智能体的市场模拟器，也缺乏多智能体强化学习集成 [9]。Grid2Op 提供了一个功能强大的 Gymnasium 环境，可从集中式运营商视角模拟输电网运行，但不适用于分散式市场仿真 [44]。

• **通用多智能体强化学习框架：** PyMARL/EPyMARL [45, 46] 和 MARLlib [47] 等库提供了出色的算法实现，但它们是与领域无关的算法库，而非环境。它们在架构上通常侧重于集中训练、分散执行范式，缺少本地能源市场研究所需的内置领域专用能源模型。Mosaik 等协同仿真框架 [41] 虽然灵活，但要整合不同的仿真器（例如用于市场逻辑、电网物理特性和智能体学习的仿真器），需要投入大量集成工作 [41, 48]。

表1 总结了所评估框架的关键特征，以及这些特征与使用多智能体强化学习研究去中心化、考虑物理因素的本地能源市场的需求之间的关系，尤其关注 DTDE 和隐式协调。

Tabl表 1：本地能源市场与多智能体强化学习仿真框架 g.

框架	主要关注点	符合 Gym 标准?	模块化市场?	GridDetail	支持 DTDE 吗?	Ref
Lemlab/Hamlet	本地能源市场（优化）	No	是（基于优化的）	各不相同	N/A	[12]
OPLEM/SIMTES	LEM/TES（优化）	No	是（基于优化）	详细	N/A	[37, 38]
CityLearn	需求响应	Yes	No	抽象化	仅执行	[15, 20]
ASSUME	批发市场（强化学习）	No	是（基于拍卖）	详细	否（集中训练、分散执行）	[5]
RL-ADN	储能系统调度（强化学习）	No	No	High	未指定	[42]
Grid2Op	输电电网运维（强化学习）	Yes	No	High	范式无关	[44]
PyMARL/MARLlib	多智能体强化学习算法	N/A	N/A	N/A	有限（集中训练、分散执行）	[45, 46, 47]
我们的框架	**去中心化本地能源市场（多智能体强化学习）**	**Yes**	**是（原生支持多智能体强化学习）**	**详细**	**是（核心功能）**	-

2.5 研究空白与贡献机遇

这项文献综述揭示了能源系统仿真工具和多智能体强化学习应用领域现状中的显著空白。市场与电网二分的局面依然存在，鲜有框架能够在基于学习的范式中成功整合这两个领域的高保真模型 [4, 35]。此外，多智能体强化学习训练方法中的集中化与去中心化困境仍未解决 [15, 49]。大多数先进的多智能体强化学习能源框架依赖集中训练、分散执行 [5, 14],；尽管这种方式便于训练，却要求有一个掌握全局信息的集中式实体，这从根本上违背了真正去中心化、保护隐私且具有韧性的本地能源市场原则 [44, 50]。

因此，研究领域存在明显空白：缺乏专为去中心化本地能源市场设计的原生支持多智能体强化学习的仿真框架，该框架应同时具备：(i) 适用于多智能体强化学习智能体的模块化市场机制；(ii) 纳入现实的物理电网约束（损耗、拥塞）；(iii) 建立在完全分散训练、分散执行范式之上，并促进对此范式的研究；(iv) 明确纳入相关机制（例如在观测和奖励中共享KPI信号），以研究并促进隐式合作。

我们提出的框架旨在填补这一空白。它为研究界提供了首个标准化、符合 Gymnasium 规范的开源工具，将这些四项基本特性有机整合在一起（参见第 1.4 节）。通过提供这一组合，我们的框架可作为工具，推进对真正去中心化、智能化、具备物理感知且高效的本地能源市场的理解与开发，使其能够支撑未来的能源电网。

3 方法论

本文详述的仿真框架是一个全面的开源工具，用于在多智能体强化学习范式下分析本地能源市场。该问题被形式化为去中心化部分可观测马尔可夫决策过程，以准确刻画此类去中心化系统中固有的不确定性下的决策与有限信息所带来的挑战。这一形式化为智能体与其环境之间的交互奠定了数学基础，从而确保以有原则的方法对本地能源市场的复杂动态进行建模。

该框架的一项核心设计原则是将其实现为符合 Gymnasium 标准的多智能体环境。遵循这一标准是一项关键贡献，因为它确保了与广泛的最新研究进展强化学习库和算法的兼容性。通过采用这一广受认可的接口，该框架降低了其他研究人员的入门门槛，有助于开展可复现实验，并允许对新的多智能体强化学习策略进行标准化基准测试。

本文档旨在引导用户从宏观理解逐步深入到具体实现细节，其中详细介绍了各个核心模块，并分步说明仿真循环，阐述各模块在每个交易时段如何交互，以及训练和推理流程的具体内容，包括所支持的多智能体强化学习算法和训练范式。

这份详尽的文档确保该框架不仅可用作验证工具，也能成为研究界未来扩展和开发的平台。对于希望从更深的技术层面了解该框架的研究人员，我们在专门的 GitHub代码库中提供了补充材料（https://github.com/salazarna/marlem）。

3.1 仿真框架架构

该框架采用模块化的三级架构，在逻辑上将环境、参与者和涌现系统动力学区分开来（见图 1）。这种关注点分离体现了一项关键设计原则，赋予框架高度的灵活性和可扩展性。研究人员可以单独隔离并研究各个组件，例如替换市场机制或智能体学习算法，而无需更改系统的其他部分。这种模块化对于框架发挥多功能研究工具的作用至关重要，可支持对不同本地能源市场配置进行系统的消融研究和比较分析。

- **第 0 层——平台市场：**这一基础层对智能体运行所处的物理与经济环境进行仿真。它并非被动的背景，而是一个具有自身复杂动态的主动组件。该层由市场模块和电网模块组成：前者提供高度可配置的平台，用于结算能源交易；后者则依据真实约束对物理电力网络进行建模。通过执行既定物理定律（例如根据电网拓扑处理拥塞和输电）及市场监管规定，该层使仿真立足于物理现实，从而构建出丰富且高保真的环境，供智能体学习如何应对。该层的保真度至关重要，因为它直接影响智能体为实现最佳性能而必须学习的策略的复杂性。详细技术规范见第 3.3 节，其中介绍市场模块；第 3.5 节则介绍电网模块。

- **第 1 层——多智能体强化学习智能体：**这一层包含自主智能体实体，它们是仿真中的主要参与者。每个智能体都被建模为一个独立且追求自身经济利益的参与者，配备自己的分布式能源资源（DERs），例如电池，并拥有独特的发电与需求曲线。智能体根据局部观测学习其市场参与策略，目标是最大化个体奖励。这一层的复杂性源于智能体的学习过程，以及它们与市场和彼此之间的策略互动。该框架所要解决的核心研究挑战，是理解智能体追求个体奖励与实现理想集体结果之间的张力。完整的智能体模型，包括电池动态和状态管理，详见第 3.4节。

- **第 2 层——隐式合作：**这一概念层体现了该框架的核心科学研究重点。它并非显式控制器或算法，而是由第 0 层与第 1 层之间经设计的互动所产生的涌现式系统级协调。这种合作通过信息结构（即智能体观测的内容）和激励结构（即奖励函数的设定）进行隐式引导。该框架提供了促进并量化这种涌现行为所需的工具，从而支持对去中心化协调的研究。实现细节见第 3.6节。

3.2 去中心化部分可观测马尔可夫决策过程的表述 ss

仿真由环境模块进行编排，该模块主要由一个 Gymnasium 多智能体环境构成，封装了整个本地能源市场。为应对去中心化本地能源市场的复杂性，我们将问题正式建模为一个去中心化部分可观测马尔可夫决策过程。该过程由元组 I，S，A，T，R，Ω，O 定义，其中 I表示参与市场的智能体集合，S 表示环境状态集合，A 表示联合动作空间，T表示状态转移函数，R 表示奖励函数，Ω 表示联合观测空间，O 表示观测函数。下标 i表示智能体 $i \in I$对应的状态、空间或函数。

这种表述尤其适用，因为它明确承认了该问题的核心挑战：每个智能体都必须在非平稳环境中，基于自身的部分观测，在不确定性下做出决策；而状态转移取决于所有智能体的联合动作。此外，这种表述为分析本地能源市场提供了可靠的数学视角：

- 部分可观测性由 Ω 和 $O(o|s, a)$ 刻画，直接体现了去中心化系统中至关重要的信息约束。智能体无法获知同伴的私有内部状态（例如电池电量、财务状况），这既反映了隐私问题，也符合分布式系统的物理现实。

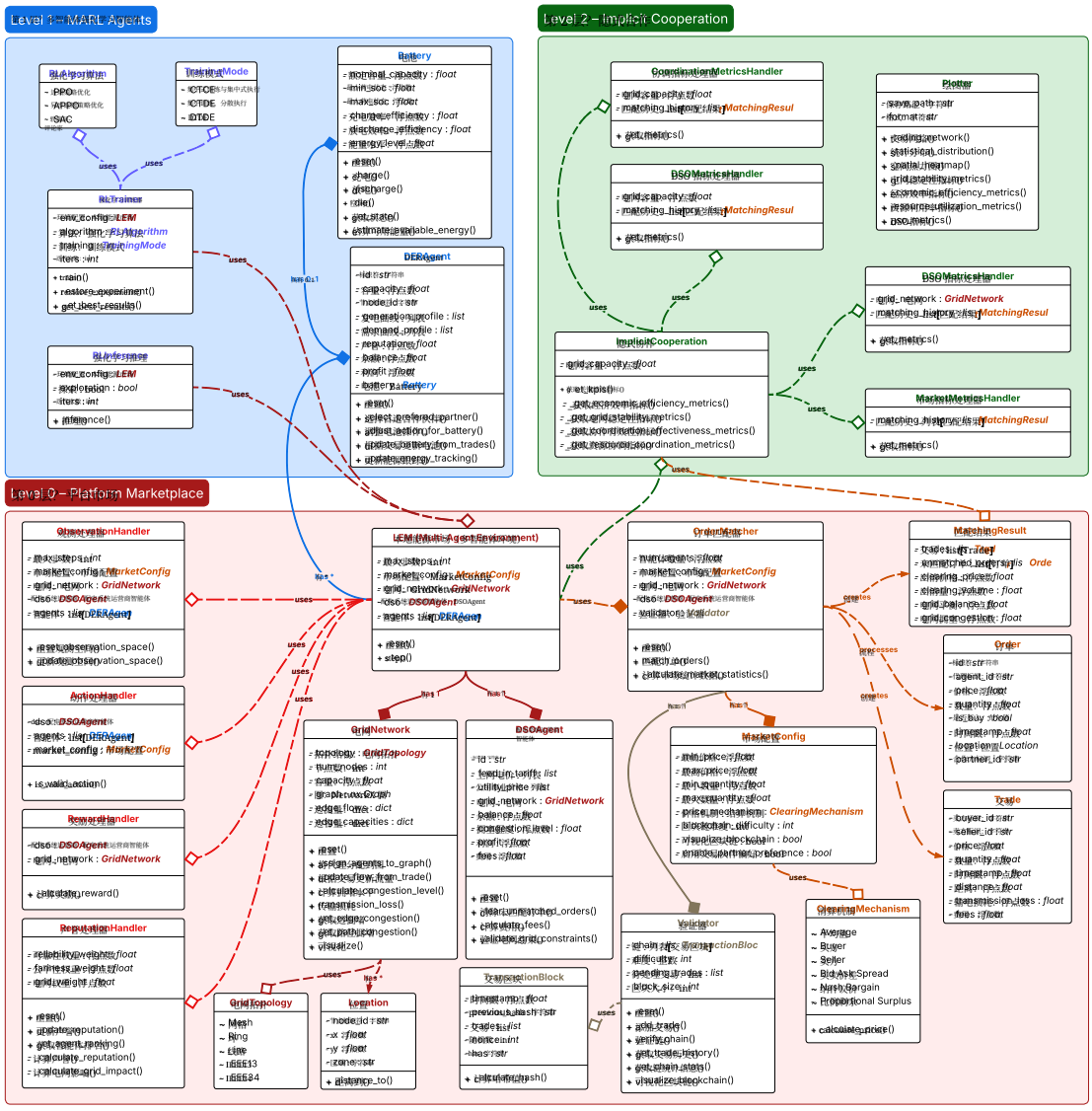


图 1：展示 MARLEM框架核心组件之间关系的高层类图。

- 非平稳性源于 $T(s'|s,a)$，其取决于联合动作 $\{a_1 \in A_1, \ldots, a_i \in A_i\}\ \forall\ i \in I$。由于每个智能体都会独立更新其策略，从任何单个智能体的角度来看，环境的动态变化似乎都在不断发生。

- $R(s,a)$旨在通过提供信号，帮助智能体将各自的动作与集体结果联系起来，从而解决信用分配问题；这对于促进隐式合作至关重要。

3.2.1 环境状态空间（S）和转移函数（T）

环境状态 $s \in S$ 是特定交易时段 $t \in \{1, \ldots, \tau\}$下整个本地能源市场的完整快照，但任何单个智能体都无法完全观测到它。它由时间状态（例如一天中的时刻）、完整的市场状态（所有

订单交易和交易历史）、物理电网状态（所有线路上的拥塞情况）以及所有智能体的私有内部状态组成 .

状态转移函数 $T(s'|s,a)$ 由环境步进方法中的确定性顺序逻辑隐式定义。给定一个联合动作$\{a_1 \in A_1, \ldots \ldots, a_i \in A_i\} \ \forall \ i \in I$，环境首先将智能体动作处理为正式订单，然后结算市场，根据由此产生的交易更新所有智能体和电网状态，最后推进仿真时钟，从而转移到新状态（$s' \in \mathcal{S}$）。这种对状态转移的过程式定义确保环境演化的一致性和可复现性（见第3.9节）。

3.2.2 联合观测空间 (Ω)

观测空间的设计对于实现隐式合作至关重要。为便于在部分可观测环境中进行学习，每个智能体 $i \in I$都会接收一个局部观测向量 $o_i \in \Omega_i$，该向量是从全局状态中精确构造的子集。它为决策提供足够丰富的信号，同时不违背去中心化和隐私原则。观测向量包括：

- **市场信号（公开信息）：** 所有市场参与者均可获取的公开信息。包括最近一次出清的价格和交易量、有关点对点与配电系统运营商交易量的匿名统计数据、配电系统运营商当前的买入价格（上网电价）和卖出价格（公用事业电价），以及每个智能体的信誉评分。这些信号为所有参与者提供共同的信息基础，减少不确定性，并帮助智能体对市场状态形成共同预期。

- **智能体特定信号（私有信息）：** 智能体的私有信息，其他智能体无法观测，从而保障隐私。这组信号包括其当前的能源发电量和需求预测、电池的精确荷电状态、累计利润，以及动态信誉评分。这些私有信息对于智能体根据自身具体情况和约束制定策略至关重要。

- **隐式协作关键绩效指标（共享信号）：** 该框架的一项核心创新特征是在观测向量中纳入一组选定的系统级关键绩效指标，例如社会福利、电网拥塞以及供需失衡（见第3.6节）。这些关键绩效指标作为共享的公共信号，为所有智能体提供对整体市场健康度的一致表征（类似于股市指数），使它们即使在没有直接通信的情况下，也能学习自身局部动作与理想全局结果之间的相关性。例如，智能体可以学会：导致观测到的电网拥塞指标上升的动作，可能会使未来奖励降低。

3.2.3 动作空间（A）

每个智能体的动作 $a_i \in A_i$ 被定义为一个连续向量，这种形式能够支持更精细的竞价策略表达能力，并避免离散动作空间。它由四个组成部分构成：

$$a_i = \langle p_{bid}, q_{bid}, \alpha, \beta \rangle \tag{1}$$

其中，p_{bid}表示出价，q_{bid}表示出价数量，$\alpha \in \{0, 1\}$表示订单类型（0 为卖出，1 为买入），而 $\beta \in \{0, 1, \ldots\ldots, |I+1|\}$是一个索引，表示首选交易伙伴（包括配电系统运营商和无偏好选项）。将 β纳入动作组成部分也是一项创新，使智能体不仅能够学习出价内容，还能学习与谁进行交易。这有助于研究超越简单基于价格交互的复杂社会动态，例如信任网络的形成、稳定交易联盟的形成或社区能源团体的形成。

为确保动作在仿真环境中有效，我们采用两层约束。首先，将智能体策略生成的价格和数量原始连续值裁剪至预设的市场边界（$p_{bid} \in [p_{min}, p_{max}]$和 $q_{bid} \in [q_{min}, q_{max}]$）。其次，在考虑电池限制的同时，对智能体请求的 q_{bid} 施加约束，以确保其满足物理可行性。框架会计算智能体在物理上最多能够出售的能源（依据其发电盈余和可放电电池容量），或购买的能源（依据其电力缺口和可充电电池容量）。

3.2.4 奖励函数（R）

奖励函数 R 是促进隐式合作的主要机制，它将系统稳定性这一高层目标转化为每个智能体可优化的具体信号。智能体 $i \in I$的奖励是一个多目标函数，旨在平衡个体经济激励与系统层面的稳定性目标：

$$R_i = R_{base,i} \cdot (1 + f_{coop} \cdot f_{contrib,i}) - \gamma_{DSO,i} - \gamma_{UD,i} \tag{2}$$

这些组成部分经过设计，可协同运作：

- $R_{base,\ i}$：这是一个表示智能体个体表现的加权和，主要受其交易经济利润的影响。该项驱动智能体采取利己行为，确保其学会有效的竞价策略，以代表自身进行运营。$R_{base,\ i}$是一个包含以下组成部分的复合函数：

  – **经济组成部分：**基于智能体从成功交易中获得的利润。该组成部分激励智能体学习有效的竞价策略。– **电网平衡组成部分：**奖励有助于减少整体电网不平衡的动作（即在盈余时买入、在电力缺口时卖出）。– **资源分配组成部分：**奖励对资源的高效使用（例如，完全执行订单，以及最大限度地减少未使用容量）。– **交易组成部分：**奖励智能体在交易时段内参与市场 $t \in \{1, \ldots\ldots, \tau\}$。– **稳定性组成部分：**奖励有助于长期市场稳定性的行为（例如，保持价格一致性并改善电网状态）。

- f_{coop}：该全市场合作因子根据系统关键绩效指标计算得出，并作为智能体基础奖励的乘法奖励。其动作能够改善整体市场健康度（例如提高社会福利、减少拥塞）的智能体将获得显著更高的奖励。同时，当市场呈现出稳定且高效的特征（例如拥塞程度低、社会福利高）时，所有参与的智能体都会获得额外奖励。这为系统层面的改进创造了共享激励，类似于一种公共池资源：市场健康度是一项所有智能体都能从其维护中受益的资源。通过这种方式，无需显式协调即可促进隐式合作的形成。

- $f_{contrib,\ i}$：该因子通过量化智能体 i 的具体动作对市场积极状态所作的贡献程度，进一步完善合作激励。它是一种有效的信用分配机制，帮助单个智能体了解自身对集体产生的具体正面或负面影响。例如，在全电网电力缺口期间出售能源、从而有助于电网稳定性的智能体，其贡献得分将高于在盈余期间出售能源的智能体。

- $\gamma_{DSO,\ i}$和$\gamma_{UD,\ i}$：这些是针对与配电系统运营商交易以及未满足需求而施加的线性惩罚，旨在鼓励智能体优先进行点对点交易，并促进形成自给自足的本地市场。

这种多层面的奖励结构是促进隐式合作的主要机制，它引导各个智能体的去中心化学习流程，最终实现涌现式的系统级协调。

3.3 本地能源市场设计与交易机制

该市场模块提供了对去中心化本地能源市场的逼真且灵活的仿真。其主要创新在于一种混合匹配机制，将由智能体驱动的交易伙伴偏好与经典的基于价格的拍卖相结合。此外，该机制还配备了可配置的运行规则，以及用于信任和安全的集成机制，这些对于此类系统在现实世界中的可行性至关重要。市场按离散的交易时段运行 $t \in \{1, \ldots, \tau\}$，并遵循从订单提交到最终验证的系统化工作流程（见算法 1）。

3.3.1 配电系统运营商的角色

配电系统运营商是一个关键组成部分，承担着对市场稳定性和运行至关重要的多种角色。首先，它作为最后保障做市商，确保每个交易时段 $t \in \{1, \ldots, \tau\}$都能实现市场出清。其次，它充当现实世界中电网运营商的代理，负责管理电网稳定性并收取电网使用费用。配电系统运营商维护一个余额（B_{DSO}），用于表示净能源头寸（见（3））。

$$B_{DSO} = \sum_{trades} (E_{sold} - E_{bought}) \tag{3}$$

给定交易的总费用 F_{total}由多项费用构成，包括拥塞费用、输电线路的单位距离使用费用（由配电系统运营商预先设定）、不平衡费用、电压支撑费用、热极限费用，以及跨区交易奖励/惩罚。F_{total}的计算公式如下：

算法1 交易时段 $t \in \{1, \ldots\ldots, \tau\}$的市场运营工作流程

输入： 所有智能体的集合 I，联合动作$\{a_1 \in A_1, \ldots\ldots, a_i \in A_i\}\ \forall\ i \in I$，当前市场状态 $s \in$S。

输出： 已执行交易的集合，更新后的市场状态 $s' \in$ S。

▷ 阶段*1*：订单收集与验证

$O_t \leftarrow \emptyset$ ▷ 初始化有效订单空集

对所有智能体 $i \in I$ **执行**

  $o_i \leftarrow$ 动作转订单(a_i) ▷ 将原始动作转换为正式订单结构

  如果 订单验证(o_i, s) 为 True **则**

    $O_t \leftarrow O_t \cup \{o_i\}$

  结束 if

结束 for

▷ 阶段*2*：点对点市场出清

$trade_{pref}$，$O_{rem} \leftarrow$偏好匹配(O_t) ▷ 阶段 1：匹配相互偏好

$trade_{price}$，$O_{unm} \leftarrow$价格匹配(O_{rem}) ▷ 阶段 2：通过连续双向拍卖清算

$trade_{p2p} \leftarrow trade_{pref} \cup trade_{price}$

对所有 交易 $\in trade_{p2p}$**执行**

  更新代理状态$(trade)$ ▷ 更新买方和卖方的能源与利润

结束循环

▷ 阶段 *3*：*DSO* 清算与最终确定

$trade_{dso} \leftarrow$ DSO 清算(O_{unm}) ▷ 阶段 3：与 DSO 清算剩余订单

对于所有 交易 $\in trade_{dso}$ **执行**

  更新代理状态$(trade)$

结束循环

$trade \leftarrow trade_{p2p} \cup trade_{dso}$ ▷ 汇总该交易时段内的所有交易

ValidateBlock$(trade)$ ▷ 模拟对交易区块进行区块链验证

$s' \leftarrow$ UpdateMarketState$(trade,\ s)$ ▷ 更新全局价格、交易量和信誉

$\zeta_{t+1} \leftarrow$ UpdateGridState$(trade, s)$ ▷ 更新受影响的边上的拥塞

返回 $trade, s'$

$$F_{total} = F_{cong} + F_{trans} + F_{imb} + F_{volt} + F_{thermal} + F_{zone} \tag{4}$$

$$F_{cong} = f_{cong} \cdot (C_t - C_{threshold}) \cdot q_{ij} \cdot p_{ij} \tag{4a}$$
$$F_{imb} = f_{imb} \cdot |B_{impact}| \cdot q_{ij} \cdot p_{ij} \tag{4b}$$
$$F_{volt} = \min(V_{drop}, V_{threshold}) \cdot q_{ij} \cdot p_{ij} \tag{4c}$$
$$F_{thermal} = \min(f_{thermal}, f_{threshold}) \cdot q_{ij} \cdot p_{ij} \tag{4d}$$
$$F_{zone} = -f_{zone} \cdot q_{ij} \cdot p_{ij} \tag{4e}$$

$$V_{drop} = f_{volt} \cdot d_{ij} \tag{4f}$$
$$f_{thermal} = \frac{C_t - T_{threshold}}{1 - T_{threshold}} \tag{4g}$$

$$B_{impact} = \begin{cases} \min\left(\frac{q_{ij}}{|B_{grid}|}, 1\right) & \text{if DSO buying/selling during excess supply/demand} \\ -\min\left(\frac{q_{ij}}{|B_{grid}|}, 1\right) & \text{if DSO buying/selling during excess demand/supply} \\ 0 & \text{for balanced grid} \end{cases} \tag{4h}$$

其中， $C_t \in [0,1]$ 表示交易时段 $t \in \{1, \ldots, \tau\}$ 的拥塞水平， $C_{threshold} \in [0,1]$ 表示拥塞干预阈值（默认值为电网容量利用率的 80%）， $B_{impact} \in [-1,1]$ 表示平衡影响评分， V_{drop} 表示电压降， $V_{threshold} \in [0,1]$ 表示电压降阈值（默认值为电压降的 5%）， d_{ij} 表示智能体 i 和 $j \in I$ 之间沿网络的距离， $T_{threshold} \in [0,1]$ 表示电网不稳定的热阈值（默认值为热极限下电网容量利用率的 20%）， $B_{grid} \in [0,1]$ 表示交易时段 $t \in \{1, \ldots, \tau\}$ 的电网平衡（余额为正表示供给过剩，余额为负表示需求过剩）， f 表示各项充电的系数，可根据电网特性进行配置； q_{ij} 和 p_{ij} 分别表示两名智能体之间（包括与配电系统运营商之间）交易的数量和价格。

这一费用结构确保智能体受到激励，做出有利于电网的交易决策，同时为配电系统运营商提供收入，以维护电网基础设施和稳定性。此外，它还允许分析各种电网费率政策及其对点对点交易激励的影响。

3.3.2 多阶段订单撮合过程

该框架的一项创新是一种用于撮合已提交订单的算法，它通过三阶段顺序过程对市场进行结算（见图 2）。这种混合方法旨在优先本地交易和基于关系的交易，同时确保市场流动性与完全结算。此外，通过允许智能体在开放匿名市场的探索性与稳定可信伙伴关系的利用潜力之间取得平衡，它使合作行为的涌现得以研究。各阶段的顺序安排是有意为之的。

- **阶段 1：基于偏好的匹配：** 匹配器首先尝试促成那些在各自报价订单中相互将对方指定为首选伙伴的智能体之间的交易。该阶段明确使人们能够研究稳定的基于信任的交易联盟如何形成，以及它们是提升还是降低整体市场效率。通过优先处理这些预先安排的交易，系统使既有关系优先于更为匿名且可能波动的公开市场，从而体现一种社会效率。

- **阶段 2：基于价格的匹配：** 所有剩余订单均通过双向拍卖处理。这是一种成熟且经济高效的匿名交易机制。买入订单按价格降序排列，卖出订单则按价格升序排列。随后，对所有买方出价不低于卖方要价的匹配订单对执行交易（$p_{buy} \geq p_{sell}$）。信誉评分作为次要排序标准，用于优先考虑更值得信赖的智能体。

- **阶段 3：DSO 出清：** 前几个阶段结束后仍未匹配的订单由 DSO 结算。DSO 确保满足所有剩余需求，并购买所有盈余。不过，DSO 与智能体进行这些交易时采用的价格对智能体而言并不那么有利：购买时采用较低的上网电价（p_{fit}），出售时采用较高的公用事业电价（$p_{utility}$）。这种价格差异（$p_{fit} < p_{P2P} < p_{utility}$）为智能体在前几个阶段通过点对点市场进行交易提供了强大的经济激励，从而促进形成自给自足的本地市场。

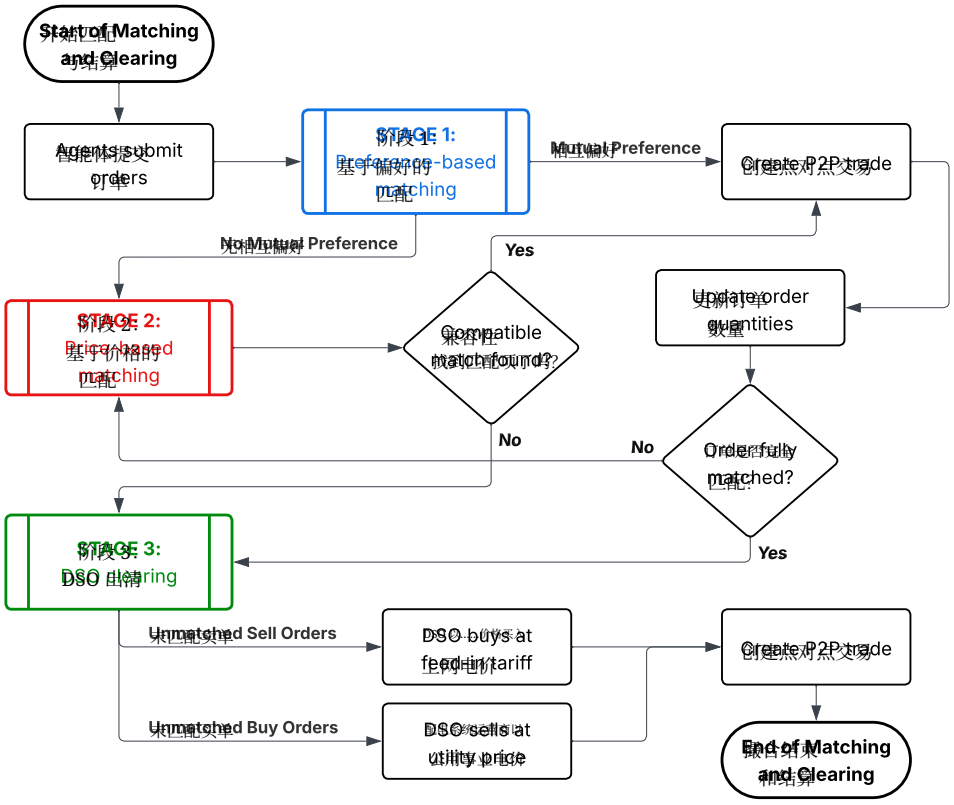


图 2：展示多阶段订单匹配和市场出清流程的示意图。

3.3.3 出清机制

该框架的一项关键特性是其市场设计的模块化。点对点交易的交易价格由可配置的结算机制决定，这使研究人员能够比较不同定价规则对主体行为和市场结果的影响。已实现的机制包括：

- **平均定价：**这是一种简单且公平的机制，将价格设定为买价和卖价的中点，从而确保结算过程透明且计算效率高（见（5））。

$$p_{ij} = \frac{p_{buy} + p_{sell}}{2} \tag{5}$$

- **按报价支付（卖方定价）：**交易按卖方的出价执行（$p_{ij} = p_{seller}$）。此规则可能会激励卖方提交更具策略性的出价，从而对价格发现和整体市场效率产生重大影响。

- **按报价支付（买方定价）：**交易按买方的报价执行（$p_{ij} = p_{buyer}$），这对卖方有利，因为 $p_{buy} \geq p_{sell}$。

- **纳什议价定价：**价格的计算旨在将经济剩余（$p_{buy} - p_{sell}$）在两个智能体之间平均分配（见（6））。该方法促进公平结果，并可能通过确保公平性来促进更长期的市场参与。

$$p_{ij} = p_{sell} + \frac{p_{buy} - p_{sell}}{2} \tag{6}$$

• **比例剩余定价：**盈余根据每个智能体对盈余的贡献按比例分配，并以参考市场价格（p_{ref}）为衡量基准。见（7）。

$$p_{ij} = p_{sell} + \frac{p_{buy} - p_{ref}}{(p_{buy} - p_{ref}) + (p_{ref} - p_{sell})} \cdot (p_{buy} - p_{sell}) \tag{7}$$

3.3.4 声誉与去中心化验证

为应对去中心化系统固有的信任与安全挑战，该框架集成了两项关键功能：

• **声誉系统：** 声誉模块 为每个智能体 $i \in I$分配动态信誉评分 $r_i \in [0,1]$，该评分是其可靠性、价格公平性和电网贡献的加权平均值。此评分在双向拍卖中作为平局判定依据，直接影响市场结果，并通过赋予可靠的智能体优先市场准入权予以奖励。声誉更新过程是迭代进行的：每个交易时段 $t \in \{1, \ldots\ldots, \tau\}$结束后，模块会分析已完成的交易，评估每个智能体对电网的影响，计算各项新的组成评分，并更新智能体的总体声誉，同时引入衰减因子，使近期行为的权重更高。信誉评分的正式定义如下：

$$r_i = w_{rel} \cdot \delta_{rel,i} + w_{fair} \cdot \delta_{fair,i} + w_{grid} \cdot \delta_{grid,i} \tag{8}$$

其中，评分 δ分别衡量 $t \in \{1, \ldots\ldots, \tau\}$的可靠性（匹配订单量与总订单量之比）、价格公平性（与市场出清价格的偏差）和电网贡献（交易是缓解还是加剧电网压力），而 w为相应的权重。

• **去中心化验证：** 验证器模块 模拟区块链账本，以在没有中央仲裁者的情况下确保交易完整性。每个交易时段结束后 $t \in \{1、\ldots\ldots、\tau\}$，所有已执行交易都会经过哈希处理（SHA-256），汇总成一个区块，并通过简化的工作量证明共识机制进行验证。验证通过后，该区块将被添加到不可篡改账本中。这一流程形成了市场活动的透明记录，这是确保可审计性和信任的关键特性。尽管工作量证明存在众所周知的能源效率低下问题，但这里仍采用它，因为它是一种广为人知且简单直接的模型，可用于模拟去中心化共识。

3.4 智能体模型与决策过程

框架中的智能体是复杂实体，代表拥有各自分布式能源资源的产消者，每个实体都具有独特特征和物理约束。代理模块 是基本构建单元，封装了智能体的内部状态、物理资产，以及支配其与市场交互的逻辑。

3.4.1 代理架构与状态管理

每个智能体都是一个自包含实体，由一组核心属性定义，包括唯一标识符、其在电网上的物理位置，以及其在每个交易时段 $t \in \{1、\ldots\ldots、\tau\}$的预测能源发电量 G_t和需求 D_t的时间序列曲线。智能体的内部状态是动态的，会跟踪其余额 B_t（正余额表示能源盈余，而负余额表示电力缺口）和利润、不断变化的信誉评分 r，以及其在市场中的累计表现。这包括跟踪已满足的总需求与被推迟的需求，从而衡量智能体的能源安全。智能体的余额根据（9）进行估算，其中 $E_{bought、t}$和 $E_{sold、t}$分别是在交易时段 $t \in \{1、\ldots\ldots、\tau\}$于市场中买入和卖出的能源数量。

$$B_t = (G_t - D_t) + (E_{bought,t} - E_{sold,t}) \tag{9}$$

我们的智能体模型的一项关键创新是集成了伙伴选择机制。智能体可以通过其强化学习策略学习选择偏好的交易伙伴，从而形成复杂的、基于声誉的交易关系。当不受强化学习策略控制时，智能体可以恢复采用基于规则的策略，根据高声誉和较短电网距离的加权评分对交易伙伴进行优先排序，促进建立既值得信任又高效的交易关系。

智能体在每个交易时段的工作流程 $t \in \{1, \ldots, \tau\}$ 是系统化的：它处理来自其策略的动作，根据物理约束对其进行调整，创建正式的市场订单，并在市场出清后，根据交易结果更新其完整的内部状态。例如，累计利润按照（10）进行更新。

$$\text{profit}_{t+1} = \text{profit}_t + \sum_{trades} (p_{sell} \cdot q_{sell} - p_{buy} \cdot q_{buy}) \tag{10}$$

3.4.2 受物理约束的决策

智能体模型的一个核心特征是明确执行物理约束，这使智能体学习到的经济策略立足于物理现实。在将来自智能体策略的原始动作提交至市场之前，会对智能体的动作施加约束，以确保其可行性，防止智能体试图出售其并不拥有的能源，或购买其无法储存的能源。在交易时段 $t \in \{1, \ldots, \tau\}$，智能体可出售的最大数量（$q_{sell,max}$）和可购买的最大数量（$q_{buy,max}$）根据其内部能量平衡和电池状态，按照（11）计算，其中 C_{nom} 是电池的额定容量，SoC_{min} 和 SoC_{max} 是其运行荷电状态（SoC）限制，而 $E_{dischargeable}$ 和 $E_{chargeable}$ 分别表示该时刻电池中物理可用的能源和储能容量。

$$q_{sell,max} = \max(0, G_t - D_t) + E_{dischargeable} \tag{11a}$$

$$q_{buy,max} = \max(0, D_t - G_t) + E_{chargeable} \tag{11b}$$

$$E_{dischargeable} = (SoC_t - SoC_{min}) \cdot C_{nom} \tag{11c}$$

$$E_{chargeable} = (SoC_{max} - SoC_t) \cdot C_{nom} \tag{11d}$$

3.4.3 电池储能动态

电池是一个可选组件，由一组决定其行为的关键物理参数定义：C_{nom}、SoC_{min} 和 SoC_{max}，以及其充电和放电效率（$\eta_{charge} \in [0,1]$ 和 $\eta_{discharge} \in [0,1]$）。

电池的能量水平（$E_{bat,t} \mid SoC_{min}.C_{nom} \leq E_{bat,t} \leq SoC_{max}.C_{nom}$）在每个交易时段 $t \in \{1, \ldots, \tau\}$ 根据智能体内部余额和市场活动所产生的净能量流而变化。充放电动态计入了储能过程中固有的能量损耗。当以一定量的能源 E_{in} 充电时，储存的能源增加。相反，当一定有用量的能源 E_{out} 由电池输出时，储存的能源会以更大的量被消耗，以计入放电效率损失（见（12））。

$$E_{bat,t+1} = \begin{cases} \min(E_{bat,t} + E_{in} \cdot \eta_{charge}, SoC_{max} \cdot C_{nom}) \\ \max(E_{bat,t} - \frac{E_{out}}{\eta_{dicharge}}, SoC_{min} \cdot C_{nom}) \end{cases} \tag{12}$$

（12）中的充放电公式确保这些操作始终受电池瞬时荷电状态及其最大充放电速率的约束。这种精细建模对于准确刻画储能作为本地能源市场灵活性来源的价值至关重要。

3.4.4 时间序列曲线

每个智能体的行为与决策情境主要由其发电与需求曲线驱动。这些时间序列数据由专用算法生成，旨在产生真实且随机的输入。该模块既可以从外部文件加载数据，从而基于现实世界数据集开展实验，也可以按程序生成数据。

- **发电曲线：**为模拟可再生能源，发电曲线被设计为模拟太阳能光伏系统的出力。发电量 G_t 被建模为智能体容量、基础太阳辐照度曲线（建模为以正午为中心的高斯函数）以及表示大气效应的随机噪声分量的函数：

$$G_t = C_{nom,i} \cdot \lambda_t \cdot (1 + \varepsilon_{noise} + \varepsilon_{cloud}) \tag{13}$$

  其中 $C_{nom,i}$ 为智能体 $i \in I$ 的额定容量（太阳能光伏系统加电池），λ_t 为交易时段 $t \in \{1, \ldots\ldots, \tau\}$ 的太阳辐照度，被建模为以正午为中心的高斯函数，ε_{noise} 为随机噪声项，ε_{cloud} 为模拟云层覆盖效应的项，可引入更大且相关性更强的波动。
- **需求曲线：**需求曲线的生成旨在模拟典型的消费模式，在早晨和傍晚形成峰值。这与太阳能发电在时间上自然错配，从而构成能源交易和储能的基本经济驱动力。

该算法也用于复现市场模块（回顾第3.3.1节）。配电系统运营商提供的上网电价和公用事业电价以阶梯函数的形式生成，以反映典型的批发市场价格变化，

需求高峰时段的价格较高。该算法确保公用事业电价始终高于上网电价，从而为智能体参与点对点交易提供经济激励。为增强真实性，所有生成的曲线都经过平滑滤波器处理，以减少突变并生成更自然的时间序列数据。

3.5 物理电网模型与约束

为捕捉市场活动与物理基础设施之间的关键相互作用，该框架纳入了电网模块。这一组件对于研究经济决策与其对电力电网稳定性和效率造成的现实影响之间的耦合关系不可或缺。它使本地能源市场立足于物理现实，促使智能体学习不仅有利可图，而且在物理上可行的策略。

3.5.1 电网拓扑与智能体分配

物理配电网络建模为加权无向图 G = V、 ξ，智能体位于节点 $v \in$ V 上，并通过边 $e \in \xi$ 相互连接。其一项关键特性是支持多种拓扑类型：

- **预定义拓扑：** 为开展标准化、可复现实验，该框架纳入了电气工程文献中的经典模型，例如 IEEE 13 节点和 IEEE 34节点测试馈线。
- **程序生成拓扑：** 为系统分析网络结构，可以创建网格、环形和线路等通过算法生成的网络拓扑。

每个智能体都被分配到一个对应特定节点 $v \in$ V的位置。随后，两个智能体 d_{ij}之间的电气距离并非按简单的欧几里得距离计算，而是其各自节点（v_i 和 v_j）之间沿网络图最短路径的长度。该距离是计算能源传输物理效应的一项关键输入。

3.5.2 物理约束建模

电网模块 对所有能源交易施加两项基本物理约束，这些约束会直接影响经济结果，并给智能体带来复杂的学习挑战。

- **输电损耗：** 电力在输电线路上传输时不可避免地会耗散。框架将此类损耗建模为（l_{ij}），对于数量为 q_{ij}、电气距离为 d_{ij} 的交易，损耗是考虑可配置损耗系数 κ 的线性函数（参见（14））。

$$\ell_{ij} = d_{ij} \cdot q_{ij} \cdot \kappa \tag{14}$$

这些损耗 l_{ij} 会在电力通过电网传输时降低 q_{ij}，并由配电系统运营商计入交易费用，从而直接激励智能体优先与附近的同伴进行本地交易。这是形成地理聚集型社区级能源市场的主要驱动因素。

- **拥塞：** 潮流 $\mathcal{F}$ 在每条输电线路 $e \in \xi$ 上相对于其最大容量进行监测。总电网容量在各边（C_e）之间分配，阻抗较低（较短）的边通常具有更高的容量。拥塞水平 ζ 在边 $e \in \xi$ 上是该边当前潮流与其最大容量之比（见（15））。

$$\zeta_e = \frac{\mathcal{F}_e}{\max(C_e)} \tag{15}$$

每个交易时段 $t \in$ {1、……、 τ}结束后，每笔已执行交易最短路径上的每条输电线路 $e \in \xi$ 的潮流 F 都会更新。至关重要的是，随后会将更新后的全网拥塞水平 ζ 反馈至智能体的观测空间，作为一项系统级关键绩效指标。这一机制使智能体能够学习协同缓解电网压力的策略，例如减少高负载输电通道上的电力输送。这体现了一种自发形成的分布式拥塞管理方式，而这对于未来电网运行的稳定性至关重要。

3.5.3 电网状态的动态更新与管理

电网状态的 电网并非静态不变，而是会随着市场活动动态演变。这通过一个系统 tic 更新过程 该过程在每个交易时段 $t \in \{1, \ldots, \tau\}$之后进行。

- **潮流更新过程：**对于每笔交易，算法首先在网络图中找出买方与卖方之间的最短路径。然后，沿该路径的每条边 $e \in \xi$上的潮流 F 都会增加交易量 q_{ij}（同时考虑输电损耗）。这确保每笔经济交易对电网造成的物理影响都能反映在电网状态中。对于违反容量约束的交易，配电系统运营商可以予以处罚或禁止。

- **电网平衡计算：**通过电网平衡指标 B_{grid}监测电网的整体运行状况，该指标量化了本地能源市场中的净能源盈余或电力缺口。其计算方式见（16），即对所有智能体的净能源头寸求和。数值 B_{grid}接近零表示市场实现了自给自足，而偏差较大则意味着高度依赖外部电网，这是衡量系统性能的关键指标。

$$B_{grid} = \sum_{i \in I} (G_i - D_i + E_{bought,i} - E_{sold,i}) \tag{16}$$

3.5.4 分区电网结构

为支持对更复杂、大规模电力系统和精细市场设计的研究，该框架支持分区电网结构。此功能可将主电网图 G 划分为一组彼此独立且连通的子图，称为区域，其中每个区域 Z = V, $\xi \subseteq$ G。虽然该功能可配置，但默认情况下，如果买方和卖方位于同一区域，框架会给予奖励（负费用）。

此分区定义主要服务于两项研究目的。首先，它能够对较大电网中地理位置或电气特性不同的区域进行建模，而各区域可能具有不同的特性。其次，它有助于实施先进市场机制，例如基于位置的边际定价或分区拥塞管理方案。例如，发生在不同区域的智能体之间的交易可能会被收取额外费用或受到约束，以模拟跨较大区域输送电力的成本和限制。这为分析去中心化能源市场的可扩展性和分层控制提供了全面的工具。

3.6 隐式合作与性能测量

本研究的一个核心假设是，在精心设计的激励机制下，自利型智能体在共享环境中运行时，无需直接通信便能学会隐式合作。隐式合作模块 是这一过程的基础。它并不强制要求智能体合作，而是通过向智能体提供有助于促进集体利益的系统级信息，对合作进行测量和激励。

隐式合作模块 会计算一组关键绩效指标，作为反馈信号，引导智能体的学习过程，使其朝着互利结果发展。这些关键绩效指标既纳入智能体的观测空间，为系统状态提供共享信号，也纳入奖励函数，直接奖励有助于提升市场健康度的动作。由此形成一个反馈回路，使个体动作受到其对集体影响的塑造。

3.6.1 协调反馈回路

促进隐式合作的机制通过一个闭环运行，将关键绩效指标融入学习过程，并在每个交易时段发挥共享信号的作用。将这些关键绩效指标纳入每个智能体的观测向量后，所有智能体都会收到关于系统状态的一致共享信号。

在交易时段 $t \in \{1, \ldots, \tau\}$结束时，电网模块 会计算物理结果（例如拥塞），而市场模块 会记录经济结果（例如价格、交易量）。随后，隐式合作模块会使用这些数据计算系统级关键绩效指标。这些关键绩效指标并非仅供事后分析记录，而是通过以下两种方式主动反馈到环境中：

1. **作为共享观测：**将经过筛选的关键绩效指标向量附加到每个智能体的局部观测中。这为所有智能体提供了关于系统集体状态的一致共享信号，使它们能够学习自身局部动作与全局结果之间的相关性。2. **作为奖励组成部分：**关键绩效指标被直接用作计算 f_{coop} 和$f_{contrib、i}$ 项的输入，这些项属于奖励函数（参见（2））。这为智能体采取能够改善这些系统级指标的动作创造了直接且切实的激励，从而有效地使个体自身利益与集体利益保持一致。

从全局状态到局部信号的持续信息流，使智能体能够学习协调一致、亲社会的行为，而无需显式通信或中央控制器。该反馈机制会根据不断变化的市场条件进行调整。

3.6.2 关键绩效指标

该框架会计算一套全面的关键绩效指标，并对其分类，以便从多个维度评估系统性能。以下是用于衡量涌现式合作的指标：

1. **经济效率关键绩效指标：** 这组指标用于衡量市场创造价值和促进有效价格发现的能力。

   • **社会福利：** 所有交易创造的经济价值总和，即消费者剩余与生产者剩余之和。它是衡量整体市场效率的主要指标。

   $$\text{Social Welfare} = \sum_{trades} p_{trade} \cdot q_{trade} \tag{17}$$

   • **市场流动性：** 交易的能源总量，反映市场活动和深度。

   $$\text{Liquidity} = \sum_{trades} q_{trade} \tag{18}$$

   • **平均买卖价差：** 买入和卖出订单价格之间的平均差额，用于衡量市场效率。

   $$\text{Spread} = \mathbb{E}[p_{ask}] - \mathbb{E}[p_{bid}] \tag{19}$$

   • **价格波动：** 时间窗口内出清价格的标准差，用于反映市场稳定性 y.

2. **Gri与稳定性关键绩效指标：** 这组指标用于评估电力电网的物理状况和运行效率。

   • **供需失衡：** 经电网容量归一化后的净能量失衡，用于衡量供给与需求的平衡程度。数值接近零表示系统运行稳定，不会给更广泛的电网造成压力。

   $$\text{Imbalance} = \frac{|\sum q_{buy} - \sum q_{sell}|}{C_{grid}} \tag{20}$$

   • **电网拥塞：** 所有输电线路的平均拥塞水平 $e \in \xi$定义见（15），用于反映基础设施承受的物理压力。拥塞水平较低表明系统运行状态良好，未超出其物理限值，从而提升可靠性。

   • **电网平衡：** 电网的整体能量平衡，计算方式为总发电量与能耗之差，如（16）所示。

3. **资源协调关键绩效指标：** 这些指标用于评估分布式能源资源的利用和协调效率。

   • **分布式能源资源自发自用：** 与配电系统运营商交易相比，通过点对点交易完成的交易量占总交易能源量的比例。该值较高，表明本地市场具有较强的自给能力和有效性，并降低了对集中式公用事业公司的依赖。

   $$\text{Self-Consumption} = \frac{\sum q_{P2P}}{\sum (q_{P2P} + q_{DSO})} \tag{21}$$

   • **灵活性利用：** 在点对点交易中实际利用的可用灵活能源比例。该指标衡量智能体利用其能源灵活性资源（发电盈余、需求缺口和电池容量）的效率。

   $$\text{Flexibility Utilization} = \frac{\sum q_{p2p}}{\sum q_{available}} \tag{22}$$

其中， $\sum q_{available}$ 表示所有智能体的可用灵活性总量，计算时包括指定交易时段 $t \in \{1、\ldots\ldots、\tau\}$的可售灵活性（发电盈余加电池放电容量）和可购灵活性（需求缺口加电池充电容量）。

4. **协调有效性关键绩效指标：**这些指标用于衡量智能体之间协调的涌现及其效果。

- **协调得分：**用于衡量协调程度，反映市场的余额。得分接近 1 表示系统达到了完美平衡。

$$\text{协调得分} = 1 - \text{失衡} \tag{23}$$

- **协调收敛性：**衡量近期时间窗口内交易量的稳定性，以判断是否已形成稳定、协调的模式。其计算方式与价格波动率指标类似。

3.7 多智能体强化学习框架

前文各节详细介绍了环境的设计，本小节则介绍负责训练智能体策略的多智能体强化学习框架。我们工作的一个重要贡献是，该框架原生支持比较不同的多智能体强化学习训练与推理范式，这对于研究学习与控制中集中化和去中心化之间的权衡至关重要。这是一个核心研究工具，能够在不同程度的信息共享和架构约束下，严谨地并列评估协调策略。整个训练与推理流程分别由专用的训练器模块和推理模块管理，并基于 Ray RLlib 等成熟库构建 [51]，以确保鲁棒性、可扩展性和可复现性。

关于多智能体强化学习框架的详细论述，包括完整的数学原理和实现细节，将留待后续论文展开；以下内容将全面概述其关键组成部分与能力。

3.7.1 训练与执行范式

本研究的一项主要目标是评估不同去中心化程度下的性能与可扩展性。为此，该框架旨在无缝实现并比较三种典型的多智能体强化学习范式，每种范式都在性能、可扩展性和现实世界可部署性之间实现了独特的平衡。

- **集中式训练、集中式执行（集中式训练与集中式执行）：**在这种范式中，单一的集中式策略可以访问完整的全局状态（所有智能体的观测拼接而成），并同时为所有智能体输出联合动作。该模型代表系统性能的理论上限，因为它体现了完全信息和完全协调执行的理想状态。尽管由于违反去中心化原则且面临指数级扩展的挑战，这种方法在现实世界系统中往往难以实现，但它为衡量理论最优方案与更实用的去中心化方法之间的性能差距提供了基准。
- **集中式训练、去中心化执行（CTDE）：**这是一种广受欢迎且实用的范式，代表了一种混合方法。智能体策略在离线环境中通过集中式模拟器进行训练，并可访问全局信息（例如其他智能体的状态和动作，以及集中式评论家）。这使学习算法能够解决信用分配问题，并学习复杂的协调策略。然而，在推理阶段，每个智能体的策略完全以去中心化方式运行，仅依赖自身的局部观测。这种方法极具吸引力，因为它能在离线环境中学习复杂的协调策略，同时生成可部署于现实去中心化场景的策略，且运行时无需中央控制器。
- **分散式训练与分散式执行（DTDE）：**这代表了完全去中心化的理想模式，与许多现实世界系统面临的约束最为接近。每个智能体仅依据自身获得的局部观测和奖励，独立学习自己的策略，无需集中式训练协调器。这种范式具有很强的可扩展性，也能避免单点故障。不过，它仍面临重大的理论挑战，包括环境的非平稳性（因为其他智能体的策略会不断变化）以及信用分配的难度。我们的框架着重通过观测和奖励函数中的共享 KPI 信号实现隐式合作，专为应对这些挑战而设计，从而使 DTDE 成为更可行、更有效的方法。

3.7.2 强化学习算法与基础设施

该框架与高性能强化学习库集成，以实现学习算法。虽然模块化设计与具体算法无关，但我们的实验主要采用适用于多智能体环境和连续动作空间的最新策略梯度和演员-评论家方法。我们重点介绍三种已实现的算法：

- **近端策略优化（PPO）：** 一种稳健的同策略算法，以稳定性和样本效率著称[52]。其核心创新是裁剪目标函数，可防止策略更新幅度过大而造成破坏，因此成为复杂控制任务中强大而可靠的基线。

- **异步近端策略优化（APPO）：** APPO 通过异策略校正（V-trace）提升分布式训练的可扩展性，使算法能够更有效地从异策略数据中学习 [52]。这在多智能体环境中尤为有价值，因为智能体可能采用不同的学习率和探索策略，从而产生标准 PPO 无法高效处理的异策略数据。

- **软演员-评论家（SAC）：** 一种通过在目标函数中加入熵奖励来鼓励探索的异策略算法 [53]。这种在探索与利用之间取得平衡的原则性方法，在连续控制任务中非常有效，往往能产生更稳健、性能更优的最终策略；对于在本地能源市场中发现新颖的协调策略，这一点尤为重要。

为管理大规模多智能体强化学习实验的复杂性，训练基础设施采用了先进技术，例如基于种群的训练 [54],，这是一种混合方法，可联合优化一组智能体的神经网络权重与超参数，并将计算资源动态分配给最有前景的训练运行。该方法使困难的超参数调优过程自动化，通常能得到更稳健、性能更高的最终策略，这在多智能体强化学习的非平稳环境中尤为有益。整个实验工作流程——从训练和检查点保存到指标收集与最终推理——均受到管理，以确保可复现性和严谨的科学比较。

3.8 分析与可视化

分析模块 与可视化模块 已集成到该框架中，提供一套用于处理、分析和绘制仿真运行期间生成的高维数据的工具。该模块对于超越原始数值输出、定性理解本地能源市场中的涌现动力学至关重要。

可视化功能经过分类，以便从不同视角呈现仿真结果：

- **社会与关系动态分析：** 为研究交易关系和市场结构的形成，框架可以生成交易网络图。在图中，智能体以图中的节点呈现，已执行交易则以连接这些节点的边表示。边的粗细与两个智能体之间的能源交易总量成正比，而边的灰度则表示智能体之间的距离，从而直观呈现市场的社会结构。这使我们能够直接观察稳定的交易联盟和首选伙伴关系如何随时间演变。

- **经济绩效分析：** 为了解市场结果的统计特性，可视化模块 可以生成价格和交易量分布图。这些直方图和核密度估计揭示了市场出清价格和交易量的分布情况，有助于深入了解不同市场机制下的价格稳定性、市场流动性以及价格发现的整体效率。

- **时空模式识别：** 为分析市场活动的物理维度，可以生成交易活动空间分布图。该可视化将电网拓扑呈现为热力图，其中不同节点或区域的颜色强度对应于交易活动量。这对于识别局部交易枢纽的形成，以及将经济活动与拥塞等物理电网现象联系起来尤其有用。

- **性能仪表板：** 为了以较高层次对不同实验场景进行比较和概览，该框架提供了一套指标仪表板。这些综合图表将分析模块 中最关键的关键绩效指标汇总到一种一目了然的格式中。系统为电网稳定性、经济效率、资源利用率和配电系统运营商绩效分别提供了专用仪表板，研究人员可借此快速评估和比较实验的多方面结果。

3.9 完整仿真工作流程

前文各节详细介绍了框架的各个组成部分；本小节最后将阐述这些模块如何协同运作，构成一个完整的端到端仿真循环。LEM 环境中的步进方法负责管理这一系列操作，确保在每个离散交易时段，数据都能在智能体、市场、电网和协调模块之间依逻辑流动 $t \in \{1, \ldots, \tau\}$。理解这一工作流程，对于了解该框架如何维持状态一致性并促进智能体学习过程至关重要。每个交易时段 $t \in \{1, \ldots, \tau\}$的整个过程可概念化为八个不同阶段的序列，从最初的智能体决策到最终的观测和奖励反馈（见图 3）。

仿真过程的详细分解如下：

1. **动作处理与验证：** 工作流程从多智能体强化学习策略接收到的原始动作开始。对于每个智能体，第一步也是最关键的一步，是将其预期动作置于物理现实的约束下。智能体根据当前发电量、需求和自身电池状态限制每个动作。随后，动作模块 执行基于规则的验证，确保动作的各个组成部分（价格、数量、买入/卖出、交易伙伴）均处于预定义的市场边界内。这两步验证确保只有在物理和经济上合理的动作才能进入下一阶段。 2. **订单创建与市场提交：** 经过验证的动作随后被格式化为正式订单。如果智能体的策略不直接控制交易伙伴选择，则可在此处调用基于规则的选择方法，根据声誉和电网距离为潜在交易伙伴评分。此阶段最终形成一组结构化订单，准备提交至市场。 3. **市场出清与交易执行：** 订单集合被传递给市场模块，该模块执行三阶段结算流程：首先尝试匹配相互偏好的交易伙伴，然后通过基于价格的双向拍卖处理剩余订单（将信誉评分纳入考量），最后由配电系统运营商清算所有剩余的未匹配订单。此阶段的输出是所有已执行交易的列表及汇总统计数据。 4. **能量跟踪与电池更新：** 利用所有已执行交易的列表更新每个智能体的物理状态（例如内部能量跟踪变量以及电池荷电状态的变化）和电网状态（例如余额与拥塞）。 5. **奖励计算与关键绩效指标更新：** 交易时段 $t \in \{1, \ldots, \tau\}$的经济和物理结果现已确定，隐式合作模块 收集计算完整系统级关键绩效指标集所需的全部数据（交易、智能体状态、电网条件）。随后，这些关键绩效指标被传递给奖励函数，由其计算每个智能体最终的综合奖励，将其个体表现与源自系统级关键绩效指标的合作奖励相结合。 6. **状态转移与环境更新：** 此阶段推进仿真的全局状态，并使环境的内部时钟递增。至关重要的是，根据声誉模块的结果，更新依赖于完整市场结果的智能体级状态，例如财务余额和声誉评分。 7. **观测更新与信息共享：** 观测模块 为每个智能体构建下一步局部观测（$o_{i,t+1} \in O_i$）。该新观测向量由更新后的环境状态构成，包括智能体的私有信息、公开市场信号，以及在第 5 阶段计算的系统级关键绩效指标。此步骤闭合信息回路，为智能体提供下一次决策所需的反馈。 8. **回合管理与终止：** 最后，环境检查终止条件（例如达到该回合的最大交易时段数 τ ）。如果回合尚未结束，环境将新生成的观测和计算出的奖励返回给多智能体强化学习算法，随后整个工作流程针对交易时段 $t + 1 \in \{1, \ldots, \tau\}$ 重复执行。

这一协调后的工作流程确保每个智能体的动作都会影响市场和电网，集体结果会被衡量并转化为共享信号（关键绩效指标），随后这些信号将用于指导智能体在下一个交易时段的决策并为其提供激励。这构成了核心反馈回路，使我们能够研究复杂去中心化系统中涌现的隐式合作。

4 实验设置

为评估所提出的LEM 仿真框架的能力并展示其效用，设计了一项计算实验。本节详述实验场景、实现细节以及用于比较的基准模型。尽管该框架支持广泛的实验，本文聚焦于一项全面的案例研究：

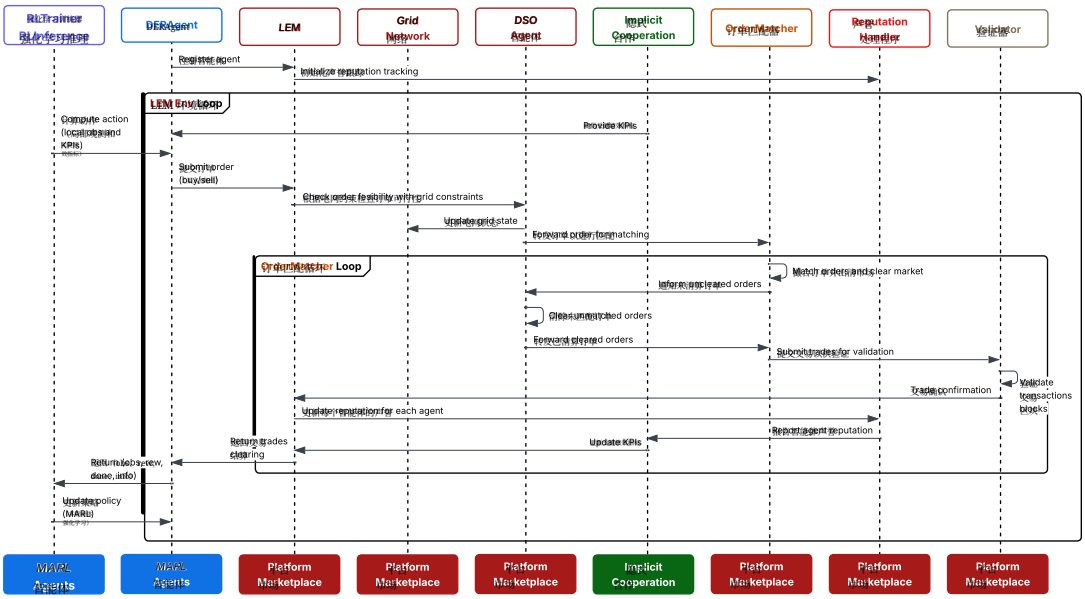


图3：MARLEM框架仿真工作流的时序图。

电池储能协调。选择本案例研究，是因为它能够展示个体智能体能力（储能）、市场动态、物理约束与隐式协调涌现之间的相互作用，从而利用该框架的全部核心模块。对已实现案例研究的完整内容感兴趣的读者（见第4.3节），或希望复现此处所述实验的读者，可找到详细配置及相应实现，见项目专用的GitHub代码库（https://github.com/salazarna/marlem）。

4.1 案例研究：电池储能协调

本案例研究展示并分析了电池储能系统在去中心化本地能源市场中的作用，并利用该框架的集成模型。

主要目标是分析不同的电池配置和部署策略如何影响智能体（多智能体强化学习或其他方式）之间隐性协调的出现，如何影响整体市场效率（例如价格波动、社会福利），以及如何通过增强灵活性促进电网稳定性。

仿真环境配置为7个智能体，在额定容量为1200 kW、采用网状拓扑的电网中按24小时日周期运行。采用买卖价差出清机制，之所以专门选择该机制，是因为买入价格与卖出价格之间的固有差异为储能套利创造了明确的经济激励，从而凸显智能体从其电池中学习（若不学习，则加以利用）所获得的价值。市场允许较宽的价格区间（即35-280美元/兆瓦时）和数量区间（即0.1-200千瓦时），以适应储能驱动的策略。

本文展示了两个不同的场景（读者也可在专门的 GitHub代码库中找到六个完整场景）。每个场景中的智能体具有相同的基础发电容量（即 60 kW，但存在一定的随机变化）和相应的曲线，但电池的部署方式有所不同：

1. **无电池：** 建立一个无任何储能灵活性的基准场景，仅迫使智能体参与实时市场。所有智能体均配置为无电池。2. **策略性电池：** 探索一种部署方案：预先专门设计智能体特征（发电／需求时序）和储能能力（比例范围为 0.5–1.2，效率高达 95%），以促进社区内的协调负荷平衡和时移（例如，将早晨发电峰值的智能体与储能配对，使其能够为晚间需求峰值的智能体提供服务）。

4.2 基准模型

为了评估 为了评估案例研究中智能体所实现的性能，以下基线对于建立 g 语境：

- **零智能智能体：**这基线作为性能的下限。在每个时间步，.它们使用均匀随机分布，在市场的价格与数量限制内有效的价格界限内提交随机报价，并遵守各自的物理约束（例如，电池荷电状态、发电可用性）。将（训练后的）多智能体强化学习性能与该基线进行比较，表明学习策略并非微不足道，并能取得超越随机水平的有意义的经济与协调结果。为清晰起见并展示该框架的能力，5 节给出的结果将使用这些零智能智能体，以说明不同配置（例如，电池场景）的影响。

- **采用CTCE范式的多智能体强化学习：**如第3.7.1节所述，该范式在所用多智能体强化学习算法的语境中充当理论上界。单一策略在可访问完整全局状态的条件下进行训练，并使用相同的底层多智能体强化学习算法（例如，近端策略优化、异步近端策略优化或软演员-评论家）同时控制所有智能体，代表该算法在完全信息与完全协调条件下所能达到的最佳性能。将去中心化执行结果（集中训练、分散执行或DTDE）与CTCE基准进行比较，可以量化归因于去中心化执行所固有的部分可观测性与独立决策的性能下降。虽然完整的多智能体强化学习训练留待未来工作（见第6节），该基准对于情境化潜在性能仍然具有相关性。

4.3 其他已实现的案例研究

除了侧重于电池储能协调外，该框架的多功能性还通过专用 GitHub代码库中提供的其他完整实现的案例研究得到了进一步展示。这些案例包括：

- **市场机制比较：** 比较六种不同的定价/出清规则，以评估其对效率和智能体策略的影响。

- **智能体异质性与市场势力：** 探讨在不同智能体规模集中度（均衡、垄断、寡头垄断）下的市场动态。

- **DSO干预策略：** 评估不同的DSO监管立场（宽松、适度、严格）对市场参与和电网稳定性的影响。

- **电网拓扑结构与拥塞效应：** 考察网络结构和容量限制对交易模式与协调的影响。

这些补充研究虽未在此详述，但凸显了该框架作为本地能源市场研究综合工具的能力。

5 结果与讨论

本节展示所提出仿真框架的分析能力和技术经济敏感性。如第 4 节所述，此处呈现的研究结果采用零智能代理基线生成。通过比较无电池 和策略性电池 场景，并使用在随机但有效的启发式方法下运行的非学习型代理，我们可以展示该框架的高保真建模能力。这使我们能够观察系统的物理机制（市场规则、电网约束和代理资产配置）如何塑造结果，而不受学习得到的多智能体强化学习策略影响。完整的多智能体强化学习训练结果构成本研究下一阶段的工作，详见第 6 节。

5.1 框架能力演示

该框架的核心能力在于量化不同物理资产配置所产生的系统性影响。通过仿真并对比这两种情景，我们可以观察到，即使智能体的主体行为并非策略性的，储能的存在及其策略性部署也会从根本上改变市场结果。

5.1.1 价格波动与市场稳定性

电池的一项主要功能是进行时间套利（低价买入、高价卖出），从而稳定价格。我们的框架能够捕捉这种经济物理交互。图 4对案例研究中两种情景下24小时仿真期间的点对点市场出清价格统计分布进行了比较。

无电池场景（见图 4a）作为基线，呈现出一个波动市场。价格分布呈多峰形态，至少存在两个明显的价格簇。这是分裂市场的统计证据，在这种市场中，智能体完全受发电量与需求之间的时间错配影响：发电量导致价格较低（例如，接近 140 美元/MWh 的价格簇），而需求导致价格较高（例如，接近 230 美元/MWh 的价格簇）。173.6 美元的平均值偏离了 157.5 美元的中位数，进一步表明该市场存在偏斜且不稳定的特征。

策略性电池场景（见图 4b）展示了市场结构的转变。价格分布变为单峰，并处于近似正态分布的稳定价格区间内。173.6 美元的平均值与 158.5 美元的中位数现在几乎相同，这一事实从定量角度证实了这一点；这正是对称、稳定且可预测的市场的特征。

这一结果表明，即使智能体不具备智能，仅仅配置策略性电池，也能发挥稳定资产的作用。电池提供了时间灵活性，使市场能够弥合高发电时段与高需求时段之间的差距。这种套利有效地削减了价格极值，使市场得以收敛至稳定的出清价格。这一发现验证了该框架的高保真技术经济模型，证明其能够捕捉物理资产对市场动态产生的涌现性系统效应。

5.1.2 电网稳定性与系统协调

该框架的隐式合作模块旨在通过测量供需不平衡，并计算式（23）中定义的协调得分，量化系统稳定性。为说明这一点，图 5 和图 6 分别展示了24小时仿真周期内的电网净失衡和点对点交易比例。

在无电池情景下，电网不平衡曲线波动剧烈，反映了发电量与需求之间的时间错配。这迫使系统持续、大量依赖配电系统运营商充当能量的吸收方或供给方，导致协调核心得分较低。

“策略性电池 ”场景呈现出更平滑的净负荷曲线，整体协调得分也更高。智能体的电池即使由零智能策略控制，也能发挥被动阻尼作用，吸收系统性扰动。“策略性电池 ”场景的设计（即，让早间发电并配备储能的智能体，与晚间有需求并配备储能的智能体相配合）形成了自然的全系统互补性。

5.1.3 市场结构与经济效率

除了标量指标之外，该框架的可视化模块还揭示了市场结构。图7 展示了储能如何改变市场的社会和空间特征。

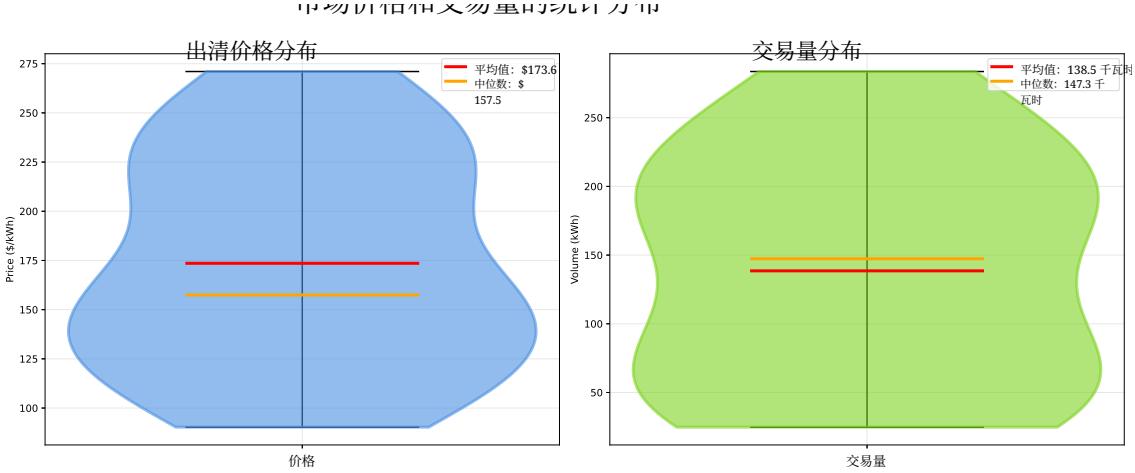
图7 展示了点对点交易网络，其中节点代表智能体，边代表它们之间的交易量。无电池场景（见图7a）呈现了这样一种市场：大多数智能体（蓝色节点）处于净需求状态，被迫从配电系统运营商（主要卖方）处购买全部能源。这表明该市场缺乏流动性，无法自给自足，点对点交易量也很低。

战略性电池场景（见图 7b）表明，智能体大多已分化为明确的角色（即，上午/下午发电的智能体是净卖方，晚间有需求的智能体是净买方），并且彼此之间的交易更加频繁。这表明，电池带来的时间灵活性促成了市场的形成。此外，配电系统运营商的角色也发生了逆转：在无电池场景中，它是主要卖方，而现在则成为主要买方。这一结果表明，本地能源市场已从依赖外部电网的能源短缺状态，转变为自给自足状态，并且如今已成为其盈余能源的净出口方。这一验证证明了该框架能够捕捉市场结构和经济流的系统性变化。

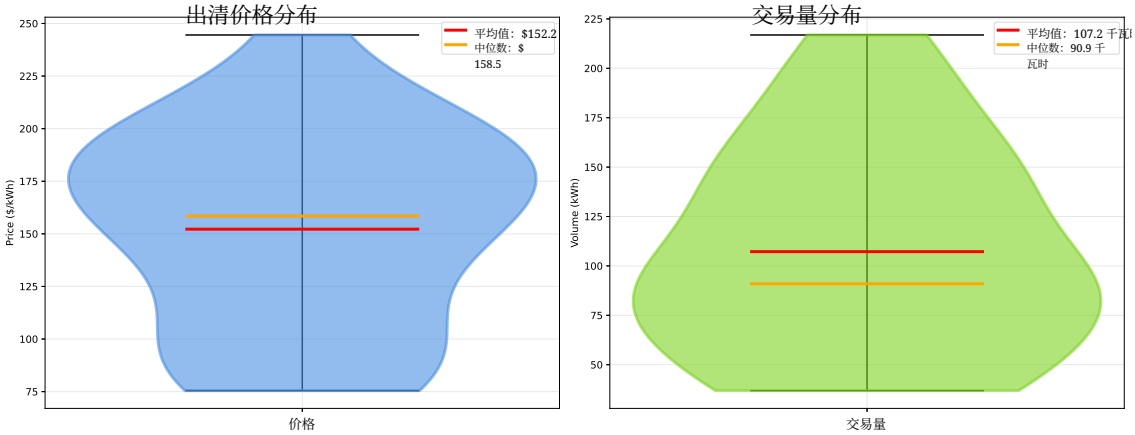
5.2 讨论

零智能体仿真的结果验证了该框架的设计，表明其底层技术经济模型具有足够高的保真度，能够捕捉物理资产配置对市场和电网性能的影响。表 2 给出了对比无电池和策略性电池场景的定量汇总，重点关注智能体总奖励，这些奖励可作为整体市场可行性和智能体盈利能力的代理指标。

策略性电池场景的平均总奖励为正，达到 +42929.5，与基线的巨大负奖励 -61556.1 相比，结果发生了逆转。这表明，策略性部署的储能资源具有经济和系统价值，



(a) 无电池场景。



(b) 策略性电池场景。

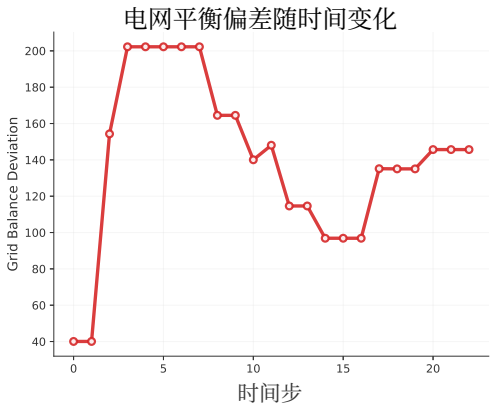
Fig图4：案例研究中市场出清价格与交易量的统计分布 .

即使以非策略方式运行，也能通过实现时间套利为市场生态系统创造价值。这项验证为量化不同系统设计和物理资产的价值提供了试验平台。

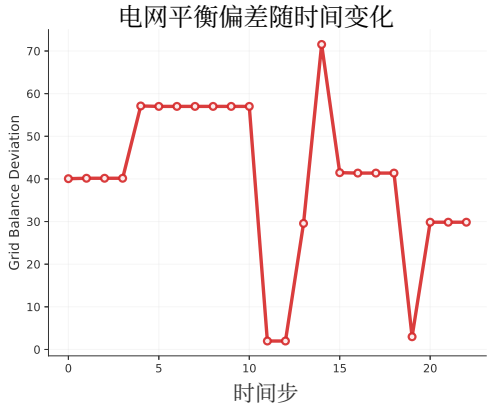
该框架已被证明对协调所需的物理前提条件（即储能）较为敏感；例如，策略性电池场景创造了清晰且互补的角色，非常适合未来的多智能体强化学习智能体学习并承担。尽管本研究保持市场机制不变，储能对价格波动的影响直接证明了该框架分析市场机制与物理资产之间耦合关系的能力。此外，本案例研究直接探讨了物理电网约束的影响。电池参数（例如荷电状态或效率）是一种时间物理约束，其存在会影响主体行为和市场结果，从而验证了该框架检验市场与电网二分的能力。

尽管所展示的7智能体研究验证了该框架的功能能力，正式的定量可扩展性分析仍是未来工作的关键组成部分。这将涉及执行智能体数量不断增加的仿真（例如10、50、 100+ 个智能体），以绘制计算时间并分析智能体数量对市场关键绩效指标的影响，从而测试该框架在负载条件下的性能。

我们必须再次强调，本文所呈现的结果仅具有示范性，并采用零智能代理基线来验证仿真框架的能力。要理解涌现行为动态以及隐式合作的全部潜力，还需要进行完整的多智能体强化学习训练（作者将在后续论文中发表这项工作）。

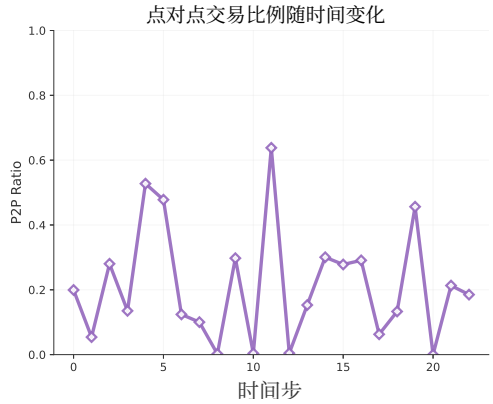


（a）无电池场景。

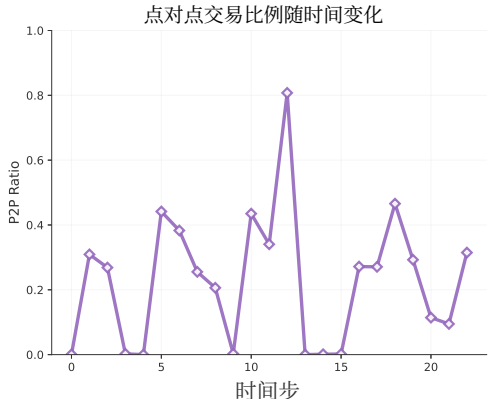


（b）策略性电池场景。

图 5：案例研究中的电网平衡偏差随时间变化。



(a) 无电池场景。



(b) 策略性电池场景。

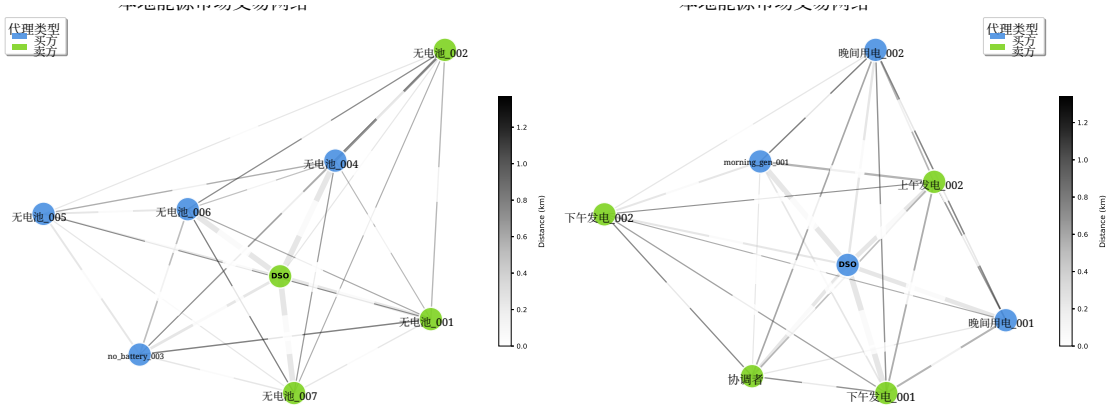
图 6：案例研究中点对点交易比例随时间的变化。

最后，与所有仿真一样，组件模型都是抽象模型，需要持续依据现实世界数据进行验证和校准，以提高其保真度。

6 结论与未来工作

本文介绍了一种新颖的开源仿真框架，专为使用多智能体强化学习研究去中心化本地能源市场而设计。针对当前能源仿真工具领域的不足，特别是缺乏一种能够在标准化、去中心化范式下集成真实市场机制、物理电网约束和先进多智能体强化学习能力的统一平台，我们的框架提供了一个稳健、可扩展且兼容 Gymnasium 的环境，可用于研究未来能源系统的复杂动态。

主要贡献在于 LEM 仿真框架本身。该框架采用去中心化部分可观测马尔可夫决策过程（Dec-POMDP）进行架构设计和建模，通过将模块化市场平台（采用混合的基于偏好与价格的匹配机制和信誉系统）与基于物理原理的电网模型（纳入输电损耗和拥塞管理）有机结合，克服了普遍存在的市场与电网二分问题。该框架旨在支持并促进对完全去中心化学习与执行范式的研究，突破以往研究中常见的集中式训练假设所带来的局限。



（a）无电池场景。　　（b）策略性电池场景。

图 7：案例研究中策略性电池场景的本地能源市场交易网络。

表 2：案例研究中无电池 和 策略性电池 场景的性能比较。

指标	无电池	策略性电池
平均奖励	-61556.1	+42929.5
最终奖励	-30827.1	+78202.6
电池比例	0.00	0.74
平均效率	-	95%
荷电状态范围	-	10-90%

第二项重要贡献是明确聚焦于推动对隐式合作的研究并促进其实现。我们通过嵌入框架的一种新颖方法实现这一目标：在智能体的部分观测中加入系统级关键绩效指标，并将这些指标纳入多目标奖励函数。这由此形成独特的反馈回路，展示了一条可行路径，使自主的自利型智能体无需明确的通信或集中式控制，也能朝着有利于集体的协同行为发展。

本研究为依托所开发框架开展的未来研究奠定了基础。接下来的首要工作是全面验证该框架所支持的隐式协作机制。这需要利用框架的全部功能开展大规模多智能体强化学习训练实验。具体而言，我们将采用已实现的最新多智能体强化学习算法（即近端策略优化、异步近端策略优化和软演员-评论家），在各种训练与推理范式下训练智能体：将集中式训练与集中式执行作为基准，采用实用的集中训练、分散执行方法，以及最重要的、符合现实世界本地能源市场原则的完全去中心化 DTDE 范式。这些实验将系统地评估不同的市场设计、电网条件和学习参数如何影响隐式协调的形成及其效果，并量化去中心化、性能与稳定性之间的权衡。

除这项核心验证外，该框架的模块化和可扩展性还为未来研究开辟了若干有前景的方向：

- **集成更复杂的分布式能源资源模型：**通过纳入更精细的模型来提升智能体模型的保真度，例如电池退化、电动汽车充放电模式、建筑热动力学，或异质的智能体目标（如风险厌恶、公平性问题）。

- **探索先进的多智能体强化学习算法：**实现并基准测试更广泛的先进多智能体强化学习算法，包括采用通信协议（以便与隐式方法进行对比）、替代价值分解技术或基于模型的方法。

- **可扩展性分析：**通过执行规模逐渐增大的仿真，开展正式的定量可扩展性分析

智能体数量（例如，10、50、 100+ 个智能体），绘制计算时间并分析智能体数量对市场关键绩效指标的影响。

- **应用于现实世界数据：**利用来自试点本地能源市场项目或真实配电网的实证数据，对仿真框架进行验证与校准，以提升其准确性及其对政策分析的实际相关性。
- **通信约束分析：**扩展该框架，以显式建模现实通信网络限制（延迟、带宽）以及潜在威胁（例如虚假数据注入）对去中心化市场运行的鲁棒性与安全的影响。

我们相信，这一LEM仿真框架将成为研究界、政策制定者以及行业从业者的有用工具，助力其设计、分析并部署可持续未来所需的智能、韧性、高效且真正去中心化的能源系统。

参考文献

[1] C. Feng, A. L. Liu，《太阳能与储能点对点能源交易：一种网络化多智能体强化学习方法》（2024年10月）。arXiv:2401.13947，doi:10.48550/arXiv.2401.13947。[2] A. Weidlich, T. Künzel, F. Klumpp，《电力市场仿真模型中灵活与非灵活发电的竞价策略：模型描述与研究发现》，载于：第九届未来能源系统国际会议论文集，ACM，德国卡尔斯鲁厄，2018年，第532–537页。doi:10.1145/3208903.3214347。[3] 加州独立系统运营商、太平洋煤气电力公司、南加州爱迪生电力公司、圣迭戈燃气电力公司，《高分布式能源资源电网中的输电与配电运行协调》，白皮书，Gridworks，美国加利福尼亚州，由CAISO、PG&E、SCE、SDG&E工作人员编写，并得到Gridworks支持（2017年6月）。URL https://gridworks.org/wp-content/uploads/2017/01/Gridworks_CoordinationTransmission.pdf[4] D. Qiu, J. Wang, J. Wang, G. Strbac，《面向双边拍卖市场自动化点对点能源交易的多智能体强化学习》，载于：第三十届人工智能国际联合会议论文集，人工智能国际联合会议组织，加拿大蒙特利尔，2021年，第2913–2920页。doi: 10.24963/ijcai.2021/401。[5] N. Harder, K. K. Miskiw, M. Khanra, F. Maurer, P. Patil, R. Qussous, C. Weinhardt, M. Klobasa, M. Ragwitz, A. Wei- dlich，《ASSUME：用于结合强化学习探索电力市场动态的基于智能体的仿真框架》，SoftwareX 30（2025）102176。doi:10.1016/j.softx.2025.102176。[6] D. Qiu, J. Wang, Z. Dong, Y. Wang, G. Strbac，《用于点对点多能源交易的平均场多智能体强化学习》，IEEE Transactions on Power Systems 38（5）（2023）4853–4866。doi:10.1109/TPWRS.2022. 3217922。[7] R. Henry, D. Ernst，《Gym-ANM：用于电力配电系统有源网络管理任务的强化学习环境》，Energy and AI 5（2021）100092。doi:10.1016/j.egyai.2021.100092。[8] F. Charbonnier, B. Peng, T. Morstyn, M. McCulloch，《集中式演练去中心化协作：面向住宅能源灵活性可扩展协调的多智能体强化学习》（2023年6月）。arXiv:2305.18875，doi:10.48550/arXiv.2305.18875。[9] R. M. Czekster，《智能电网建模与仿真工具》（2020年11月）。arXiv:2011.07968，doi:10.48550/arXiv:2011.07968。[10] D. P. Chassin, K. Schneider, C. Gerkensmeyer，《GridLAB-D：开源电力系统建模与仿真环境》，载于：2008年IEEE/PES输电与配电会议暨展览会，IEEE，美国伊利诺伊州芝加哥，2008年，第1–5页。doi:10.1109/TDC.2008.4517260。[11] R. D. Zimmerman, C. E. Murillo-Sanchez, R. J. Thomas，《MATPOWER：用于电力系统研究与教育的稳态运行、规划与分析工具》，IEEE Transactions on Power Systems 26（1）（2011）12–19。doi:10.1109/TPWRS.2010.2051168。[12] M. Zade, S. D. Lumpp, P. Tzscheutschler, U. Wagner，《满足社区本地能源市场中的用户偏好——基于拍卖的出清方法》，Applied Energy 306（2022）118004。doi:10.1016/j.apenergy.2021. 118004。

[13] N. T. Mbungu, R. M. Naidoo, R. C. Bansal, V. Vahidinasab, 《微电网应用的最优智能能源协调概述》, IEEE Access 7（2019年）163063–163084。doi:10.1109/ACCESS.2019.2951459。[14] J. Van Tilburg, L. C. Siebert, J. L. Cremer, 《MARL-iDR：面向基于激励的住宅需求响应的多智能体强化学习》, 载于：2023年IEEE贝尔格莱德PowerTech会议，IEEE，贝尔格莱德，塞尔维亚，2023年，第1–8页。doi:10.1109/PowerTech55446.2023.10202941。[15] K. Nweye, K. Kaspar, G. Buscemi, T. Fonseca, G. Pinto, D. Ghose, S. Duddukuru, P. Pratapa, H. Li, J. Mohammadi, L. Lino Ferreira, T. Hong, M. Ouf, A. Capozzoli, Z. Nagy, 《CityLearn v2：电网交互型社区的能源灵活、韧性、以居住者为中心且碳感知的管理》, 《建筑性能模拟期刊》18（1）（2025年）17–38。doi:10.1080/19401493.2024.2418813。[16] G. Wen, J. Fu, P. Dai, J. Zhou, 《DTDE：一种新型协作式多智能体强化学习框架》, 《创新》2（4）（2021年）100162。doi:10.1016/j.xinn.2021.100162。[17] A. D. Domınguez-Garcıa, C. N. Hadjicostis, 《分布式能源资源辅助服务供给协调：架构与算法》。[18] D. Li, N. Lou, Z. Xu, B. Zhang, G. Fan, 《通过隐式共识生成实现多智能体强化学习中的高效通信》。[19] S. Li, J. K. Gupta, P. Morales, R. Allen, M. J. Kochenderfer, 《面向多智能体强化学习的深度隐式协调图》（2021年）。[20] J. R. Vázquez-Canteli, Z. Nagy, S. Dey, G. Henze, 《CityLearn：规范化需求响应与城市能源管理领域的多智能体强化学习研究》。[21] S. Keren, C. Essayeh, S. V. Albrecht, T. Mortsyn, 《能源网络中的多智能体强化学习：计算挑战、进展与开放问题》。[22] D. Biagioni, X. Zhang, D. Wald, D. Vaidhynathan, R. Chintala, J. King, A. S. Zamzam, 《PowerGridworld：用于电力系统多智能体强化学习的框架》, 载于：第十三届ACM未来能源系统国际会议论文集，ACM，线上会议，2022年，第565–570页。doi:10.1145/3538637.3539616。[23] C. Zhu, M. Dastani, S. Wang, 《带通信的多智能体深度强化学习综述》, 《自主智能体与多智能体系统》38（1）（2024年）4。doi:10.1007/s10458-023-09633-6。[24] A. G. Givisiez, K. Petrou, L. F. Ochoa, 《输电系统运营商—配电系统运营商协调模型及求解技术综述》, Electric Power Systems Research 189（2020年）106659。doi:10.1016/j.epsr.2020.106659。[25] T. Alazemi, M. Darwish, M. Radi, 《可再生能源整合的输电系统运营商/配电系统运营商协调：系统性文献综述》, 《能源》15（19）（2022年）7312。doi:10.3390/en15197312。[26] M. Klein, U. J. Frey, M. Reeg, 《模型中的模型——能源系统分析中的基于智能体的建模与仿真》, 《人工社会与社会模拟期刊》22（4）（2019年）6。doi:10.18564/jasss.4129。[27] S. Trimarchi, F. Casamatta, L. Gamba, F. Grimaccia, M. Lorenzo, A. Niccolai, 《能源商品市场基于智能体的模型及其与强化学习模型自然融合的综述》, 《能源》18（12）（2025年）3171。doi:10.3390/en18123171。[28] A. Akhatova, L. Kranzl, F. Schipfer, C. B. Heendeniya, 《城市区域能源系统脱碳的基于智能体建模——系统性文献综述》, 《能源》15（2）（2022年）554。doi:10.3390/en15020554。[29] S. Li, 《结构化协作式多智能体协调》, 博士论文，斯坦福大学，加利福尼亚州斯坦福，航空航天工程系（2022年5月）。URL https://purl.stanford.edu/cr844qy0850[30] J. R. Vázquez-Canteli, Z. Nagy, 《面向需求响应的强化学习：算法与建模技术综述》, Applied Energy 235（2019年）1072–1089。doi:10.1016/j.apenergy.2018.11.002。[31] R. May, P. Huang, 《用于研究和优化点对点产消者能源市场的多智能体强化学习方法》, Applied Energy 334（2023年）120705。doi:10.1016/j.apenergy.2023.120705。[32] C. Sabillon, A. A. Mohamed, A. Golriz, B. Venkatesh, 《综合配电交易型能源市场平台》, 《IET发电、输电与配电》15（16）（2021年）2344–2355。doi:10.1049/gtd2.12182。[33] L. Qiu, 《面向城市社区协调式智能电网与建筑能源管理的多智能体强化学习》。

[34] S. Oh, J. Jung, A. Onen, C.-H. Lee, 《从聚合商视角设计的基于强化学习的需求响应策略》, Frontiers in Energy Research 10（2022年）957466。doi:10.3389/fenrg.2022.957466。[35] H. Markgraf, M. Althoff, 《面向价格型需求响应的安全多智能体强化学习》, 载于：2023 IEEE PES 欧洲创新智能电网技术大会（ISGT EUROPE）, IEEE, 格勒诺布尔, 法国, 2023年, 页码 1–6。doi:10.1109/ISGTEUROPE56780.2023.10407281。[36] J. Guerrero, A. Chapman, G. Verbic, 《低压网络中受网络约束的去中心化点对点能源交易》（2018年9月）。arXiv:1809.06976, doi:10.48550/arXiv.1809.06976。[37] C. Essayeh, T. Morstyn, 《OPLEM：本地能源市场开放平台》, Applied Energy 373（2024年）123848。doi:10.1016/j.apenergy.2024.123848。[38] L. S. Melo, F. L. Tofoli, D. Issicaba, M. E. P. Monteiro, G. C. Barroso, R. F. Sampaio, R. P. S. Leão, 《用于评估交易型能源系统的联合仿真平台》, Electric Power Systems Research 223（2023年）109693。doi:10.1016/j.epsr.2023.109693。[39] L. Thurner, A. Scheidler, F. Schäfer, J.-H. Menke, J. Dollichon, F. Meier, S. Meinecke, M. Braun, 《Pandapower——一种用于便捷建模、分析和优化电力系统的开源 Python 工具》, IEEE Transactions on Power Systems 33（6）（2018年）6510–6521。arXiv:1709.06743, doi:10.1109/TPWRS.2018.2829021。[40] K. Wehrle, M. Güneş, J. Gross（编）, 《网络仿真的建模与工具》, Springer Berlin Heidelberg, 柏林, 海德堡, 2010年。doi:10.1007/978-3-642-12331-3。[41] C. Steinbrink, M. Blank-Babazadeh, A. El-Ama, S. Holly, B. Lüers, M. Nebel-Wenner, R. P. Ramírez Acosta, T. Raub, J. S. Schwarz, S. Stark, A. Nieße, S. Lehnhoff, 《使用 mosaik 进行 CPES 测试：协同仿真的规划、执行与分析》, Applied Sciences 9（5）（2019年）923。doi:10.3390/app9050923。[42] S. Hou, S. Gao, W. Xia, E. M. Salazar Duque, P. Palensky, P. P. Vergara, 《RL-ADN：用于主动配电网中储能系统优化调度的高性能深度强化学习环境》, Energy and AI 19（2025年）100457。doi:10.1016/j.egyai.2024.100457。[43] A. Green, energy-py：面向能源系统的强化学习, https://github.com/ADGEfficiency/energy-py, 访问日期：2024年（2024年）。[44] A. Marot, B. Donnot, G. Dulac-Arnold, A. Kelly, A. O'Sullivan, J. Viebahn, M. Awad, I. Guyon, P. Panciatici, C. Romero, 《学习运行电力网络挑战赛：回顾性分析》（2021年10月）。arXiv:2103.03104, doi:10.48550/arXiv.2103.03104。[45] M. Samvelyan, T. Rashid, C. S. de Witt, G. Farquhar, N. Nardelli, T. G. J. Rudner, C.-M. Hung, P. H. S. Torr, J. Foerster, S. Whiteson, 《星际争霸多智能体挑战赛》（2019年12月）。arXiv:1902.04043, doi:10.48550/arXiv.1902.04043。[46] G. Papoudakis, F. Christianos, L. Schäfer, S. V. Albrecht, 《合作任务中多智能体深度强化学习算法的基准测试》, 载于：神经信息处理系统会议数据集与基准专题论文集（NeurIPS）, 2021年。网址：http://arxiv.org/abs/2006.07869[47] S. Hu, Y. Zhong, M. Gao, W. Wang, H. Dong, X. Liang, Z. Li, X. Chang, Y. Yang, 《Marllib：可扩展且高效的多智能体强化学习库》, Journal of Machine Learning Research（2023年）。[48] C. Steinbrink, A. A. Van Der Meer, M. Cvetkovic, D. Babazadeh, S. Rohjans, P. Palensky, S. Lehnhoff, 《使用 MOSAIK 和 HLA 进行智能电网联合仿真：比较研究》, Computer Science - Research and Development 33（1-2）（2018年）135–143。doi:10.1007/s00450-017-0379-y。[49] Z. Tian, S. Zou, I. Davies, T. Warr, L. Wu, H. B. Ammar, J. Wang, 《通过行动学习隐式通信》, AAAI 人工智能大会论文集 34（05）（2020年）7261–7268。doi:10.1609/aaai.v34i05.6217。[50] T. Fonseca, L. L. Ferreira, B. Cabral, R. Severino, K. Nweye, D. Ghose, Z. Nagy, 《EVLearn：扩展 cityLearn 框架以支持电动汽车仿真》, Energy Informatics 8（1）（2025年）16。doi:10.1186/s42162-024-00445-w。[51] E. Liang, R. Liaw, P. Moritz, R. Nishihara, R. Fox, K. Goldberg, J. E. Gonzalez, M. I. Jordan, I. Stoica, 《RLlib：分布式强化学习的抽象》（2018年6月）。arXiv:1712.09381, doi:10.48550/arXiv.1712.09381。[52] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, 《近端策略优化算法》（2017年8月）。arXiv:1707

.06347, doi:10 ·48550/arXiv .1707.06347.

[53] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, 《软演员-评论家：使用随机演员进行离策略最大熵深度强化学习》（2018年8月）。arXiv:1801.01290, doi:10.48550/arXiv.1801.01290。[54] M. Jaderberg, V. Dalibard, S. Osindero, W. M. Czarnecki, J. Donahue, A. Razavi, O. Vinyals, T. Green, I. Dunning, K. Simonyan, C. Fernando, K. Kavukcuoglu, 《神经网络的基于种群的训练》（2017年11月）。arXiv: 1711.09846, doi:10.48550/arXiv.1711.09846。