ORIGINAL RESEARCH **Open access**

Multi-Agent Reinforcement Learning for Transactive Energy Management in Local Energy Markets

Fatima El Idrissi[1*], Samira Bennani[1], Youssef Benali[2]

Abstract

Local energy markets (LEMs) are increasingly recognized as mechanisms for orchestrating the generation, storage, and utilization of power from distributed resources. When integrated with automated systems, they offer a promising framework for energy management that promotes fairness and generates system-wide cost reductions. Because energy markets are inherently both cooperative and competitive, they require multi-agent automation featuring adaptive energy-trading agents. Nevertheless, the specific dynamics of how agents interact with their environment can sometimes lead to unintended market behaviors. Consequently, creating market mechanisms optimized for reinforcement learning agents necessitates careful consideration of these interactions. This paper presents the autonomous local energy exchange (ALEX), an experimental platform integrating a double auction mechanism with multi-agent learning. Instead of depending on fixed external pricing, participants establish their own internal price signals and execute energy management strategies based on market interactions. The primary contribution of this study is the investigation of how specific market features align with autonomously learning agents. Through simulation experiments, the impacts of various market characteristics are assessed to formulate an optimal market design. Findings indicate that maintaining market truthfulness preserves demand-response capabilities, whereas weak budget balancing supplies a robust reinforcement signal to guide learning agents. The subsequent agent behavior is evaluated against two benchmark models: time-of-use rates and net billing. Pricing derived from ALEX demonstrates greater responsiveness to variations in community net load than time-of-use pricing. Furthermore, by accounting for renewable energy utilization more precisely, the proposed approach decreased energy bills by a median of 38.8% relative to net billing, thereby verifying its enhanced capacity to support demand response.

Keywords Demand response, Reinforcement learning, Transactive energy, DER integration, Local energy market, Distributed energy resources (DER)

*Correspondence:
Fatima El Idrissi
fatima.idrissi@gmail.com
[1] Department of AI Applications in Smart Grids, Faculty of Engineering, Mohammed V University, Rabat, Morocco
[2] Department of Intelligent Renewable Energy Systems, Faculty of Sciences and Engineering, University of Casablanca, Casablanca, Morocco

Introduction

Demand response (DR) strategies are increasingly utilized to enhance the value of distributed energy resources (DER), including rooftop solar panels, while offsetting the drawbacks of their fluctuating availability. These DR approaches are generally categorized as either direct or indirect. Indirect DR typically relies on incentive signals—often monetary—to modify consumer habits [1-3]. Conversely, direct DR provides grid operators with immediate authority to balance the network. With the ongoing expansion of DER, traditional centralized DR strategies are expected to face significant scalability limitations [1, 4, 5]. Consequently, the growing unpredictability and intermittency of energy supply and demand

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

will require control mechanisms that are both more granular and highly resilient. Although various efforts have sought to address these issues [6-8], there is a growing consensus that decentralized alternatives require thorough investigation [9]. Within this landscape, transactive energy (TE) is emerging as a favored structural model for decentralized DR [4, 5]. According to the U.S. Department of Energy Gridwise Architecture Council, TE is characterized as "a system of economic and control mechanisms that allows the dynamic balance of supply and demand across the entire electrical infrastructure using value as a key operational parameter" [10]. When a TE system operates within a restricted geographical area, it is frequently called a local energy market (LEM). Since a universally accepted formal definition of a LEM remains absent [11, 12], this study utilizes the framework proposed by Mengelkamp *et al*. [11], which describes it as "a market platform for trading locally generated (renewable) energy among residential customers within a geographically and socially close community. Supply security is ensured through connections to a superimposed energy system".

Sustaining high levels of DR participation, particularly within a LEM setting, necessitates automated operations. Initially, rule-based systems crafted by domain experts were evaluated for this function. Today, however, learning-based techniques are favored primarily due to their enhanced scalability and robustness. Yet, many current methodologies that apply learning methods to LEMs fail to align the market structure with the underlying automation algorithm. Because most conventional LEM frameworks were developed for either human traders or rigid rule-based automation [13], this misalignment presents a notable challenge. It leads to the inefficient use of DER, as the market environment is not properly calibrated to maximize the benefits of the chosen automation technique. This issue is particularly pronounced with reinforcement learning (RL) agents, which are highly capable of rapidly discovering and exploiting vulnerabilities within cooperative–competitive multi-agent environments [14]. In their literature review, Mengelkamp *et al*. explicitly point out this gap, asserting that "a comprehensive comparison of the impacts of different trading designs (especially market mechanisms) should be carried out. Specifically, the impact of different allocation mechanisms on the market objectives and agent behavior needs to be evaluated" [11]. Subsequent research from the same authors [15] echoes this perspective, observing that the overall behavior of the system is shaped by both market and agent design. This paper contends that LEM mechanisms must be specifically customized to harness the advantages of their respective automation algorithms while concurrently suppressing any inherent vulnerabilities.

To our current understanding, prior literature has not presented explicit experimental designs intended to uncover the specific prerequisites an LEM mechanism must meet to function seamlessly with independently learning RL actors. This manuscript seeks to initiate the closure of this research gap. We concentrate specifically on two core inquiries:

• Which characteristics are essential for a LEM settlement mechanism to successfully accommodate RL-based automation?

• Can the ensuing market dynamics efficiently facilitate DR in a LEM featuring a substantial integration of DER?

To address these questions, we investigate three distinct settlement frameworks that encompass recognized standards for auction settings. By evaluating the policies formulated by the agents under each framework, we identify the optimal market structure. Subsequently, the resulting LEM exchanges are modeled as a dynamic pricing signal, and their economic performance is benchmarked against established pricing methods.

Background and Related Work

This segment delivers a comprehensive examination of pertinent literature, with a specific emphasis on research integrating RL into LEMs. For a more expansive perspective on LEM frameworks, readers are directed to the comprehensive overview authored by Mengelkamp *et al*. [11], the game-theoretic analysis conducted by Pilz *et al*. [12], and the evaluation of LEM settlement and operational mechanisms presented by Khorasany *et al.* [13].

Reinforcement learning for local energy markets

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Numerous researchers have explored the integration of RL with dynamic pricing for centralized energy management. For instance, Kim *et al*. [2] and Lu *et al*. [3] formulate RL-driven dynamic pricing strategies from the viewpoint of service retailers. Both studies tackle the challenges of forecasting how participants will react to pricing structures by reducing the need for precise customer-side data. They establish a Markov decision process grounded in models of customer preferences and behaviors. A Q-learning agent is subsequently trained to concurrently reduce consumer expenses and enhance the service provider's profitability. These two methodologies diverge in how they construct the reward function, a critical determinant in RL performance. Lu *et al*. [3] apply a weighted sum encompassing both the retailer and the customers, whereas Kim *et al*. [2] employ a simulated utility function. Even though these proposed frameworks effectively execute dynamic pricing without the need for scheduling, they inherently depend on capturing consumer preferences and behaviors through utility functions. To resolve certain limitations of these methods, Liu *et al*. [9] implement deep reinforcement learning (DRL) within a resource-sharing, consumer-centric economic model.

Zhang *et al*. [16] train an RL agent to oversee a community-shared battery system, trading its available resources within a TE market to optimize economic outcomes. The battery's economic performance serves as the reward function. The study demonstrates that favorable economic gains are attainable, even after accounting for the battery's operational expenses.

Foruzan *et al*. [17] analyze the dynamics of autonomous Q-learning agents trading energy via a LEM within a microgrid environment. Each agent optimizes its individual profit. The DERs under management encompass battery energy storage systems, rooftop solar panels, wind turbines, and diesel generators. Random models are utilized to approximate the stochastic nature of participant behavior. The researchers evaluate various microgrid setups and conduct a comprehensive hyperparameter analysis of the RL algorithm with respect to return, self-sufficiency, and fairness.

Zhou *et al.* [18] merge Q-learning with a fuzzy rule-based framework to train agents for energy resource trading within a peer-to-peer LEM, where prices are explicitly linked to the supply-demand ratio. The researchers assess the efficacy of multiple community structures containing varying quantities of battery energy storage systems and renewable generation assets. They show that this configuration typically yields lower energy costs than established TOU and net-billing benchmarks.

Chen *et al.* [19] utilize a DQN variant to facilitate the automated interactions of prosumers equipped with battery energy storage systems operating within a LEM. The action space for these RL agents comprises four discrete choices encompassing buy/sell and charge/discharge operations. The resulting learned policy outperforms intuitive, rule-based strategies. Furthermore, it outperforms a strictly random policy representing a zero-intelligence agent, a benchmark initially introduced by Ghode *et al*. [20] to evaluate agent proficiency in automated trading environments. In a separate study, Chen *et al*. [21] explore the role of energy brokers powered by Q-learning, acting as a consensus mechanism for LEM settlements, utilizing profit as the agent's reward. Through multiple ablation and sensitivity analyses, the researchers demonstrate that these brokers effectively learn to optimize their own profit alongside market efficiency. Additionally, a recent paper by Jogunola *et al*. [22] enhances the DQN agent by applying prioritized experience replay [23] to further elevate economic returns.

Bose *et al.* [24] concentrate on the emergent interactions among participants in a static LEM architecture across varying degrees of DER integration. The study illustrates that the presence of RL-based agents in this setting can foster partial energy self-sufficiency. Furthermore, they reveal that the market's DER penetration level directly influences both the degree of self-sufficiency and the complexity of the agents' interactions. Mengelkamp *et al*. [25] evaluate three distinct adaptations of the Erev–Roth RL algorithm designed to automate participation in a LEM. Their findings indicate that these adaptations further enhance the self-sufficiency of the LEM when compared to the foundational Erev–Roth algorithm [26].

Mengelkamp *et al*. [16] benchmark a peer-to-peer LEM against a closed-book, double-auction LEM structured with settlement rounds. The study compares the performance of zero-intelligence agents and "intelligent" agents, adapted from Nicolaisen *et al*. [27], across both LEM frameworks. They determine that all evaluated market scenarios provide

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

comparable economic benefits, with the peer-to-peer LEM operated by intelligent agents achieving marginally superior results. Nevertheless, they observe that applying an identical strategy across diverse markets generates distinct pricing trends. Consequently, the researchers conclude that ensuring optimal system performance requires the simultaneous co-development of both the agent strategy and the market design.

Harrold *et al.* [28] apply rainbow DRL to master energy arbitrage within a microgrid context. Similarly, Lee *et al.* [29] utilize dynamic pricing and DRL to optimize the revenue of multiple electric vehicle charging stations. Even though these investigations are not directly connected to the methodology introduced in this paper, they serve as excellent references for prospective research concerning the integration of energy storage into the proposed LEM.

ALEX: Autonomous Local Energy Exchange

The need to model consumer behavior through utility functions and heavy reliance on forecasting may limit the scalability and robustness of traditional DR techniques. Moreover, the anticipated surge of online DERs in the near future introduces potential infrastructure requirements that could pose further obstacles to the successful deployment of DR systems. Specifically, the computational and communication costs associated with centralized control mechanisms could become prohibitive, particularly given the high temporal resolutions required to fully capture the stochastic usage patterns of electric vehicles and the intermittent nature of rooftop solar panels.

Addressing these challenges, this article introduces a distributed, multi-agent approach integrated with a double auction market mechanism. In formulating this architecture, named ALEX (autonomous local energy exchange), the following assumptions were established:

• Market participants are self-interested and, consequently, prioritize their personal economic well-being during the decision-making process.

• Participants are open to delegating certain decision-making processes related to indirect DR interactions to automated systems (e.g., employing RL agents).

• Every participating unit is equipped with a smart meter, and an adequate volume of high-resolution historical data is accessible to train the RL agents.

• The overarching large-scale electricity grid to which the customers are connected functions as an infinite bus.

Core concept

In principle, ALEX operates as a behind-the-meter DR strategy designed for a localized community, utilizing a double auction market to facilitate coordination. The primary market participants comprise the consumers residing within the community. Nevertheless, this framework could be broadened to incorporate temporarily present entities, such as electric vehicles. The market utilizes double auctions with a predefined settlement frequency Δt. During each interval $[t, t + \Delta t)$, the participants can convey their energy trading intentions by submitting bids:

$$\text{bid}_t = (q_t^{\text{bid}}, p_t^{\text{bid}}), q \in [0 \ldots q_{\max}^{\text{ask}}], p \in [p_{\min}, \ldots, p_{\max}], \quad (1)$$

and asks

$$\text{ask}_t = (q_t^{\text{ask}}, p_t^{\text{ask}}), q \in [0 \ldots q_{\max}^{\text{ask}}], p \in [p_{\min}, \ldots, p_{\max}], \quad (2)$$

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Where bidt and askt indicate the respective intent to purchase or sell energy. These are structured as tuples detailing the desired quantity q and the desired price p for the energy to be exchanged. Quantities are measured in Wh and are constrained between 0 and a designated maximum. The value of $q_{\max}^{\text{ask}}$ must be established to accommodate the anticipated peak generation derived from historical data. In the same way, $q_{\max}^{\text{bid}}$ ought to be configured to handle the projected peak load demand. Correspondingly, prices in $ fall within a defined window bounded by pmin and pmax. Bids and asks are resolved pairwise at the conclusion of every settlement round, generating a settlement signal mt for every participant:

$$m_t = (q_t^{\text{settlement}}, p_t^{\text{settlement}}) \quad (3)$$

It is crucial to emphasize that within this framework, participants actively determine the price signals and formulate energy management choices based on their market interactions. In contrast, most alternative DR methodologies merely require agents to react to external price signals.

The settlement signal mt is structured as a list of tuples containing the resolved quantities and their corresponding prices. Notably, participants are provided with information only about their individual settlements and therefore lack access to data about other participants' behavior.

Once the internal trades are finalized, any remaining energy generation or demand within the community is reconciled with the broader electricity grid using retail rates. For this article, we assume net billing, a widespread model where surplus energy is sold back to the grid at a rate of $p_{\text{sell}}^{\text{grid}}$, while deficient energy is procured from the grid at a cost of $p_{\text{buy}}^{\text{grid}}$, which incorporates associated fees. This behind-the-meter configuration provides the community with a margin of profitability by circumventing certain fees. Consequently, this dynamic naturally constrains the internal market pricing boundaries as follows:

$$p_{\min} = p_{\text{sell}}^{\text{grid}} <= p_{\text{market}} <= p_{\text{buy}}^{\text{grid}} = p_{\max}, \quad (4)$$

We propose the hypothesis that if participant interactions are primarily governed by supply and demand, ALEX will effectively serve as a decentralized, indirect DR mechanism. The ensuing pricing structure naturally generates economic incentives for all market participants to balance supply and demand. We validate this assertion through the experimental procedures.

ALEX as a stochastic game

To evaluate the characteristics of the suggested framework, the bidding strategies and auction mechanisms of the participants may be represented through an appropriate mathematical model. The participant interactions within ALEX can be formulated mathematically as a discounted stochastic game:

$$\Gamma := (n, L, S, A, P, R) \ \forall t \in [0...T], \lambda \in (0...1), \quad (5)$$

Where n indicates the total player count, L denotes the roster of players with a magnitude of |L| = n, S stands for the state space, A signifies the action space, P embodies the transition probabilities between states, R is the reward function, t designates the ongoing time step throughout the evaluation interval [0... T], and λ acts as the discount factor.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

The components S and A can each be separated into n distinct elements, Si and Ai, respectively, illustrated as follows:

$$S = S^1 \times \cdots \times S^n, \quad (6)$$

$$A = A^1 \times \cdots \times A^n. \quad (7)$$

The superscript i designates a particular participant Li, whereas the subscript denotes the time step t. The general action space A is denoted differently from the distinct group of actions executed at time t, aligning with the standard reinforcement learning (RL) terminology .

The probabilities of state transitions, corresponding to any collective action set executed during time step t, are specified as follows:

$$\forall a_t : \ P(S_{t+1}|S_t, a_t) := S_t \rightarrow S_{t+1} \quad (8)$$

Similar to the reinforcement learning environment, the payoff or reward within the stochastic game during time step t is formulated as:

$$R_t := S \times A \rightarrow r, \quad (9)$$

Which translates the tuple (St, at) into a real value $r \in \mathbb{R}$. In a similar vein, every agent seeks to optimize its personal return Gt (A.1). Consequently, each actor applies its uniquely formulated policy π^i (A.4) and (A.5) to select the action group a_t^i, relying on the states observed in S_t^i.

During every time increment, the agents possess the ability to engage with the marketplace through the submission of bids (1) alongside asks (2). This generates the subsequent definition for an action:

$$a_t^i = (\text{bid}_t^i, \text{ask}_t^i, e_t^i), \quad (10)$$

Wherein the supplementary variable, e_t^i, is allocated for prospective extensions of the framework, such as incorporating nonmarket operations like thermal load regulation or battery storage management.

Lastly, the observed states for individual agents are delineated as follows:

$$S^i = (d_t^i, g_t^i, m_{t-1}^i), \quad (11)$$

Where d_t^i and g_t^i represent the energy demand and power generation at time t, respectively, while m_{t-1}^i denotes the financial settlements acquired during time t-1.

The transition probabilities P emerge from the joint actions of all participating entities. Nevertheless, as a consequence of the pairwise settlement approach, the market structure and the observation space, participant Li does not have complete access to P. This characteristic guarantees that the formulated model functions as an authentic stochastic game.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

The existence of at least one stable Nash equilibrium is assured within Γ, provided that the parameters n, S, and A remain finite. This requisite can be satisfied by restricting the pricing values p to a sensible level of decimal precision (for instance, the 4 or 5 significant figures typically applied in banking sectors). The action space A is inherently constrained by the stipulation established earlier in **Eq. 4**. Consequently, the state space S is necessarily finite, ensuring that any deployment of ALEX will inherently possess at least one stable Nash equilibrium.

Automation using reinforcement learning

Given that participation in the formulated stochastic game Γ necessitates strategic proficiency, creating automated participant interactions (most commonly for prosumers, though theoretically applicable to any grid-integrated entity) with the LEM offers a logical solution to the challenges associated with precisely simulating consumer behavior. The suggested methodology focuses on training RL agents to execute market interactions and energy management activities, thereby mitigating suboptimal human decision-making. From a theoretical standpoint, RL is exceptionally well-suited for this objective because the stochastic game detailed in the preceding subsection parallels a Markov decision process when adhering to the established criteria outlined in Ref. [30]. The architecture established in this segment is designed to remain algorithm-agnostic. Nonetheless, the ensuing experimental validations utilize independent Q-learning.

Throughout this article, the reward function, R^i, is defined in the following manner:

$$R_t^i = (\text{profit}_t^{i,\text{LEM}} + \text{profit}_t^{i,\text{grid}}) - (\text{cost}_t^{i,\text{LEM}} + \text{cost}_t^{i,\text{grid}}), \quad (12)$$

Where,

$$\text{cost}_t^{i,\text{LEM}} = q_t^{i,\text{settled-bids}} \times p_t^{i,\text{settled-bids}}, \quad (13)$$

$$\text{profit}_t^{i,\text{LEM}} = q_t^{i,\text{settled-asks}} \times p_t^{i,\text{settled-asks}}, \quad (14)$$

$$\text{cost}_t^{i,\text{grid}} = q_t^{i,\text{grid-buy}} \times p_t^{i,\text{grid-buy}}, \quad (15)$$

$$\text{profit}_t^{i,\text{grid}} = q_t^{i,\text{grid-sell}} \times p_t^{i,\text{settled-sell}}, \quad (16)$$

And,

$$q_t^{i,\text{settled-bids}} + q_t^{i,\text{grid-buy}} = d_t^i \ (\text{load demand}), \quad (17)$$

$$q_t^{i,\text{settled-asks}} + q_t^{i,\text{grid-sell}} = g_t^i \ (\text{generation}). \quad (18)$$

Additional metrics, including social welfare expenditures, have been intentionally omitted because the primary objective at present is to evaluate system dynamics and participant actions based exclusively on pure economic performance. Nonetheless, such variables could be incorporated into subsequent research efforts. Because the RL agents are trained on high-frequency smart meter records, explicit models of consumer behavior can be bypassed. This omission is justified

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

because an ample volume of data is better at capturing nuanced and individualized customer behavioral trends, all while preserving scalability.

Experiments and Discussion

Suitability of settlement mechanism

A standard market actor can be modeled as a prosumer household. Such a residence might incorporate any mixture of generation, storage, and controllable loads. Throughout the experimental trials, the generation and load sources are extracted directly from smart meter data. Because the central objective of this article is to investigate the performance of the LEM across varying proportions of active supply and demand, techniques for profile shaping, such as load shifting or battery storage, are not considered.

Experimental design

This phase of the study directly addresses the initial research inquiry: Which specific attributes must a LEM settlement system possess to successfully accommodate RL-driven automation? Furthermore, the findings derived from this trial will illustrate how distinct market characteristics shape the behavioral strategies acquired by the RL algorithms. Because the operational setting—encompassing the market rules and the interactive tactics of competing entities—is equally as critical as the chosen learning protocol, defining an optimal market architecture is a mandatory prerequisite for future investigations and the deployment of advanced agent models, action options, and strategic frameworks. The boundaries of this trial are deliberately restricted to amplify the settlement system's impact on the resulting agent behaviors, thereby ensuring robust convergence limits even with ALEX functioning as a partially observable and nonstationary domain. During this phase, three distinct settlement procedures, each exhibiting unique market traits, undergo evaluation:

1. Average-Price (M1): Transactions finalize whenever the submitted bid meets or exceeds the requested ask price. The ultimate settlement value is calculated as the midpoint of these two figures.

2. Exact-Match (M2): Both purchasers and suppliers select their respective bid and ask figures from a predetermined menu of prices. A transaction is only executed if the bid and ask values are identical.

3. Exact-Price (M3): Transactions proceed as long as the bid is equal to or higher than the ask price. In this scenario, the purchaser acquires the resource from the auctioneer at their specific bid rate, whereas the supplier unloads their resource to the auctioneer at their designated ask rate.

Every double auction framework is characterized by a specific set of attributes [31]: individual rationality1, economic efficiency2, budget balancing3 and truthfulness4. While a perfect system would fulfill all four criteria, achieving this in a real-world scenario is practically impossible [31]. Given that the architectural foundation of ALEX, combined with the application of RL agents, inherently secures both economic efficiency and individual rationality, the three evaluated settlement procedures can be distinguished strictly based on their truthfulness and their approach to budget balancing. This distinction is detailed in **Table 1.**

Across all examined settlement procedures, three specific conditions reflecting diverse community supply-to-demand proportions are assessed: excess generation (10:1), excess consumption (1:10), and exact equilibrium (where supply equals demand). The assessment of each framework relies on the behavioral strategies formulated by the agents, alongside the ensuing market dynamics driven by the stabilized bid, ask, and settlement figures, all conducted under an identical training regimen. The primary objective is to pinpoint a market framework that not only respects fundamental supply and demand principles but also aligns effectively with the learning trajectories of RL agents. A system meeting these criteria should logically yield the subsequent outcomes:

Table 1. Settlement mechanism properties.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Mechanism	Market characteristics			
	Individual Participation Incentive	Economic Efficiency	Budget Balance	Truthful Reporting
M1	Yes	Yes	Strong	False
M2	Yes	Yes	Strong	True
M3	Yes	Yes	Weak	True

- Excess supply case: Generating units vie for available consumption, forcing ask prices downward, which subsequently pulls bid prices lower.
- Excess demand case: Consuming entities vie for limited generation, forcing bid prices upward, which correspondingly elevates ask prices.
- Equal supply and demand case: Both bid and ask valuations stabilize near the median of the permissible pricing spectrum.
- For all cases: The average bid, ask, and transaction prices must exhibit minimal dispersion.

A cohort consisting of n = 4 learning actors is established. A pair of these actors, denoted by $di > gi$, function in the capacity of purchasers, whereas the other pair, indicated by $di < gi$, operate as suppliers. Providing two entities for each role sustains competitive tension across both market facets and is intended to thwart monopolistic tendencies. Constant, time-invariant (steady-state) energy consumption patterns are applied to all agents, ensuring the aggregated supply and demand align perfectly with the predefined ratios outlined earlier.

$$\frac{g^{\mathrm{LEM}}}{d^{\mathrm{LEM}}} = \frac{\sum_i g^i}{\sum_i d^i}. \quad (19)$$

Implementing steady-state load profiles simplifies the objective for the agents, restricting their primary function to discovering equilibrium pricing. Such an arrangement compresses every agent's observation space down to a solitary point and anchors the variables to the residual load capacity. Consequently, this heightens the experimental clarity by isolating the acquisition of pricing strategies from other variables. From the perspective of any individual actor, this shifts the paradigm into a partially observable, nonstationary multi-armed bandit problem, wherein the total number of arms aligns with the quantity of discrete pricing options available. Every participating entity can utilize a standalone tabular Q-learning algorithm, incorporating an ϵ-greedy exploration strategy alongside a designated learning rate α. This approach preserves broad convergence assurances despite the inherently complex nature of the generated environment [32].

Throughout this specific trial, the parameter α is fixed at 0.1, γ at 0.98, and ϵ at 0.1. The metrics for ϵ and α undergo an annealing process beginning at the 100th episode to optimize the trade-off between exploratory actions and convergence velocity, applying a reduction multiplier of 0.98 with each successive episode. Within this straightforward configuration, should the agents be unable to formulate strategies that mirror the aforementioned standards, the corresponding market frameworks will be deemed unviable for future application.

The experimental procedures were executed utilizing the T-REX simulation platform. Computational tasks were processed on a dedicated workstation equipped with a Ryzen 9 3900X CPU alongside 32 GB of 3200MHz DDR4 RAM. Under these exact hardware specifications, a single episode required roughly 5 minutes to complete. Exhaustive details regarding the experimental setup are accessible via the project's GitHub repository [33].

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Results and discussion

It is vital to remember that within the ALEX framework, the participating entities are responsible for simultaneously establishing price signals and executing energy management choices via their market engagements. Consequently, analyzing both the behavioral outputs of the agents and the final settlement figures is paramount. **Figure 1** illustrates the density plots representing the acquired strategies for bid and ask valuations, alongside the final transaction prices, specifically for the M1 settlement structure. In the unbalanced scenarios, the bid, ask, and settlement values generally converge tightly toward the anticipated margins of the pricing spectrum. Nevertheless, the perfectly balanced scenario exposes a severe flaw. Even though the transaction prices aggregate accurately near the median of the available range, the submitted bid and ask offers wildly diverge toward opposite extremes. This erratic behavior stems directly from the M1 mechanism's deficiency in truthfulness.

The agents lack any inherent motivation to tender bid and ask figures that genuinely reflect their perceived true value of the energy commodity (which would closely mirror the settlement price). Because the M1 framework dictates that the final price is merely the mathematical average of a coupled bid and ask, bidding at the extremes dramatically boosts the probability of securing an advantageous financial outcome. Simultaneously, this tactic maximizes profitability if an adversarial agent opts for a completely honest bidding approach. Consequently, adopting this divergent tactic clearly represents the most mathematically optimal strategy under these conditions. This extreme price divergence creates substantial complications, particularly if researchers were to implement continuous, unrestricted action spaces for pricing: excessively inflated or deflated bid/ask figures could destabilize the entire settlement process. Due to this significant vulnerability, the M1 architecture is definitively disqualified.

Figure 2 displays the performance metrics associated with the M2 settlement approach. In stark contrast to the first framework, M2 completely fails to demonstrate definitive convergence across all tested scenarios. One plausible rationale for this failure is that M2 fulfills the rigorous prerequisites of a theoretically perfect double auction market—a conceptual model that the Myerson–Satterthwaite theorem proves is unfeasible to execute in reality [31]. Alternatively, the stringent requirements for budget balancing might severely restrict the volume of finalized transactions, thereby yielding highly infrequent reward signals across a nonstationary learning landscape. Consequently, even though a Nash equilibrium theoretically exists within the mathematical model, the agents are highly unlikely to locate it owing to inadequate reinforcing feedback. Even if M2 could theoretically achieve convergence over an extended training duration, its inability to stabilize under the same simulation constraints applied to M1 and M3 renders it functionally inferior. RL systems are frequently required to refine their strategies using live data streams, a process significantly more sluggish than accelerated computational simulations. At present, substantial archives of historical smart meter records remain scarce. Therefore, deploying this framework in a live, physical environment would mandate on-the-fly learning capabilities. Taking these practical limitations into account, the M2 market structure is rejected as impractical for real-world implementation.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

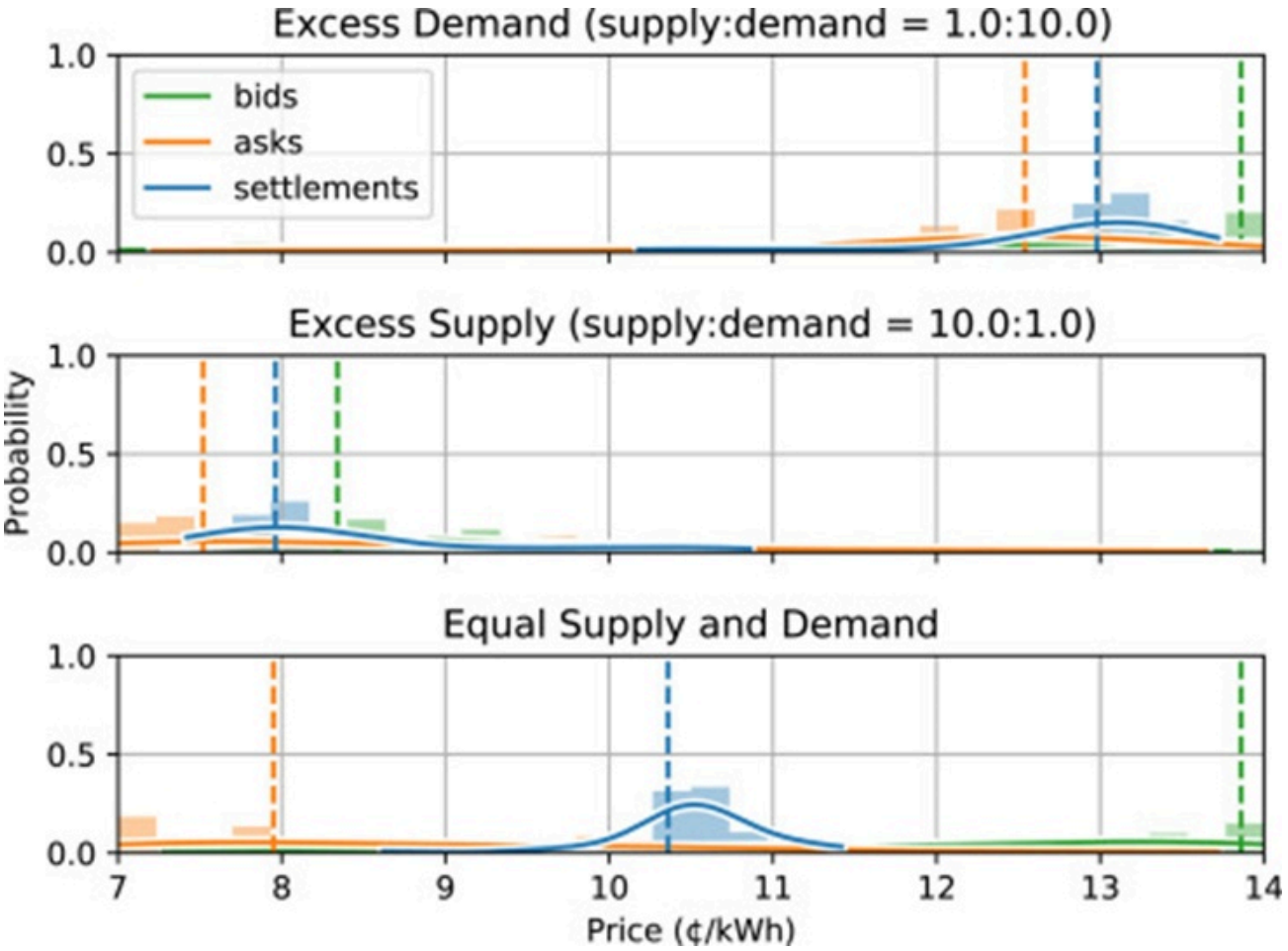


Figure 1. Validation policies for bid, ask, and resulting settlement prices for agents operating under M1 for episodes 70 to 100. The histograms show the probabilities of the discrete action policies for the agents and the resulting settlement prices. Price modes are highlighted by vertical dashed lines. Probability density functions of the histograms are overlaid on top, which are approximated with the Gaussian KDE function in the scikit-learn Python package with default parameters.

The experimental outcomes about the M3 settlement structure are depicted in **Figure 3**. Much like the M1 scenario, the resulting bid, ask, and transaction figures aggregate tightly along the anticipated margins during unbalanced conditions, exhibiting even denser clustering. Crucially, and in direct contrast to M1, the balanced scenario mirrors this stable dynamic, accurately localizing prices near the exact midpoint of the available spectrum. Based on these observations, one can definitively deduce that the M3 framework inherently promotes truthfulness. The participating actors possess minimal motivation to establish bid or ask valuations that stray significantly from the eventual settlement rates, which are expected to closely reflect the authentic commodity value based on the prevailing supply-to-demand ratio. Because it successfully meets all predefined evaluative benchmarks, the M3 architecture is officially validated for subsequent advanced analysis.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

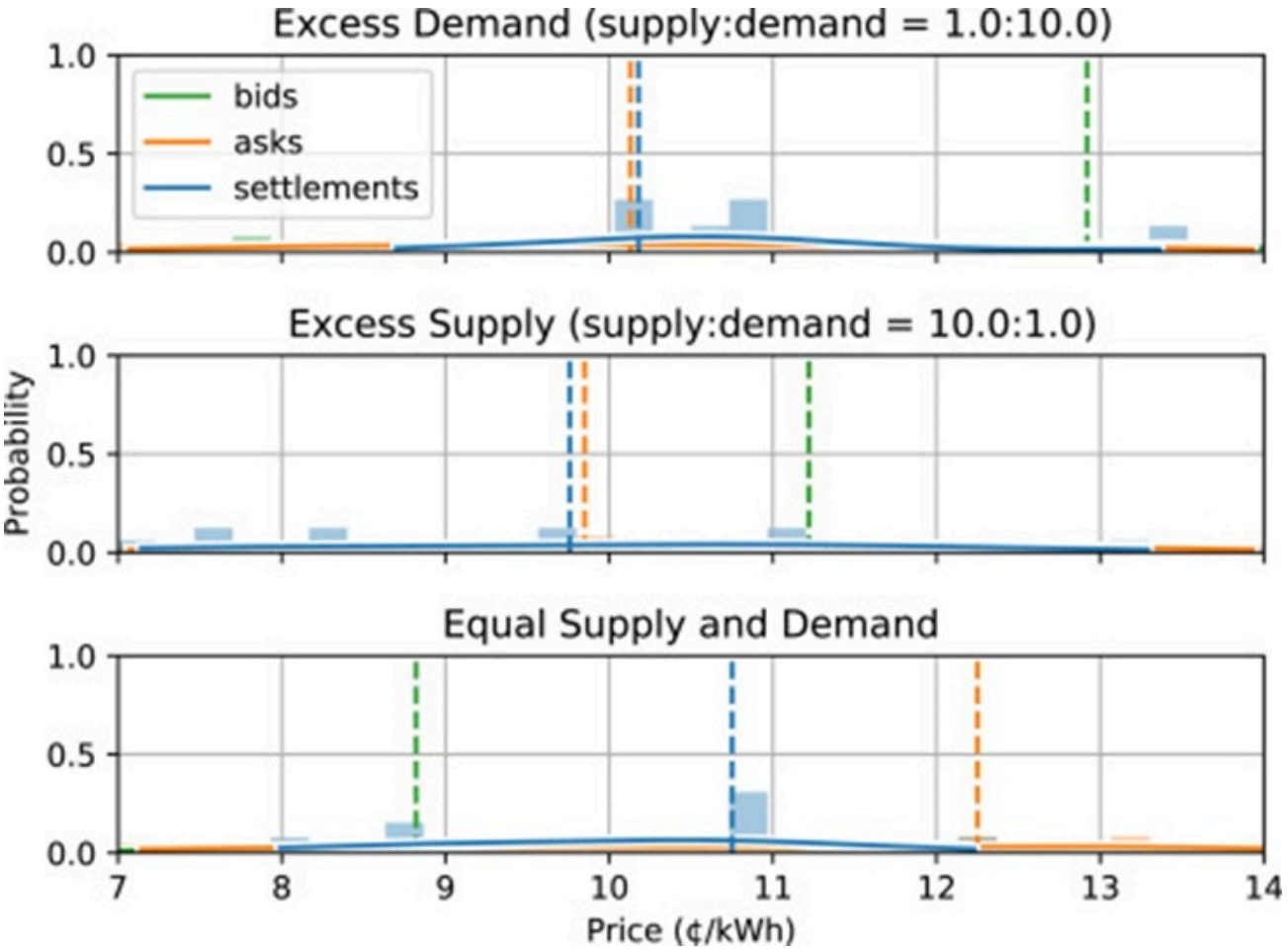


Figure 2. Validation policies for bid, ask, and resulting settlement prices for agents operating under M2 for episodes 70 to 100. The histograms show the probabilities of the discrete action policies for the agents and the resulting settlement prices. Price modes are highlighted by vertical dashed lines. Probability density functions from the histograms are overlaid and approximated with the Gaussian KDE function in the scikit-learn Python package with default parameters.

The preliminary trials utilized exceptionally severe supply-to-demand proportions. Analyzing the resulting settlement trends under moderately balanced conditions will generate a more comprehensive understanding of the system's behavioral dynamics. As a result, intermediate proportions of 1.5:1 and 1:1.5 have been incorporated into the M3 testing protocol. Furthermore, the simulation timeline is elongated by an additional 100 episodes, maintaining the identical annealing process .

As demonstrated by the data presented in **Figure 4**, the final transaction rates stabilize marginally below the median point when the supply-to-demand ratio is 1.5:1, and climb slightly above the median when the ratio shifts to 1:1.5. This empirical evidence firmly validates the prevailing influence of fundamental supply and demand principles, demonstrating that settlement metrics reliably track the availability and scarcity of the resource. Additional empirical trials encompassing a wider variety of supply and demand configurations will be executed to construct a robust mathematical model representing these pricing dynamics.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

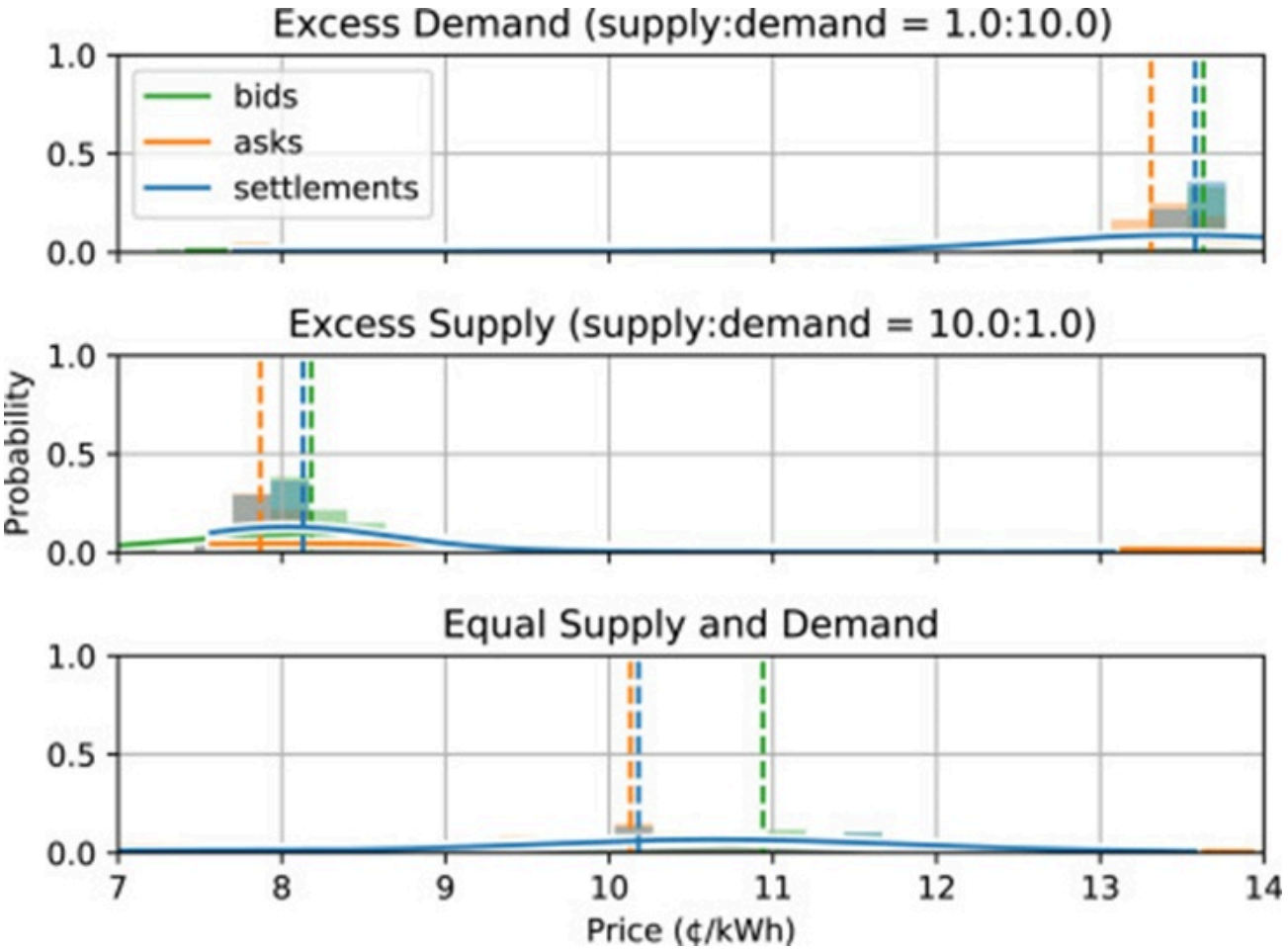


Figure 3. Validation policies for bid, ask, and resulting settlement prices for agents operating under M3 for episodes 70 to 100. The histograms show the probabilities of the discrete action policies for the agents and the resulting settlement prices. Price modes are highlighted by vertical dashed lines. Probability density functions of the histograms are overlaid on top and approximated with the Gaussian KDE function in the scikit-learn Python package with default parameters.

To summarize, these investigatory trials indicate that successful RL agent optimization necessitates a relaxed approach to budget balancing, a condition that reliably generates robust and frequent reinforcement signals. Furthermore, intrinsic truthfulness is absolutely critical to guarantee that the developed behavioral strategies accurately embody core supply and demand principles. Within any LEM characterized by these twin attributes, the autonomous self-interest of the actors naturally optimizes the overall value transferred across the network, thereby ensuring macroeconomic efficiency upon reaching convergence. Although strict, zero-sum budget balancing remains elusive in this design, such an outcome might actually prove highly beneficial in real-world applications: allowing the central auctioneer to retain a marginal surplus provides the necessary capital to fund the upkeep of the market's physical and digital infrastructure.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

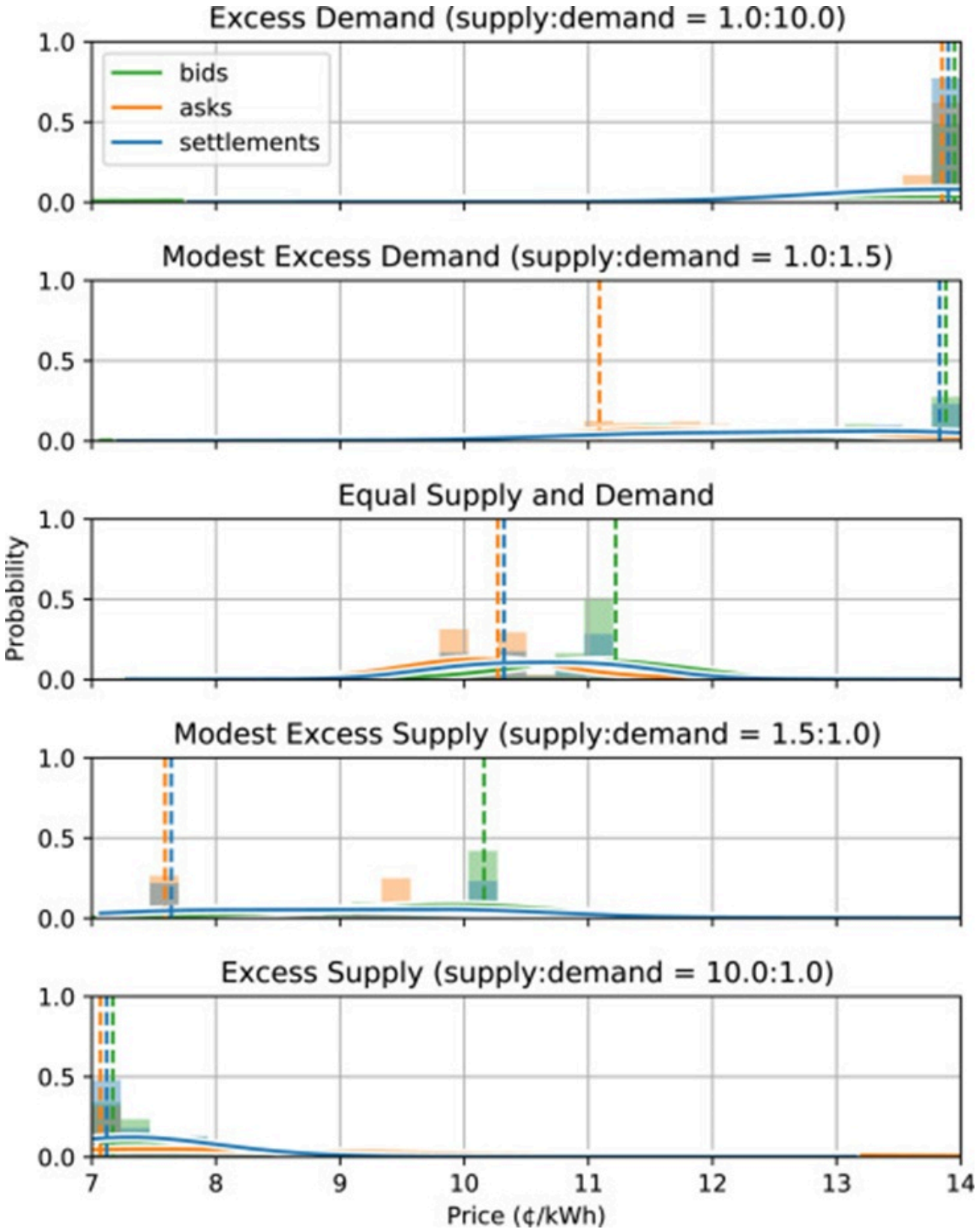


Figure 4. Validation policies for bid, ask, and resulting settlement prices for agents operating under M3 for episodes 100 to 200. The exploration factor and learning rate are annealed starting from episode 100, with a multiplier of 0.98 applied at the beginning of each episode. The histograms show the probabilities of the discrete action policies for the agents and

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

the resulting settlement prices. Price modes are highlighted by vertical dashed lines. Probability density functions of the histograms are overlaid on top and approximated with the Gaussian KDE function in the scikit-learn Python package with default parameters.

Economic study

Experimental design

This subsequent phase of testing concentrates entirely on the secondary research inquiry: Can the emergent market dynamics adequately facilitate Demand Response (DR) initiatives within a LEM heavily saturated with Distributed Energy Resources (DER)? This evaluation is executed by juxtaposing the newly devised methodology against traditional tariff frameworks, specifically net billing.5 and time-of-use (TOU) schedules. Based on the insights gained during the preceding trial, the pricing metrics at the point of market equilibrium are heavily dictated by foundational supply and demand concepts **(Figure 4).** By running supplementary computational runs using varied supply-to-demand thresholds, researchers can construct an empirical pricing algorithm using standard interpolation techniques. This algorithm could potentially serve as a streamlined method for establishing localized transaction rates, circumventing the complexity of deploying a live, continuous auction environment. In practice, the exact supply-to-demand ratio governing a specific local network could be extracted directly from aggregate smart meter telemetry.

The economic analysis relies on a simulated residential microgrid encompassing a total of ten individual actors. Because comprehensive smart home datasets originating from Canada were unavailable, researchers utilized openly accessible consumption profiles derived from the SunDance database [34, 35]. To construct this digital community, ten specific energy consumption signatures were randomly selected. The identification numbers associated with these specific households are: 10011, 1001625, 1002714, 10068, 100703, 1001420, 1003173, 1001230, 100114, and 100196. Every included entity functions as a prosumer, actively engaging in the energy exchange process to secure maximum financial advantages. The entire microgrid is theoretically mapped onto a single electrical bus located downstream of a unified community smart meter. Consistent with the constraints of the initial experiments, researchers entirely abstained from any load shaping interventions. To visually demonstrate the shifting supply and demand dynamics across this trial group, the total compiled supply and consumption metrics spanning one specific summer day (June 1, 2015) are graphically represented in **Figure 5**.

This analytical trial quantifies the fluctuations in aggregate electricity expenditures resulting from the activation of a localized energy trading ecosystem, which dynamically calculates transaction rates contingent upon regional supply and demand metrics. Because these localized rates constantly fluctuate over time and remain intimately tied to the prevailing supply-to-demand balance, the pricing outputs generated by the LEM can be directly contrasted against standard TOU baseline tariffs.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

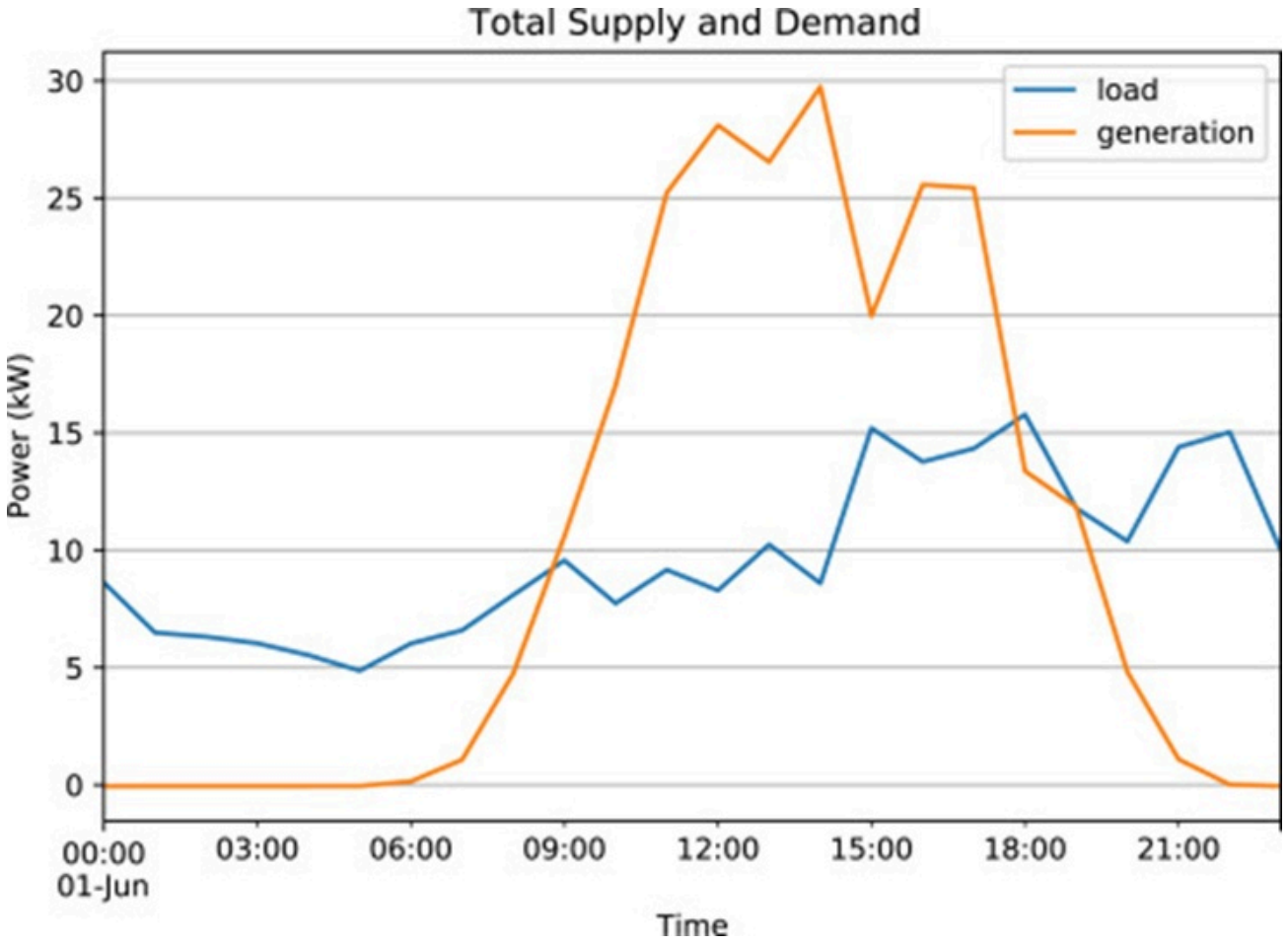


Figure 5. Total supply and demand profile of the residential community test over one summer day on June 1, 2015.

Results and discussion

The comprehensive market model is derived from the preceding experiments, supplemented by four additional supply-demand proportions (1:1.1, 1:2, 1.1:1, 2:1). **Figure 6** displays the resulting price behavior. In authentic applications, where each participant possesses unique DER capabilities and demand patterns, ALEX agents may formulate distinct pricing trajectories. The primary objective here is to evaluate the economic utility of an ALEX-governed trading system and measure its performance against conventional tariffs that lack the capacity for individualized pricing.

The following expression characterizes the pricing curve:

$$P(s,d) = \begin{cases} P^{\text{grid}}_{\text{NB,load}} & P(s,d) \geq P^{\text{grid}}_{\text{NB,buy}} \\ P^{\text{grid}}_{\text{NB,gen}} & P(s,d) \leq P^{\text{grid}}_{\text{NB,sell}} \\ -0.0254\frac{s}{d} + 0.1426 C^{\text{TOU}}_{H,M,L} & \text{if buying} \\ -0.0280\frac{s}{d} + 0.1299 C^{\text{TOU}}_{H,M,L} & \text{if selling,} \end{cases} \quad (20)$$

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Where s denotes energy supply, d denotes energy demand, $P^{\text{grid}}_{\text{NB,load}}$ reflects the cost of procuring electricity from the grid under net billing, $P^{\text{grid}}_{\text{NB,gen}}$ signifies the revenue per unit when injecting energy into the grid under net billing, and $C^{\text{TOU}}_{\text{H,M,L}}$ represents the scaling factors utilized for TOU adjustments.

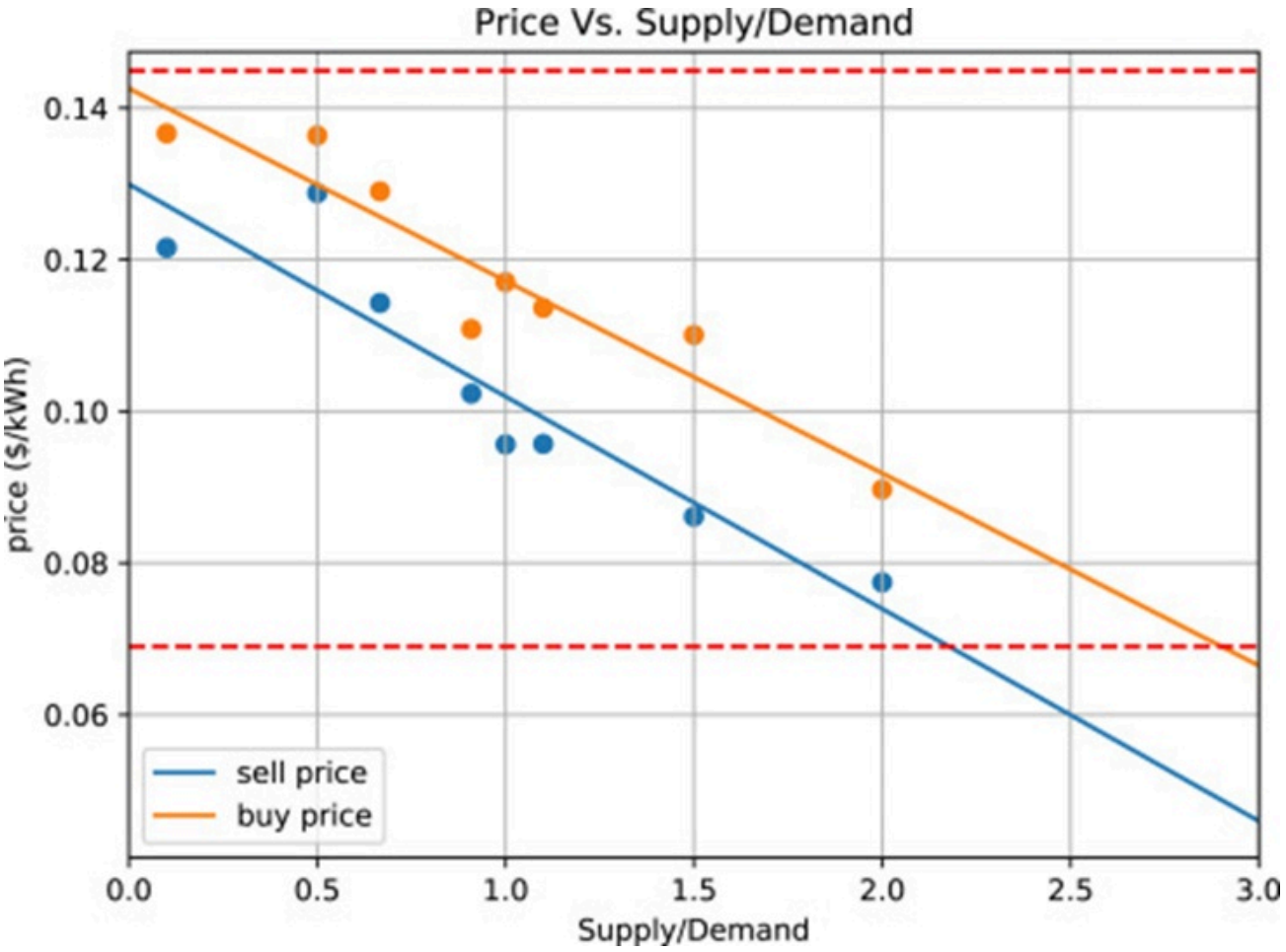


Figure 6. Pricing model developed for the test local market. The dotted lines show the price boundaries defined by (4). Linear regression between the price points leads to a well-fit, generalized mathematical model.

Applying this model to specific energy profiles yields a tailored pricing schedule for a local energy community, as illustrated in **Figure 7** based on the data points from **Figure 5.**

The internal prices generated by this model closely mirror the community's supply-demand ratio throughout the temporal cycle. For example, during the nocturnal hours when photovoltaic production is absent, the rate for peer-to-peer energy exchange is 0.1449, equivalent to the consumer's cost for grid-supplied energy. As daylight proceeds, specifically by 7:00 AM, the emergence of solar energy initiates a commensurate decline in peer-to-peer pricing. By 9:00 AM, generation significantly outpaces demand, driving the peer-to-peer rate down to 0.069, mirroring the grid injection price under net billing standards.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

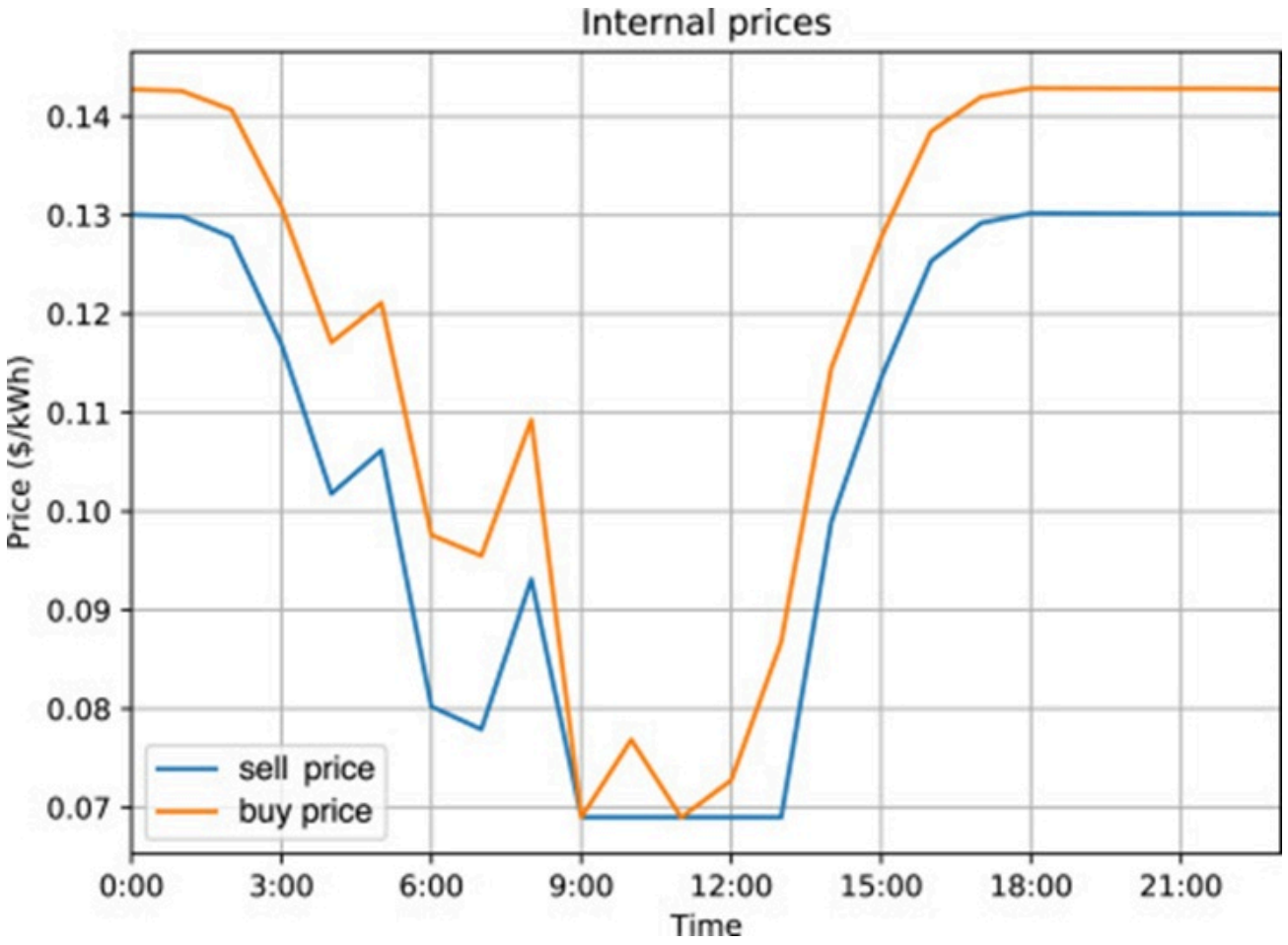


Figure 7. Internal prices of ALEX used to conduct transactions.

The fiscal performance of all community participants under this pricing regime is evaluated and contrasted with net billing. **Figure 8** depicts this comparison, with summarized data provided in **Table 2**.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

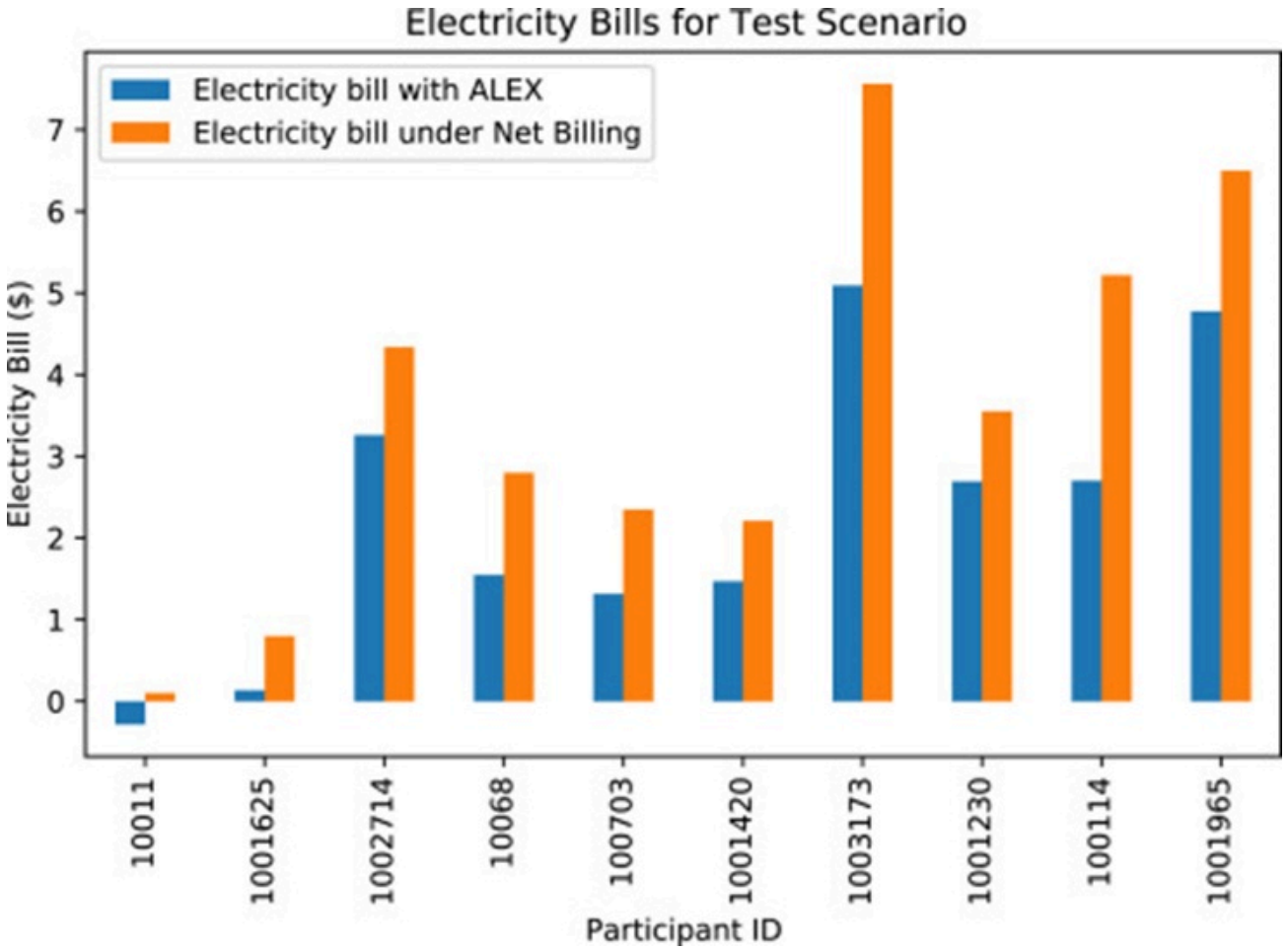


Figure 8. Electricity bill comparison between net billing and ALEX.

Data in **Table 2** demonstrate that adopting a localized market framework yields significant financial gains for the community, reflected in a 35.9% decrease in total aggregate expenses. Individual bill reductions show a mean of 74.51% and a median of 38.8%. This financial improvement is attributed to the optimized allocation of local energy assets, which the local market price captures more precisely. The implementation of a behind-the-meter community architecture also provides financial advantages to participants who may not possess the capital to invest in personal DER infrastructure; they gain access to surplus generation from neighbors, which is more economical than grid-based procurement. Conversely, suppliers are incentivized to prioritize peer-to-peer transactions over grid injection to capture higher margins. Effectively, this framework democratizes the financial advantages associated with DER adoption. While different in scope, recent research by Jogunola *et al*. [22] noted financial benefits of 35% through energy storage integration, rising to 55% with combined PV and storage. Although the current iteration employs only PV, the anticipated addition of energy storage in subsequent research is expected to yield performance levels equal to or exceeding those cited in Ref. [22].

Table 2. ALEX vs. Net Billing (NB).

Participant ID	Electricity cost ($)		Difference (%)	Energy purchased from utility grid (kWh)		Difference (%)	Energy acquired through local trading (kWh)	Energy exported to utility grid (kWh)		Differenc (%)

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

	NB	ALEX	%	NB	ALEX	%	NB	ALEX	NB	ALEX
10011	0.10	−0.28	383.2	6.83	2.35	65.59	–	4.48	12.88	0
1001625	0.80	0.13	83.2	10.73	2.28	78.75	–	8.45	10.97	0
1002714	4.34	3.26	24.8	32.45	13.10	59.63	–	19.35	5.31	0
10068	2.80	1.55	44.7	22.37	4.08	81.76	–	18.29	6.41	0
100703	2.35	1.32	44.0	22.47	5.94	73.56	–	16.53	13.17	0
1001420	2.21	1.47	33.6	17.63	11.96	32.16	–	5.67	4.98	0
1003173	7.56	5.09	32.7	61.73	22.00	64.36	–	39.73	20.03	0
1001230	3.55	2.69	24.2	24.89	8.03	67.74	–	16.86	0.80	0
100114	5.22	2.70	48.2	49.55	12.65	74.47	–	36.9	28.40	0
1001965	6.50	4.78	26.5	50.78	19.07	62.45	–	31.71	12.45	0
Aggregate	35.43	22.70	35.9	299.44	101.46	66.12	–	198.00	115.40	0

The local market price remains inherently dynamic and tied to supply-demand fluctuations. This logic parallels the intent of TOU pricing, which accounts for grid-wide supply-demand volatility rather than localized imbalances. Consequently, we contrast these two approaches to assess their relative merits. Ontario TOU serves as the Canadian benchmark and is frequently referenced as an alternative in jurisdictions that currently lack such mechanisms, including Alberta.

Figure 9 contrasts the two pricing schedules, highlighting significant divergence in their respective profiles. While the local market price recedes toward midday as generation capacity expands, the TOU rate trends upward, suggesting an environment where load exceeds generation. Although this may be accurate for industrial and commercial consumers absent from our local test group, it underscores the lack of correlation between TOU and the specific supply-demand reality of our test site. This systemic disconnect, likely a driver of the low TOU participation, supports the need for granular, localized price signals to optimize DER integration and overall system performance. ALEX offers a scalable, distributed mechanism for generating such highly relevant signals at a reduced cost.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

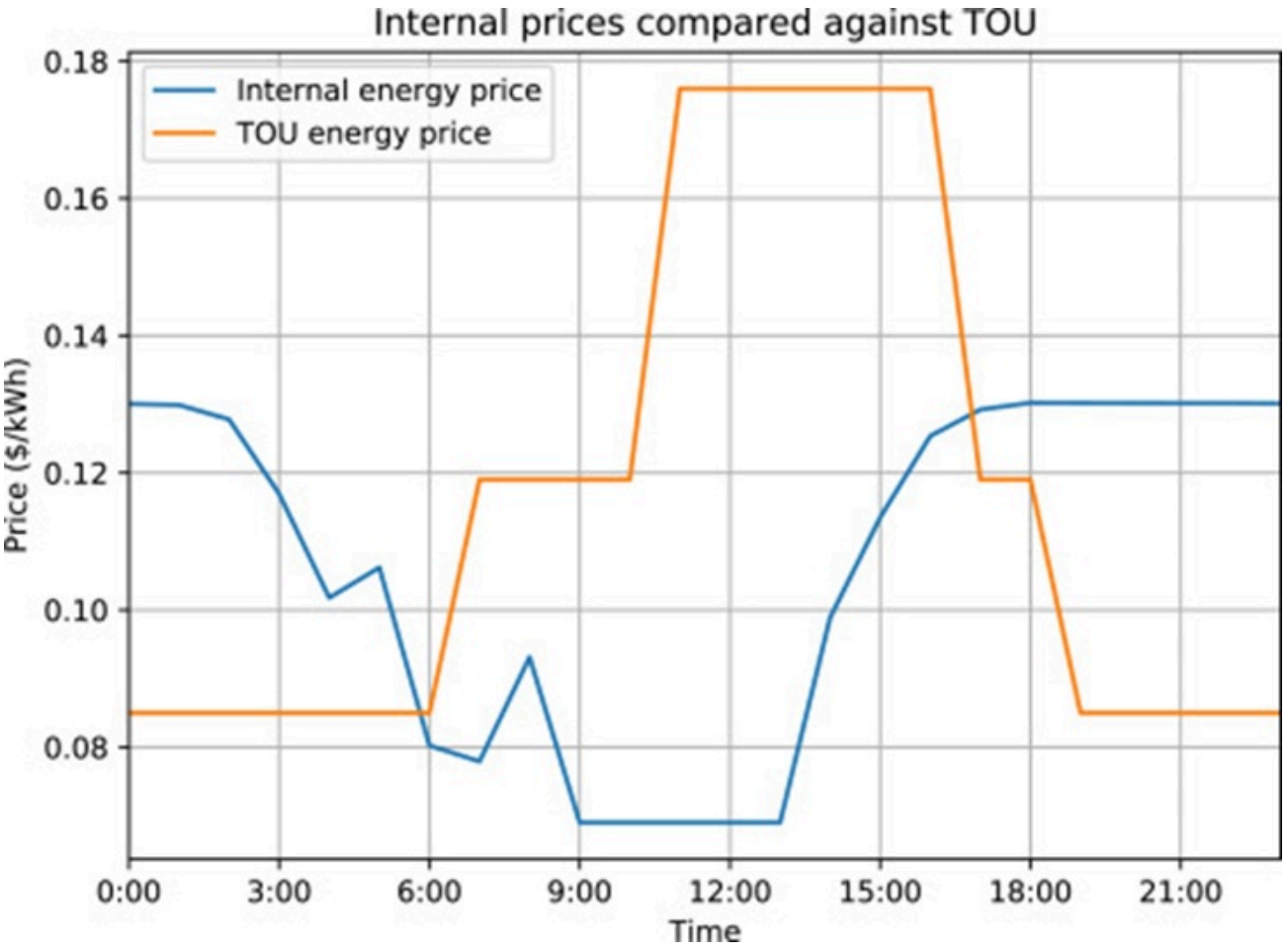


Figure 9. Local market pricing schedule compared against Ontario summer TOU prices.

Conclusion

Demand Response (DR) mechanisms are essential tools for managing Distributed Energy Resources (DERs). As the energy mix grows in complexity and the prevalence of these resources increases, automated management and coordination become mandatory. This research explores the potential for utilizing multi-agent reinforcement learning to automate Local Energy Markets (LEMs). ALEX, our proposed LEM framework, supports arbitrary closed-book, double auction settlement protocols, facilitating the identification of market traits that compel independent Q-learning agents to align with supply-demand laws. After validating an appropriate settlement mechanism, we contrasted the resulting market dynamics with standard DER integration strategies.

The initial experiment trained agent cohorts across three distinct market configurations. Findings indicate that in double auction settings, truthfulness is a prerequisite for achieving a collective policy that respects the law of supply and demand. Furthermore, weak budget balancing provides the necessary reward density to facilitate effective agent training. When combined with individual rationality, these mechanisms maximize economic efficiency while mitigating the risks of weak budget balancing.

The second experiment evaluated the LEM against conventional net billing and TOU benchmarks. Because the consensus pricing in ALEX is fundamentally rooted in supply-demand dynamics, it provides a significantly more

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

responsive and accurate signal than TOU-based pricing, likely enhancing DER utilization efficiency. These gains provide positive systemic effects that extend upstream. The community-level economic impact was substantial, with an aggregate bill reduction of 35.9% compared to net billing, while individual participants observed mean and median reductions of 74.51% and 38.8%, respectively.

These results establish a foundation for ongoing investigation. Future initiatives will incorporate battery energy storage systems and more robust RL algorithms to improve profile-shaping capabilities, aiming to further decrease costs and strengthen grid stability. We will continue benchmarking ALEX against alternative LEM implementations to determine the optimal automation approach. As this framework matures, subsequent studies will evaluate its systemic impacts, including efficiency improvements in local markets incorporating electric vehicles and storage, as well as the broader effects of inter-ALEX energy trading across the main grid.

Acknowledgements

The authors gratefully acknowledge support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada/Alberta Innovates (AI) grant ALLRP 561116–20 and by the Canada First Research Excellence Fund under the Future Energy Systems program at the University of Alberta.

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 Feb 2024 Revised: 04 May 2024 Accepted: 21 May 2024
Published online: 15 November 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Fattahi A, Deihimi M. A review of demand-side management: Reconsidering theoretical framework. Renew Sustain Energy Rev. 2017;80:367-79.
https://doi.org/10.1016/j.rser.2017.05.207.

2. Kim BG, Zhang Y, van der Schaar M, Lee JW. Dynamic pricing and energy consumption scheduling with reinforcement learning. IEEE Trans Smart Grid. 2016;7(5):2187-98. https://doi.org/10.1109/TSG.2015.2495145.

3. Lu R, Hong S, Zhang X. A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. Appl Energy. 2018;220:220-30. https://doi.org/10.1016/j.apenergy.2018.03.072.

4. Chen S, Liu CC. From demand response to transactive energy: state of the art. J Mod Power Syst Clean Energy. 2017;5(1):10-9.

5. Abrishambaf O, Lezama F, Faria P, Vale Z. Towards transactive energy systems: An analysis on current trends. Energy Strategy Rev. 2019;26:100418.

6. Yu M, Lu R, Hong S. A real-time decision model for industrial load management in a smart grid. Appl Energy. 2016;183. https://doi.org/10.1016/j.apenergy.2016.09.021.

7. Huang X, Hong SH, Li Y. Hour-ahead price based energy management scheme for industrial facilities. IEEE Trans Ind Inf. 2017;13(6):2886-98. https://doi.org/10.1109/TII.2017.2711648.

8. de Sá Ferreira R, Barroso LA, Lino PR, Carvalho MM, Valenzuela P. Time-of-use tariff design under uncertainty in price-elasticities of electricity demand: A stochastic optimization approach. IEEE Trans Smart Grid. 2013;4(4):2285-95. https://doi.org/10.1109/TSG.2013.2241087.

9. Liu Y, Zhang D, Deng C, Wang X. Deep reinforcement learning approach for autonomous agents in consumer-centric electricity market. In: 2020 5th IEEE International Conference on Big Data Analytics (ICBDA). IEEE; 2020:37-41. https://doi.org/10.1109/ICBDA49040.2020.9099946.

10. Forfia D, Knight M, Melton R. The view from the top of the mountain: Building a community of practice with the GridWise transactive energy framework. IEEE Power Energy Mag. 2016;14(3):25-33.

11. Mengelkamp E, Diesing J, Weinhardt C. Tracing local energy markets: A literature review. it - Inf Technol. 2019;61(2-3):101-10. https://doi.org/10.1515/itit-2019-0016.

12. Pilz M, Al-Fagih L. Recent advances in local energy trading in the smart grid based on game-theoretic approaches. IEEE Trans Smart Grid. 2019;10(2):1363-71. https://doi.org/10.1109/TSG.2017.2764275.

13. Khorasany M, Mishra Y, Ledwich G. Market framework for local energy trading: a review of potential designs and market clearing approaches. IET Gener Transm Distrib. 2018;12(22):5899-908. https://doi.org/10.1049/iet-gtd.2018.5309.

14. Baker B, Kanitscheider I, Markov T, Wu Y, Powell G, McGrew B, et al. Emergent tool use from multi-agent autocurricula. arXiv:1909.07528. 2020.

15. Mengelkamp E, Staudt P, Garttner J, Weinhardt C. Trading on local energy markets: A comparison of market designs and bidding strategies. In: 2017 14th International Conference on the European Energy Market (EEM). IEEE; 2017:1-6.

16. Zang H, Kim J. Reinforcement learning based peer-to-peer energy trade management using community energy storage in local energy market. Energies. 2021;14(14). https://doi.org/10.3390/en14144131.

17. Foruzan E, Soh LK, Asgarpoor S. Reinforcement learning approach for optimal distributed energy management in a microgrid. IEEE Trans Power Syst. 2018;33(5):5749-58.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

https://doi.org/10.1109/TPWRS.2018.2823641.

18. Zhou S, Hu Z, Gu W, Jiang M, Zhang XP. Artificial intelligence based smart energy community management: A reinforcement learning approach. CSEE J Power Energy Syst. 2019;5(1):1-10. https://doi.org/10.17775/CSEEJPES.2018.00840.

19. Chen T, Su W. Local energy trading behavior modeling with deep reinforcement learning. IEEE Access. 2018;6:62806-14. https://doi.org/10.1109/ACCESS.2018.2876652.

20. Gode DK, Sunder S. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. J Polit Econ. 1993;101(1):119-37.

21. Chen T, Su W. Indirect customer-to-customer energy trading with reinforcement learning. IEEE Trans Smart Grid. 2019;10(4):4338-48. https://doi.org/10.1109/TSG.2018.2857449.

22. Jogunola O, Tsado Y, Adebisi B, Nawaz R. Trading strategy in a local energy market, a deep reinforcement learning approach. In: 2021 IEEE Electrical Power and Energy Conference (EPEC). IEEE; 2021:347-52. https://doi.org/10.1109/EPEC52095.2021.9621459.

23. Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. arXiv:1511.05952. 2016.

24. Bose S, Kremers E, Mengelkamp EM, Eberbach J, Weinhardt C. Reinforcement learning in local energy markets. Energy Inf. 2021;4(1):7. https://doi.org/10.1186/s42162-021-00141-z.

25. Mengelkamp E, Gärttner J, Weinhardt C. Intelligent agent strategies for residential customers in local electricity markets. In: Proceedings of the Ninth International Conference on Future Energy Systems (e-Energy '18). Association for Computing Machinery; 2018:97-107.

26. Erev I, Roth AE. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. Am Econ Rev. 1998;88(4):848-81.

27. Nicolaisen J, Petrov V, Tesfatsion L. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. IEEE Trans Evol Comput. 2001;5(5):504-23.

28. Harrold DJ, Cao J, Fan Z. Data-driven battery operation for energy arbitrage using rainbow deep reinforcement learning. Energy. 2022;238:121958.

29. Lee S, Choi DH. Dynamic pricing and energy management for profit maximization in multiple smart electric vehicle charging stations: A privacy-preserving deep reinforcement learning approach. Appl Energy. 2021;304:117754.

30. Oliehoek FA, Amato C. A Concise Introduction to Decentralized POMDPs. Springer; 2016.

31. Myerson RB, Satterthwaite MA. Efficient mechanisms for bilateral trading. J Econom Theory. 1983;29(2):265-81.

32. Hu J, Wellman MP. Nash Q-learning for general-sum stochastic games. J Mach Learn Res. 2003;4(Nov):1039-69.

33. Zhang S. TREX-Publication-resources. GitHub Repository. 2021. https://github.com/sd-zhang/publications-resources

34. Chen D, Irwin D. SunDance: Black-box Behind-the-Meter Solar Disaggregation. In: E-Energy '17: Proceedings of the Eighth International Conference on Future Energy Systems. 2017:45-55.

35. Barker S, Mishra A, Irwin D, Cecchet E, Shenoy P, Albrecht J. Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes. In: Proc SustKDD. 2012.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press