# T-REX: A Model-Free Open-Source Framework for Integrated Design and Simulation of Transactive Energy Systems

Shida Zhang[a], Daniel May[a], Mustafa Gül[b], Petr Musilek[a,c,*]

[a] *Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada*
[b] *Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada*
[c] *Applied Cybernetics, University of Hradec Králové, Hradec Králové, Czech Republic*



**Abstract**

The energy transition is necessary to mitigate climate change. However, it may negatively affect the electric power system if renewable and distributed energy resources are not integrated properly. Transactive energy (TE) is an energy coordination framework meant to address this problem using economic and control mechanisms to dynamically balance energy across the entire electrical infrastructure. Transactive energy schemes have been designed, most commonly, using model-based co-simulators. However, these approaches are severely affected by the lack of comprehensive empirical data from deployment projects. As a result, they suffer from low flexibility and poor generalization. In this context, a transactive renewable energy exchange simulation framework, T-REX, is designed to solve this problem via an agent-based, model-free approach. This design philosophy provides benefits that would be difficult to achieve using other existing state-of-the-art simulators. The flexibility of T-REX is showcased by demonstrating the ease of implementing both contemporary transactive control co-simulations using powerflow-centric approaches, as well as agent-based, economy-driven approaches. We also propose and investigate a novel implementation of a transactive energy system, named Autonomous Local Energy Exchange (ALEX). The overall concept of ALEX is first introduced, along with the steps necessary for its implementation. T-REX is then used to develop and evaluate various settlement mechanisms for ALEX, with the goal of finding at least one mechanism suitable for training reinforcement learning agents to behave in ways that truly reflect the price theory. This is a crucial first step in applying machine learning methods to the transactive energy framework. The simple experiments presented in this contribution demonstrate the capabilities of T-REX as a platform to empirically test market designs. More importantly, they clearly show that certain properties of the market can greatly influence the agent behaviour.

*Keywords:* Artificial intelligence, electricity markets, power system economics, simulation software, solar energy, market design, microgrid, transactive energy

*Corresponding author

## 1. Introduction

Sustainability considerations, empowered by recent technological breakthroughs in renewable power generation, battery storage technology, and advanced data logistics systems are changing the nature of the electric grid. The proliferation of distributed energy resources (DER) is an especially important tool in the effort to continuously curb global carbon emissions, but not without consequences. The highly intermittent nature of renewable energy sources (RES) brings operational challenges that the current grid infrastructure was not designed to handle. As a result, phenomena such as voltage fluctuations, reverse power flow, and degraded power quality [1, 2, 3] are expected to be more frequent if RES and DER are not properly integrated into the grid. If these problems are not appropriately mitigated, operational detriments may overcome environmental benefits, making RES and DER less appealing and slowing their adoption.

Transactive energy (TE) is a framework designed to capture the nature of the newly evolved grid and to accelerate the engineering and deployment of flexible, robust, and adaptive energy management systems. The GridWise Architecture Council (GWAC) defines TE as "a set 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." [4]. This definition provides a high level view, but it is too broad and vague to develop specific solutions and standards from. To properly implement TE, it is necessary to test out different combinations of devices, control algorithms, policies, etc. to identify which deliver the best value overall. Unfortunately, TE's conceptual novelty results in an insufficient amount of data to build reliable models from. Worse yet, attempting to gather TE implementation-specific data through real life deployments can be expensive, it is in itself risky without sufficient data to justify, and it still leaves the question of how generalizable the collected data is. This chicken-and-egg problem is one of the reasons behind the need for TE simulation software that would provide cybernetic testing grounds to investigate solution approaches and generate virtual data. The lack of empirical data makes it especially important for TE simulators to accurately reflect real-life scenarios for the results to be reliable. This article presents T-REX, short for transactive renewable energy exchange, a TE simulation framework that enables model-free approaches to address the weaknesses of existing TE simulators. Namely, it facilitates high time-resolution, agent-based simulations without using aggregator-based modelling.

The rest of the article is organized as follows. Section 2 provides a background on TE, existing TE simulators, and transactive control (TC). Section 3 details the design philosophy and architecture of T-REX [5]. Section 4 first demonstrates the flexibility of T-REX by showcasing the ease of setting up both a classic TC co-simulation, as well as an agent-based model-free simulation. Our proposed TE implementation, Autonomous Local Energy Exchange (ALEX) is introduced next. Section 5 uses T-REX to empirically test and select a market design that is both conducive to reinforcement learning Barto and Sutton [6] and capable of training agent behaviours to truthfully reflect price theory. Finally, conclusions and directions for future work are presented.

## 2. Related Work

To provide insight into the lessons learned and challenges yet to be addressed, the first part of this section reviews the successes and failures of recent developments and trends in demand response (DR) and TE. The latter part investigates the tools available to apply this knowledge and to drive future TE paradigms.

Surveys by Chen et al. [7] and Abrishambaf et al. [8] provide high level views of DR programs and current TE trends. Chen et al. classify DR programs into two categories, both aiming to positively affect the balance of supply and demand through influencing customer behavior [7]:

- Incentive based programs generally implement direct load control (DLC) on certain appliances, such as air conditioners. This allow utility companies to directly limit the usage of these devices when loads are high, in return for financial compensation.
- Price-based approaches, such as time-of-use (TOU) pricing, try to match electricity prices with the overall load on the grid. Their goal is to dissuade customers from using electricity when prices are high, and to encourage usage when prices are low.

Chen et al. observe that customers generally prioritize comfort over responding to price signals in price-based DR programs and prefer certainty over unknown inconveniences of incentive-based DR programs [7]. They conclude that technologies that enable automated response will be the key to making DR and DR-like programs successful.

TE as a framework has the potential to be more flexible and adaptable to customer needs than DR. Abrishambaf et al. [8] provide a comprehensive review of TE research trends and highlight the challenges that need to be addressed. They find that most research in TE focuses on mathematical models and formulation over considering practical implementation details. They also note that most implemented TE systems demonstrate wide gaps between expectations and results. This further emphasizes the need for proper simulation tools for technical verification during the design phase to prevent implementation failures. We conclude that the ideal simulation platform must be suited to confront the practical challenges of implementing TE. These include infrastructure requirements, networking and communication performance, physical and cybersecurity considerations, and electrical grid conditions.

### *2.1. TE Simulators*

The design and development of a TE simulator is one of the primary focuses of this article. This section therefore investigates available TE simulation tools with a focus on whether they can adequately address the problems and challenges identified by the surveys [7, 8]. A report by NIST [9] is a great starting point for review. It details four simulators designed for the NIST Transactive Energy Modelling and Simulation Challenge for the Smart Grid. To provide a wider coverage, we extend our review beyond the NIST TE Challenge and include the following TE simulators: FNCS [10], TESP [11], IESM [12], SEPSS [13], C2WT-TE [14], and TE-SAT [15].

Although their software architectures differ, there is a common thread between all reviewed TE simulators: they primarily focus on maintaining or achieving the optimal power flow within the simulated scenario through co-simulation, an approach generally known as transactive control (TC). The two most popular sub-simulators are GridLab-D [16, 17] for electricity grid simulation and Matpower [18] for powerflow analysis. The simulation components are generally model-based instead of using metered data, and automation is performed with expert-designed controllers where applicable. Network layer models are sometimes included to emulate the effects of degraded data to gauge system performance under edge cases.

While these simulators attempt to bring together existing tools into a more cohesive environment, they leave several identified challenges unaddressed. As previously mentioned, the lack of empirical data to build accurate models makes model-based approaches questionable in general. The lack of TE-specific data further extends to hinder the methods presented alongside these simulators. The generally low temporal and spatial resolution means that they cannot represent the complex, minute-by-minute, customer-to-customer interactions required to realistically capture the transient properties that truly define TE. This lack of granularity is further emphasized through aggregation-based modelling methods that typically approximate the net interactions of hundreds, if not thousands of customers at distribution nodes. Without the ability to accurately capture the individuality and nuances of these interactions at high time and spatial resolutions, fine-grained control opportunities that ultimately affect the aggregated net load at the distribution nodes cannot be realistically estimated. This is a critical weakness that prevents the effective engineering of modern, highly targeted, peer-to-peer smart grids.

The identified weakness can be addressed through an agent-based, model-free, and meter data-driven solution that can operate at high spatial and temporal resolution. T-REX has been designed from the ground up for simulating customer-to-customer interactions at the very edge of the grid, using real metered data at subminute resolution. Its modern software design approach allows advanced artificial intelligence (AI) and machine learning (ML) frameworks to be easily integrated for agent design and control. Its hardware-scalable architecture takes advantage of modern computing infrastructure and allows long-term, high-resolution, large-scale simulations to be completed within a reasonable amount of time.

### *2.2. Transactive Control Implementations*

In addition to the simple demonstrations shown alongside the TE simulators reviewed above, several authors attempted to design and implement more realistic TC systems. Hu et al. [19] proposed an aggregator-based optimization approach that generates charging/discharging schedules for electric vehicles (EVs). A simulation was performed on a Danish distribution network to show the decreased frequency of line congestion and voltage violation. Similarly, Nazir et al. [20] used an aggregator-based model incorporated into a model predictive control (MPC) framework to both calculate optimal price signals and to control thermostatically controlled loads (TCLs) and storage devices to decrease power oscillations at substation feeders. Soarez et al. [21] introduced a comparable aggregator-based approach using a dual decomposition algorithm. To validate the efficacy of the algorithm, they performed a field test involving six houses. Unfortunately, the results were inconclusive due to the lack of flexibility at the test site. This highlights one of the biggest problems with the practical aspects of TE: field tests are rare and expensive. It means that the proposed implementations are often untested. Proposals that have the opportunity to be field-tested are often used to test a hypothesis rather than to validate simulated results, which is an inefficient use of the time and resources available. A proper TE simulator must provide as realistic test environment as possible to maximize the efficiency and reliability of implementations to be validated through field testing.

## 3. Design of a Model-Free TE Simulation Framework

The model-free approach is an enticing way to address the lack of sufficient TE-specific data to build accurate models from. By carefully selecting, designing, configuring, and assembling simulation components, it is possible to generate a large amount of self-feeding data using only a small amount of metered data to initialize the system. This approach greatly reduces assumptions and biases in model building. As a result, it should produce more realistic and accurate outcomes.

### *3.1. System Architecture*

Since the goal of a TE framework is to find an optimal combination of technologies, algorithms, and policies, the architecture of T-REX must be modular by design. Functionally independent modules are tied through a data fabric, which enables simulators of any type or specificity to be composed with ease. Figure 1 shows the simplified architecture diagram of the current simulation framework.

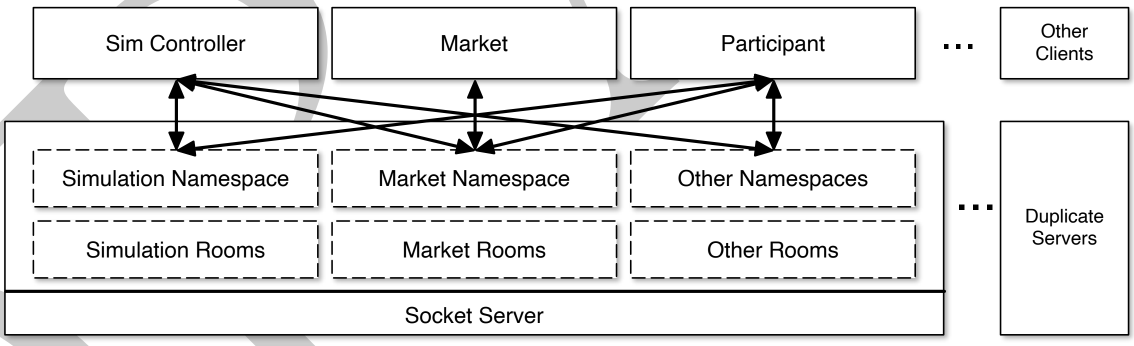


Figure 1: Simplified T-REX V3 Architecture Diagram

The key advantage of this design is horizontal scalability over hardware, which takes advantage of modern computing architecture, i.e. high-core count CPUs and compute clusters designed for highly parallel applications. This not only yields faster simulations when compared to a single-threaded approach, but also enables agent-based, interaction-focused approaches that are unfeasible using other existing TE simulators.

#### 3.1.1. Data Fabric

The foundation of T-REX is the data fabric, which is built using a real networking system instead of a networking model. The implementation is based on `socket.io` [22], an established, powerful and widely used real-time messaging engine. This guarantees compatibility, scalability, and reliability. The asynchronous, highly parallel design eliminates the need for pseudo-random sequence queues that are typically a major source of bottle-necks. It also enables standard networking performance and penetration testing techniques to truly evaluate performance and cybersecurity aspects of the TE systems. If `socket.io` is used in the deployment, then a high level of consistency and re-usability can be achieved, greatly streamlining the implementation pipeline.

#### 3.1.2. Modules

The functional modules of T-REX are built as `socket.io` clients. As in the deployment case, interactions between modules are facilitated by passing messages using the `socket.io` API. Although designers are free to use payloads of any permissible size, format, and endpoints, care should be taken to preserve genericity and minimize bandwidth usage. There have been three main classes of clients implemented, as shown in the architecture diagram and described below:

- Participant modules, which are in charge of energy trading and managing energy resources that are directly accessible. Participants are, for example, households and self-driving EVs.
- Non-participant modules, e.g., the TE market. The market facilitates the discovery and exchange of energy between participants.
- Simulation-only modules, e.g., the simulation controller, or a powerflow calculation module. The simulation controller augments the deployment environment to form a simulation model. It can also perform advanced functions such as training curricula for ML applications.

With a few restrictions pertaining to the simulation mode, the number of modules of each type is unlimited. The functions are also not restricted to the list described above. For example, a traffic module can function in parallel with multiple markets to guide self-driving EV participants to find optimal paths to carry passengers in conjunction with charge and discharge locations to maximize profit. Other modules that do not use traffic data to make decisions simply do not know about its existence.

## 4. Implementing TE in T-REX

This section demonstrates two use cases for T-REX. They implement: (i) a powerflow-driven TC co-simulation, and (ii) a novel, agent-based, economy-driven TE market called Autonomous Local Energy Exchange (ALEX), that is not easily modelled in existing TE frameworks. It is important that T-REX supports both implementations, so that outcomes can be compared in a fair and consistent manner. All simulations are composed using a JSON configuration file. The configuration files for each of the TE implementations shown in this section can be found in the GitHub repository [23] for this article.

### *4.1. Backward Compatibility: Transactive Control Co-Simulation*

To reiterate, the premise behind TC can be summarized in the following steps:

1. A price, or a sequence of time-based prices, is generated ahead of time based on the predicted optimal power flow.

2. The prices are sent to the end-users so that their behaviours may change with the prices. Automated price-reactive agents may be used to automatically control certain appliances to shape loads based on the prices.

3. Repeat from step 1.

Simulation controller aside, the rest of this simulation setup can be composed using the following modules:

- A non-participant powerflow module. The current implementation is based on OpenDSS [24] and its Python API [25]. Other compatible powerflow software can also be integrated.

- TC Market with a submodule that generates prices based on the received powerflow data.

- Participants containing load profiles, controllable devices, and price-reactive logic.

Each module component has more specific parameters that can be configured. It is important to stress that the parameters should be configured so that each module is aware of the context of the simulation. In the case of TC co-simulation, the "powerflow" section must be available and include all relevant network parameters. The number of agents should match the number of loads in the network, except for static loads.

Figure 2 shows a highly simplified version of the sequence flow diagram of the TC co-simulation implemented in T-REX. Due to the asynchronous nature of T-REX, many independent asynchronous functions and parallel loops have been omitted from the diagram, and only an approximation of the main flow path is shown.

### *4.2. ALEX: Autonomous Local Energy Exchange*

Since TC can be considered a superset of DR, the same issues mentioned in Section 2.2 apply. Because active participation cannot be guaranteed in most DR programs (even with relatively simple price schedules), the much more fine-grained pricing used in TC will likely make little difference unless automation and other technologies are used to eliminate humans from the control loop. Fundamentally, nonparticipation is caused by a misalignment of values and interests between human

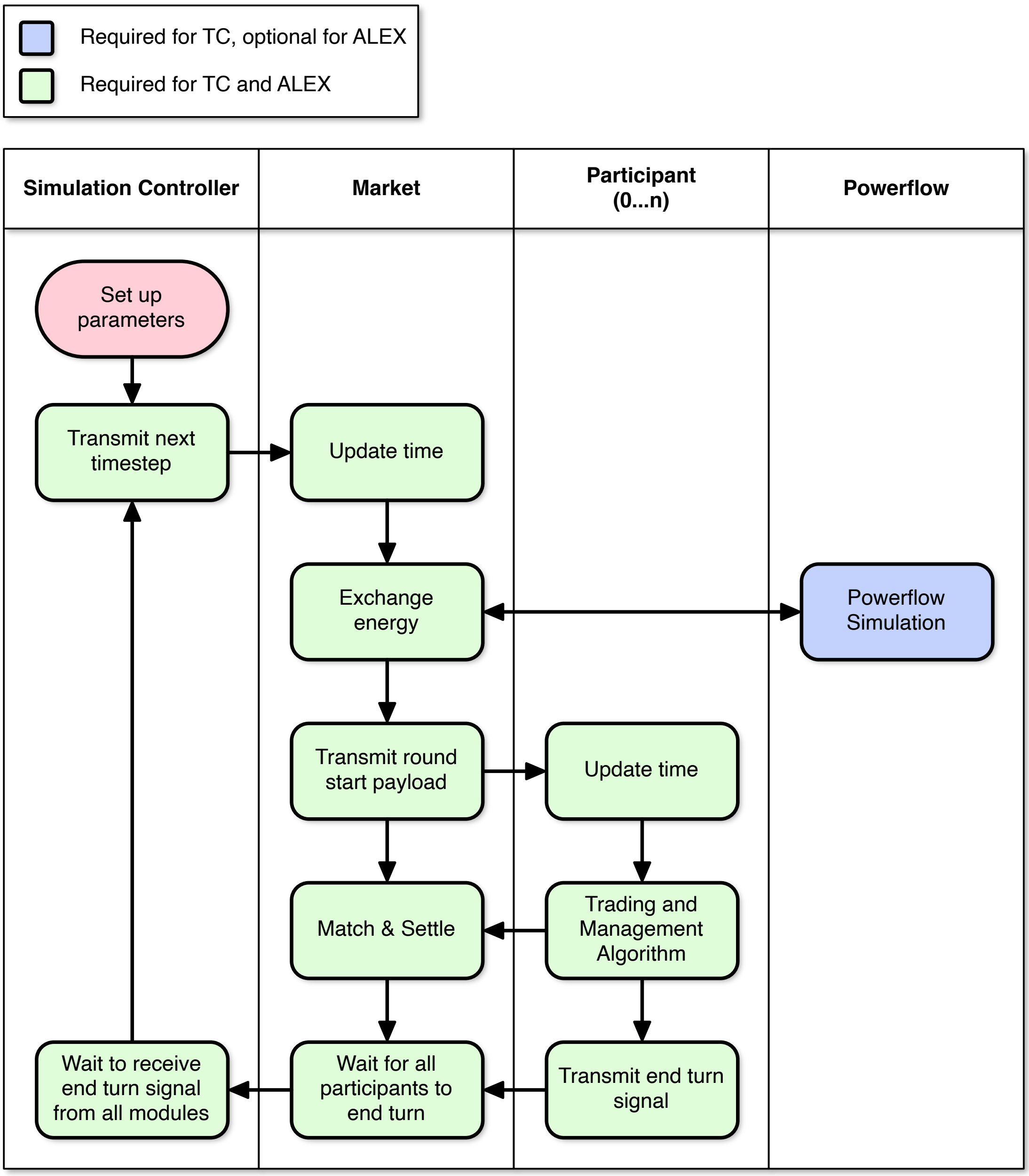


Figure 2: Simplified swimlane diagram of TE schemes implemented in T-REX

users and price setters. A well-known example of this is stated by Chen et al. [7]. For the customers, the change of electricity bill is negligible in comparison to the efforts required to make behaviour changes to fully participate in DR programs.

ALEX aims to address these issues by leveraging basic market economy principles to determine prices for control and coordination, and reinforcement learning (RL) for automation. RL is highly suited for integrating both stationary and mobile DERs onto the grid due to its adaptability to environmental changes, and robust performance in partially observable environments, which are necessary during the grid transition period. In the right environment, RL agents can develop highly complex strategies to maximize returns. To exploit this fact and to maximize the effectiveness of RL, an appropriate price determination mechanism is required. This mechanism should be implemented so that the true value of energy commonly accepted by all participants, at a given time and location, is determined by the energy resources accessible and the intended behaviour of the agents. Once the prices are determined, agents can use learned knowledge to choose the optimal mix of resources that maximizes profits. When the balance of supply and demand is accurately reflected by the value of energy, the act of competing for maximum profit creates a distributed competitive-collaborative process that constantly moves the power balance closer to the optimum, as an emergent side effect. This effectively re-aligns the goals of grid operators (optimal powerflow) and customers (maximum profit or savings with minimal effort). Interested read may refer to the seminal work of Sutton and Barto [6] that provides a comprehensive background of RL.

Similar proposals of market-based control in favor of TC are not a new concept [26]. However, it is understandable that the sense of losing direct control by relegating a portion of powerflow optimization as an emergent property of the market function can be concerning. This further justifies the need for a flexible and robust TE simulator that can accurately simulate alternative TE approaches (such as ALEX) to develop alternative system configurations, evaluate their efficacy, and verify their security.

The composition of ALEX, similar to the TC co-simulation, is a testament to the high-level abstractions offered by the modular architecture of T-REX. The appropriate market module and participant logic suitable for ALEX must be supplied accordingly. The key difference from the TC co-simulation is that the supply/demand balance of power is achieved through the market mechanism alone. Therefore, powerflow has no bearing on the determination of prices. As a result, powerflow calculations are optional and can be run in parallel with the rest of the simulation. However, because it is still critical for grid planning and design, powerflow can be simulated after-the-fact using the data from the energy transaction ledger. The main simulation sequence is largely unchanged from Figure 2, with the powerflow module removed to increase the simulation speed.

To reiterate, implementing ALEX requires the following steps to be taken:

1. A market mechanism must be designed and validated to be conducive to RL. As the learning environment, the market settlements should produce reward signals that reflect price theory.

2. Trading agents should be designed to learn strategies that can leverage all available energy resources to maximize net profit.

The design and implementation of the market and participant modules are detailed next to develop some high-level insights. The compatibility of the settlement mechanism with RL is elaborated on in Section 5.

### *4.2.1. Market*

Price theory states that the price for any specific good or service is based on the balance of supply and demand. In a market-based TE approach, the role of the market is to efficiently facilitate the exchange of energy so that the price can appropriately and accurately reflect the balance of supply and demand at the time of exchange. ALEX adapts and adjusts an existing market design to fit three key considerations:

1. **Suitability for electricity grids with high penetration of DER and RES.** This means that, from a high level perspective, a market (or a collection of markets) must be able to effectively target localization and the transient nature of RES.

2. **Technical constraints and requirements of deployment:** Data acquisition, transportation, and cost must be minimized.

3. **Machine learning considerations for agents** (see section 4.2.2): Related to the point above, ML will play an important role in trading and managing of energy resources in place of humans. For this reason, the market should be conducive to learning. One way to achieve this is to compose the market with a small set of explicit rules. The rules should provide a strong feedback signal, and they should be flexible enough to offer large action spaces.

With these considerations in mind, the final market is a modified form of double auctions [27][28]. The rules, explicitly implemented in the code, are described below:

1. It is assumed, for the time being, that the grid is an infinity bus and it can be interacted with through net-billing. We therefore adapt retail electricity prices in Alberta, where buying energy from the grid costs \$0.1449/kWh, and selling earns \$0.069/kWh.

2. The local market has two energy pools: one for dispatchable sources, such as battery energy storage systems (BESS), and one for non-dispatchable sources, such as photovoltaics (PV). This is intended to distinguish the source of energy, and to allow for the value of dispatchability to emerge.

3. Auctions settle for energy to be delivered during the one-round period from the end of the current round. However, the delivery period can be parametrically adjusted during run-time for future design explorations.

4. During the current round, participants submit bids and asks for energy to be delivered during or beyond the next delivery period.

5. A modified double auction system is used to settle trades: bids/asks are settled pairwise, with bids sorted from the highest to lowest, and asks in reverse to ensure pareto equality.

6. Bid/ask quantities can be partially settled.

7. A bid/ask quantity must be an integer multiple of 1 Wh. This is in consideration of future hardware integration, to allow direct use of the watt-pulse function of most smart-meters.

8. During the delivery period, if a seller is in short supply, it must financially compensate for the shortage at net metering prices. If batteries are available, the seller has the option to compensate by discharging its batteries, for all or part of the shortage during this period.

9. During the delivery period, if a buyer settled for more energy than used, the buyer must still pay the seller for the unused energy at the settlement price.

Settlement methods are designed to be modular and excluded from these overarching rules. The design of the settlement mechanisms is part of the case study introduced in this article, described in Section 5.

#### *4.2.2. Participants*

The market participants are individual customers located at the very edge of the grid, which are typically human participants. However, the difficulties of accurately modeling individual human behaviour impede the degree of real-world carry-over of such human-in-the-loop approaches. We instead propose an AI focused approach, where AI agents must learn to compensate for nonoptimal human behaviour using the energy resources available, while maximizing profits. Such AI agents can be trained on easily accessible, highly accurate, and granular real-time meter data. Sufficient data mass captures nuanced and individualized human behavioural patterns and displaces the need to model human behavior inside the training loop.

Recent successes of OpenAI and Google in game environments [29, 30] have demonstrated the ability of deep reinforcement learning (DRL) agents to excel in complex, strategy-rich, and partially observable environments. However, a major roadblock for transferring such technology to the application of TE is the lack of a suitable training environment for such agents. T-REX and ALEX are purposely designed to fill this gap. The following section describes the use of T-REX to empirically select the most suitable market settlement mechanism for RL.

## 5. Case Study: Settlement Mechanism Compatibility for Reinforcement Learning

As previously mentioned, ALEX requires both a market that can produce settlement prices reflecting the balance of supply and demand, and RL agents that can learn the appropriate behaviours to maximize income. The goal of the experiments performed in this case study is to develop an appropriate settlement mechanism that drives agent behaviour to naturally reflect price theory. This is a prerequisite first step to allow for introducing more complex components such as DRL agents and controlled residential BESS in later developments.

Section 5.1 details the experimental setup, agent design, and settlement mechanisms tested in this study. The experimental results are described in Section 5.2.

| Market | Budget Balancing | Truthfulness | Strictness |
|---|---|---|---|
| M1 | Yes | No | Low |
| M2 | Yes | Yes | High |
| M3 | Weak | Yes | Low |

Table 1: Summarized settlement mechanism properties

### *5.1. Experimental Design*

Within ALEX, agents interact with each other over a localized energy market using a settlement mechanism. The experiments described in this section investigate the equilibrium states of three settlement mechanisms. The mechanisms are named and described as follows:

1. Average-Price (M1): Trades are settled if the bid price is greater than or equal to ask price. The settlement price is the average of the bid and ask prices.

2. Exact-Match (M2): Sellers and buyers choose bid and ask prices from a list of available prices. Trades are settled if the bid price equals the ask price.

3. Exact-Price (M3): Trades are settled if the bid price is greater than or equal to the ask price. The buyer buys from the auctioneer at the bid price, and the seller sells to the auctioneer at the ask price.

According to Myerson and Satterthwaite [31], a double auction mechanism can be defined by the following properties: individual rationality, budget balancing, truthfulness, and Pareto efficiency. An ideal mechanism would fulfill all four, but is not realistically possible. Since ALEX's setup guarantees both individual rationality and Pareto efficiency, M1-3 can be differentiated by their budget balancing, truthfulness, and strictness of settlement. These properties are summarized in Table 1.

To quantify a mechanism's ability to reflect price theory, three scenarios with different supply/demand ratios are evaluated: over-supply (10:1), over-demand (1:10), and balanced (5:5). Agents are given a synthetic flat energy profile corresponding to this ratio. It is necessary to include at least two agents of each type to prevent either a supply or demand monopoly situation, which would result in prices settling into an arbitrary equilibrium due to M1-3's individual rationality properties.

Each participant is represented by a simple RL agent that aims to maximize the return - exponentially weighted, cumulative sum of rewards that depends on its actions $a$ and observed states $s$, as follows

$$G = \sum_{i}^{T} \gamma^i r_{t_i} \tag{1}$$

where $G$ is the return, $\gamma$ is the discount factor, $t$ is the current step, and $T$ is the end of the episode. The logical choice of the reward function is the net profit of each action, and the available actions are bid/ask prices. To accommodate M2 and to further simplify the experiment, the action space is selected as a table of 100 evenly spaced, discrete prices between \$0.07/kWh and \$0.14/kWh. The current experiments are designed with tabular RL agents and synthetic data. The development of a reliable training pipeline for DRL agents using real metered data in such a complex, multi-agent, partially observable environment is a nontrivial task that is out of the scope of this article.

Keeping the market in constant equilibrium makes it easy to judge whether the price theory is properly reflected. It also greatly simplifies the RL problem. By using flat energy profiles for all participants, the problem is reduced to a stateless, probabilistic (multi-arm) bandit [6]. This allows many tabular RL algorithms to be used. Q-learning has been selected, as it has strong convergence guarantees within such settings [6]. Q-learning is an established off-policy RL algorithm, learning the value of a state-action tuple $Q(a_t, s_t) = \mathbb{E}\, G_t$ for the value-maximizing target policy $\pi$ by exploring the environment with a behavioral policy $b$. It is usually a $\epsilon$-greedy policy, following $\pi$ with a probability $1-\epsilon$ and taking a random exploratory action with probability $\epsilon$. This leads to the following learning update

$$Q_{\text{new}}(s_t, a_t) = (1-\alpha)Q_{\text{old}}(s_t, a_t) + \alpha\big(r_t + \gamma \max_a Q(s_{t+1}, a)\big), \tag{2}$$

where $\alpha$ is the learning rate, and $\gamma$ is the discount factor [6].

The following default hyperparameters are used for training: $\alpha = 0.1$, $\gamma = 0.98$, and $\epsilon = 0.1$. Training is performed over 100 episodes, with the length of each episode being 14 days at 1-minute steps, for a total of 2,016,000 learning steps.

### *5.2. Results and Discussion*

A successful combination of market design and converged agents is expected to produce the following results:

- For the excess supply case, the generators compete for demand. This drives ask prices low with bid prices following.
- For the excess demand case, the consumers compete for supply. This drives bid prices high with ask prices following.
- For the equal supply and demand case, the bid and ask prices should be around the middle of the available price window.

Figures 3, 4, and 5 show the price concentrations of bids, asks, and settlements for each settlement mechanism and different ratios of supply and demand. The scatter plots are produced by aggregating the quantity bids, asks, and settlements over the final 10 episodes of the exploration policy, and then normalized to remove the effects of supply/demand imbalance for ease of visualization. For all three plots, normalized quantities below 0.15 can be safely attributed to exploratory actions and considered noise.

Of the three settlement mechanisms tested, one can observe that only M3 shows the aforementioned desired behaviours for all scenarios. While M1 and M3 show similar behaviour in the unbalanced case, M1 clearly fails in the balanced case. M1's lack

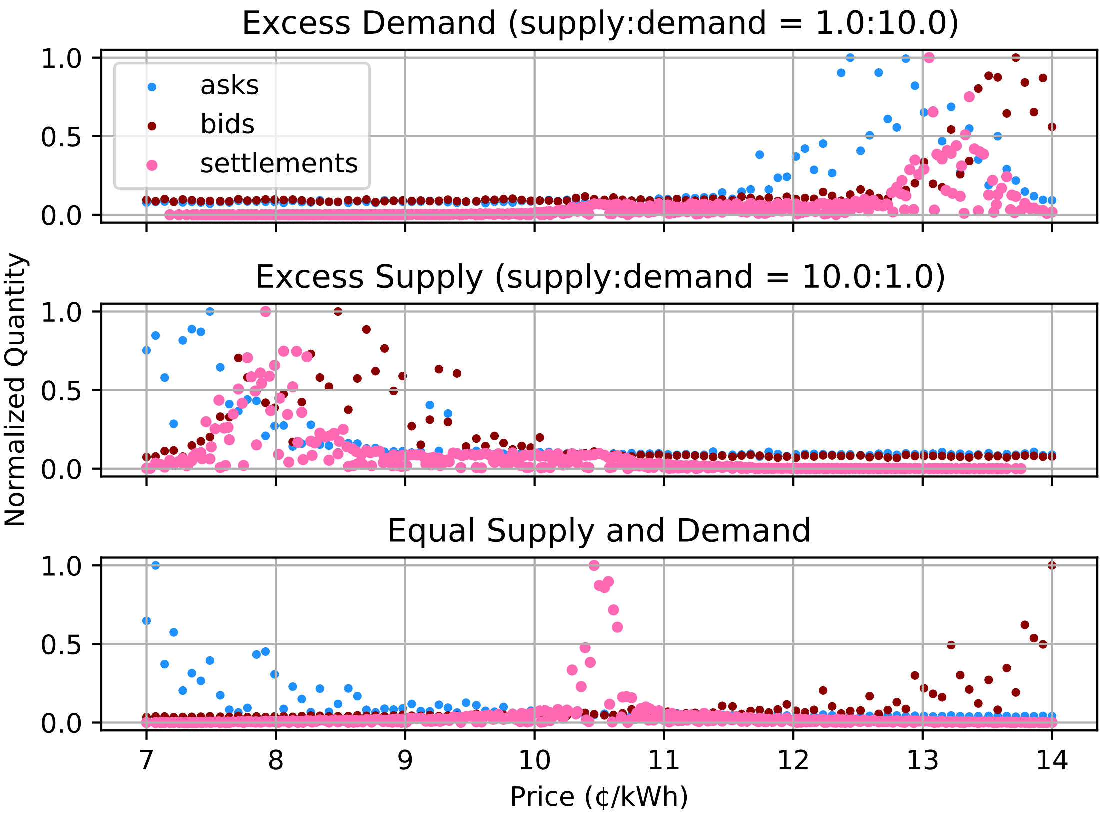


Figure 3: Price concentrations of M1 training policy, which does not have the truthfulness property, clearly makes agent actions diverge to the extremities of their action spaces for the balanced scenario. This shows that truthfulness is a mandatory property.

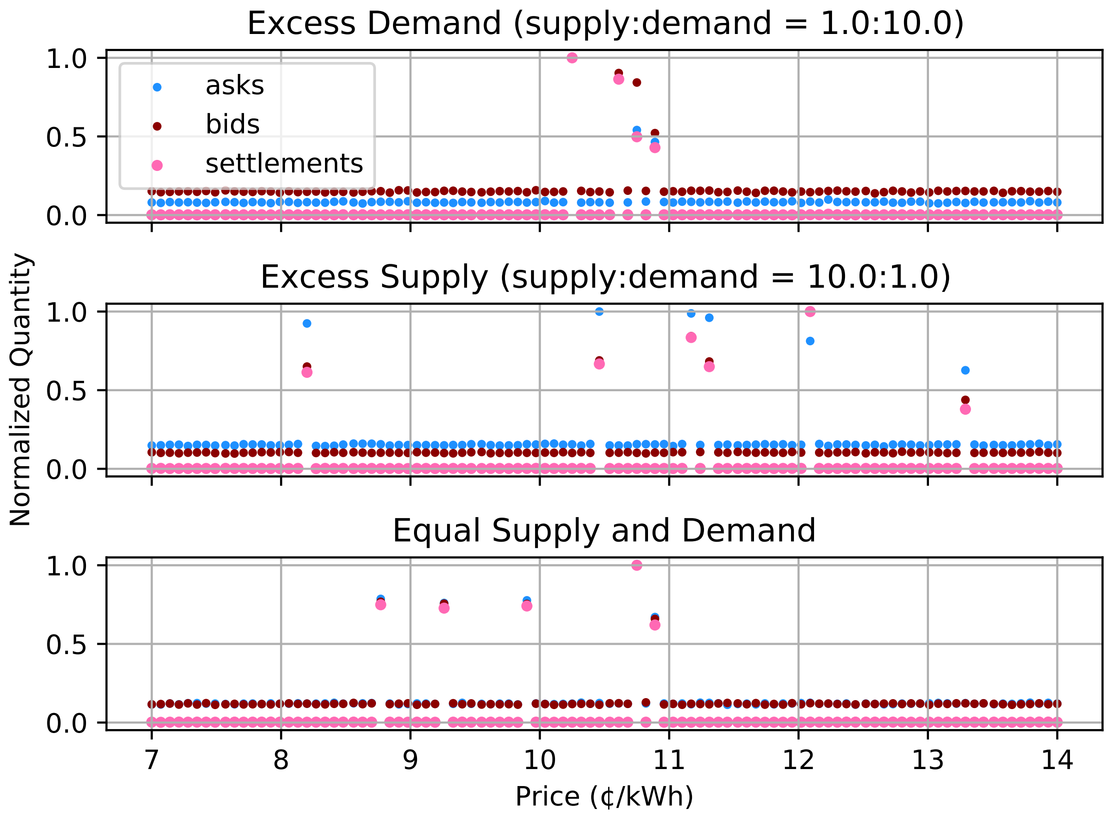


Figure 4: Price concentrations of M2 training policy, which clearly shows that the strictness of the settlement mechanism is non conducive to learning. Actions failed to converge where expected for the same amount of training.

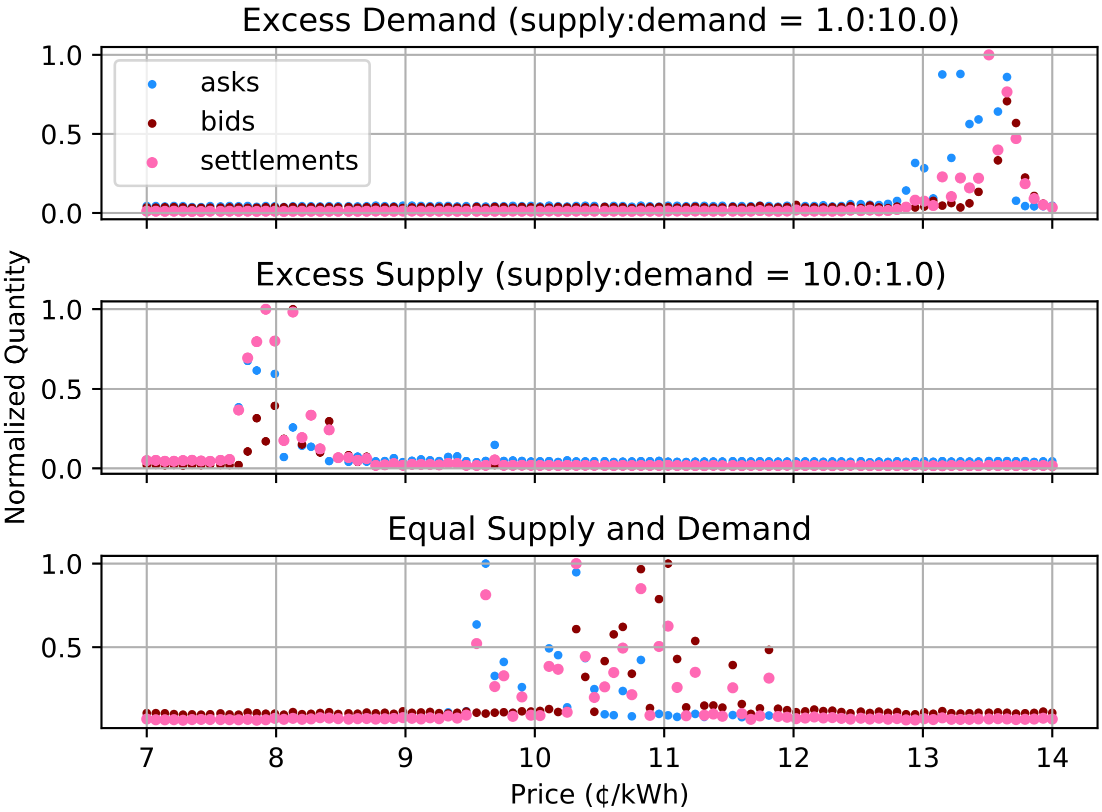


Figure 5: Price concentrations of M3 training policy, which clearly shows the expected behaviours in accordance to price theory. It is notable that, in contrast to Figure 3, bid and ask prices converged to one side of action space for both buyers and sellers, especially in the unbalanced cases.

of truthfulness incentivizes buyers to bid higher and sellers to bid lower to maximize the chances of making a trade. M2's failure to produce the desired results can be attributed to its strictness, which gives the agents much less opportunity to learn, and thus prevents the emergence of competent market participation. The key takeaway is that truthfulness is the most important property for ALEX, and likely for automated transactive markets in general. To further validate M3, training was extended by 100 episodes, while annealing both learning and exploration rates with a multiplier of 0.98 per episode. In this specific setting, annealing guarantees convergence given a sufficiently large number of episodes [6].

Figure 6 shows the aggregated and normalized results for the final 20 episodes using the greedy policy. More episodes are used in comparison to the training policy plots to show more variance over episodes and to compensate for the sparsity of the greedy policy. It is clear that the action and settlement prices generally correspond to the level of supply demand balance.

Another interesting result can be seen from the equal supply and demand case, where the bid and ask prices are identical. This shows that agents in M3 are able to achieve strong budget balancing, even though it is not strictly a property of M3. Recall that M2, which was explicitly designed to be strongly budget balanced, failed to achieve this goal. M2's high strictness greatly hindered the training of the RL agents, due to the lack of strength and frequency of the reward signal.

The extent to which the settlement mechanism impacts the observed behavior was not expected. Truthfulness is clearly the most important attribute for this application, as it is the property that drives agent actions to expected price ranges in accordance with price theory. Overly strict mechanisms greatly hinder learning and produce nonoptimal behaviours given the same training time. Finally, the experiments have shown that some theoretically undesirable properties, such as weak budget balancing, can be overcome by agent rationality. This shows the great importance of experimentation to truly understand how market design can drive participant behaviour.

## 6. Conclusion and Future Work

This article introduces the Transactive Renewable Energy Exchange (T-REX), a transactive energy simulation platform. In contrast to the contemporary model-based co-simulators, T-REX adopts an agent-based, model-free, and data-driven simulation approach that eliminates the model-related challenges faced by other existing approaches. By designing for modern software frameworks and hardware architecture, T-REX offers much more fine-grained and realistic simulations while keeping superior simulation time. Direct integration with modern ML libraries such as `Tensorflow` makes T-REX a perfect environment to apply modern AI approaches in one of the most urgent real-world applications.

The comparative advantages of T-REX are presented by first showing the ease of implementing both a contemporary TC co-simulation and an agent-based, economy-driven TE approach. The later is then exemplified by proposing a novel model-free, economy-driven, and ML-centric TE approach called ALEX, including the steps

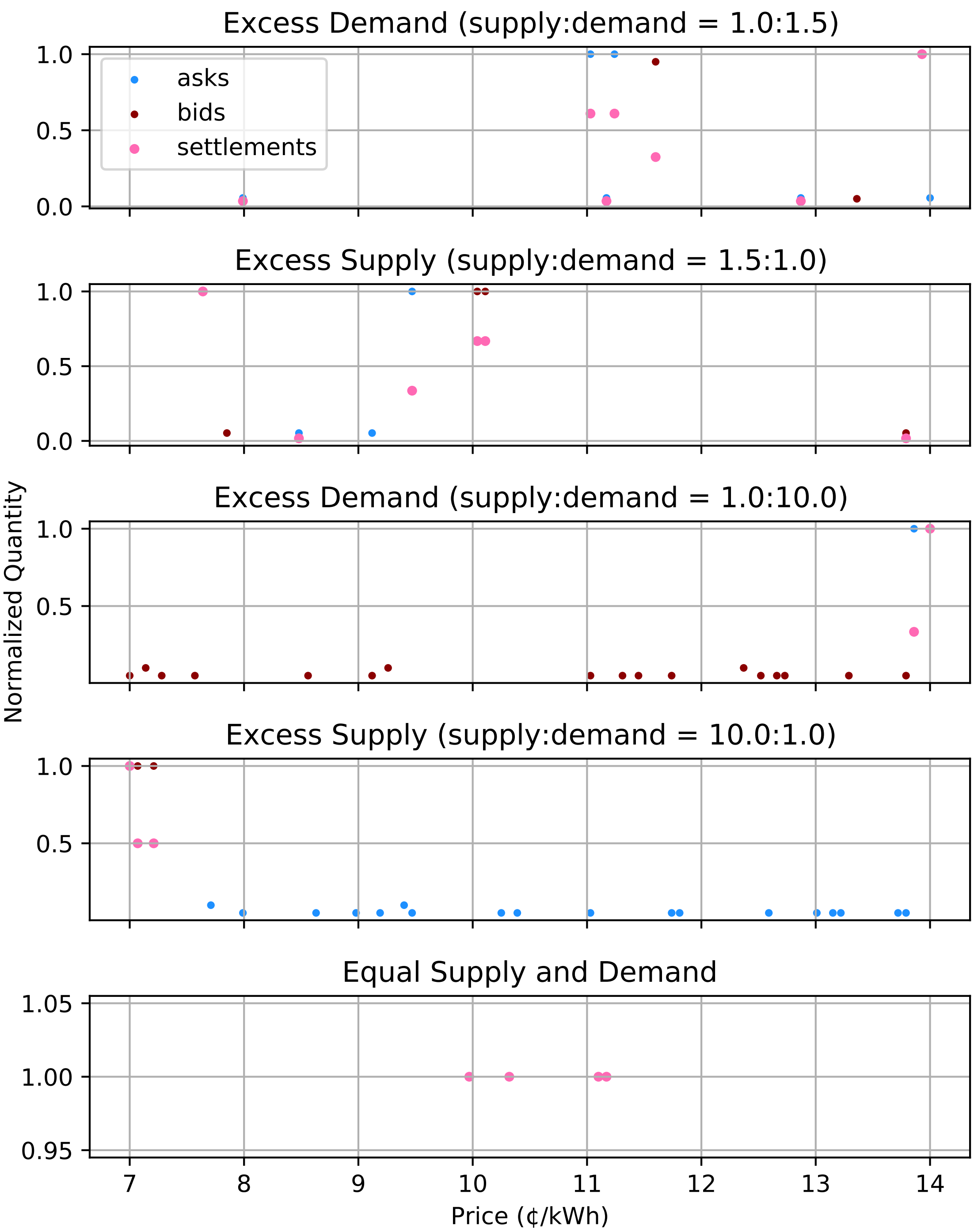


Figure 6: Price concentrations of M3 greedy policy after annealing both learning and exploration for an additional 100 episodes. Less extreme supply and demand ratios have been tested to verify that the range of settlement prices generally corresponds to the balance of supply and demand, as expected.

needed for its successful implementation. T-REX's capability as a design tool is demonstrated through experiments designed to select the market settlement mechanism most compatible with RL. The experiments reveal that the most compatible market mechanism must satisfy the truthfulness property. Furthermore, too much strictness can greatly hinder learning and it should be relaxed where applicable. These findings are crucial to lay the ground for future work. The two immediate areas of interest are deeper research into advanced agent design in more realistic scenarios, as well as training agents to use BESS. The second topic is particularly interesting, as the combination of ALEX and the temporal shifting properties of battery storage could allow a natural emergence of ancillary services.

## References


[1] S. M. Ismael, S. H. A. Aleem], A. Y. Abdelaziz, A. F. Zobaa, State-of-the-art of hosting capacity in modern power systems with distributed generation, Renewable Energy 130 (2019) 1002 – 1020. URL: http://www.sciencedirect.com/science/article/pii/S0960148118307936.

[2] T. O. Olowu, A. Sundararajan, M. Moghaddami, A. I. Sarwat, Future challenges and mitigation methods for high photovoltaic penetration: A survey, Energies 11 (2018).

[3] M. P. Chapelsky C, Gerasimov K, Der impacts to urban utilities study summary (2019). URL: https://www.epcor.com/products-services/power/Documents/micro-generation-research-solar-energy-electricity-grid-2019.pdf.

[4] D. Forfia, M. Knight, R. Melton, The view from the top of the mountain: Building a community of practice with the gridwise transactive energy framework, IEEE Power and Energy Magazine 14 (2016) 25–33.

[5] S. Zhang, TREX-Core, https://github.com/trex-ai/TREX-Core, 2020.

[6] A. Barto, R. Sutton, Reinforcment Learning, an Introduction, 2 ed., MIT Press, 2018.

[7] S. Chen, C.-C. Liu, From demand response to transactive energy: state of the art, Journal of Modern Power Systems and Clean Energy 5 (2017) 10–19.

[8] O. Abrishambaf, F. Lezama, P. Faria, Z. Vale, Towards transactive energy systems: An analysis on current trends, Energy Strategy Reviews 26 (2019) 100418.

[9] D. Holmberg, M. Burns, S. Bushby, A. Gopstein, T. McDermott, Y. Tang, Q. Huang, A. Pratt, M. Ruth, F. Ding, Y. Bichpuriya, N. Rajagopal, M. Ilic, R. Jaddivada, H. Neema, NIST transactive energy modeling and simulation challenge phase II final report, Technical Report NIST SP 1900-603, National Institute of Standards and Technology, Gaithersburg, MD, 2019.

[10] S. Ciraci, J. A. Daily, J. C. Fuller, A. R. Fisher, L. D. Marinovici, K. Agarwal, Fncs: A framework for power system and communication networks co-simulation (2014).

[11] Q. Huang, T. E. McDermott, Y. Tang, A. Makhmalbaf, D. J. Hammerstrom, A. R. Fisher, L. D. Marinovici, T. Hardy, Simulation-Based Valuation of Transactive Energy Systems, IEEE Transactions on Power Systems 34 (2019).

[12] S. Mittal, M. Ruth, A. Pratt, M. Lunacek, D. Krishnamurthy, W. Jones, A System-of-Systems Approach for Integrated Energy Systems Modeling and Simulation: Preprint (2015).

[13] M. D. Ilić, R. Jaddivada, X. Miao, Scalable Electric Power System Simulator, in: 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2018.

[14] H. Neema, J. Sztipanovits, M. Burns, E. Griffor, C2WT-TE: A model-based open platform for integrated simulations of transactive smart grids, in: 2016 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), 2016. doi:10.1109/MSCPES.2016.7480218.

[15] H. Neema, J. Sztipanovits, D. J. Hess, D. Lee, TE-SAT: Transactive Energy Simulation and Analysis Toolsuite, in: 2020 IEEE Workshop on Design Automation for CPS and IoT (DESTION), 2020.

[16] D. P. Chassin, J. C. Fuller, N. Djilali, Gridlab-d: An agent-based simulation framework for smart grids, Journal of Applied Mathematics 2014 (2014).

[17] D. P. Chassin, K. Schneider, C. Gerkensmeyer, GridLAB-D: An open-source power systems modeling and simulation environment, in: 2008 IEEE/PES Transmission and Distribution Conference and Exposition, 2008, pp. 1–5. doi:10.1109/TDC.2008.4517260.

[18] R. D. Zimmerman, C. E. Murillo-Sánchez, R. J. Thomas, Matpower: Steady-state operations, planning, and analysis tools for power systems research and education, IEEE Transactions on Power Systems 26 (2011).

[19] J. Hu, G. Yang, C. Ziras, K. Kok, Aggregator operation in the balancing market through network-constrained transactive energy, IEEE Transactions on Power Systems 34 (2019) 4071–4080.

[20] M. S. Nazir, I. A. Hiskens, A dynamical systems approach to modeling and analysis of transactive energy coordination, IEEE Transactions on Power Systems 34 (2019) 4060–4070.

[21] A. Soares, O. De Somer, D. Ectors, F. Aben, J. Goyvaerts, M. Broekmans, F. Spiessens, D. van Goch, K. Vanthournout, Distributed optimization algorithm for residential flexibility activation—results from a field test, IEEE Transactions on Power Systems 34 (2019) 4119–4127.

[22] D. Arrachequesne, Socket.io, 2021. URL: `https://github.com/socketio/socket.io`.

[23] S. Zhang, Trex-ae2021-resources, `https://github.com/sd-zhang/trex-ae-2021`, 2021.

[24] EPRI, Epri distribution system simulator, 2021. URL: `https://sourceforge.net/projects/electricdss/`.

[25] D. Krishnamurthy, Opendssdirect.py, `https://github.com/dss-extensions/OpenDSSDirect.py`, 2017.

[26] E. F. Bompard, B. Han, Market-Based Control in Emerging Distribution System Operation, IEEE Transactions on Power Delivery 28 (2013).

[27] D. Friedman, J. Rust, The Double Auction Market: Institutions, Theories, and Evidence, Reading, Mass., 1993.

[28] D. Friedman, A simple testable model of double auction markets, Journal of Economic Behavior & Organization (1991).

[29] Dota 2 with Large Scale Deep Reinforcement Learning, arXiv:1912.06680 [cs, stat] (2019). URL: `http://arxiv.org/abs/1912.06680`, arXiv: 1912.06680.

[30] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, D. Silver, Grandmaster level in StarCraft II using multi-agent reinforcement learning, Nature 575 (2019). Number: 7782 Publisher: Nature Publishing Group.

[31] R. B. Myerson, M. A. Satterthwaite, Efficient mechanisms for bilateral trading, Journal of Economic Theory 29 (1983) 265–281.