Real-Time Price Elasticity Reinforcement Learning for Low Carbon Energy Hub Scheduling Based on Conditional Random Field

Weiqi Hua, Minglei You, and Hongjian Sun
Department of Engineering, *Durham University*, Durham, U.K.
{weiqi.hua, minglei.you, hongjian.sun}@durham.ac.uk

***Abstract*—Energy hub scheduling plays a vital role in optimally integrating multiple energy vectors, e.g., electricity and gas, to meet both heat and electricity demand. A scalable scheduling model is needed to adapt to various energy sources and operating conditions. This paper proposes a conditional random field (CRF) method to analyse the intrinsic characteristics of energy hub scheduling problems. Building on these characteristics, a reinforcement learning (RL) model is designed to strategically schedule power and natural gas exchanges as well as the energy dispatch of energy hub. Case studies are performed by using real-time digital simulator that enables dynamic interactions between scheduling decisions and operating conditions. Simulation results show that the CRF-based RL method can approach the theoretical optimal scheduling solution after 50 days training. Scheduling decisions are particularly more dependent on received price information during peak-demand period. The proposed method can reduce 9.76% of operating cost and 1.388 ton of carbon emissions per day, respectively.**

***Index Terms*—Energy hub, price elasticity, real-time digital simulator (RTDS), reinforcement learning (RL), conditional random field, carbon reduction.**

I. Introduction

Scheduling multiple energy carriers/vectors, e.g., coupled electricity and gas networks, has received much attention in recent years [1]–[4]. Motivation behind the scheduling is that the electricity, natural gas, and distributed renewable energy sources can be systematically and optimally integrated to improve operating performance, e.g., carbon reduction [1], cost minimization, reliability [2], and stability [3]. Model-based optimization is an efficient tool to solve the scheduling problem, under which the scheduling objective is designed according to certain criteria, subject to technical constraints [4]. Although designing energy hub scheduling as an optimization problem yields a theoretically optimal solution, there are still several major challenges. First, the scale of energy hub may vary depending on network structures and types of energy sources. The model-based optimization problem with predefined parameters may not perform well if the practical conditions cannot be accurately evaluated by model parameters. Second, the scheduling requires accurate prediction of energy hub inputs and outputs, e.g., renewable generation and energy demand. An off-line optimization may lose system fidelity as this method fails to dynamically adapt with the uncertainties of supply and demand. Third, the intrinsic characteristics of energy dispatch responding to price signal, i.e. price elasticity of power supply, is not captured by existing optimization methods.

With respect to the first two challenges, model-free scheduling becomes an alternative solution [5]. The reinforcement learning (RL), as a typical model-free solution, can dynamically optimize a control policy by updating a state-action value function (Q-function) through interacting with systems [6]. From implementation and operational perspectives of energy hub, the RL outperforms the model-based optimization in terms of following aspects: 1) Instead of requiring predefined parameters and assumptions, RL is simply based on historical data, which enables the model to be more scalable and compatible for various scales of energy hubs. Both computational and economical burdens of optimization tool can be addressed; 2) When the problem is formulated as a multi-objective optimization, objective functions do not always share the same dimension, such as carbon emissions and operating cost, whereas this dimensional difference can be eliminated from historical data by using RL. In the literature [7], [8], the RL was implemented for solving scheduling problems to assist or replace model-based methods. Nonetheless, a common method to obtain Q-function for power system scheduling was based on off-line historical data. This off-line method cannot dynamically adapt with system operating conditions and update scheduling decisions.

In order to dynamically balance supply and demand, real-time pricing scheme was studied in [9]. Given real-time pricing scheme, energy hub operators face a challenging decision of either importing electricity and natural gas from main grid or energy dispatch within the energy hub. If this decision-making is cleverly designed, it can enable energy hub to operate with a minimal cost, by exploring the intrinsic characteristics of energy hub operators, e.g., price elasticity. Although the price responding strategies were well studied in existing literature [10], [11], little progress has been made on the analysis of price elasticity.

In this paper, a linear conditional random field (CRF) method [12] is adopted as a logistic regression approach to capture the temporal variations of scheduling behaviours influenced by energy prices. This linear CRF provides a statistical model for RL to make optimal decisions through dynamically interacting with system operations. Compared with naive Bayes methods [13], the logistic regression of linear CRF is capable of modelling the energy hub scheduling due to the discrete decisions of importing energy from main grids and inside dispatch. Additionally, the linear CRF can be accommodated by the RL through modelling the Q-function as an expectation of operating cost with the merit of minimal assumption for system model, such as statistical distribution of energy prices.

In brief, this paper focuses on solving energy hub scheduling problems by addressing several limitations in existing studies:

- The scalability of energy hub scheduling has to be addressed to better adapt with dynamic system conditions.
- The dependency of scheduling decisions on the energy prices and their variations has not been studied by existing scheduling methods.
- The impacts of price elasticity variation on the scheduling decisions have not been carefully studied.

To fill these research gaps, this paper makes the following contributions:

- We propose a new CRF method to analyse dynamic price elasticity of energy sources. Dependency of scheduling decisions on energy prices and decision transient are captured.
- RL model is developed to provide scalability guarantee with minimal assumptions on model structures. Testbed is developed using RTDS to dynamically interact between the learning model and system operation.
- Carbon emissions caused by electricity and gas exchanges and transmission loss within energy hub are considered during scheduling to reduce total carbon emissions.
- Simulation results show that the proposed CRF method successfully describes the dependency of scheduling decisions on energy prices and the dependency of transient action through real-time recursively updating weighting factors. Both costs and carbon emission can be reduced using this proposed method.

The remainder of this work is summarized as follows. In Section II, the system model for energy hub scheduling is introduced considering carbon mitigation. The CRF-based RL for price elasticity modelling is presented in Section III. Section IV conducts case studies to evaluate the proposed method. Section V draws the conclusions.

II. System Model

In this section, the overall energy hub model is illustrated and the technical constraints involving carbon mitigation are introduced as a preliminary for the proposed algorithm.

A. Energy Hub Components

To study the energy hub scheduling, You's model [3] is considered, which optimizes dispatch of multiple sources in energy carriers with the objective of reducing operating costs and improving system stability. In our proposed method, this model is extended to involve the effects of dynamic price elasticity on scheduling decisions and the carbon mitigation during the operating process. We will first briefly discuss energy hub components in You's model, and readers can refer to [3] for further details.

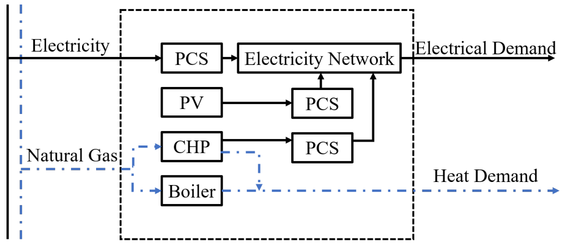

Fig. 1. Energy hub with electricity and natural gas networks.

The energy hub consists of power conditioning system (PCS), combined heat and power (CHP) unit, boiler, photovoltaic (PV) panels, and electricity network as shown in Fig. 1. The PCS is installed at PV and CHP as a voltage source inverter to improve grid voltage stability. The CHP converts natural gas into electricity and recovers the generated heat for supplying heat load. The boiler converts natural gas into heat through heating contained fluid with higher efficiency than CHP. The PV is a renewable energy source commonly used in consumers' domain to convert solar energy into direct current (DC) electricity. The electricity network transmits the electricity to demand side within the energy hub.

B. Technical Constraints

During the energy hub operations, some basic constraints should be taken into account from technical perspective as

$$(Q_e^{\text{PCS}})^2 + (P_e^{\text{PCS}})^2 \leq (S_{e,\max}^{\text{PCS}})^2, \tag{1}$$

$$P_e^{\text{CHP}} = \eta_e^{\text{CHP}} \times P_g^{\text{CHP}}, \tag{2}$$

$$P_h^{\text{CHP}} = \eta_h^{\text{CHP}} \times P_g^{\text{CHP}}, \tag{3}$$

$$0 \leq P_e^{\text{CHP}} \leq S_{e,\max}^{\text{CHP}}, \tag{4}$$

$$-\sqrt{(S_{e,\max}^{\text{CHP}})^2 - (P_{e,\max}^{\text{CHP}})^2} \leq Q_e^{\text{CHP}} \leq \sqrt{(S_{e,\max}^{\text{CHP}})^2 - (P_{e,\max}^{\text{CHP}})^2}, \tag{5}$$

$$0 \leq P_h^{\text{CHP}} \leq S_{h,\max}^{\text{CHP}}, \tag{6}$$

$$P_h^{\text{Boiler}} = \eta_h^{\text{Boiler}} \times P_g^{\text{Boiler}}, \tag{7}$$

$$0 \leq P_h^{\text{Boiler}} \leq S_{h,\max}^{\text{Boiler}}, \tag{8}$$

$$P_e^{\text{PV}} = \eta_e^{\text{PV}} \times P_r, \tag{9}$$

$$-\sqrt{(S_{e,\max}^{\text{PV}})^2 - (P_{e,\max}^{\text{PV}})^2} \leq Q_e^{\text{PV}} \leq \sqrt{(S_{e,\max}^{\text{PV}})^2 - (P_{e,\max}^{\text{PV}})^2}, \tag{10}$$

where P_e^{PV}, P_e^{CHP}, and P_e^{PCS} are electricity outputs of PV, CHP, and PCS, respectively, P_h^{CHP} and P_h^{Boiler} are heat outputs of CHP and boiler, respectively, η_e^{CHP}, η_h^{CHP}, η_h^{Boiler} and η_e^{PV} are energy transforming efficiency factors, P_g^{CHP} and P_g^{Boiler} are gas input for CHP and boiler, respectively, P_r is solar irradiance, $P_{e,\max}^{\text{CHP}}$ and $P_{e,\max}^{\text{PV}}$ are maximum real power output for CHP and PV, respectively, $Q_e^{\text{PCS}}, Q_e^{\text{PV}}$, and Q_e^{CHP} are reactive power for PCS, PV, and CHP, respectively, and $S_{e,\max}^{\text{PCS}}$, $S_{e,\max}^{\text{CHP}}$, $S_{h,\max}^{\text{CHP}}$, $S_{h,\max}^{\text{Boiler}}$, and $S_{e,\max}^{\text{PV}}$ are maximum capacities.

The PCS controls active and reactive power as a voltage source inverter through injecting or ejecting reactive power constrained by real power and apparent power capacity as (1). Electrical and thermal outputs of CHP are given by (2)-(3), in which the real power, reactive power and heat are constrained by (4)-(6). The heat power output of boiler is given by (7) and constrained by maximum capacity as (8). The PV output is given by (9) and constrained by (10) when PCS converts DC to alternating current (AC).

C. Carbon Emissions Evaluation

Based on You's model, carbon emissions during energy hub operation are considered into the scheduling. The carbon emissions are primarily caused by the electricity and natural gas exchanges from main energy networks and the transmission line loss within the energy hub, which results in an additional cost due to purchasing the carbon tax. Thus, the goal is to evaluate these portions of carbon emissions and introduce an additional constraint for carbon mitigation. The carbon emissions caused by electricity and natural gas exchanges can be quantified through using carbon intensity I which is defined as the amount of carbon emissions per unit of power/heat flow with the unit of tCO_2/MWh [14] as

$$I = \frac{R}{P} \tag{11}$$

where I is carbon intensity, P is power or heat flow, and R is carbon emission rate which quantifies the amount of carbon emission per unit of time with the unit of tCO_2/h.

With respect to carbon emissions caused by the transmission line loss within the energy hub, the topology structure of power networks needs to be considered. Given that power flow distribution is based on proportional sharing principle [14], the carbon emission distribution follows the same principle. Denote i and j as inflow and outflow branches of bus z, respectively. The carbon emission rate of outflow branch can be expressed as the sum of carbon emissions of inflow branch and bus-connected source

$$R_j = \sum_{i \in z} P_{i,j} e_i + \sum_{s \in z} P_s e_s, \tag{12}$$

where e_i and e_s are carbon intensities in branch i and bus-connected source s, respectively, and $P_{i,j}$ is the share of power flow in jth branch coming from ith branch P_i. According to proportional sharing principle, $\frac{P_{i,j}}{P_j} = \frac{P_i}{\sum_{i \in z} P_{i,j} + \sum_{s \in z} P_s}$. Hence, the additional constraint for carbon mitigation is that the total carbon emission rate from natural gas R_g, electricity R_e, and transmission loss R_j during the scheduling horizon T is less than carbon emission limit as

$$\sum_{t=1}^{T} R_g t + R_e t + R_j t \leq E_{\text{max}}, \tag{13}$$

where R_e and R_g are carbon emission rates for electricity and natural gas, respectively, and E_{max} is carbon emission limit.

III. Proposed Algorithm

This section describes the proposed algorithm of CRF-based RL for price elasticity modelling. The action space A for energy hub scheduling consists of: decisions for electricity exchange from main grid A_e, decisions for natural gas exchange from gas network A_g, decisions for dispatching gas to CHP for producing electricity $A_{e,\text{CHP}}$ and heat $A_{h,\text{CHP}}$, and decisions for dispatching gas to boiler A_{Boiler}. The control actions for energy hub scheduling are binary variables, representing the corresponding components to be switched on or off: $a_e, a_g, a_{e,\text{CHP}}, a_{h,\text{CHP}}, a_{\text{Boiler}} \in \{0, 1\}$. The action vector $\mathbf{a}$ is subsequently defined as $\mathbf{a} = (a_e, a_g, a_{e,\text{CHP}}, a_{h,\text{CHP}}, a_{\text{Boiler}})$. The state space S of energy hub scheduling consists of: the price of electricity S_e and the price of natural gas S_g. The state vector is subsequently defined by electricity price π_e and natural gas price π_g as $\mathbf{s} = (\pi_e, \pi_g)$.

At the beginning of each scheduling interval t, market operator announces the electricity and gas prices $\mathbf{s}^t = (\pi_e^t, \pi_g^t)$ to energy hub operator. Energy hub operator then decides and dispatches the scheduling results $\mathbf{a}^t = (a_e^t, a_g^t, a_{e,\text{CHP}}^t, a_{h,\text{CHP}}^t, a_{\text{Boiler}}^t)$ at the end of scheduling interval t. The goal of proposed algorithm is to introduce the dynamic price elasticity of energy sources in energy hub so that given observed states $\{\mathbf{s}^1, ..., \mathbf{s}^t\}$ and past actions $\{\mathbf{a}^1, ..., \mathbf{a}^{t-1}\}$, the probability of energy hub operator' s decisions for energy exchanges from main networks and inside dispatch $\mathbf{a}^t$ at t can be analysed. With this analysis, the RL is performed to obtain an optimal control policy for energy hub scheduling.

A. CRF for Elasticity Modelling

In our proposed algorithm, the linear CRF [12] is adopted to describe the price elasticity of energy sources and action transient dependency. Unlike the Hidden Markov Model [15] to assume that the observed state variables are independent with each other, in linear CRF, the action at current time t is dependent on all the observed states $\{\mathbf{s}^1, ..., \mathbf{s}^t\}$. This is more suitable for considering practical scheduling problem, because energy hub operator may make strategically decisions responding to price signal variations so as to minimize total daily operating cost. The linear CRF obeys the Markov property [12], which means that conditioned on $\mathbf{s}^t$, action $\mathbf{a}^t$ at time t is independent of action $\mathbf{a}^k$ at time $k, (k \neq t)$, given $\mathbf{a}^{t+1}$ and $\mathbf{a}^{t-1}$, as $p(\mathbf{a}^t \mid \mathbf{a}^1, ...\mathbf{a}^t, \mathbf{s}^1, ...\mathbf{s}^t) = p(\mathbf{a}^t \mid \mathbf{a}^{t-1}, \mathbf{a}^{t+1}, \mathbf{s}^1, ..., \mathbf{s}^t)$. Hence, the conditional probability $p(\mathbf{a}^t \mid \mathbf{s}^t)$ is modelled as

$$p(\mathbf{a}^t \mid \mathbf{s}^t) = \frac{1}{Z(\mathbf{s}^t)} \prod_t e^{\mu^t \Phi^t(\mathbf{a}^t, \mathbf{s}^t)} \prod_{t-1} e^{\lambda^{t,t-1} \Psi^{t,t-1}(\mathbf{a}^t, \mathbf{a}^{t-1})}, \tag{14}$$

where $\Phi^t(\mathbf{a}^t, \mathbf{s}^t) := \mathbf{a}^t \mathbf{s}^t$ is the state feature function to describe the dependency of action $\mathbf{a}^t$ on state $\mathbf{s}^t$ at time t, $\Psi^{t,t-1}(\mathbf{a}^t, \mathbf{a}^{t-1}) := \mathbf{a}^t \mathbf{a}^{t-1}$ is the transient feature function to describe the dependency of action $\mathbf{a}^t$ at time t on action $\mathbf{a}^{t-1}$ at time $t-1$, μ^t and $\lambda^{t,t-1}$ are weighting factors to describe the strength of these dependencies, and $Z(\mathbf{s}^t) := \sum_{\mathbf{a}^t} \prod_t e^{\mu^t \Phi^t(\mathbf{a}^t, \mathbf{s}^t)} \prod_{t-1} e^{\lambda^{t,t-1} \Psi^{t,t-1}(\mathbf{a}^t, \mathbf{a}^{t-1})}$ is a normalization factor. Motivated by the state feature function and transient feature function, the weighting factors μ^t and $\lambda^{t,t-1}$ can be defined as

$$\mu^t := \frac{f_{\mathbf{s}^t}(\mathbf{a}^t)}{t}, \lambda^{t,t-1} := \frac{f_{\mathbf{a}^{t-1}}(\mathbf{a}^t)}{t}, \tag{15}$$

where $f_{\mathbf{s}^t}(\mathbf{a}^t)$ is the total amount of time in which the action $\mathbf{a}^t$ is performed as on ($\mathbf{a}^t$=1) given state $\mathbf{s}^t$ and $f_{\mathbf{a}^{t-1}}(\mathbf{a}^t)$ is the total amount of time in which the action $\mathbf{a}^t$ is performed as on ($\mathbf{a}^t$=1) given that action $\mathbf{a}^{t-1}$ is performed as on ($\mathbf{a}^{t-1}$=1). μ^t and $\lambda^{t,t-1}$ can be updated at each time step when receiving new pieces of information $\gamma_{\mathbf{s}^t}(\mathbf{a}^t)$ and $\gamma_{\mathbf{a}^{t-1}}(\mathbf{a}^t)$ as $f_{\mathbf{s}^t}(\mathbf{a}^t) = f_{\mathbf{s}^{t-1}}(\mathbf{a}^{t-1}) + \gamma_{\mathbf{s}^t}(\mathbf{a}^t)$ and $f_{\mathbf{a}^{t-1}}(\mathbf{a}^t) = f_{\mathbf{a}^{t-2}}(\mathbf{a}^{t-1}) + \gamma_{\mathbf{a}^{t-1}}(\mathbf{a}^t)$. Hence, weighting factors μ^t and $\lambda^{t,t-1}$ can be updated recursively as

$$\mu^t = \frac{f_{\mathbf{s}^{t-1}}(\mathbf{a}^{t-1}) + \gamma_{\mathbf{s}^t}(\mathbf{a}^t)}{t-1} \cdot \frac{t-1}{t} = \mu^{t-1} + \frac{1}{t}[\gamma_{\mathbf{s}^t}(\mathbf{a}^t) - \mu^{t-1}], \tag{16}$$

$$\lambda^{t,t-1} = \lambda^{t-1,t-2} + \frac{1}{t}[\gamma_{\mathbf{a}^{t-1}}(\mathbf{a}^t) - \lambda^{t-1,t-2}]. \tag{17}$$

B. Reinforcement Learning

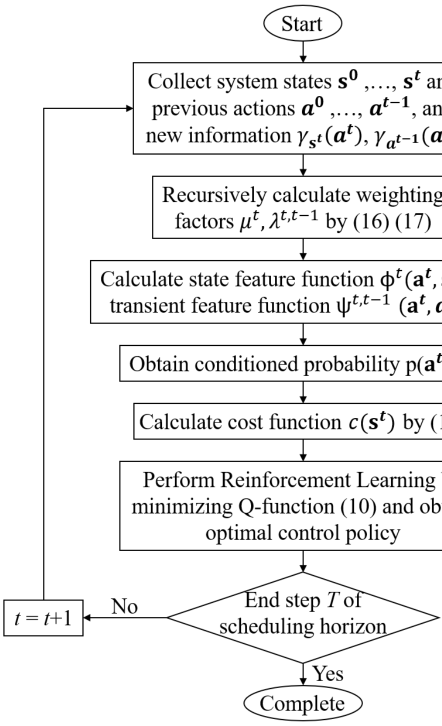

Fig. 2. Flowchart of the proposed algorithm.

Unlike You's model [3] which only considers the cost of importing energy from main network, our research also involves the revenue from providing electricity and heat services inside the energy hub as a compensation of cost, and assumes that the electricity and gas prices of providing these services are the same as the energy prices of importing. Hence, considering aforementioned CRF-based price elasticity, minimal cost of current state can be defined as the expectation with respect to $p(\mathbf{a}^t \mid \mathbf{s}^t)$ when the probability of importing power from main grids becomes minimum $p(\mathbf{a}^t = 0 \mid \mathbf{s}^t)$, and the probability of inside dispatch of energy hub becomes maximum $p(\mathbf{a}^t = 1 \mid \mathbf{s}^t)$ as

$$\begin{aligned} c(\mathbf{s}^t) =& \mathbb{E}\{P_e a_e^t t\pi_e + P_g a_g^t t\pi_g - [(P_e^{\text{PCS}} + P_e^{\text{PV}} + \\ & a_{e,\text{CHP}}^t P_e^{\text{CHP}}) t\pi_e + (a_{h,\text{CHP}}^t P_h^{\text{CHP}} + a_{\text{Boiler}}^t P_h^{\text{Boiler}}) t\pi_g] \mid \mathbf{s}^t\} \\ =& (P_e t\pi_e + P_g t\pi_g) \times p(\mathbf{a}^t = 0 \mid \mathbf{s}^t) \\ & - [P_e^{\text{CHP}} t\pi_e + (P_h^{\text{CHP}} + P_h^{\text{Boiler}}) t\pi_g] \times p(\mathbf{a}^t = 1 \mid \mathbf{s}^t) \\ & - (P_e^{\text{PCS}} + P_e^{\text{PV}}) t\pi_e. \end{aligned} \tag{18}$$

The objective of RL is to find a control policy $h : S \to A$ to minimize the total cost from state $\mathbf{s}^1$ to current state $\mathbf{s}^t$, which can be defined by Q-function as

$$Q^h(\mathbf{a}^t, \mathbf{s}^t) = \mathbb{E}\{c(\mathbf{s}^0) + \xi c(\mathbf{s}^1) + \xi^2 c(\mathbf{s}^2) + ... + \xi^t c(\mathbf{s}^t)\} \tag{19}$$

where ξ is discounting factor. The Q-function is a discounted cumulative reward following policy h. Considering the objective of cost minimization, the optimal policy h^* is obtained from minimized Q-function which is subject to constraints (1)-(13) as

$$h^* \in \arg \min_{\mathbf{a}^t, P_e, P_g, Q_e^{\text{CHP}}, Q_e^{\text{PV}}} Q^*(\mathbf{a}^t, \mathbf{s}^t) \tag{20}$$

In the conclusion, the flowchart of proposed algorithm is presented in Fig. 2 to visually represent the process of CRF-based RL.

IV. Numerical Results

The performance of proposed method is evaluated through numerical tests using real-time simulations. We used a modified 4-bus medium voltage distribution system [3] as shown in Fig. 3. The electricity network and gas network are coupled at bus 3 with a CHP and a boiler installed. The PV is installed at bus 2. The data of hourly electricity load and PV generation are obtained from Gridwatch [16] and the data of hourly heat load is obtained from [17]. Both electricity and heat data are scaled to energy hub model by proportion. For comparison, we used the same parameters setting with You's model [3] and compared the performances on cost and carbon emissions. The U.K. real-time price data was obtained from [18].

A. Simulator Set Up

In order to provide a real-time scheduling and performance evaluation, the proposed energy hub scheduling model was implemented on RTDS as shown in Fig. 4. Compared with off-line simulators [19], [20] for establishing mathematical representation of power system operation, the RTDS is capable of interacting with real power system components. It operates continuously to provide an ideal environment for testing energy hub. The smart meters and controllers were modelled by Data Acquisition and Actuator module, which provides an interface between control commands and RTDS. After performing CRF-based RL by Matlab, the scheduling results were transmitted from Matlab back to RTDS by Giga-Transceiver Analogue Input Card whereas the real-time operating signals

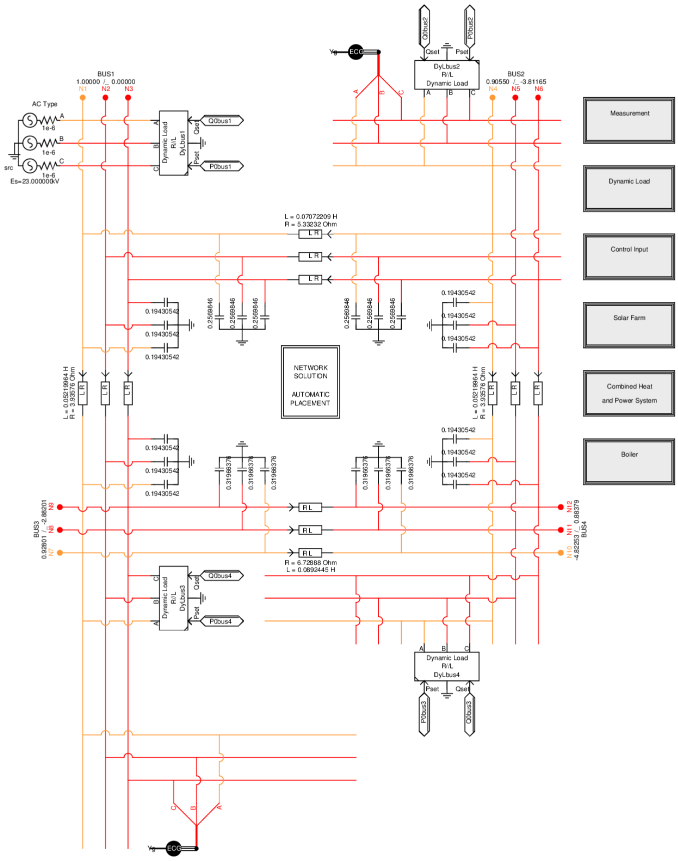

Fig. 3. 4-bus energy hub system on RTDS.

were transmitted from RTDS to Matlab by Giga-Transceiver Analogue Output Card.

Fig. 4. Interactions Setup of Matlab/PC, RTDS and DAA.

B. Evaluation of Reinforcement Learning

1) Benchmark: To evaluate the effects of RL, a benchmark is designed as the minimization of daily costs with the same decision variables and constraints of (18) as below

$$\min \sum_{t=1}^{T} P_e a_e^t t \pi_e + P_g a_g^t t \pi_g - [(P_e^{\text{PCS}} t + P_e^{\text{PV}} t + a_{\text{CHP}} P_e^{\text{CHP}} t)\pi_e + (a_{\text{CHP}}^t P_h^{\text{CHP}} t + a_{\text{Boiler}}^t P_h^{\text{Boiler}} t)\pi_g], \tag{21}$$

The optimization is solved by Matlab optimization toolbox. This benchmark optimization yields a theoretically optimal scheduling with predefined cost coefficients and carbon intensities. The goal of our proposed algorithm is to obtain a learning policy that approximates the optimal solution.

2) Learning Results: The simulations consist of 50 days of historical hourly data. The weighting factors μ^t and $\lambda^{t,t-1}$ are updated at each new time step recursively according to (16) and (17) to describe the strength of state-action dependency and state-transient dependency. The probability distribution of weighting factors μ^t and $\lambda^{t,t-1}$ at each interval for electricity and gas prices until day 50 is presented in Fig. 5. Each column represents the distribution of the dependency for each control action on various price levels from low to high. Each row represents the difference of the dependency for various control actions responding to the same price level. The dark blue colour represents a lower value of dependency whereas the bright yellow colour represents a higher value of dependency. It can be seen that the probability of μ^t is higher during the peak price period corresponding to peak demand, which means that the action $\mathbf{a}^t$ is more dependent on received price information. By contrast, $\lambda^{t,t-1}$ is relatively independent of the price fluctuation, and thus presents homogeneous distribution because it is only relevant to the transient between states.

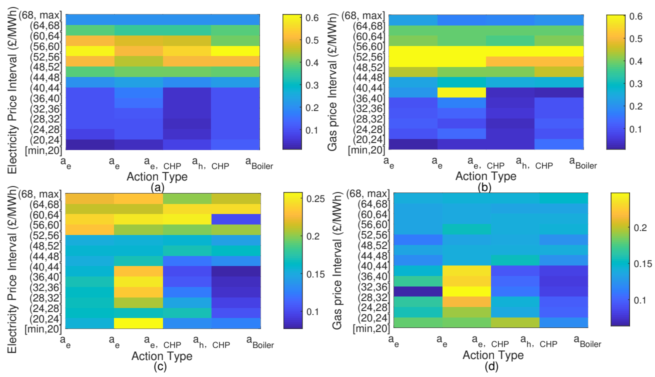

Fig. 5. Probability distribution of weighting factors of μ^t and $\lambda^{t,t-1}$ at each interval for (a) μ^t with π_e, (b) μ^t with π_g, (c) $\lambda^{t,t-1}$ with π_e, and (d) $\lambda^{t,t-1}$ with π_g. The x-axis represents action type and the y-axis represents price interval.

The results of RL after 10 days, 30 days, and 50 days for various electricity prices are presented in Fig. 6. The outputs of CHP electricity, power exchange, and corresponding real-time electricity prices are selected as examples to compare the learning results with benchmark optimization as (21). It can be seen that with historical data accumulation, the learning results are approaching to optimal solutions irrespective of price information.

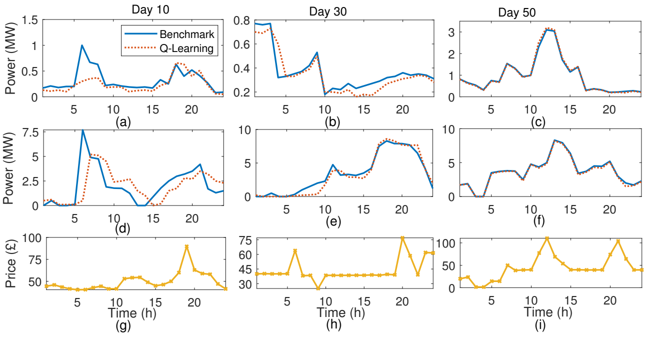

Fig. 6. Illustration of learning process for electricity generated by CHP (a)(b)(c), power exchange (d)(e)(f) with electricity prices (g)(h)(i). The day number is indicated at the top of each column.

C. Evaluation of Cost and Carbon Reduction

We compare our proposed algorithm with the cost minimization formulation in You's work [3] as

$$\min_{P_e, P_g, Q_e^{\text{CHP}}, Q_e^{\text{PV}}} c_e(P_e) + c_g(P_g), \tag{22}$$

where P_e and P_g are electricity and natural gas importing from main energy networks, respectively, and c_e and c_g are corresponding costs. The cost minimization problem is subject to the same constraints (1)-(13). The scheduling outputs and corresponding average carbon intensity are presented in Fig. 7. Our proposed model realises 1.388 ton of daily carbon reduction, i.e., from 6.956 ton to 5.568 ton per day. This is primarily due to the peak electricity demand reduction by using increasing proportion of gas, as the energy hub operator is more sensitive to the peak-time price with considered price elasticity. In addition, the daily costs are reduced by 9.76 % from £ 3,012 to £ 2,718 considering the revenue from internal energy supply.

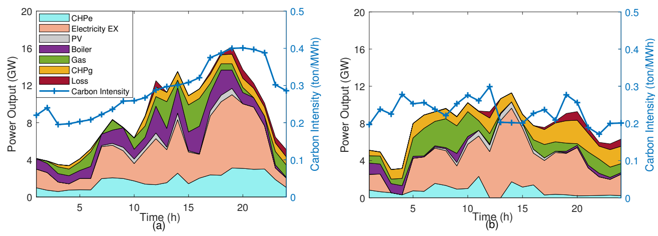

Fig. 7. Comparison of energy scheduling and carbon intensity for (a) Scheduling from (22) and (b) proposed algorithm.

V. Conclusion

To improve the scalability of energy hub scheduling, this paper proposed a new CRF-based RL method that is different from the existing model-based optimization methods. In addition to the consideration of intrinsic price elasticity, the proposed method explored the use of state action feature and action transient feature; Hence the temporal variations of scheduling behaviours influenced by energy prices can be incorporated in the learning process. The learning was implemented in a real-time digital simulation environment, capable of dynamically adapting the system operating conditions. Simulations show that the weighting factors can accurately capture the dependency features. During peak demand period, the scheduling decisions were more dependent on price signal. The RL can approximate the theoretical optimal scheduling after 50 days of training. The carbon emissions and operating cost can be significantly reduced using this proposed method.

VI. Acknowledgments

This work was supported by the European Regional Development Fund (ERDF) project "Solid Wall Insulation Innovation" (https://swiiproject.co.uk). The authors gratefully acknowledge Durham County Council and Durham Energy Institute for their strong and timely support.

References

[1] M. Majidi, S. Nojavan, and K. Zare, "A cost-emission framework for hub energy system under demand response program," *Energy*, vol. 134, pp. 157–166, 2017.
[2] Z. Zhou, C. Sun, R. Shi, Z. Chang, S. Zhou, and Y. Li, "Robust energy scheduling in vehicle-to-grid networks," *IEEE Network*, vol. 31, no. 2, pp. 30–37, March 2017.
[3] M. You, W. Hua, M. Shahbazi, and H. Sun, "Energy hub scheduling method with voltage stability considerations," in *2018 IEEE/CIC Int. Conf. ICCC Workshops*. IEEE, 2018, pp. 196–200.
[4] P. S. Georgilakis and N. D. Hatziargyriou, "Unified power flow controllers in smart power systems: models, methods, and future research," *IET Smart Grid*, vol. 2, no. 1, pp. 2–10, 2019.
[5] L. Maharjan, M. Ditsworth, M. Niraula, C. C. Narvaez, and B. Fahimi, "Machine learning based energy management system for grid disaster mitigation," *IET Smart Grid*, vol. 2, pp. 172–182(10), June 2019.
[6] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
[7] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement learning of heuristic ev fleet charging in a day-ahead electricity market," *IEEE Trans. Smart Grid*, vol. 6, no. 4, 2015.
[8] E. C. Kara, M. Berges, B. Krogh, and S. Kar, "Using smart devices for system-level management and control in the smart grid: A reinforcement learning framework," in *Int. Conf. SmartGridComm*. IEEE, 2012.
[9] B. P. Bhattarai and et al., "Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions," *IET Smart Grid*, vol. 2, pp. 141–154(13), June 2019.
[10] W. E. Elamin and M. F. Shaaban, "New real-time demand-side management approach for energy management systems," *IET Smart Grid*, vol. 2, pp. 183–191(8), June 2019.
[11] M. Zaery, E. M. Ahmed, and M. Orabi, "Low operational cost distributed prioritised coordinated control for dc microgrids," *IET Smart Grid*, vol. 2, pp. 233–241(8), June 2019.
[12] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.
[13] L. Jiang, H. Zhang, and Z. Cai, "A novel bayes model: Hidden naive bayes," *IEEE Trans. Knowledge and Data Eng.*, vol. 21, Oct 2009.
[14] C. Kang, T. Zhou, Q. Chen, J. Wang, Y. Sun, Q. Xia, and H. Yan, "Carbon emission flow from generation to demand: A network-based model," *IEEE Trans. Smart Grid*, vol. 6, pp. 2386–2394, Sep. 2015.
[15] A. M. Gonzalez, A. M. S. Roque, and J. Garcia-Gonzalez, "Modeling and forecasting electricity prices with input/output hidden markov models," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 13–24, Feb 2005.
[16] [Online]. Available: https://www.gridwatch.templar.co.uk/
[17] [Online]. Available: http://www.ukenergywatch.org/
[18] [Online]. Available: https://www.energybrokers.co.uk/electricity
[19] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. power syst.*, vol. 26, 2011.
[20] O. Anaya-Lara and E. Acha, "Modeling and analysis of custom power systems by pscad/emtdc," *IEEE trans. power delivery*, vol. 17, no. 1, pp. 266–272, 2002.