Learn to Coordinate City-Scale Virtual Power Plants: A Reasoning-Guided Evolutionary Framework for Hierarchical Heterogeneous Device Scheduling

Jinwei Zeng
Department of Electronic Engineering
BNRist, Tsinghua University
Beijing, China

Guozhen Zhang
TsingRoc.ai
Beijing, China

Minbo Ma
Hongyuan Su
Department of Electronic Engineering
BNRist, Tsinghua University
Beijing, China

Yu Zheng
Massachusetts Institute of Technology
Cambridge, MA, USA

Jian Yuan
Department of Electronic Engineering
BNRist, Tsinghua University
Beijing, China

Yong Li
Department of Electronic Engineering
BNRist, Tsinghua University
Beijing, China

Abstract

As intermittent renewables increasingly penetrate power systems, virtual power plants (VPPs) have emerged as a critical component of power systems, aggregating geographically distributed devices to respond to price signals and mitigate supply-demand imbalances. Consequently, coordinating heterogeneous resources across hierarchical multi-region pricing to fulfill committed bids while maximizing arbitrage has become a critical challenge. Existing operations-research and reinforcement-learning approaches rely on handcrafted formulations or learned policies that struggle to scale or lack interpretability, limiting trustworthiness in safety-critical energy systems. In contrast, Large language models enable evolving optimization by reasoning over structured decisions and generating executable programs, making the process more interpretable and effective. Building on this paradigm, we propose *VPPEvolve*, a reasoning-guided evolutionary framework for hierarchical VPP scheduling. VPPEvolve tackles three key challenges through three complementary designs: (i) an evolvable chained program representation that formalizes hierarchical VPP dynamics through executable structures for system-level bid allocation and device-level dispatch; (ii) a spatio-temporal profiling and reflection module that bridges the semantic gap between volatile numerical signals and structured reasoning space; and (iii) a device attribution-aware inspiration module that enables LLM-informed evolution by explicitly attributing device-level contributions to improve heterogeneous coordination. Extensive experiments on two city-scale datasets confirm consistent gains in economic profit and bid-tracking stability, with post-hoc analyses and deployment confirming a white-box paradigm that reduces grid-side stress. Codes and data are available at: https://github.com/JinweiZzz/VPPEvolve.

Keywords

Virtual Power Plant Scheduling, Large Language Models, Evolutionary Algorithms

ACM Reference Format:
Jinwei Zeng, Guozhen Zhang, Minbo Ma, Hongyuan Su, Yu Zheng, Jian Yuan, and Yong Li. 2026. Learn to Coordinate City-Scale Virtual Power Plants: A Reasoning-Guided Evolutionary Framework for Hierarchical Heterogeneous Device Scheduling. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26), August 09–13, 2026, Jeju Island, Republic of Korea.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3770855.3818440

1 Introduction

As renewable energy resources with inherent intermittency and temporal variability continue to penetrate the power grid, the spatio-temporal mismatch between power generation and load demand has become increasingly pronounced [26, 35]. Virtual power plants (VPPs), referred to as the aggregation of geographically distributed renewable resources, flexible storage units, and controllable load devices, have therefore emerged as a critical component of current power systems [13, 22]. By leveraging their flexible and coordinated dispatch capabilities, VPPs can effectively mitigate supply-demand imbalances, relieve grid burden, and enhance overall grid stability [5, 9, 18].

In this context, effective coordination and dispatch of VPP devices are critical for incentivizing operator participation and fully realizing their potential in alleviating grid-side stress. **From a system perspective, VPP operators coordinate distributed heterogeneous devices across multiple pricing regions to pursue economic arbitrage while satisfying pre-committed power output targets [6, 34] (Figure 1).** Early VPP scheduling largely relied on expert-driven heuristics due to the strong controllability and reliability requirements of energy systems [27]. As device scale and heterogeneity increase, data-driven optimization approaches have emerged for VPP scheduling. Traditional operations research methods, including integer programming [10–12, 23] and genetic algorithms [4], offer effective solutions but often struggle with scalability and lack transparent reasoning processes. More recent studies employ multi-agent reinforcement learning to model non-linear device interactions [14, 15, 17, 21, 28, 30]. However, most existing approaches are developed for devices within a single pricing region, limiting their transferability to hierarchical settings where devices operate under different locational electricity prices. Moreover, reinforcement learning methods still suffer from limited interpretability and robustness, restricting their deployment in safety-critical energy systems [20]. Therefore, developing a framework that jointly achieves interpretability, robustness, and scalable

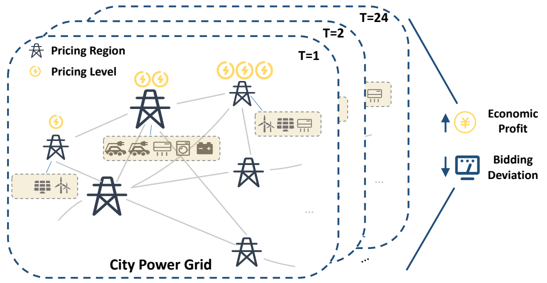

Figure 1: Illustration of hierarchical VPP scheduling task.

decision-making remains a critical yet underexplored direction for hierarchical multi-region VPP scheduling.

As VPP scheduling inherently requires deep understanding of operational constraints and heterogeneous device characteristics, while safety-targeted energy system applications demand interpretable decision-making, leveraging the domain knowledge and reasoning capabilities of large language models (LLMs) emerges as a promising direction. Recent advances in LLM-driven evolutionary frameworks, which translate optimization problems into executable decision logic that can be iteratively refined through reasoning and evaluation feedback [8, 19, 25], further demonstrate the feasibility of learning scalable and interpretable scheduling strategies. However, building an evolutionary framework for hierarchical heterogeneous VPP scheduling presents three key challenges. First, the problem involves hierarchical coordination across multiple regions with diverse devices, which is difficult to express as a code-based solution representation that LLMs can effectively reason about and iteratively evolve. Second, the system exhibits strong spatio-temporal complexity, manifested through dynamic numerical signals such as electricity prices and photovoltaic/wind generation, which need to be transformed into structured representations that LLMs can reason over. Third, heterogeneous devices possess distinct operational constraints and roles, making it non-trivial to incorporate domain knowledge to reason about device interactions and achieve coordinated and complementary scheduling strategies.

To address these challenges, we propose ***VPPEvolve***, a reasoning-guided evolutionary framework for hierarchical heterogeneous VPP scheduling. To convert the complex, cross-level decision process into an executable code-based form, we reformulate the hierarchical scheduling process as a chained executable program, where a high-level allocation function distributes bids across regions and a device-level dispatch function coordinates heterogeneous resources to satisfy region-level targets. To bridge raw spatio-temporal observations with semantic reasoning, we introduce a spatio-temporal profiling and reflection module that encodes region-level and global patterns into program-level feedback, guiding adaptive evolutionary updates under dynamic conditions. We further develop a device attribution-aware inspiration module that analyzes how candidate programs exploit different device types and provides structured guidance for evolutionary crossover, enabling coordinated scheduling under heterogeneous constraints. Extensive experiments on two city-scale datasets demonstrate that VPPEvolve achieves an average 11.4% profit improvement and 27.6% reduction in bid deviation. Interpretability analyses and real-world deployment further highlight the transparency of the solutions produced by VPPEvolve and demonstrate its effectiveness in alleviating grid burden and enabling more sustainable energy system operations. Our main contributions can be summarized as follows:

- We study hierarchical virtual power plant scheduling and, to the best of our knowledge, present the first LLM-driven evolutionary framework that jointly addresses interpretability, robustness, and performance requirements in this setting.
- We propose a reasoning-guided evolutionary framework featuring an evolvable chained program abstraction, a spatio-temporal reflection mechanism, and a device attribution-aware inspiration module, enabling reasoning-driven coordination of heterogeneous devices under dynamic spatio-temporal conditions.
- We demonstrate, through extensive city-scale experiments and real-world deployment, that VPPEvolve produces interpretable and high-performing scheduling strategies, alleviating grid-side stress and mitigating spatio-temporal imbalances.

2 Preliminaries

2.1 Virtual Power Plant

With the large-scale integration of renewable energy resources such as wind and photovoltaic power, the power system is transitioning from a deterministic system dominated by controllable conventional generators to a new system characterized by significant uncertainty and temporal variability [7]. The spatiotemporal mismatch between uncertain renewable generation and urban electricity demand further increases the need for flexible scheduling in urban power grids. In this context, virtual power plants that aggregate diverse controllable resources are emerging as key participants in modern power systems [24].

A virtual power plant (VPP) refers to an operational entity that aggregates and coordinates geographically distributed, heterogeneous energy resources across multiple pricing regions, enabling their participation in electricity markets and power system operations [22]. A VPP typically consists of the following six types of components [16, 33] (can be extended in our framework):

- **Distributed Generators.** Distributed generators represent renewable energy sources whose power output is determined by environmental conditions and is only partially controllable through curtailment. **Wind power** and **photovoltaic power** (PV) are the most common ones. At each time step, operators determine the grid injection ratio to regulate how much renewable generation is delivered to the grid.
- **Energy Storage Devices.** Energy storage devices include **stationary battery storage systems** and **electric vehicles** (EVs), both of which are modeled as storage-like resources with state-of-charge (SoC) dynamics. These devices can flexibly shift energy across time by charging or discharging electricity, subject to power limits and energy capacity constraints. At each time step, available EVs and storage batteries can be scheduled to charge or discharge subject to capacity limits and current state-of-charge constraints. EVs are further associated with energy demand requirements that must be satisfied by the end of the scheduling horizon, reflecting mobility-driven operational constraints.

item **Flexible Loads.** Flexible loads represent electricity consumption that can be temporally shifted or modulated without significantly affecting user comfort [29]. In this work, we consider two common types of flexible loads: **air conditioning systems** (AC) and **washing machines**. At each time step, the operator determines the activation levels of AC units and washing machines. For AC systems, indoor temperature is constrained to remain within a predefined comfort range $[H_{\min}, H_{\max}]$. For washing machines, each washing task must be completed within a specified time window $[t_{\text{in}}, t_{\text{ter}}]$.

Under current electricity market mechanisms, the power grid guides market participants to adjust their generation and consumption through dynamic electricity price signals. The primary objective of a VPP participant is to maximize economic arbitrage while minimizing deviations from its committed power output targets (referred to as **bid**), which are determined through prior agreements between the VPP operator and the electricity market.

2.2 Problem Formulation

Based on the scenario description, we formulate the problem as a sequential decision-making process (Figure 1). At each session, the decision scenario is as follows:
Available input: Device-level parameters, including operational constraints and capacities; Electricity prices and forecasted renewable generation (photovoltaic and wind) for all regions of the next session; current states of all devices; and the cleared bid target specifying the aggregated power output requirement of the VPP.
Output: Device-level dispatch actions that determine the operational decisions of each device for the next session, subject to device constraints.
The overall objective is to maximize aggregated economic profit from all sessions while minimizing deviations from the cleared bid targets.

3 Related Works

3.1 Virtual Power Plant (VPP) Scheduling

Due to the strong heterogeneity of energy devices and their tightly coupled operational constraints, early energy scheduling systems relied on domain knowledge and expert experienceâĂŞdriven rules to construct robust and effective scheduling strategies [27]. As power systems evolved in scale and complexity, more advanced methods have been explored, which generally fall into three categories: mixed-integer optimization, meta-heuristics, and reinforcement learning. Mixed-integer optimization methods formulate VPP scheduling as a profit- and deviation-aware mathematical program, where the operational constraints of heterogeneous devices are explicitly modeled [10–12, 23]. While effective, these approaches typically produce black-box optimal schedules with limited interpretability, which limits their trustworthiness in stability-critical energy systems. Meta-heuristic methods, such as genetic algorithms [32], provide flexible model-free search but often fail to exploit the rich temporal and structural knowledge of VPP scheduling and also lack interpretability. On the other hand, reinforcement learning has shown potential in learning implicit dispatch strategies, but existing works mostly focus on simplified single-region settings where all devices follow the same pricing policy [14, 15, 17, 20, 28, 30]. With the expansion of VPP deployments, the spatial scope of controllable devices has grown significantly, often spanning multiple regions with heterogeneous locational electricity prices. Consequently, the decision space evolves into a region-device hierarchy, which increases optimization complexity and poses challenges for scalable and stable learning for reinforcement learning methods. Moreover, limited robustness and the lack of transparent decision rationales further restrict the applicability of reinforcement learning in real-world power system operations. Overall, robust and interpretable solutions that can effectively incorporate domain knowledge to enhance scheduling performance remain largely unexplored.

Table 1: Comparison of VPP scheduling options. (Know.: human knowledge; Robust.: robustness; Inter.: interpretability; Perf.: performance.)

Method	Know.	Robust.	Inter.	Perf.
Human Experts	✓	✓	✓	✗
Mixed-integer Optimization	✗	✓	✗	–
Genetic Algorithm	✗	✓	✗	–
Reinforcement Learning	✗	✗	✗	–
LLM-based Evolutionary Framework	✓	✓	✓	–

3.2 LLM-based Evolutionary Optimization

Considering the broad domain knowledge and human-like reasoning capacity of large language models, recent studies have begun to explore their use in evolutionary optimization. Among these efforts, AlphaEvolve provides a representative framework that translates candidate solutions into executable code representations and iteratively evolves programs using language models, where high-performing solutions are reused as inspirations for subsequent generations and a fitness evaluator guides selection and mutation [25]. This code-centric evolutionary paradigm has demonstrated strong effectiveness across diverse domains, including tensor decomposition [25], open-ended mathematical discovery [25], hardware circuit design [25], game strategy planning [3], geospatial model discovery [19], and the vehicle routing problem [8]. A key advantage of these LLM-driven evolutionary methods is that they naturally embed human knowledge into the optimization process and yield explicit, interpretable strategies in the form of executable programs, which often exhibit stronger robustness than purely data-driven approaches. Despite these promising advances, the application of this paradigm to energy system scheduling remains largely unexplored. Energy systems are characterized by strong spatio-temporal dynamics and significant device-level heterogeneity. How to guide LLM-based evolutionary methods to effectively capture and exploit these structural properties in multi-region virtual power plant scheduling, therefore, remains a non-trivial and critical challenge.

4 Method: *VPPEvolve*

To fully exploit the strengths of large language modelâĂŞdriven evolutionary frameworks in incorporating human knowledge and producing interpretable solutions, we design an evolutionary framework tailored to the hierarchical virtual power plant scheduling problem, termed *VPPEvolve*. Firstly, to transform the complex hierarchical VPP scheduling task into an evolvable optimization problem, we propose an **evolutionary framework** that formulates the scheduling process as an executable, code-based chained solution and orchestrates the overall evolutionary process. To further address the pronounced spatio-temporal complexity and device heterogeneity modeling challenges of the scenario, we propose a **spatio-temporal profiling and reflection module** that guides evolution by translating numerical spatio-temporal signals into structured and reasonable representations, together with a **device attribution-aware inspiration module** that explicitly models the heterogeneity of different device types through attribution-based comparison, thereby providing informed evolutionary guidance. This section presents detailed illustrations of the framework and the two modules.

4.1 Evolutionary Framework with Chained Program Representation

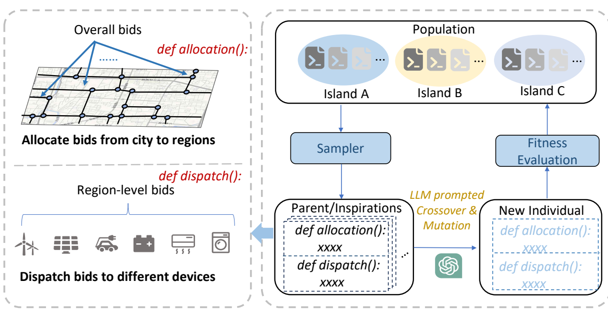

Figure 2: Illustration of VPPEvolve's evolutionary framework with chained program representation.

To leverage the natural capability of large language models to understand, generate, and modify executable code, we reformulate hierarchical virtual power plant (VPP) scheduling as a code-based solution space, enabling it to be directly optimized through language-model-driven evolution. However, city-scale hierarchical VPP scheduling involves a large number of heterogeneous devices. Constructing independent decision functions for each device type or individual unit would dramatically expand the search space, reduce evolutionary efficiency, and obscure shared coordination patterns across devices.

Meanwhile, we observe that hierarchical VPP scheduling naturally follows a multi-level decision structure: system-level power output targets must first be allocated across regions, while region-level scheduling depends on the operational states of heterogeneous devices within each region. Motivated by this structure, we represent the scheduling process as a chained decision framework (Figure 2). Specifically, the aggregated city-level power target is first distributed to regions according to locational price signals and aggregated device profiles, after which device-level dispatch decisions are generated based on real-time device states. Under this formulation, the evolutionary objects are represented as two chained executable decision functions. The inputs and outputs of these functions are defined as follows:

Program Specification: *Allocation():*

Function: Calculate an allocation score for each region, which is subsequently normalized via a softmax function to distribute the total bid across regions according to the resulting probabilities.
Inputs: Regional electricity price; expected photovoltaic and wind generation; device capacities and operating states (storage, electric vehicles, air-conditioning, and washing machines); time index; and region identifier. **Output:** Allocation score for each region.

Program Specification: *Dispatch():*

Function: Calculate device-level dispatch actions for each region given the allocated bid quantity.
Inputs: Same as the input of Allocation().
Output: Device-level dispatch actions, including renewable utilization ratios, charging or discharging ratios for storage and electric vehicles, air-conditioning load ratio, and the number of activated washing machines.
Note: Device-level dispatch actions are executed via a greedy scheduling strategy that sequentially allocates power to individual devices in order of increasing degradation cost, subject to device capacity, until the region-level bid target is satisfied.

Regarding the evolutionary paradigm, we follow the AlphaEvolve framework to iteratively evolve scheduling programs based on the formulated code template, as illustrated in the right part of Figure 2. Specifically, the process resembles a genetic evolution procedure. We maintain a population of previously generated programs along iterations, which is divided into multiple islands to encourage exploration and diversity. At each generation, a sampler selects a set of top-performing and diverse programs. An explorationâĂŞexploitation mechanism designates one program as the parent while treating others as inspiration programs. A large language model is then prompted to perform code-level evolution on the parent program, allowing it to incorporate structural or logical elements from the inspiration programs while preserving executability. The newly generated program is subsequently executed in the hierarchical VPP environment, where scheduling decisions are simulated and evaluated. Fitness is computed based on economic profit and bidding deviation, with a weighted combination serving as the overall fitness score. The evaluated program is inserted back into the population, replacing weaker individuals within its parent island, and the evolutionary process repeats. At fixed intervals, inter-island migration exchanges programs across islands to promote population

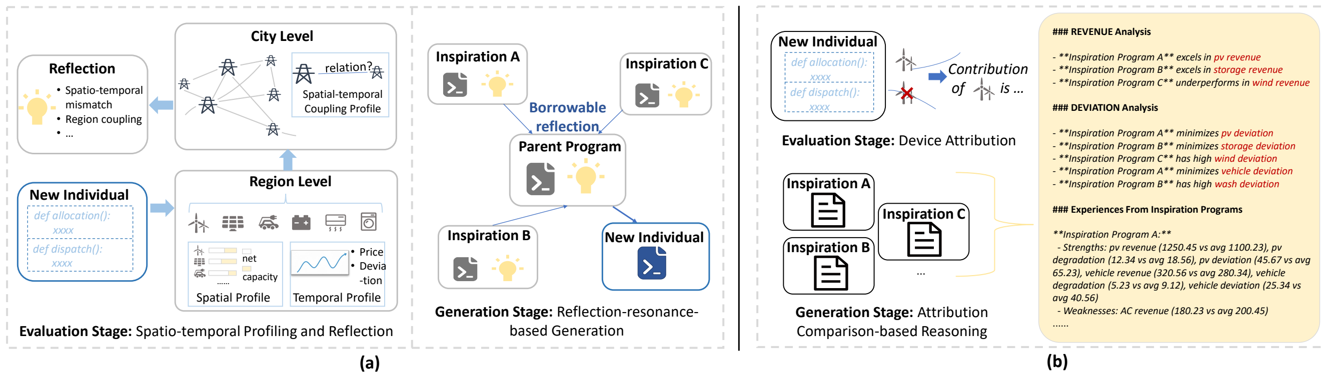

Figure 3: (a) Illustration of the spatio-temporal profiling and reflection module. (b) Illustration of the device attribution-aware inspiration module.

diversity. Through this iterative procedure, VPPEvolve progressively refines both region-level allocation logic and device-level dispatch logic, enabling the emergence of coordinated and interpretable scheduling strategies.

4.2 Spatio-temporal Profiling and Reflection

As the VPP scheduling task requires coordinating heterogeneous devices across multiple regions and time steps, capturing the underlying spatio-temporal dependencies is essential. However, these dependencies are primarily expressed as complex numerical signals, such as dynamic price fluctuations and renewable generation profiles, which large language models are not inherently well-suited to interpret directly. Without proper structuring, raw numerical inputs can limit the effectiveness of reasoning-driven evolution. To address this challenge, we explicitly profile the spatio-temporal characteristics of each region during the learning process and transform numerical observations into structured semantic representations. This enables the language model to perform expert-like reasoning over spatio-temporal mismatches and guide targeted evolutionary improvements (Figure 3**a**).

Specifically, at the evaluation stage of each candidate program, we construct a program summary template that characterizes current spatio-temporal performance from three complementary perspectives. First, **a region-level spatio-temporal profile** to capture key temporal patterns such as high-price and low-price periods, deviation peak and valley hours, as well as spatial summaries including device capacities and net energy contributions for each region. Second, **a global spatio-temporal profile** to aggregate these statistics at the system level, providing an overall view of price-deviation dynamics and the distribution of total device capacities and utilizations. Third, **a coupling analysis** is introduced to explicitly characterize structural relationships across regions and time. It distinguishes homogeneous patterns that admit shared strategies from complementary patterns where different regions or time segments can mutually compensate. The resulting spatio-temporal profiles are summarized into concise textual representations and provided to a large language model to generate reflective feedback, highlighting the strengths and weaknesses of the current strategy in terms of price exploitation, deviation tracking, and spatio-temporal coordination.

During the generation stage of a new program, we employ a reflection-resonanceâĂŞbased generation mechanism to guide the utilization of inspiration programs. Specifically, the reflection associated with the parent program serves as a diagnostic signal to identify its spatio-temporal weaknesses and coordination gaps. Based on these signals, inspiration programs that exhibit complementary behaviors along the corresponding dimensions are preferentially selected and emphasized during code evolution. Guided by this reflection-aware resonance, the language model performs program-level mutation and recombination on the parent program, selectively borrowing structures from inspirations that address the identified deficiencies. This design enables a more directed evolutionary process than implicit inspiration learning.

4.3 Device Attribution-aware Inspiration

In nowadays complex virtual power plant setting, device-level constraints and capabilities exhibit substantial heterogeneity across both regions and device types. Renewable generators, storage units, electric vehicles, and flexible loads differ significantly in capacity limits, operational flexibility, and cost structures, leading to highly diverse dispatch behaviors under the same system-level objectives. As a result, programs that achieve similar aggregate performance may rely on fundamentally different device-level strategies, while naively combining inspirations without accounting for such heterogeneity can obscure critical trade-offs and hinder effective evolution.

To explicitly account for such heterogeneity, we introduce an inspiration comparison module that enables quantitative, device-aware comparison across inspiration programs (Figure 3**b**). Specifically, during the evaluation stage, we compute term attribution for each program by aggregating the net energy contributions of different device types across regions and scenarios. Overall performance metrics, including economic profit and bidding deviation, are then proportionally attributed to each device type based on the magnitude of its aggregated net contribution. However, because different device types operate on inherently different scales and exhibit distinct operating ranges, absolute attribution values are not directly comparable. To address this issue, we adopt a relative comparison mechanism that evaluates device-level strategies by contrasting attribution patterns across inspiration programs, enabling more robust and interpretable assessments. The resulting comparative attribution summaries are incorporated into the evolutionary generation process as structured guidance for selecting and combining inspiration programs during code evolution. In this way, rather than relying solely on scalar fitness values or code-level similarity, the evolutionary process reasons about strategy differences at a finer granularity by explicitly characterizing device-level contributions to key performance metrics.

5 Experiments

5.1 Experimental Setup

5.1.1 Baselines. To validate the superiority of our method, we compare it against representative virtual power plant scheduling approaches, including integer programming-based optimization methods, meta-heuristic search methods, and reinforcement learning-based methods.

Integer programmingâĂŞbased optimization methods:

- **MILP [11].** A mixed-integer optimizationâĂŞbased scheduling approach that maximizes VPP profit by jointly optimizing device dispatch and market transactions under multiple locational marginal prices.
- **MIQCP [12].** MIQCP extends MILP by incorporating quadratic objective functions and constraints, enabling risk-aware and variance-sensitive optimization.

Meta-heuristic search methods:

- **Genetic Algorithm [4, 32].** Genetic algorithms are widely used population-based methods for black-box optimization, iteratively improving candidate solutions through selection, crossover, and mutation to explore complex, non-convex solution spaces.

Reinforcement learning-based methods:

- **VppDDPG [15].** VppDDPG models single-region virtual power plant scheduling as a sequential decision-making problem and employs reinforcement learning to learn a device-level dispatch policy that maximizes cumulative profit while satisfying device constraints. We adapt this method to our multi-region scenario by introducing an additional actor that allocates the total bid among regions, after which the original algorithm is applied to perform intra-region scheduling.
- **MAATC [14].** MAATC treats each region as an individual agent and employs a global transformer-based critic to evaluate how well multiple regions coordinate to maximize overall profit while satisfying the bidding quantity objective.
- **HierarchicalRL [2].** Given the hierarchical nature of the hierarchical VPP scheduling problem, we include a benchmark hierarchical reinforcement learning method that learns a two-level policy. Specifically, a high-level policy allocates region-level bidding targets, and a low-level policy executes device-level dispatch over multiple fine-grained time steps.

5.1.2 Metrics. Consistent with the general objective of virtual power plant scheduling [31], which is to maximize economic profit under the condition of satisfying the committed bidding quantity, we introduce two metrics for evaluation:

- **Profit.** The arbitrage profit of VPP devices, calculated as the electricity revenue from dispatched generation minus the corresponding generation cost.
- **Deviation.** The absolute difference between the actual total generation of the VPP and the committed generation amount.

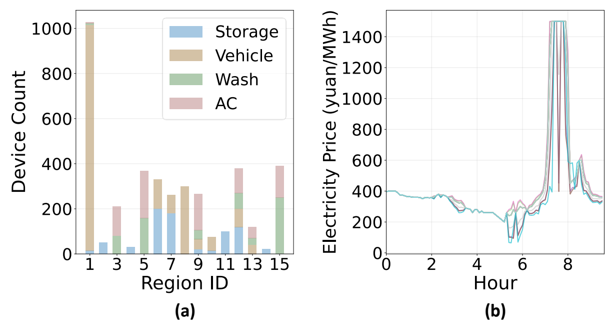

Figure 4: Dataset characteristic visualizations. (a) Device compositions of the 15 regions in the Shanxi dataset. (b) Varying electricity price curves for the 15 regions in the Shanxi dataset

5.1.3 Datasets. To comprehensively evaluate the performance of our method, we conduct experiments on two city-scale datasets: a real-world dataset collected from **Shanxi** Province and a synthetic dataset constructed for **Beijing**. Shanxi Province is among the first regions in China to implement locational marginal pricing [1]. We collect four consecutive days of operational data from a real VPP operator in Shanxi, including regional electricity prices, device configurations, and both forecasted and realized wind and photovoltaic generation. We report results averaged over the four days to mitigate the impact of daily fluctuations and ensure robust evaluation.

In addition, we construct a synthetic VPP dataset for Beijing. Based on a reconstructed power grid topology, we identify 25 transmission regions and infer region-level photovoltaic capacities and wind potentials. We further generate diverse flexible resources, including electric vehicles, energy storage systems, air conditioners, and washing machines, to represent urban-scale flexibility, while wind generation is not considered. As locational marginal pricing (LMP) has not yet been implemented in Beijing, regional electricity prices are derived from system-level price data by introducing spatial differentials according to regional load statistics. Detailed construction procedures are provided in the Appendix A. Statistical comparisons between datasets are summarized in Table 3, and Figure 4 illustrates the device composition and electricity price curves, highlighting significant spatio-temporal heterogeneity.

5.1.4 Implementation. In our main implementation, we adopt *gpt-o4-mini* as the base reasoning model. To account for forecast uncertainty, we follow a predict-execute evaluation protocol: scheduling decisions are generated using forecasted information during the

Table 2: Average performances in two datasets under planning and execution stages. Bold denotes the best results, and underlined denotes the second-best results. The unit of profit and deviation is million yuan and MWh respectively.

	Shanxi				Beijing			
	Planning Stage		Execution Stage		Planning Stage		Execution Stage	
Method	Profit↑	Deviation↓	Profit↑	Deviation↓	Profit↑	Deviation ↓	Profit↑	Deviation↓
MILP	23.43	32.13	16.48	26.83	0.589	1.087	0.799	1.273
MIQCP	15.44	29.59	13.15	29.52	0.649	1.365	0.452	1.451
Genetic Algorithm	12.64	82.94	10.15	56.27	0.415	1.891	0.372	1.928
VppDDPG	16.75	52.38	14.39	44.72	0.893	0.556	0.826	0.202
MAATC	13.05	66.44	11.72	64.39	0.528	1.472	0.492	1.528
HierarchicalRL	21.29	30.98	18.94	25.43	0.923	0.545	0.896	0.175
Ours	**28.82**	**25.54**	**21.77**	**18.19**	**0.986**	**0.484**	**0.904**	**0.075**
Impr. (%)	**23.0%**	**13.69%**	**14.9%**	**28.46%**	**6.8%**	**11.19%**	**0.9%**	**57.14%**

Table 3: Statistics of our two city-scale datasets.

	Shanxi	Beijing
Number of regions	15	25
Range	4 days	1 day
Interval	1h	1h
Number of ACs	1445	2797
Number of storage devices	900	1851
Number of AVs	2014	4097
Number of washing machines	956	1555

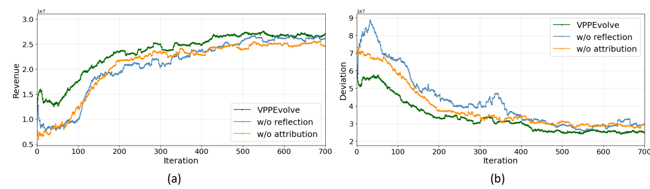

Figure 5: Training curves of ablation variants in Shanxi dataset.

planning stage and are executed and evaluated under realized renewable generation during the execution stage. The evolutionary process runs for up to 1,000 iterations with a population size of 50 and an early stopping mechanism. The initial program is a naive rule-based policy (see Appendix C). At each iteration, three top-performing programs and two diversity-promoting programs are selected as inspiration programs. During new program generation, the exploitation ratio, referring to the probability of selecting top programs as parents, is set to 0.6.

To maintain diversity, we employ an island-based migration strategy with three islands, where the top programs of each island migrate to the next island every 30 iterations. All baseline methods are tuned via grid search over their hyperparameters to ensure fair comparison. The prompt template, initial program, and best program are provided in Appendix C. The datasets, code, and baseline implementations are publicly available at https://github.com/JinweiZzz/VPPEvolve.

5.2 Performance Comparisons

Table 2 summarizes the performance of VPPEvolve and all baselines on the Shanxi and Beijing datasets. The planning stage reports expected profit and deviation under predicted wind and PV outputs, while the execution stage evaluates the actual deployment results under true generation. As shown, VPPEvolve consistently achieves the best performance across both datasets and stages, yielding an average improvement of 11.4% in profit and a reduction of 27.6% in deviation compared with the strongest baselines. These results demonstrate the robustness and effectiveness of our framework under realistic uncertainties and diverse city-scale environments.

A closer examination of the baselines reveals distinct limitations. The genetic algorithm performs relatively poorly, likely due to its inability to leverage structured knowledge, leading to inefficient exploration in highly constrained and high-dimensional scheduling spaces. Mixed-integer programming methods achieve moderate performance; however, the non-convex nature of the hierarchical VPP scheduling problem and the growing combinatorial complexity restrict their ability to effectively explore high-quality solution regions. Reinforcement learning-based approaches remain competitive due to their strong nonlinear representation capacity, yet the hierarchical and spatio-temporal decision structure introduces additional challenges for stable policy learning and convergence. In contrast, VPPEvolve integrates reflection-guided evolutionary updates with the domain knowledge embedded in large language models, enabling more efficient exploration and improved coordination among heterogeneous devices. Overall, the experimental results indicate that VPPEvolve not only incorporates human knowledge, robustness, and interpretability, but also surpasses existing optimization paradigms in performance, further validating the effectiveness and practical applicability of our framework.

5.3 Ablation Studies

To evaluate the contribution of each component, we conduct ablation studies by removing the spatio-temporal profiling and reflection module (w/o reflection) and the device attribution-aware inspiration module (w/o attribution) (Figure 5). As illustrated in the figure, removing either module consistently results in both slower evolutionary progress and inferior final performance, indicating that each component contributes jointly to evolution efficiency and solution quality. Without these mechanisms, the evolution

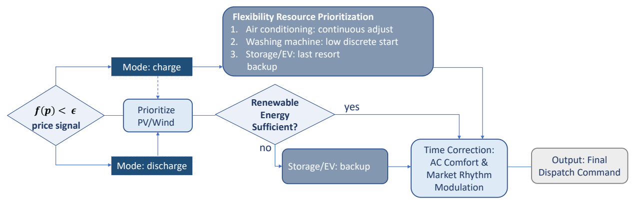

Figure 6: Interpreted decision logics of our learned best program.

process lacks structured guidance from spatio-temporal reasoning and device-level attribution, leading to less effective exploration and weaker coordination among heterogeneous resources. Overall, the full VPPEvolve framework achieves faster convergence and stronger optimization outcomes, demonstrating that the integration of structured reflection and attribution-aware reasoning is essential for improving both evolutionary efficiency and scheduling performance in hierarchical VPP scenarios.

5.4 Interpretability Analysis

A strength of LLM-based evolutionary optimization lies in its explicit code-based executable solution representation, which makes direct strategy interpretation feasible. Therefore, we extract the best program learned by VPPEvolve on the Shanxi dataset and conduct an interpretability analysis, with the decomposed decision process visualized in Figure 6. Overall, the learned strategy follows a hierarchical and price-aware decision paradigm that closely aligns with practical VPP operational principles. The program first determines an operational mode for device groups in each region based on relative electricity price signals and learned thresholds, and subsequently allocates region-level power output responsibilities.

Under charging-oriented operation, the policy establishes a flexible resource prioritization scheme that sequentially activates different device types according to their operational flexibility. Air-conditioning loads are adjusted continuously, discrete appliances such as washing machines are selectively triggered under sufficiently low prices, and storage charging is reserved as a last-resort mechanism. The controller further evaluates whether photovoltaic and wind resources should contribute additional output so that aggregated generation progressively matches charging demand. In contrast, under discharging-oriented operation, renewable resources are prioritized as the primary energy source, while backup flexibility is activated through storage systems and electric vehicles when renewable output becomes insufficient. Their participation is modulated by device states and risk-aware weighting, indicating that the learned policy treats them as adaptive balancing buffers rather than dominant energy suppliers.

Before generating the final dispatch command, the program applies a time-aware correction layer that captures diurnal demand patterns and market periodicity, suggesting that the evolutionary process internalizes temporal structures without explicit hard constraints. Overall, the learned program exhibits a coherent and interpretable structure characterized by price-driven mode selection, renewable-first coordination, risk-sensitive backup activation, and temporally modulated refinement. While the high-level decision logic remains transparent and expert-like, the underlying functional expressions are evolved nonlinear formulations discovered through efficient search, enabling improved performance beyond manually designed rules.

Figure 7: Interface of the deployed system. A demonstration video is provided in the project repository.

6 Deployment

Our algorithm has been deployed on TsingRoc.ai's virtual power plant platform, as illustrated in Figure 7 (also demonstration video in our project repository), which presents a city-scale interface under both an actual operational scenario and a VPPEvolve-enabled optimization scenario. TsingRoc.ai is a leading provider of electricity trading and virtual power plant solutions, operating real-world energy assets. The deployed system supports real-time monitoring of VPP operations across 25 pricing regions in Beijing and includes grid burden visualization to demonstrate grid-side impacts.

The upper panel shows the baseline operational state, where VPP devices are sparsely coordinated, while the lower panel illustrates the scenario after deploying VPPEvolve. The colors of the grid topology indicate the current burden level of each transmission line, reflecting the spatio-temporal imbalance intensity. With VPPEvolve, grid imbalance is significantly alleviated, achieving a 66.78% reduction in grid ramping magnitude and approximately 48 kt reduction in daily grid carbon emissions.

7 Conclusion

In this work, we present VPPEvolve, a reasoning-guided evolutionary framework for multi-regional virtual power plant scheduling that combines executable program representations with language-model-driven reasoning. By integrating hierarchical program evolution, spatio-temporal profiling, and device attribution awareness, VPPEvolve enables scalable and interpretable coordination of heterogeneous resources. Experiments on city-scale datasets show consistent improvements in economic performance and bid-tracking stability, while interpretability analyses reveal structured, expert-like decision patterns learned through nonlinear program evolution. Real-world deployment further demonstrates the practical effectiveness of VPPEvolve in alleviating grid burden under spatio-temporal variability. Overall, VPPEvolve provides a scalable and transparent paradigm for reasoning-driven optimization in complex energy systems.

References

[1] Shanxi Provincial Energy Administration. 2025. Electricity Market Rules Framework V15.0. Government technical document. https://sxb.nea.gov.cn/dtyw/jggg/202501/P020250122517294268027.pdf

[2] Pierre-Luc Bacon, Jean Harb, and Doina Precup. 2017. The option-critic architecture. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.

[3] Nikolas Belle, Dakota Barnes, Alfonso Amayuelas, Ivan Bercovich, Xin Eric Wang, and William Wang. 2025. Agents of Change: Self-Evolving LLM Agents for Strategic Planning. arXiv:2506.04651 [cs.AI]

[4] Carmen LT Borges and Djalma M Falcao. 2006. Optimal distributed generation allocation for reliability, losses, and voltage improvement. *International Journal of Electrical Power & Energy Systems* 28, 6 (2006), 413–420.

[5] Khalil Gholami, Mohammad Taufiqul Arif, Md Enamul Haque, Ali Arefi, and SM Muyeen. 2025. Comprehensive review of cutting-edge virtual power plant advancements for flexibility enhancement in future power grids. *Energy strategy reviews* 62 (2025), 101959.

[6] Adrian Gligor, Piotr Cofta, Tomasz Marciniak, and Cristian-DragoÈŹ Dumitru. 2020. Challenges for the Large-Scale Integration of Distributed Renewable Energy Resources in the Next Generation Virtual Power Plants. In *Proceedings*, Vol. 63. MDPI, 20.

[7] Jose M Gonzalez, James E Tomlinson, Eduardo A Martínez Ceseña, Mohammed Basheer, Emmanuel Obuobie, Philip T Padi, Salifu Addo, Rasheed Baisie, Mikiyas Etichia, Anthony Hurford, et al. 2023. Designing diversified renewable energy systems to balance multisector performance. *Nature Sustainability* 6, 4 (2023), 415–427.

[8] AndrÃľ Hottung, Federico Berto, Chuanbo Hua, Nayeli Gast Zepeda, Daniel Wetzel, Michael RÃűmer, Haoran Ye, Davide Zago, Michael Poli, Stefano Massaroli, Jinkyoo Park, and Kevin Tierney. 2025. VRPAgent: LLM-Driven Discovery of Heuristic Operators for Vehicle Routing Problems. arXiv:2510.07073 [cs.AI]

[9] Junhui Huang, Hui Li, and Zhaoyun Zhang. 2025. Review of Virtual Power Plant Response Capability Assessment and Optimization Dispatch. *Technologies* 13, 6 (2025), 216.

[10] Evaggelos G Kardakos, Christos K Simoglou, and Anastasios G Bakirtzis. 2015. Optimal offering strategy of a virtual power plant: A stochastic bi-level approach. *IEEE Transactions on Smart Grid* 7, 2 (2015), 794–806.

[11] Meenakshi Khandelwal, Parul Mathuria, and Rohit Bhakar. 2018. Profit based self scheduling of virtual power plant under multiple locational marginal prices. In *2018 20th National Power Systems Conference (NPSC)*. IEEE, 1–6.

[12] Meenakshi Khandelwal, Parul Mathuria, and Rohit Bhakar. 2022. Virtual Power Plant (VPP) scheduling with uncertain multiple Locational Marginal Prices. *IET Energy Systems Integration* 4, 4 (2022), 436–447.

[13] Doyob Kim. 2022. Unlocking the Potential of Distributed Energy Resources. IEA report. Typeset in France by IEA (May 2022)..

[14] Xiangyu Li, Fengji Luo, and Chaojie Li. 2024. Multi-agent deep reinforcement learning-based autonomous decision-making framework for community virtual power plants. *Applied Energy* 360 (2024), 122813.

[15] Chengyang Liu, Rebecca Jing Yang, Xinghuo Yu, Chayn Sun, Gary Rosengarten, Ariel Liebman, Ron Wakefield, Peter SP Wong, and Kaige Wang. 2023. Supporting virtual power plants decision-making in complex urban environments using reinforcement learning. *Sustainable Cities and Society* 99 (2023), 104915.

[16] Jiaqi Liu, Hongji Hu, Samson S Yu, and Hieu Trinh. 2023. Virtual power plant with renewable energy sources and energy storage systems for sustainable power grid-formation, control techniques and demand response. *Energies* 16, 9 (2023), 3705.

[17] Xiao Liu, Sinan Li, and Jianguo Zhu. 2022. Optimal coordination for multiple network-constrained VPPs via multi-agent deep reinforcement learning. *IEEE Transactions on Smart Grid* 14, 4 (2022), 3016–3031.

[18] Fengzhang Luo, Xin Yang, Wei Wei, Tianyu Zhang, Liangzhong Yao, Lingzhi Zhu, and Minhui Qian. 2020. Bi-level load peak shifting and valley filling dispatch model of distribution systems with virtual power plants. *Frontiers in Energy Research* 8 (2020), 596817.

[19] Peng Luo, Xiayin Lou, Yu Zheng, Zhuo Zheng, and Stefano Ermon. 2025. GeoEvolve: Automating Geospatial Model Discovery via Multi-Agent Large Language Models. arXiv:2509.21593 [cs.AI]

[20] Hannah Markgraf, Michael Eichelbeck, and Matthias Althoff. 2024. Empowering Safe Reinforcement Learning in Power System Control with CommonPower. In *ICLR 2024 Workshop on Tackling Climate Change with Machine Learning*.

[21] Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, Adrian Kelly, Aidan OâĂŹSullivan, Jan Viebahn, Mariette Awad, Isabelle Guyon, Patrick Panciatici, and Camilo Romero. 2021. Learning to run a power network challenge: a retrospective analysis. In *NeurIPS 2020 Competition and Demonstration Track*. PMLR, 112–132.

[22] Thomas Morstyn, Niall Farrell, Sarah J Darby, and Malcolm D McCulloch. 2018. Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants. *Nature energy* 3, 2 (2018), 94–101.

[23] Furquan Nadeem, Arup Kumar Goswami, and Prashant Kumar Tiwari. 2022. Security Constrained two Stage Scheduling of Virtual Power Plant under Distributed Locational Marginal Prices. In *2022 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE)*. IEEE, 1–6.

[24] Femke JMM Nijsse, Jean-Francois Mercure, Nadia Ameli, Francesca Larosa, Sumit Kothari, Jamie Rickman, Pim Vercoulen, and Hector Pollitt. 2023. The momentum of the solar energy transition. *Nature Communications* 14, 1 (2023), 6542.

[25] Alexander Novikov, NgÃćn VÅľ, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog. 2025. AlphaEvolve: A coding agent for scientific and algorithmic discovery. arXiv:2506.13131 [cs.AI]

[26] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. 2022. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–96.

[27] Md Sayeed Salam, Khalid M Nor, and Abdul Razak Hamdan. 2025. An experience in developing an expert system for scheduling problem in electric power system. *WIT Transactions on Information and Communication Technologies* 16 (2025).

[28] Zhonghao Sun and Tianguang Lu. 2024. Collaborative operation optimization of distribution system and virtual power plants using multi-agent deep reinforcement learning with parameter-sharing mechanism. *IET Generation, Transmission & Distribution* 18, 1 (2024), 39–49.

[29] José R Vázquez-Canteli and Zoltán Nagy. 2019. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied energy* 235 (2019), 1072–1089.

[30] Jianing Wang, Chunlin Guo, Changshu Yu, and Yanchang Liang. 2022. Virtual power plant containing electric vehicles scheduling strategies based on deep reinforcement learning. *Electric power systems research* 205 (2022), 107714.

[31] Dongliang Xiao, Hongbo Sun, Daniel Nikovski, Shoichi Kitamura, Kazuyuki Mori, and Hiroyuki Hashimoto. 2020. CVaR-constrained stochastic bidding strategy for a virtual power plant with mobile energy storages. In *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, 1171–1175.

[32] G Zames. 1981. Genetic algorithms in search, optimization and machine learning. *Inf Tech J* 3, 1 (1981), 301.

[33] Alireza Zare, Miadreza Shafie-khah, Pierluigi Siano, and George Cristian Lazaroiu. 2026. A systematic review of Virtual Power Plant configurations and their interaction with electricity, carbon, and flexibility markets. *Renewable and Sustainable Energy Reviews* 226 (2026), 116448.

[34] Yun Zhang, Jie Pan, Shuiming Jiang, Hui Zhu, Feng Chen, Quanyan Shu, and Jian Zhao. 2025. Virtual power plant load aggregation optimization scheduling strategy for port adjustable resources. *Frontiers in Energy Research* 13 (2025), 1617362.

[35] Dongsheng Zheng, Xizhe Yan, Dan Tong, Steven J Davis, Ken Caldeira, Yuanyuan Lin, Yaqin Guo, Jingyun Li, Peng Wang, Liying Ping, et al. 2025. Strategies for climate-resilient global wind and solar power systems. *Nature* (2025), 1–3.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009