Hierarchical Flexibility Offering Strategy for Integrated Hybrid Resources in Real-time Energy Markets

Majid Majidi, Mehdi Hosseini, Masood Parvania
Department of Electrical and Computer Engineering, The University of Utah
Salt Lake City, UT 84112 USA
E-mails: {majid.majidi, mehdi.hosseini, masood.parvania}@utah.edu

Abstract

This paper proposes a hierarchical model for determining the energy flexibility offering strategy of integrated hybrid resources (IHRs) in power distribution systems to participate in real-time energy markets. The proposed model utilizes the scalability, fast response time, and uncertainty observation of deep reinforcement learning (DRL) to overcome the scalability issue of operating numerous flexible resources and deliverability of energy flexibility to the real-time markets in the presence of the network constraints. To that end, the power distribution system is divided into multiple IHRs, where different types of flexible loads, energy storage systems, and solar plants with controllable inverters are operated through local IHR controllers, trained by deep deterministic policy gradient (DDPG) algorithm. Active power request and reactive power capacity of IHRs are then transmitted to a central flexibility controller, where a quadratic optimization model ensures the deliverability of the energy flexibility to the real-time energy market by satisfying the distribution network constraints. The proposed model is implemented on the 123-bus test power distribution system, demonstrating the capability of DRL-based hierarchical model for scalable operation of IHRs in order to offer deliverable energy flexibility to the real-time energy market.

1. Introduction

Federal Energy Regulatory Commission (FERC) order no. 2222 has provided an unprecedented opportunity for distributed energy resources (DERs) to participate and provide services in the wholesale electricity markets [1]. Integrated hybrid resource (IHR) offers an economic and salable structure for aggregating and coordinating the operation of a large group of DERs to provide the services in the wholesale markets [2–4]. IHRs can facilitate local energy management in power distribution systems to serve load, while supplying the distributed energy flexibility at scale to the energy markets without compromising the operational privacy of DERs and flexible loads [2]. In addition, IHRs can increase the efficiency in control and dispatch of DERs and flexible loads, while reducing the complexity in power distribution system operation integrated with large number of flexibility resources [5].

1.1. Scope and Literature Review

The technical literature includes multiple studies for energy and flexibility management of hybrid resources in power distribution systems. Application of hybrid energy systems consisting of co-generation systems and renewable generation units for congestion management in power distribution systems is studied in [6]. A multi-stage robust optimization model is presented in [7] to assess the role of energy storage systems (ESS) in mitigating renewable generation uncertainty in power distribution systems. Coordinated operation of flexibility resources (i.e., flexible loads, inverter-based solar plants) to provide distributed energy flexibility and regulation capacity in power system is studied in [8, 9]. A local energy market framework in developed in [10] to facilitate energy management of DERs and local energy trading in power distribution systems.

In addition, data-driven methods have been utilized for energy flexibility management of DERs in power system operation. Application of deep reinforcement learning (DRL) for energy management and control of DERs and building loads is studied in [11]. DRL is implemented in a solar-electric vehicle (EV) hybrid energy system in [12] to control charging services of EV, thereby increasing the self-consumption of solar generation. The authors in [13] developed a cooperative framework based on DRL for energy management of EV charging stations integrated with solar plants and ESS. DRL-based models are developed to enable control and local energy trading of different sorts of DERs in [14, 15]. The authors in [16] developed a DRL-based model for local control and dispatch of DERs to improve resilient operation of the power distribution

URI: https://hdl.handle.net/10125/102967
978-0-9981331-6-4

system. The application of DRL for feasible control and management of DERs in the power distribution system is investigated through an analytical approach in [17]. A DRL-based hierarchical framework is proposed in [18] for resilient operation of IHRs in power distribution systems. The authors in [19] developed a data-driven flexibility design for aggregation of flexible loads, respecting their uncertainties and non-convexity in modeling their operation.

As discussed in [6]–[10], most of the existing energy flexibility management models in literature are developed based on centralized optimization-based solution approaches. However, with large-scale integration of DERs with different operational technologies and integration of numerous EV charging stations, the computational efficiency of these approaches to model the distributed energy flexibility and ensure its deliverability to energy markets is still challenging. Moreover, energy flexibility management under the unpredictable behaviour of DERs, such as EVs and solar plants in real-time operation requires scalable data-driven approaches to efficiently learn and decide on the energy control and dispatch of flexibility resources within the IHR. Data-driven and learning-based approaches are more compatible with the distributed operation of distribution systems than optimization-based approaches. Running an optimization-based operational algorithm requires a processing power that might not be readily available anywhere outside the substations. Learning-based operational models, on the other hand, only require a trained regressor or neural network with fixed weights and biases, which can be implemented on field-programmable gate arrays and placed anywhere on pole-mounted control boxes without requiring a separate processing unit, as implemented in [20, 21]. Although some studies in literature considered implementation of data-driven approaches for energy flexibility management of DERs, power distribution network model in these studies is not taken into account, thus challenging the deliverability of the flexibility to the electricity market.

1.2. Contributions

This paper proposes a novel hierarchical model for determining the offering strategy of IHRs in distribution systems to participate in the upstream real-time energy market. The structure of proposed hierarchical model is illustrated in Fig. 1, where the distribution system consists of multiple IHRs that integrate flexible and inflexible loads, ESS and solar plants, which are managed and controlled by the IHR

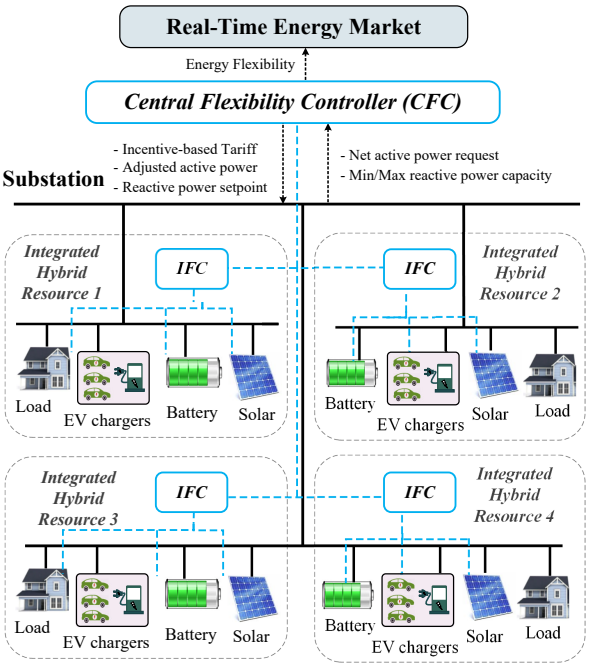

Figure 1. Hierarchical model for determining the energy flexibility offering strategy of integrated hybrid resources.

Flexibility Controller (IFC). The IFC utilizes a DRL model to dispatch flexible loads, ESS and solar plants and offer the aggregated deliverable energy flexibility to the Central Flexibility Controller (CFC). The deliverable energy flexibility of IHR is defined as the difference of net active power consumption of the IHR with and without controlling the flexibility resources. A queuing model is adopted to model the aggregated energy flexibility of flexible loads, such as EVs, subject to their quality of service constraints. In addition to the energy flexibility, the IFC determines the physical bounds of net reactive power capacities of the IHR. Once these data are determined by the local IFCs, the CFC performs an optimal power flow model to adjust the net active and reactive power consumption of the distributed IHRs and offer the deliverable energy flexibility to the real-time energy market such that the technical constraints of power distribution system are satisfied. The hierarchical model is implemented on the IEEE 123-bus test distribution system, modified by adding multiple solar plants, ESS and flexible loads, using the 5-minute real-time market price, solar generation and load data of California Independent System Operator (CAISO) and the results are presented and discussed to verify the performance of the proposed model in ensuring deliverable energy flexibility to the market.

The rest of this paper is categorized as follows: the hierarchical model for offering deliverable energy

flexibility of IHRs into the real-time energy market is formulated in Section 2. The solution approach for the local controllers based on DRL is presented in Section 3. Simulation results are presented in Section 4 to test and validate the efficiency of the proposed model, and the paper is concluded in Section 5.

2. Problem Formulation

The hierarchical model for determining the energy flexibility offering strategy of IHRs is mathematically formulated in this section. The proposed hierarchical model includes a feeder-level IFC that utilizes DRL to maximize the revenue from offering the IHR's energy flexibility, and a distribution-level CFC that aggregates all the IHRs energy flexibility and ensures its deliverability to the real-time energy market, as shown in Fig. 2. The IFC uses DRL to learn the behavior of flexible loads (e.g., EV charging), as well as uncertainties of renewable generation, and once trained, generates fast and scalable control signals for numerous resources. Once the net active power in each IHR is determined, the available capacity for generating reactive power is also determined and sent to the central controller. The CFC receives these signals from IFCs, and runs a distribution-level optimal power flow model, where each IHR is treated as a single bus with a net active power and a capacity range for generating reactive power. The CFC determines the required reactive power in each IHR, to maintain voltage within the desired limits throughout the network. This architecture requires that voltage variation within each IHR be limited to a small δ as:

$$|V_{it} - V_{jt}| < \delta, \quad \forall t, \forall i, j, \in \mathcal{I}_z, \forall z \in \mathcal{Z}, \qquad (1)$$

where $\mathcal{Z}$ is the set of all IHRs and $i, j, \in \mathcal{I}_z$ represent any pair of buses in IHR $z \in \mathcal{Z}$. The CFC may also curtail the net active power requested by each IFC to meet distribution system constraints. In the last operational step, CFC sends the adjusted net active and reactive powers to each IFC, which accordingly adjusts the power setpoints of individual DERs. The operation of IFC and CFC, and the mathematical problem that must be solved for each of them is explained next.

2.1. IHR Flexibility Controller

The objective of IFC is to maximize the revenue gained from offering the IHR energy flexibility to the CFC, as formulated in (2):

$$\max \sum_{t \in \mathcal{T}} (\lambda_t^{NF} p_{z,t}^{NF} - \lambda_t^{F} p_{z,t}^{F}), \qquad (2)$$

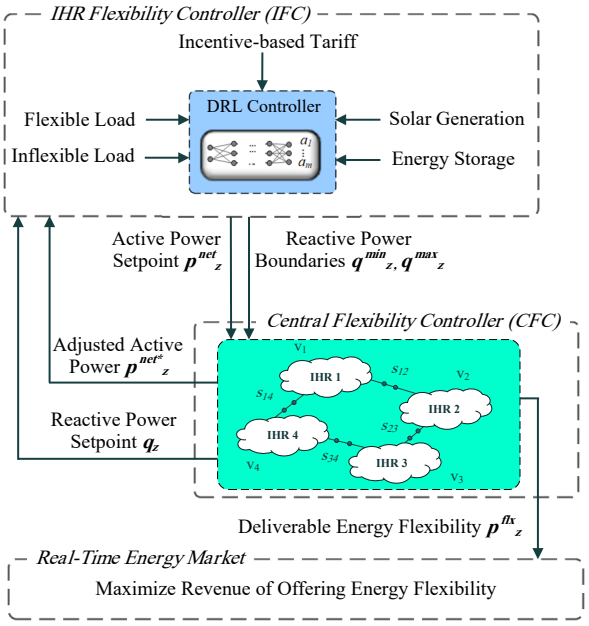

Figure 2. Hierarchical IHR energy flexibility control using IFC and CFC.

where λ_t^F and λ_t^{NF} are the incentive-based and fixed electricity tariffs, and $p_{z,t}^F$ and $p_{z,t}^{NF}$ are net IHR active power demand with and without offering flexibility at time $t \in \mathcal{T}$, and are obtained by the active power balance constraints in the IHR:

$$p_{z,t}^{NF} = p_{z,t}^{D} - \sum_{i \in \mathcal{M}_z} \overline{p}_{i,t}^{PV}, \quad \forall t, \qquad (3a)$$

$$p_{z,t}^{F} = p_{z,t}^{D} - \sum_{i \in \mathcal{M}_z} p_{i,t}^{PV} + \sum_{i \in \mathcal{E}_z} (p_{i,t}^{CH} - p_{i,t}^{DCH}) + \sum_{i \in \mathcal{C}_z} p_{i,t}^{LF}, \forall t, \qquad (3b)$$

where $p_{z,t}^D$ denotes the inflexible load, $p_{i,t}^{LF}$ represents flexible load, $p_{i,t}^{PV}$ denotes active power output of solar plants, and $p_{i,t}^{CH}/p_{i,t}^{DCH}$ shows the charge/discharge power of ESS in IHR z. In (3), the IHR can offer energy flexibility by curtailing its solar generation, charging/discharging its ESS, or shifting its flexible loads.

The operation of solar plants equipped with controllable inverters is modeled in (4a)-(4c):

$$p_{i,t}^{PV\,2} + q_{i,t}^{PV\,2} \le \overline{S}_i^{PV\,2}, \quad \forall i \in \mathcal{M}_z, \forall t, \qquad (4a)$$

$$0 \le p_{i,t}^{PV} \le \overline{p}_i^{PV}, \quad \forall i \in \mathcal{M}_z, \forall t, \qquad (4b)$$

$$\underline{pf}_i^{PV} \le \frac{p_{i,t}^{PV}}{\sqrt{p_{i,t}^{PV} + q_{i,t}^{PV}}}, \quad \forall i \in \mathcal{M}_z, \forall t, \qquad (4c)$$

where the apparent power of the solar plants is constrained to the nominal apparent power of the inverter $\overline{S}_i^{PV}$ in (4a), in which $q_{i,t}^{PV}$ denotes the reactive power output of solar plants. Active power supplied by the solar plants to the IHR operation is constrained to its maximum limit $\overline{p}_i^{PV}$ in (4b) and the power factor of solar plants is ensured to be greater than the minimum acceptable power factor $\underline{pf}_i^{PV}$ in (4c).

The next set of equations represents the operational constraints of ESS in IHRs:

$$E_{i,t} = E_{i,t-1} + \eta^C p_{t,i}^{CH} - \frac{p_{i,t}^{DCH}}{\eta^D}, \forall i \in \mathcal{E}_z, \forall t > 1, \quad (5a)$$

$$E_{i,t} = E_i^{ini}, \quad \forall i \in \mathcal{E}_z, t = 1, \quad (5b)$$

$$\underline{E_i} \leq E_{i,t} \leq \overline{E_i}, \quad \forall i \in \mathcal{E}_z, \forall t \quad (5c)$$

$$0 \leq p_{i,t}^{CH} \leq I_{i,t}^{CH} \overline{p}_i^{CH}, \quad \forall i \in \mathcal{E}_z, \forall t, \quad (5d)$$

$$0 \leq p_{i,t}^{DCH} \leq I_{i,t}^{DCH} \overline{p}_i^{DCH}, \quad \forall i \in \mathcal{E}_z, \forall t, \quad (5e)$$

$$I_{i,t}^{CH} + I_{i,t}^{DCH} \leq 1, \quad \forall i \in \mathcal{E}_z, \forall t, \quad (5f)$$

$$(p_{i,t}^{CH} - p_{i,t}^{DCH})^2 + {q_{i,t}^E}^2 \leq {\overline{s}_i^E}^2, \quad \forall i \in \mathcal{E}_z, \forall t. \quad (5g)$$

The evolution of the state of charge for ESS is presented in (5a) and the initial energy of ESS is defined in (5b), where $E_{i,t}$ and E_i^{ini} are the stored energy and the initial energy of ESS, and η^C/η^{DCH} represent the charging/discharging efficiencies of ESS. Stored energy in ESS is confined to the minimum and maximum limits $\underline{E_i}$, $\overline{E_i}$ in (5c) and charge and discharge power of ESS are constrained to their respective minimum and maximum limits $\overline{p}_i^{CH}$, $\overline{p}_i^{DCH}$ in (5d)-(5e). Simultaneous charge and discharge of ESS is avoided by the auxiliary variables $I_{i,t}^{CH}$, $I_{i,t}^{DCH}$ in (5f). The ESS apparent power is constrained to the nominal capacity $\overline{s}_i^E$ in (5g), where $q_{i,t}^E$ is the ESS reactive power.

A queuing model is adopted from [8] and presented in (6) to model the aggregate energy flexibility of shiftable loads, which is particularly ideal for modeling the charging flexibility of a fleet of EVs with specific time limitations and energy demands.

$$Q_{i,t}^F = Q_{i,t-1}^F + A_{i,t}^{LF} - p_{i,t}^{LF}, \forall i \in \mathcal{C}_z, \forall t > 1, \quad (6a)$$

$$Q_{i,t}^F = Q_i^{ini,F}, \quad \forall i \in \mathcal{C}_z, t = 1, \quad (6b)$$

$$0 \leq p_{i,t}^{LF} \leq \overline{p}_i^{LF}, \quad \forall i \in \mathcal{C}_z, \forall t, \quad (6c)$$

$$Q_{i,t}^F = 0, \quad \forall i \in \mathcal{C}_z, t = t_i^D. \quad (6d)$$

The queuing model in (6a) states that the queue backlog at each time $Q_{i,t}^F$ equals to the unsupplied load at the previous time $Q_{i,t-1}^F$ plus the submitted load request $A_{i,t}^F$ minus the delivered power at each time $p_{i,t}^F$. Initial queue backlog value is defined in (6b) and the delivered power to the loads is limited to its maximum limit $\overline{p}_i^{LF}$ in (6c). A deadline-based policy is imposed (6d) to ensure that all the load request of flexible loads is supplied by their respective deadline t_i^D.

The net IHR active power is determined by solving (2) subject to constraints (3)-(6). Note that each IHR may contain hundreds of DERs and flexible loads; solving such integer problem using mathematical optimization over a time horizon and for multiple uncertainty scenarios can be time-consuming and even intractable. This issue is resolved using DRL-trained IFCs that can make continuous or discrete decisions within milliseconds. Details regarding the DRL approach is explained in details in Section 3. Once the IHR operation problem is solved and the net active power in the IHR is determined, the minimum and maximum reactive power capacities of IHR, $\underline{q}_{z,t}^F$ and $\overline{q}_{z,t}^F$, are obtained through (7):

$$\overline{p}_i^E = \max(\overline{p}_i^{CH}, \overline{p}_i^{DCH}), \quad \forall i, \quad (7a)$$

$$\underline{q}_i^E = -\sqrt{\overline{s}_i^{E^2} - \overline{p}_i^{E^2}}, \ \overline{q_i^E} = \sqrt{\overline{s}_i^{E^2} - \overline{p}_i^{E^2}}, \forall i, \quad (7b)$$

$$\underline{q}_i^{PV} = -\sqrt{\overline{s}_i^{PV^2} - \overline{p}_i^{PV^2}}, \ \overline{q_i^{PV}} = \sqrt{\overline{s}_i^{PV^2} - \overline{p}_i^{PV^2}}, \forall i, \quad (7c)$$

$$\overline{q}_{z,t}^F = q_{z,t}^D + \sum_{i \in \mathcal{C}_z} q_{i,t}^{LF} - \sum_{i \in \mathcal{E}_z} \underline{q}_i^E - \sum_{i \in \mathcal{M}_z} \underline{q}_i^{PV}, \forall z, \forall t, \quad (7d)$$

$$\underline{q}_{z,t}^F = q_{z,t}^D + \sum_{i \in \mathcal{C}_z} q_{i,t}^{LF} - \sum_{i \in \mathcal{E}_z} \overline{q}_i^E - \sum_{i \in \mathcal{M}_z} \overline{q}_i^{PV}, \forall z, \forall t. \quad (7e)$$

In (7a)-(7e), $\overline{p}_i^E$ is the maximum active power of ESS, $\underline{q}_i^E$,$\overline{q}_i^E$ are the minimum and maximum reactive power output of ESS, $\underline{q}_i^{PV}$,$\overline{q}_i^{PV}$ are the minimum and maximum reactive power output of solar plants, and $q_{z,t}^D$,$q_{i,t}^{LF}$ are the reactive power consumption of inflexible and flexible loads.

2.2. Central Flexibility Controller

The IHRs are represented in CFC as single buses characterized by a requested active power consumption, and minimum and maximum reactive power capacities. The CFC model is formulated as a quadratic optimization problem, where the objective is to maximize the revenue of offering deliverable energy flexibility to the real-time energy market, while the non-flexible power demand is met. The CFC objective

function is formulated as:

$$\max \sum_{t\in\mathcal{T}}(P_t^{G,0}-P_t^{G,F})\lambda_t^{RT}-\sum_{t\in\mathcal{T}}\sum_{z\in\mathcal{Z}}p_{z,t}^{C}\lambda^{C}, \quad (8)$$

where $P_t^{G,0}$ and $P_t^{G,F}$ are the active power consumption of the distribution system with and without flexible IHRs, λ_t^{RT} is the real-time market price, and $p_{t,z}^{C}$ is the active load request curtailment with the unit cost of λ^{C}. Any positive/negative value for $(P_t^{G,0}-P_t^{G,F})$ represents the reduction/increase in the active power consumption of the distribution system with introducing the operational energy flexibility in IHRs. The CFC maximizes its objective, subject to the power flow constraints of the distribution network, to ensure the deliverability of the energy flexibility offered to the market. However, since IHRs are treated as single buses, the CFC problem is much smaller and faster than running a large-scale optimization problem for the whole network including all DER and flexible loads. The CFC problem constraints are formulated in (9a)-(9i):

$$P_t^{G,F}=\sum_{1z\in\mathcal{L}}P_{1z,t}+g_1V_{1,t}^{SQ}, \quad \forall t, \quad (9a)$$

$$Q_t^{G,F}=\sum_{1z\in\mathcal{L}}Q_{1z,t}+b_1V_{1,t}^{SQ}, \quad \forall t, \quad (9b)$$

$$P_{zz'',t}+p_{z,t}^{F}-p_{z,t}^{C}= \sum_{z'z\in\mathcal{L}}\Big(P_{z'z,t}-r_{z'z}I_{z'z,t}^{SQ}\Big)+g_zV_{z,t}^{SQ}, \forall z\in\mathcal{B},\forall t, \quad (9c)$$

$$Q_{zz'',t}+q_{z,t}^{F}= \sum_{z'z\in\mathcal{L}}\Big(Q_{z'z,t}-x_{z'z}I_{z'z,t}^{SQ}\Big)+b_zV_{z,t}^{SQ}, \forall z\in\mathcal{B},\forall t, \quad (9d)$$

$$\underline{q}_{z,t}^{F}\le q_{z,t}^{F}\le\overline{q}_{z,t}^{F}, \forall z,\forall t, \quad (9e)$$

$$V_{z,t}^{SQ}-V_{z',t}^{SQ}=-2\left(r_{z'z}P_{z'z,t}+x_{z'z}Q_{z'z,t}\right) +\left(r_{z'z}^2+x_{z'z}^2\right)I_{z'z,t}^{SQ}, \quad \forall(z'z)\in\mathcal{L},\forall t, \quad (9f)$$

$$\underline{V}_z^{SQ}\le V_{z,t}^{SQ}\le\overline{V}_z^{SQ}, \quad \forall z\in\mathcal{B},\forall t, \quad (9g)$$

$$I_{z'z,t}^{SQ}\le\overline{I}_{z'z,t}^{SQ}, \quad \forall(z'z)\in\mathcal{L},\forall t, \quad (9h)$$

$$V_{z,t}^{SQ}I_{z'z,t}^{SQ}\ge P_{z'z,t}^2+Q_{z'z,t}^2, \quad \forall(z'z)\in\mathcal{L},\forall t. \quad (9i)$$

The active and reactive power balance constraints for the slack bus in the distribution system are presented in (9a)-(9b), where $P_{1z,t}$, $Q_{1z,t}$ are the active and reactive power flows from substation bus to the bus z, $V_{1,t}^{SQ}$ is the squared voltage on the substation bus, g_1,b_1 are the shunt conductance and susceptance at the substation bus and $\mathcal{L}$ is the set of lines in the distribution system. The energy balance constraints for the feeder buses are presented in (9c)-(9d), where $P_{zz'',t}$, $P_{z'z,t}$ and $Q_{zz'',t}$, $Q_{z'z,t}$ are respectively the active and reactive power flows in lines zz'' and $z'z$, $V_{z,t}^{SQ}$ is the squared voltage on bus z, $I_{z'z,t}^{SQ}$ is the squared current flow in line $z'z$, and $r_{z'z}$, $x_{z'z}$ are the resistance and reactance of the line $z'z$ and g_z, b_z are the shunt conductance and susceptance at the bus z. The reactive power consumption of IHRs $q_{z,t}^{F}$ is adjusted in (9e) and the voltage drop on distribution buses is modeled in (9f). The voltage level constraints on distribution buses is presented in (9g) and current flow limit is expressed in (9h), where $\underline{V}_z^{SQ}$, $\overline{V}_z^{SQ}$ are the minimum and maximum squared voltage limits and $\overline{I}_{z'z,t}^{SQ}$ is the squared current flow limit. Finally, the complex power flow in the distribution system is constrained in (9i). Solving this problem results in the maximum revenue of the distribution system from offering energy flexibility to the real-time energy market; it also provides the exact reactive power setpoint and adjusted active power setpoint for each IHR. The final result of the hierarchical operation is $(P_t^{G,0}-P_t^{G,F})$, which shows the amount of deliverable energy flexibility to the real-time energy market.

2.3. Revising IFC setpoints

The CFC sends two signals, $(p_{z,t}^{F}-p_{z,t}^{C})$ and $q_{z,t}^{F}$, back to each IFC, so they adjust power generation and consumption in their own IHR. If $p_{z,t}^{C}$ is zero, i.e., the originally requested power by IFC is validated by CFC, no changes will be made in active power setpoint. However, if some of the requested active power is curtailed ($p_{z,t}^{C}>0$), then individual local loads are curtailed starting from ESS, and then flexible loads, and finally non-flexible loads. For reactive power dispatch, the IFC divides the $q_{z,t}^{F}$ signal received from CFC among available DER, including solar plants and ESS, based on their reactive generation capacity. The reactive power dispatch of DER units within IHR z are given by:

$$\hat{q}_{i,t}^{DER}=(-q_{z,t}^{F}+q_{z,t}^{D}+\sum_{i\in\mathcal{C}_z}q_{i,t}^{LF})\times\frac{\overline{q}_{i,t}^{DER}}{\sum_{i'\in\mathcal{I}_z}\overline{q}_{i',t}^{DER}}, \quad \forall i\in\mathcal{I}_z,\forall t, \quad (10)$$

where $\hat{q}_{i,t}^{DER}=\{\hat{q}_{i,t}^{PV},\hat{q}_{i,t}^{E}\}$ is the adjusted reactive power for solar plants and ESS.

3. Using Deep Reinforcement Learning for Training IFCs

In this section, a DRL-based IFC is proposed that is trained on historical behavior of flexible loads and renewable generation, and once trained, can generate fast and scalable control signals for numerous active elements. In order to use DRL, first the IHR energy flexibility control problem is formulated as a Markov Decision Process (MDP), described next.

3.1. Formulating the IHR energy flexibility control as a Markov Decision Process

An MDP is a sequential optimization problem consisting of multiple states and a range of possible actions to take in each state. Taking a certain action in a certain state results in a reward (or cost) and transition to another state, which is determined by a transition probability matrix. The operation of IFC is formulated as an MDP by defining the following:

State: A system state denoted by $s \in \mathcal{S}$ is a set of system parameters that affect decision-making. In this particular problem, state parameters include time coordinates (day of the year, day of the week, and time of the day represented by d_t^y, d_t^w and h_t, respectively), current energy reserve of all ESS within the IHR, available renewable power, non-flexible demands, the queue backlog for flexible loads, their deadlines, and real-time and forecasted energy price:

$$s_t = (d_t^y, d_t^w, h_t, \mathbf{E}_t, \mathbf{p}_t^{PV}, p_t^D, \mathbf{Q}_t, \mathbf{t}^D, \lambda^F, \tilde{\lambda}^F), \quad (11)$$

where $\tilde{(.)}$ denotes the average predicted values for the next 24 hours, and vectors are shown by bold letters. For instance, $\mathbf{E}_t = [E_{it},\ \forall i \in \mathcal{I}_z]$ represents the energy reserve of all ESS within the IHR.

Actions: Actions denoted by $a_t \in \mathcal{A}$ are operational decisions for scheduling energy flexibility and dispatching DER units, and are defined by:

$$a_t = (\mathbf{p}_t^{PV}, \mathbf{p}_t^{CH} - \mathbf{p}_t^{DCH}, \mathbf{p}_t^{LF}), \quad (12)$$

where each element of a_t is a vector of actions for a certain type of DER or flexible load in the IHR.

Reward function: Defined as $R : \mathcal{S} \times \mathcal{A} \to \mathcal{R}$, is the reward due to making operation action a_t in system state s_t. The reward function helps DRL find the desired sequence of actions, and should be defined carefully. In the current problem, it is defined based on the IFC objective in (2), but with added terms to help DRL converge faster and search for action trajectories more efficiently:

$$\begin{aligned} R_t = R(s_t, a_t) = \\ (\tilde{\lambda}_t^F - \lambda_t^F)\Big(c_1 \sum_{i \in \mathcal{E}_z} (p_{it}^{CH} - p_{it}^{DCH}) + c_2 \sum_{i \in \mathcal{C}_z} p_{it}^{LF}\Big) \\ + c_3 \sum_{i \in \mathcal{M}_z} (\overline{p}_{it}^{PV} - p_{it}^{PV}) \\ + c_4 \sum_{i \in \mathcal{C}_z} p_{i,t}^{LF}\Big(\frac{Q_{i,t}^F}{t^D - t}\Big) - c5\|a_t\|, \end{aligned} \quad (13)$$

where ESS are rewarded for charging when the real-time energy price is below average forecast price, and for discharging otherwise. Similarly, flexible loads are encouraged to shift to low-price periods. In the second term, solar plants are rewarded for producing power at full available power. Further, to encourage the DRL algorithm to meet the deadline of flexible loads, another term is added in the last line, which rewards supplying flexible loads if the accumulated queue ($Q_{i,t}^F$) is large and the remaining time to deadline ($t^D - t$) is short. A regularization term is also added to the reward function to avoid unnecessary large action.

Transition function: Defined as $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ is a transition probability matrix that specifies the probability of reaching a new state s_{t+1}, given the current state s_t and the taken action a_t. It is hard to obtain the transition function as the change of state parameters are uncertain and hence their prediction might not be exact. DRL algorithms do not explicitly use this function, but learn it implicitly through voluminous observations of historical or simulated transitions between states.

Using this MDP formulation can maximize the long-term revenue of IHR from offering energy flexibility to the CFC. To that end, instead of (2), the following objective function is maximized:

$$\max_{\pi}\ \mathbb{E}_{s,a\sim\mathcal{P}}\left[\sum_{t=1}^{\infty} \gamma^t [R(s_t, a_t)] \Big| s_0 = s\right], \quad (14)$$

where future rewards are discounted by $\gamma \in [0, 1]$, and the expectation $\mathbb{E}$ is over transition probability matrix $\mathcal{P}$. In (14), π is the action policy and $\pi(s) = a$.

3.2. Training IFC Using Deep Deterministic Policy Gradient (DDPG)

The DDPG algorithm is used here to train the IFC. DDPG is an actor-critic DRL algorithm, which means it makes decisions by training two separate deep neural networks: a critic network that evaluates the long-term reward of available actions, and an actor network that estimates the action value with the highest long-term reward. The long-term value of action a_t in state s_t

is estimated by $Q(s_t, a_t)$, or the Q-value, which is recursively calculated by the Bellman equation:

$$Q(s_t, a_t) = R_t + \gamma \mathbb{E}\Big[\max_{a_{t+1} \in \mathcal{A}_{t+1}} Q(s_{t+1}, a_{t+1})\Big]. \quad (15)$$

The Q-value of each action is estimated by the critic network as $Q(s_t, a_t; \theta^Q)$, where θ^Q is the critic's weight vector. The actor network estimates the optimal action vector for each system state as $a_t = \mu(s_t; \theta^\mu)$. With properly trained networks, (15) is re-written as:

$$Q(s_t, a_t) \approx R_t + \gamma \mathbb{E}\Big[Q(s_{t+1}, \mu(s_{t+1}; \theta^\mu); \theta^Q)\Big]. \quad (16)$$

The DRL training starts by observing a system state s_t, followed by the actor network taking action $a_t = \mu(s_t; \theta^\mu)$, resulting in the next state s_{t+1} in the power system. For better exploration of the action space and also to avoid overfitting, a random perturbation, typically in the form of a Ornstein-Uhlenbeck process, is added to a_t during the training process, resulting in $a_t = \mu(s_t; \theta^\mu) + \mathcal{N}_t$. Also, the resulting action vector $a_t = (\mathbf{p}_t^{PV}, \mathbf{p}_t^{ch} - \mathbf{p}_t^{dch}, \mathbf{p}_t^{L,F})$ must not violate constraints (3)-(6). More specifically, the solar generation must be contained within available solar power, and ESS charging and discharging must meet the maximum and minimum energy thresholds of ESS. To that end, the elements of a_t are modified as follows:

$$\hat{p}_{it}^{PV} = \min\{p_{it}^{PV}, \overline{p}_{it}^{PV}\}, \ \forall i \in \mathcal{M}_z, \quad (17a)$$

$$\hat{p}_{i,t}^{ch} = \min\{p_{i,t}^{ch}, \frac{\overline{E}_i - E_{i,t-1}}{\eta^c}\}, \ \forall i \in \mathcal{E}_z, \quad (17b)$$

$$\hat{p}_{i,t}^{dch} = \max\{p_{i,t}^{dch}, \eta^d(E_{i,t-1} - \underline{E}_i)\}, \ \forall i \in \mathcal{E}_z. \quad (17c)$$

In every step of DDPG algorithm, s_t, a_t, s_{t+1}, and the resulting reward R_t are observed and stored in a *replay memory* shown by U. Then, a random sample from the last B observation is taken from U and a training iteration is run for the critic network by minimizing the following loss function:

$$L(\theta^Q) = \mathbb{E}\Big[\big(Q(s_t, a_t; \theta^Q) - y\big)^2\Big], \quad (18a)$$

$$y = R_t + Q(s_{t+1}, a_{t+1}; \theta^Q). \quad (18b)$$

Also, the following loss gradient, obtained by the gradient chain rule, is used in the actor network to update the weights:

$$\nabla_{\theta^\mu} L(\theta^\mu) = \mathbb{E}\Big[\nabla_a Q(s_t, \mu(s_t; \theta^\mu); \theta^Q).\nabla_{\theta^\mu} \mu(s_t; \theta^\mu)\Big]. \quad (19)$$

3.3. Enhancing the Scalability of DRL

In the standard DRL, the collective action in (12) should be taken simultaneously for all DERs and flexible loads. Assuming the size of the action vector is ND and the action space for each of them is $\mathcal{A}_n, n \in \{1, ..., ND\}$, the cardinality of the collective action is $\prod_n A_n$, which is directly proportional to the number of required Q-value calculations in DRL training. Therefore, making decisions for numerous elements can be computationally expensive and result in an inefficient training process. Instead, one can break down the action space into ND single actions and create intermediate system states after taking each action. That would create $ND - 1$ intermediate states as $(s_t, a_t^1), (s_t, a_t^1, a_t^2), ..., (s_t, a_t^1, ..., a_t^{ND-1})$. In this technique, also known as multi-agent rollout, actions are taken sequentially rather than collectively, and the decision for the dispatch of each DER unit is taken after the decision of its predecessor units are determined. This reformulation reduces the complexity of Q-value computation from $\prod_n |\mathcal{A}_n|$ to $\sum_n |\mathcal{A}_n|$. The reward function is also modified accordingly to reflect the reward of taking each individual action. For instance, the reward of taking an action for the energy storage on bus i is given by:

$$R(s_t, a_t) = (\tilde{\lambda}_t^F - \lambda_t^F)(p_{it}^{CH} - p_{it}^{DCH}) + m\|a_t\|. \quad (20)$$

4. Numerical Study

In this section, the proposed hierarchical model for determining the energy flexibility offering strategy of IHRs is tested on the IEEE 123-bus test distribution system, modified by adding multiple solar plants, ESS, and flexible loads. The test system is then divided into 6 IHRs such that (1) is satisfied in each of them in any loading conditions. Each IHR Consists of flexible and non-flexible loads, and DERs (i.e., ESS and solar plants), as shown in Fig. 3. Increasing the number of IHRs leads to more local control points for the CFC, hence improving the performance of the hierarchical model in satisfying the network constraints.

Combination of flexible loads with DERs a is key point in flexible operation of the IHRs and disregarding each one of these sources affects the IHRs potential to provide energy flexibility into the distribution system operation. Characteristics of the DERs and EV charging stations within each IHR in the proposed study is listed in Table 1. As shown in the table, each IHR contains a charging station with 50 charging plugs, with maximum power rating of 6.6 kW. The 5-minute real-time market price, solar generation and load data of CAISO from 01/01/2021 to 12/28/2021 are used

Table 1. Properties of DERs and EV charging stations within each IHR

	Max. inflexible load (kW)	ESS energy capacity (kWh)	ESS power rate (kW)	Max. solar generation (kW)	Number of EV chargers
IHR 1	400	350	70	70	50
IHR 2	360	300	60	60	50
IHR 3	715	650	130	130	50
IHR 4	830	750	150	150	50
IHR 5	745	250	50	50	50
IHR 6	320	600	120	120	50

Table 2. Total revenues of each IHR in $ from offering energy flexibility to the CFC

	IHR 1	IHR 2	IHR 3	IHR 4	IHR 5	IHR 6	Total
DRL-trained model	4602.5	4691.2	4939.2	6014.1	4221.3	4387.7	28856.0
Greedy model	2205.5	2227.1	1977.8	2133.8	2412.0	2149.6	13105.9
Non-flexible model	2046.7	2141.4	1810.0	2011.2	2340.5	1963.4	12313.2

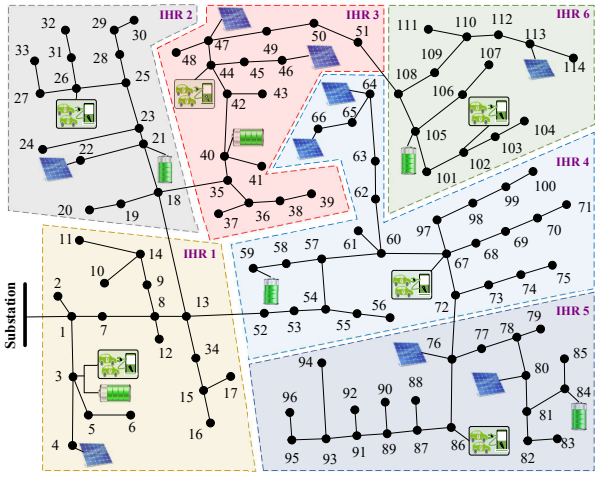

Figure 3. Modified 123-bus test system with DERs and EV charging stations, divided into 6 IHRs.

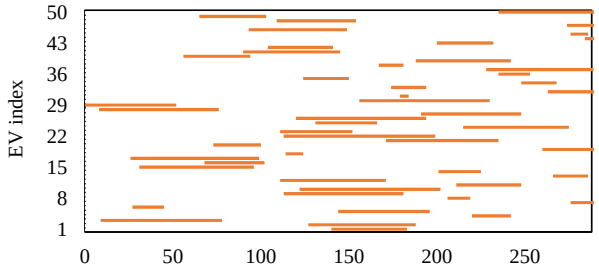

Figure 4. Synthesized arrival and departure times for charging for a fleet of 50 EVs in IHR 1 over a 24-hour period (288 5-min intervals).

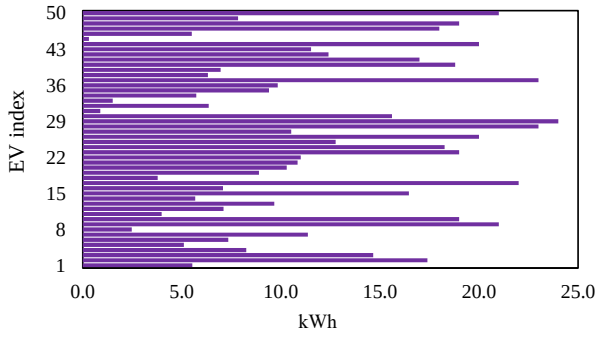

Figure 5. Synthesized charging demand in kWh for a fleet of 50 EVs in IHR 1 over a 24-hour period.

for training the DDPG algorithm [22]. The fixed electricity tariff is considered to be 8.8 (cent/kWh) and the incentive-based electricity tariff is assumed to be 1.5 times the real-time market price. EV charging load is considered as flexible loads, characterized by specific arrival times and departures (deadline), and energy demands. In order to consider the uncertainty of EV owners behaviour within the training of IFCs, the arrival time for each EV is selected randomly during the day. The time between arrival and departure is then sampled from a normal distribution with $\mu = 4hr, \sigma = 2hr$. The maximum energy demand during this period is $\overline{e}^{EV}$ = (departure - arrival) $\times$ 6.6kW. The actual energy demand is randomly sampled from the uniform distribution $e^{EV} \sim U[0, \overline{e}^{EV}]$. The arrival time and maximum duration of charging for 50 EVs in IHR 1 is shown for a sample day in Fig. 4; also Fig. 5 depicts the energy demand of EVs for the same day.

The MDP formulation of the problem, as explained in Section 3, is implemented using Gym toolkit, which allows for defining episode restart and step through functions. DDPG algorithm is then implemented using Keras package in Python. For both actor and critic networks, a feed-forward neural network is designed with two hidden layers, each containing 50 neurons and followed by ReLU activation function. Other DDPG parameters are tuned through cross-validation and set as follows: $\gamma = 0.99$, replay memory size $B = 400$, and

learning rate $= 0.001$. The coefficients of the reward function (13) are: $c_1 = 1.5, c_2 = 10, c_3 = 1.5, c_4 = 0.5, c_5 = 0.1$. For training and testing the DRL-based IFC, the collected data, including annual profiles of price, load, and solar generation, as well as synthesized EV data, are divided int [0.8,0.2] for training and testing.

Separate IFCs are trained for each IHR and are tested on the test portion of the data. Table 2 shows the total revenue of each IHR from offering energy flexibility to the CFC, and compares the results with the greedy and non-flexible methods, indicating the superiority of the proposed DRL-trained model. In the greedy method, DERs generate or discharge active power as long as they can, and EVs charge their batteries immediately after arrival. To showcase the impact of incentive-based tariff on the energy flexibility, Fig. 6 shows how the flexibility of EVs is utilized in different times of a sample 24-hour period (288 5-min intervals). At each time, if loads are shifted from that time to other times, the energy flexibility is positive and it is negative otherwise. In Fig. 6, most positive energy flexibility is taken during the peak price period between intervals 90 and 150. Note that due to time constraints of EV charging, the energy flexibility profile cannot exactly match the price profile, but it follows it closely for the most part.

The scheduled EV charging demand is shown in Fig. 7, which shows how the peak EV charging demand between intervals 100 and 160 is shifted to intervals 180 and later, to avoid charging during the peak energy tariffs. Similar results are shown in Fig. 8 for energy flexibility offering of ESS. Note that unlike EVs, ESS do no have a time constraint and can follow energy tariff more closely. The only limitation of ESS is their energy capacity and distribution system operation constraints.

As discussed in Section 2, one of the main tasks of the CFC is to determine reactive power setpoints for each IHR, using which IFCs generate reactive power setpoints for inverter-based DERs and charging stations.

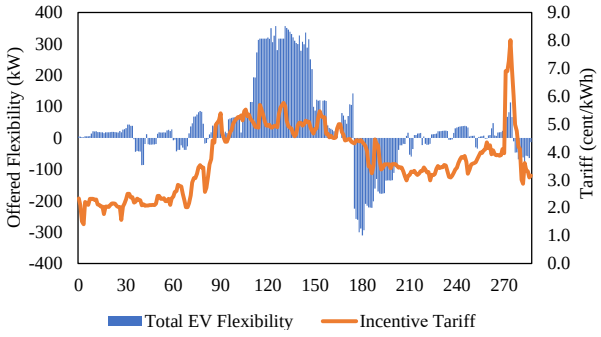

Figure 6. Total energy flexibility offered by EV charging stations from all IHRs in a sample 24-hour period.

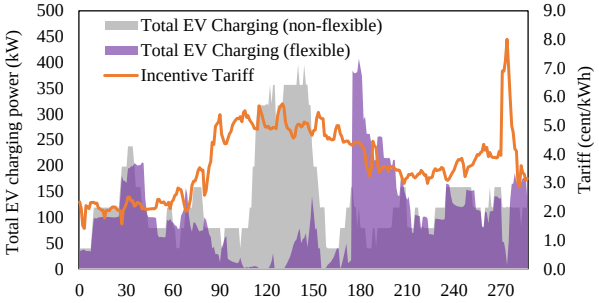

Figure 7. Shift of EV charging schedule due to incentive tariff in a sample 24-hour period.

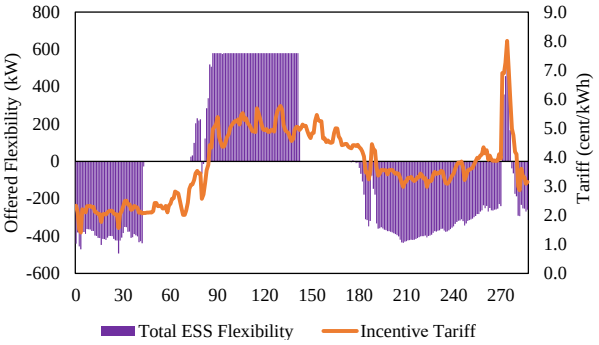

Figure 8. Total energy flexibility offered by ESS from all IHRs in a sample 24-hour period.

Fig. 9 shows how the required reactive power is supplied by different types of inverter-based devices in each IHR. When solar plants are active, they contribute as much as ESS in reactive power generation, since their power ratings are similar. However, solar plants can only supply reactive power when they also produce active power, i.e., during daytime. Note that in Fig. 9, the 24-hour period starts at a random time of the day, and not at 12am. Besides solar plants and ESS, EV charging stations also contribute to reactive power generation, but to a much less extent due to their lower power rating.

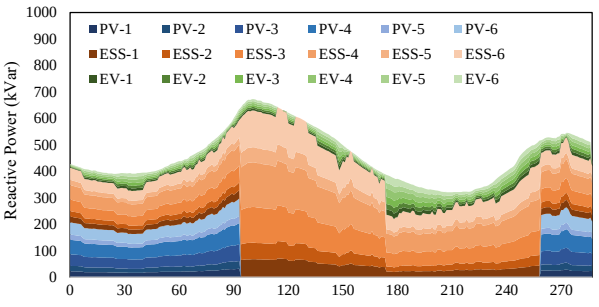

Figure 9. Reactive power generation in a sample 24-hour period (i.e., PV-1 representing reactive power generation in IHR 1 by all solar plants).

5. Conclusion

In this paper, an intelligent hierarchical model was presented for determining the energy flexibility offering strategy of IHRs to participate in real-time energy markets. The proposed model utilizes DRL for training IHR flexibility controllers, and includes a quadratic optimization central controller to maintain the grid constraints and manage the reactive power dispatch. The scalability issues of using standard DRL for operation of large number of DER and flexible loads is overcame by breaking down the action space using multi-agent rollout technique. As a result, the associated real-time energy flexibility control problem, which is intractable by integer optimization, is solved by the proposed intelligent method in a fraction of a second. The simulations were conducted on the 123-bus system and the results showed that the proposed method outperforms greedy operational algorithms by a large margin, and effectively shifts flexible loads based on the incentive-based tariff in the distribution system, while maintaining their quality of service constraints.

References

[1] "Ferc order no. 2222: A new day for distributed energy resources." Department of Energy - Federal Energy Regulatory Commission, 2020.

[2] "Unlocking the flexibility of hybrid resources." Energy Systems Integration Group - Hybrid Resources Task Force, 2022.

[3] "Informational report of california independent system operator corporation." Department of Energy - Federal Energy Regulatory Commission, 2021.

[4] M. Majidi, L. Rodriguez-Garcia, and M. Parvania, "Risk-based operation of power networks with hybrid energy systems," in *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6, IEEE, 2022.

[5] L. Yin and S. Li, "Hybrid metaheuristic multi-layer reinforcement learning approach for two-level energy management strategy framework of multi-microgrid systems," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104326, 2021.

[6] J. Hu, X. Liu, M. Shahidehpour, and S. Xia, "Optimal operation of energy hubs with large-scale distributed energy resources for distribution network congestion management," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 3, pp. 1755–1765, 2021.

[7] Z. Guo, W. Wei, L. Chen, M. Shahidehpour, and S. Mei, "Distribution system operation with renewables and energy storage: A linear programming based multistage robust feasibility approach," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 738–749, 2021.

[8] K. Oikonomou, M. Parvania, and R. Khatami, "Deliverable energy flexibility scheduling for active distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 655–664, 2019.

[9] K. Oikonomou, M. Parvania, and R. Khatami, "Coordinated deliverable energy flexibility and regulation capacity of distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106219, 2020.

[10] Y. Xiao, X. Wang, P. Pinson, and X. Wang, "A local energy market for electricity and hydrogen," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3898–3908, 2017.

[11] S. Touzani, A. K. Prakash, Z. Wang, S. Agarwal, M. Pritoni, M. Kiran, R. Brown, and J. Granderson, "Controlling distributed energy resources via deep reinforcement learning for load flexibility and energy efficiency," *Applied Energy*, vol. 304, p. 117733, 2021.

[12] M. Dorokhova, Y. Martinson, C. Ballif, and N. Wyrsch, "Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation," *Applied Energy*, vol. 301, p. 117504, 2021.

[13] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for pv/ess-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3493–3503, 2019.

[14] Y. Ye, Y. Tang, H. Wang, X.-P. Zhang, and G. Strbac, "A scalable privacy-preserving multi-agent deep reinforcement learning approach for large-scale peer-to-peer transactive energy trading," *IEEE Transactions on Smart Grid*, 2021.

[15] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, "Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach," *Applied Energy*, vol. 292, p. 116940, 2021.

[16] M. M. Hosseini and M. Parvania, "Resilient operation of distribution grids using deep reinforcement learning," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2100–2109, 2021.

[17] M. M. Hosseini and M. Parvania, "On the feasibility guarantees of deep reinforcement learning solutions for distribution system operation," *IEEE Transactions on Smart Grid*, in press, 2022.

[18] M. M. Hosseini, L. Rodriguez-Garcia, and M. Parvania, "Hierarchical intelligent operation of integrated hybrid resources for distribution grid restoration," *IEEE Transactions on Sustainable Energy*, in press, 2022.

[19] S. Taheri, V. Kekatos, H. Veeramachaneni, and B. Zhang, "Data-driven modeling of aggregate flexibility under uncertain and non-convex device models," *IEEE Transactions on Smart Grid*, 2022.

[20] E. Won, "A hardware implementation of artificial neural networks using field programmable gate arrays," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 581, no. 3, pp. 816–820, 2007.

[21] K. Ramanaiah and S. Sridhar, "Hardware implementation of artificial neural networks," *i-Manager's Journal on Embedded Systems*, vol. 3, no. 4, p. 31, 2014.

[22] "California ISO Open Access Same-Time Information System," 2022. [Online]. Available: `http://oasis.caiso.com`.