Recursive Learning Based Smart Energy Management With Two-Level Dynamic Pricing Demand Response

Huifeng Zhang, *Senior Member, IEEE*, Jiapeng Huang, Dong Yue, *Fellow, IEEE*, Xiangpeng Xie, *Senior Member, IEEE*, Zhijun Zhang, *Member, IEEE*, and Gerhard P. Hancke, *Life Fellow, IEEE*

***Abstract*— Due to dynamic characteristic of demand response and stochastic nature of power generation, it brings great challenge to smart energy management. In this paper, a demand response model is created with two-level dynamic pricing transaction among grid operator, service provider and customers, which also involves customers' active participation with load shifting issue. To effectively control system load on the demand side, an improved deep reinforcement learning approach is proposed with a recursive least square (RLS) technique to deal with the dynamic pricing demand response problem, which accelerates the on-line training and optimization efficiency. On the power generation side, a probabilistic penalty-based boundary intersection (PBI) based multi-objective optimization algorithm is improved to optimize the economic cost, emission rate and statistic voltage stability index (SVSI) simultaneously with generated stochastic scenarios, which can ensure energy conservation and environmental protection, as well as system security. The case results reveal that the proposed two-level optimization strategy successfully deals with energy management with dynamic pricing demand response.**

***Note to Practitioners*—This paper is motivated by solving stochastic energy management issue of isolated power system with dynamic pricing demand response. Those existing methods merely focus on the load demand or power generation side, and the methods for demand response issue lacks efficient on-line learning ability, while this work proposes a recursive least square based deep reinforcement learning approach to tackle with the two-level dynamic pricing demand response issue, scenario based PBI multi-objective optimization is proposed to solve the power dispatch issue on power generation side, and the numerical analysis results suggest that the proposed optimization strategy can deal with the whole energy management issue well. The future work will focus on the dynamic power-load coordination in the energy management issue.**

***Index Terms*— Demand response, stochastic, energy management, reinforcement learning, multi-objective optimization.**

Manuscript received 8 May 2024; revised 8 July 2024; accepted 16 August 2024. This article was recommended for publication by Associate Editor Q. Wei and Editor Q. Zhao upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Fund under Grant 62473202, Grant 61973171, Grant 62293500, Grant 62293505, Grant 52077106, and Grant 62233010; in part by the Basic Research Project of Leading Technology of Jiangsu Province under Grant BK20202011; and in part by the National Natural Science Fund of Jiangsu Province under Grant BK20211276. *(Corresponding authors: Huifeng Zhang; Dong Yue.)*

Huifeng Zhang, Jiapeng Huang, Dong Yue, and Xiangpeng Xie are with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China (e-mail: zhanghuifeng_520@163.com; 1222056206@njupt.edu.cn; medongy@vip.163.com; xiexiangpeng1953@163.com).

Zhijun Zhang is with Nanyang Technological University, Singapore 639798 (e-mail: mezhijun@gmail. com).

Gerhard P. Hancke is with the College of Automation, Nanjing University of Posts and Telecommunications, Jiangsu 210023, China, and also with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa (e-mail: g.hancke@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2024.3446849.

Digital Object Identifier 10.1109/TASE.2024.3446849

I. Introduction

DUE to the increasing penetration of intermittent energy resources, demand response (DR) plays a crucial role in enhancing power system reliability by reducing peak load stress and minimizing potential supply interruptions from power generators [1], [2]. Existing research has extensively explored price-incentive-based DR models to reduce consumption during peak periods [3], [4], [5], [6], [7], [8], [9], [10], [11]. Literature [3] proposes a demand response management system with service provider and customers, mutual optimal solution can be made by service provider for the utility. In [5], a deep reinforcement learning (DRL) based energy management algorithm is proposed with a neural network for electricity price forecasting. In [8], a demand response algorithm for smart facility energy management based on deep reinforcement learning is proposed, using long short-term memory units and multi-layer perceptions to effectively minimize power costs while maintaining satisfaction. Literature [9] considers the uncertainty of resident's behavior, real-time electricity price and outdoor temperature, and proposes a real-time DR strategy for optimal scheduling of home appliances. Literature [11] presents a dynamic pricing demand response model with considering grid operator, service provider and customers, and utilizes a Q-learning method to solve it. Although these price-incentive-based DR models can manage demand response to some extent, most rely on one-stage approaches that fail to consider the dynamic adjustments in customer behavior resulting from discrepancies between power generation and system load. This limitation makes one-stage approaches inefficient at dynamically controlling load demand.

Afterwards, some existing energy management approaches are presented to deal with both DR and power dispatch issues [12], [13], [14], [15], [16], [17], [18], [19], [20]. Reference [12] proposes an efficient online algorithm within an AC optimal power flow framework for real-time load

scheduling in micro-grids, providing competitive scheduling decisions based only on past and current inputs. In [14], a hierarchical optimization mechanism for demand response aggregators is proposed that integrates self-scheduling and load decomposition to optimize market participation by identifying customer behavior. Literature [17] proposes a distributed demand response method for multi-energy residential communities, using coordination decision-making and information transmission mechanisms to optimize coordination. Literature [20] proposes an extension to the classical two-stage stochastic programming model to capture the interactions between local generation and demands with uncertain renewable generation and heterogeneous energy management settings. However, those existing one-stage DR models lack of considering customers' dynamic adjustment behavior caused by the deviation between power generation and load demand, and those solution approaches rely on accurate model for dealing with dynamic and uncertain issues [21]. Hence, the wholesale price is taken as the function of power balance deviation on the basis of formulated model in literature [11], which can balance power generation and load requirement combined with load shifting mechanism. Besides, stochastic power generation model with multiple objective requirements is also created under above dynamic pricing mechanism, and a two-stage optimization strategy with adjusting both system load and power generation is proposed for dealing with smart energy management issue. The main contributions are summarized as follows:

(1) To well control load requirement on the demand side, a two-level dynamic pricing demand response model is formulated with load shifting mechanism, it can make optimal wholesale price, retail price and load shifting schemes, which can better trade-off the total economic benefit and customers' dissatisfaction.

(2) Due to dynamic and uncertain characteristics of DR, an improved deep deterministic policy gradient (DDPG) algorithm is proposed to enhance on-line learning efficiency with a recursive least square approach, which can accelerate real-time training efficiency with considering its Markov decision process with forgetting factors.

(3) To deal with stochastic multi-objectives optimization issue on power generation side, a decomposition based multi-objective algorithm is improved under a probabilistic PBI framework with random drift mechanism to adaptively search Pareto optimal schemes, which can reduce the computational complexity and speed up the optimization efficiency.

The remainder of the paper is organized as follows: The DR problem formulation is presented in section II, the stochastic optimal model of power generation is presented in section III, the proposed methodology is shown in section IV, and the simulation results and conclusion are presented in section V and section VI.

II. Upper-Layer Optimization Model Based on Two-Level Dynamic Pricing Demand Response

The electricity market plays a crucial role in managing demand-side loads by fostering dynamic interactions among the three principal stakeholders: grid operators, service providers, and customers. Grid operators oversee the high-voltage national grid, determining wholesale electricity

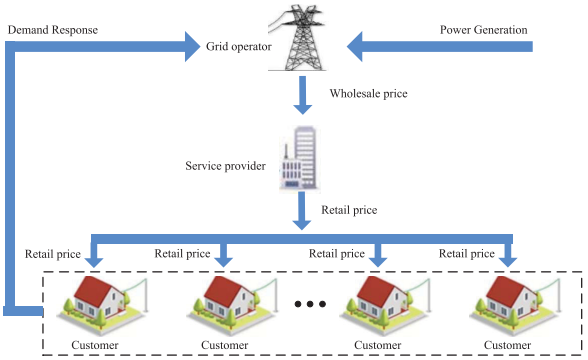


Fig. 1. The structure of dynamic pricing demand response.

prices based on real-time data regarding total load consumption and power generation. This pricing mechanism balances supply and demand to enhance system stability and efficiency. Service providers act as intermediaries, purchasing electricity at wholesale prices and reselling it to customers at retail prices. Their pricing strategies are crafted to respond to market conditions, maximize returns, and promote efficient energy use among customers. Dynamic pricing encourages customers to modify their energy consumption based on price signals, resulting in energy savings and cost reductions. Customers actively engage in the dynamic pricing demand response (DR) program, optimizing their energy usage to reduce costs and help balance energy demand [11]. A transparent and responsive pricing model aligns customer behavior with the energy system's broader goals, promoting peak shaving and valley filling to flatten the load curve and enhance supply-demand equilibrium. Figure 1 illustrates the collaborative relationship among grid operators, service providers, and customers, showing how cooperation in the electricity market improves energy management and system stability. This model incentives all parties to adopt practices that advance their respective goals while enhancing the overall efficiency and sustainability of the energy system.

A. Customers' Model

Customer's load can be classified as load demand and load consumption, load demand is customer's required load before consumption, and load consumption is the load that is actually consumed. Each load can be classified as critical load and controllable load, which means:

$$\begin{cases} l_{t,n} = l_{t,n}^{critic} + l_{t,n}^{con}, \quad t = 1, 2, \ldots, T; \ n = 1, 2, \ldots, N_c \\ L_{t,n} = L_{t,n}^{critic} + L_{t,n}^{con}, \quad t = 1, 2, \ldots, T; \ n = 1, 2, \ldots, N_c \end{cases} \tag{1}$$

where $l_{t,n}$ and $L_{t,n}$ represent load consumption and load demand of nth customer at tth period, $l_{t,n}^{con}$ and $L_{t,n}^{con}$ denote controllable load consumption and controllable load demand of nth customer at the tth period, $l_{t,n}^{critic}$ and $L_{t,n}^{critic}$ are critical load consumption and critical load demand of nth customer at the tth period. T and N_c represent total time period length and customer number. Critical loads denote the load demand that must be critically satisfied, while controllable loads mean all system loads other than critical loads. The load consumption of critical loads must equal to load demand, which means that

$l_{t,n}^{critic} = L_{t,n}^{critic}$. Load consumption of controllable loads at customers can be described by:

$$\begin{cases} l_{t,n}^{con} = L_{t,n}^{con}[1 + \xi_t(\rho_{t,n}^{retail} - \rho_t^g)/\rho_t^g] \\ \rho_{t,n}^{retail} \geq \rho_{g,t} \end{cases} \tag{2}$$

where $\xi_t < 0$ denotes the elastic parameter at the tth time period, but it ensures $l_{t,n}^{con}$ positive, which means that it satisfies $\xi_t > \rho_t^g/(\rho_t^g - \rho_{t,n}^{retail})$ when $\rho_{t,n}^{retail} > \rho_t^g$. $\rho_{t,n}^{retail}$ represents the retail price of the nth customer at the tth period, ρ_t^g is the wholesale price from the grid operator at the tth period. The dissatisfaction cost function $C_{t,n}^{dis}$ of each customer at the tth period can be described as:

$$\begin{cases} C_{t,n}^{dis} = \dfrac{\alpha_n}{2}(L_{t,n}^{con} - l_{t,n}^{con})^2 + \beta_n(L_{t,n}^{con} - l_{t,n}^{con}) \\ D_{min} \leq L_{t,n}^{con} - l_{t,n}^{con} \leq D_{max} \end{cases} \tag{3}$$

where $\alpha_n > 0$ and $\beta_n > 0$ are parameters related to the customer (CU). Specifically, α_n represents the preference value among different CUs, while β_n is a predetermined constant. The parameter α_n reflects the CU's attitude towards reducing electricity demand: The larger α_n value indicates a greater inclination of the CU to reduce demand to increase satisfaction, and vice versa. D_{min} and D_{max} denote lower bound and upper bound of the deviation between load demand and load consumption, respectively. The quadratic term in the equation encourages the system to minimize the difference between demand and consumption, thereby promoting a more balanced and stable energy consumption pattern. The linear term represents an additional penalty when the actual controllable consumption is less than the controllable load demand. The electricity purchasing cost $C_{t,n}^{pur}$ from service provider can be described as:

$$C_{t,n}^{pur} = \rho_{t,n}^{retail} * l_{t,n} \tag{4}$$

The main goal of each customer is to minimize the cost as follows:

$$C_n = \min \sum_{t=1}^{T}(C_{t,n}^{dis} + C_{t,n}^{pur}) \tag{5}$$

B. Service Provider and Grid Operator Model

The main goal of the service provider is to maximize economic profit B by alternating dynamic retail price as follows:

$$\begin{cases} B = \max \sum_{t=1}^{T}\sum_{n=1}^{N_c}(\rho_{t,n}^{retail} - \rho_t^g) * l_{t,n} \\ \kappa_{min}\rho_{t,min}^g \leq \rho_{t,n}^{retail} \leq \kappa_{max}\rho_{t,max}^g \end{cases} \tag{6}$$

where κ_{min} and κ_{max} represent the predetermined coefficients of retail price bounds, $\rho_{t,min}^g$ and $\rho_{t,max}^g$ denote the lower bound and upper bound of the wholesale price at the tth period. The grid operator decides the wholesale price ρ_t^g for the service provider with consideration of energy supply-demand balance at the current period, it will increase as demand surplus and decrease as power surplus, which means that it can be expressed as:

$$\begin{cases} \rho_t^g = G(l_t, l_t - \hat{P}_t) \\ l_t = \sum_{n=1}^{N_c} l_{t,n} \end{cases} \tag{7}$$

where $G(\cdot)$ represents the relationship function among power generation, load demand and electricity price, l_t is the total load consumption at tth period, $\hat{P}_t$ denotes the predicted total power output at tth period. The power supply $\hat{P}_t$ can be described as $\sum_{b=1}^{N_{Bus}} P_{G,t}^b$, where $P_{G,t}^b$ denotes power generation of bth bus at tth period. The service provider also offers load shifting scheme for customers to save economic cost with invariant total load of arbitrary customer, the controllable part of load consumption can be described as:

$$\begin{cases} l_{t,n}^{con} = \sum_{t' \in \Xi_t^+} \Delta_{t',t,n} - \sum_{t' \in \Xi_t^-} \Delta_{t,t',n} \\ 0 \leq \Delta_{t',t,n} \leq \Delta_{max,t,n} \\ 0 \leq \Delta_{t,t',n} \leq \Delta_{t,max,n} \\ \sum_{t \in T}\sum_{t' \in \Xi_t^+} \Delta_{t',t,n} = \sum_{t \in T}\sum_{t' \in \Xi_t^-} \Delta_{t,t',n} \end{cases} \tag{8}$$

where $\Delta_{t',t,n}$ represents load shifting from t'th period to tth period at nth customer, Ξ_t^- and Ξ_t^+ denote period set of shifting load from and to tth period, $\Delta_{t,max,n}$ and $\Delta_{max,t,n}$ are the maximum limits of load shifting from and to tth period.

C. Objective Function

As a dynamic pricing demand response program is to maximize profit of the service provider and save economic cost of customers, the total objective can be described as follows:

$$C_L = \max[wB - (1-w)\sum_{n=1}^{N_c} C_n] \tag{9}$$

where $0 < w < 1$ represents the weight parameter. With consideration of customer' quantification cost, benefit income cannot be measured in the same way, discount weight w is applied to describe total benefit of service providers.

III. Scenario Based Optimal Operation of Hybrid Energy System in Lower-Layer Model

A. Multiple Objective Requirements

On the power supply side, hybrid energy system including stable power generator, wind power and energy storages is to coordinate all energy resources to meet load demand requirement. With consideration of the stochastic characteristics of wind power, a scenario based technique is utilized to create the optimal operation model of hybrid energy system.

1) Minimization of Generation Cost: Since power generation cost is mainly caused by stable power generators and energy storage units, it can be described as [22]:

$$\begin{cases} F_1 = \sum_{s=1}^{N_s}\sum_{t=1}^{T} Pr(s)\sum_{b=1}^{N_{Bus}}(f_{The}^{b,s} + f_{ES}^{b,s}) \\ f_{The}^{b,s} = \sum_{i=1}^{N_b^c}[\alpha_{i0}^b + \alpha_{i1}^b P_{c,i,t}^{b,s} + \alpha_{i2}^b P_{c,i,t}^{n,s\,2} \\ +|\alpha_{i3}^b sin(\alpha_{i4}^b(P_{c,i}^{b,min} - P_{c,i,t}^{b,s}))|] \\ f_{ES}^{b,s} = \sum_{l=1}^{N_b^e} \rho_{ES}^{b,l}|P_{e,l,t}^{b,s}| \end{cases} \tag{10}$$

where $Pr(s)$ represents the probability of the sth scenario, $f_{The}^{b,s}$ and $f_{ES}^{b,s}$ denote generation cost and charging/discharging cost of power generator and energy storage units at bth bus in the sth scenario, α_{i0}^b, α_{i1}^b, α_{i2}^b, α_{i3}^b and α_{i4}^b are cost coefficients of the ith stable power generator at the bth bus, $P_{c,i,t}^{b,s}$ denotes power output of the ith stable power generator at the bth bus at the tth period in the sth scenario, $P_{c,i}^{b,min}$ is the lower

bound of power output of the ith stable power generator at the bth bus, $P^{b,s}_{e,l,t}$ denotes the charging/discharging output of the lth energy storage unit at the bth bus, $\rho^{b,l}_{ES}$ denotes the charging/discharging cost parameter of the lth energy storage unit at the bth bus, N_s and N_{Bus} are the number of scenarios and buses, N^c_b and N^e_b denote the number of stable power generator and energy storages at the bth bus.

2) Minimization of Emission Rate: As emission pollution is also generated by stable power generator, another objective is to minimize the emission rate of stable power generator, which can be expressed as follows:

$$\begin{cases} F_2 = \sum_{s=1}^{N_s} \sum_{t=1}^{T} Pr(s) \sum_{b=1}^{N_{Bus}} f^{b,s}_{emi} \\ f^{b,s}_{emi} = \sum_{i=1}^{N^c_b} [\beta^b_{i0} + \beta^b_{i1} P^{b,s}_{c,i,t} + \beta^b_{i2} {P^{b,s}_{c,i,t}}^2 \\ + \beta^b_{i3} exp(\beta^b_{i4} P^{b,s}_{c,i,t})] \end{cases} \tag{11}$$

where $f^{b,s}_{emi}$ represents emission rate at the bth bus in the sth scenario, β^b_{i0}, β^b_{i1}, β^b_{i2}, β^b_{i3} and β^b_{i4} are the coefficients of emission rate of the ith stable power generator at the bth bus.

3) Minimization of statistic voltage stability index (SVSI): To ensure voltage stability and keep power system away from voltage collapse, the SVSI is utilized here to evaluate the voltage stability index, it can be generally described as follows [23]:

$$F_3 = \max_k \{\hat{L}_{1,t}, \hat{L}_{2,t}, \cdots, \hat{L}_{k,t}, \cdots, \hat{L}_{N_{D,t}}\} \tag{12}$$

where N_D denotes the number of load buses, the voltage stability index $\hat{L}_{k,t}$ of the kth bus can be expressed as follows:

$$\hat{L}_{k,t} = \sum_{s=1}^{N_s} Pr(s) |1 - \frac{\sum_{i=1}^{N_{Bus}} B_{ki} V^s_{i,t}}{V^s_{k,t}}| \tag{13}$$

where $V^s_{k,t}$ represents the voltage at the kth bus, and B_{ki} denotes the element of matrix $\boldsymbol{B}$, which can be calculated as:

$$\boldsymbol{B} = \boldsymbol{Y}^{-1}_{LL} \boldsymbol{Y}_{LG} \tag{14}$$

where $\boldsymbol{Y}_{LL}$ and $\boldsymbol{Y}_{LG}$ represent sub-matrices of Jacobian matrix. $\boldsymbol{Y}_{LL}$ contains admittance elements between load buses, while $\boldsymbol{Y}_{LG}$ has admittance between load and generator buses.

B. Constraint Limits

Here, $P^{b,s}_{c,i,t}$, $P^{b,s}_{e,l,t}$, $\theta^s_{bj,t}$ and $Q^{b,s}_{G,t}$ are taken as decision variables to optimize generation cost, emission rate and SVSI with satisfying following constraint limits. Specifically, $\theta^s_{bj,t}$ is the voltage angle between bus i and bus j at the tth period, $Q^{b,s}_{G,t}$ denotes the reactive power supply at the tth period of the bth bus.

1) Power Generation Limits: As power generators have limited adjustment capacity, stable power output must satisfy following conditions:

$$\begin{cases} P^b_{c,i,min} \le P^{b,s}_{c,i,t} \le P^b_{c,i,max} \\ DN^b_{c,i} \le P^{b,s}_{c,i,t} - P^{b,s}_{c,i,t-1} \le U P^b_{c,i} \end{cases} \tag{15}$$

where $P^b_{c,i,min}$ and $P^b_{c,i,max}$ represent the minimum and maximum output of the ith stable power generator at the bth bus, $DN^b_{c,i}$ and $U P^b_{c,i}$ denote ramp down and up limits of stable output. $P^{b,s}_{e,l,t}$ can be considered as $P^{b,s}_{cha,l,t}$ when an energy storage unit is in a charging state, otherwise, it is taken as $P^{b,s}_{dis,l,t}$ when it is in a discharging state. The charging/discharging process must also satisfy:

$$\begin{cases} E^{b,s}_{l,t+1} = E^{b,s}_{l,t} + P^{b,s}_{e,l,t} * \Delta T \\ P^{b,s}_{e,l,t} = \eta^b_{cha,l} P^{b,s}_{cha,l,t}, if\ it\ is\ charging \\ P^{b,s}_{e,l,t} = -\eta^b_{dis,l} P^{b,s}_{dis,l,t}, if\ it\ is\ discharging \\ E^b_{l,min} \le E^{b,s}_{l,t} \le E^b_{l,max} \\ 0 \le P^{b,s}_{cha,l,t} \le P^b_{cha,l,max} \\ 0 \le P^{b,s}_{dis,l,t} \le P^b_{dis,l,max} \\ E^{b,s}_{l,0} = E^b_{l,initial} \end{cases} \tag{16}$$

where $E^{b,s}_{l,t}$ represents the storage state of the lth energy storage unit at the tth period at the bth bus of the sth scenario, $P^{b,s}_{e,l,t}$ is charging/discharging output, ΔT denotes time period length, $P^{b,s}_{cha,l,t}$ and $P^{b,s}_{dis,l,t}$ denote output of energy storage unit at the charging and discharging state, $\eta^b_{cha,l}$ and $\eta^b_{dis,l}$ are the efficiency factor of energy storage unit at the charging and discharging state, $E^b_{l,min}$ and $E^b_{l,max}$ denote the minimum and maximum bound of the lth energy storage unit at the bth bus, $P^b_{cha,l,max}$ and $P^b_{dis,l,max}$ represent the maximum charging and discharging limits of the lth energy storage unit at the bth bus, $E^b_{l,initial}$ denotes the initial storage unit state of the lth energy storage unit at the bth bus.

2) Power Flow Limits: The power supply $P^{b,s}_{G,t}$ (the sth scenario of $P^b_{G,t}$) at each bus can be described as:

$$P^{b,s}_{G,t} = \sum_{i=1}^{N^c_b} P^{b,s}_{c,i,t} + \sum_{j=1}^{N^w_b} P^{b,s}_{w,j,t} + \sum_{l=1}^{N^e_b} P^{b,s}_{e,l,t} \tag{17}$$

where $P^{b,s}_{w,j,t}$ denotes wind output of the jth wind generator at the bth bus at the tth period of the sth scenario, N^w_b is the number of wind generators at the bth bus. Some power flow limits must be satisfied as follows:

$$\begin{cases} P^{b,s}_{G,t} - P^b_{D,t} = V^s_{b,t} \sum_{j=1}^{N_b} V^s_{j,t} (G_{bj} cos\theta^s_{bj,t} + B_{bj} sin\theta^s_{bj,t}) \\ Q^{b,s}_{G,t} - Q^b_{D,t} = V^s_{b,t} \sum_{j=1}^{N_b} V^s_{j,t} (G_{bj} sin\theta^s_{bj,t} + B_{bj} cos\theta^s_{bj,t}) \end{cases} \tag{18}$$

where $P^b_{D,t}$ and $Q^b_{D,t}$ represents active power and reactive power demand at the tth period of the bth bus, G_{bj} and B_{bj} denote transfer conductance and transfer susceptance between bus b and bus j, N_b is the number of adjacent buses to bth bus. The summation of load $P^b_{D,t}$ is total load consumption $\sum_{n=1}^{N} L_{n,t}$, which means $\sum_{b=1}^{N_{Bus}} P^b_{D,t} = \sum_{n=1}^{N} L_{n,t}$. At the same time, these variables have the following limits:

$$\begin{cases} V_{b,min} \le V^s_{b,t} \le V_{b,max} \\ P^b_{G,min} \le P^{b,s}_{G,t} \le P^b_{G,max} \\ Q^b_{G,min} \le Q^{b,s}_{G,t} \le Q^b_{G,max} \\ |S^s_{b,t}| \le S^{max}_b \end{cases} \tag{19}$$

where $V_{b,min}$ and $V_{b,max}$ represent the minimum and maximum voltage limits of bus b, which is mainly to the control voltage to approximate reference voltage. $P^b_{G,min}$ and $P^b_{G,max}$ denote the minimum and maximum bounds of power supply at bus b, $Q^b_{G,min}$ and $Q^b_{G,max}$ are the minimum and maximum reactive power at bus b, $S^s_{b,t}$ denotes the apparent power flow at bus

b, it can be calculated by $\sqrt{{P_{G,t}^{b,s}}^2 + {Q_{G,t}^{b,s}}^2}$, and its maximum limit is labeled $S_{b,t}^{max}$.

C. Probabilistic Analysis on Intermittent Energy Resources

Wind power can be the primary reason for the stochastic characteristics of power system, while wind power generation is strongly related to wind speed, the relationship between them can be described as follows:

$$P_{w,j,t}^{b,s} = \begin{cases} P_{w,j,max}^{b} * \dfrac{v_j^b - v_{j,in}^b}{v_{j,rate}^b - v_{j,in}^b}, & v_{j,in}^b \le v_j^b < v_{j,rate}^b \\ P_{w,j,max}^{b}, & v_{j,rate}^b \le v_j^b < v_{j,out}^b \\ 0, & v_j^b < v_{j,in}^b or v_j^b \ge v_{j,out}^b \end{cases} \tag{20}$$

where v_j^b represents wind speed of the jth wind generator at the bth bus, $v_{j,in}^b$, $v_{j,rate}^b$ and $v_{j,out}^b$ denote cut-in, rated and cut-out wind speed of the jth wind generator at the bth bus. As it is stated in the literature [24], wind speed follows the Weibull distribution function. The probability of the generated scenario is mainly calculated with the distribution function of wind power, which can be calculated as follows:

$$Pr(s) = \int \int_{\Omega} \prod_{j=1}^{N_b^w} \prod_{b=1}^{N_{Bus}} f_{dis}(P_{w,j,t}^{b,s}) dP_{w,1,t}^{b,s} \cdots dP_{w,N_b^w,t}^{N_{Bus},s} \tag{21}$$

where Ω represents the security level domain of generated scenarios, $f(\cdot)$ denotes the probability density function (PDF) of $F_{dis}(\cdot)$. With consideration of the stochastic characteristics, suppose total wind power at the tth period in the sth scenario is labeled as $P_{w,t}^s$, $P_{w,t}^s$ can be divided into two parts: stable output $\overline{P_{w,t}^s}$ and fluctuating output $\widetilde{P_{w,t}^s}$. Here, the fluctuating output $\widetilde{P_{w,t}^s}$ can be divided into several intervals with different probabilities, which is mainly on the basis of different levels of power supply security. Since energy storage can be an excellent energy resources for supplementing power disturbance due to its quick response ability, the uncertainty risk levels can be divided as:

$$\Omega = \begin{cases} \Omega_1, 0 \le \widetilde{P_{w,t}^s} < \dfrac{1}{N} \displaystyle\sum_{l=1}^{N_b^e} \sum_{b=1}^{N_{Bus}} P_{l,max}^b \\ \Omega_2, \dfrac{1}{N} \displaystyle\sum_{l=1}^{N_b^e} \sum_{b=1}^{N_{Bus}} P_{l,max}^b \le \widetilde{P_{w,t}^s} < \dfrac{2}{N} \sum_{l=1}^{N_b^e} \sum_{b=1}^{N_{Bus}} P_{l,max}^b \\ \cdots \\ \Omega_N, \dfrac{N-1}{N} \displaystyle\sum_{l=1}^{N_b^e} \sum_{b=1}^{N_{Bus}} P_{l,max}^b \le \widetilde{P_{w,t}^s} < \sum_{l=1}^{N_b^e} \sum_{b=1}^{N_{Bus}} P_{l,max}^b \end{cases} \tag{22}$$

where $\Omega_i (i = 1, 2, \cdots, N)$ represents the ith uncertainty risk level, N is the number of uncertainty risk levels, $P_{l,max}^b$ denotes the maximum and minimum charging/discharging output.

IV. The Proposed Two-Stage Optimization Strategy for Smart Energy Management

Due to the dynamic characteristics of demand response and stochastic power generation, a two-stage optimization strategy is proposed to address smart energy management within the dynamic pricing mechanism of power systems. This strategy can be viewed as a hierarchical strategy consisting of an upper and lower stage, designed to optimize both the load side and the generation side. The upper layer (first stage) involves optimizing load-side management, including interactions between consumers and service providers. The grid operator determines the wholesale electricity price based on deviation of total load consumption and generation. Subsequently, the service provider sets the retail electricity price to maximize returns while promoting efficient energy use among customers. The lower stage (second stage) focuses on optimizing power generation. In the upper stage, a learning-based approach is used to solve the dynamic pricing problem on the load side. The pricing strategy is optimized based on demand response, which, in turn, affects customer behavior and load demand patterns. The dynamic pricing mechanism aims to balance supply and demand by encouraging customers to shift their consumption to off-peak hours, thereby smoothing the overall load curve. In the lower stage, a scenario-based PBI approach is used to optimize power dispatch, considering the intermittent nature of energy resources. This optimization considers probabilistic characteristics and stochastic scenarios to ensure economic cost-effectiveness, reduce emissions, and improve voltage stability. Power dispatch optimization relies on load demand derived from the dynamic pricing layer to ensure that the generation mix aligns with the expected load pattern. Separating the optimization into these two stages allows for strategic optimization of each aspect, resulting in a comprehensive strategy that maximizes economic efficiency and system reliability.

A. Improved Deep Reinforcement Learning Approach for Dynamic Pricing Demand Response

Given the inherent uncertainty and flexibility within the electricity market, employing deep reinforcement learning for dynamic pricing decision-making presents substantial advantages. This approach eliminates the need to predefined an environmental model dictating retail pricing behavior. Instead, it facilitates an empirical understanding of the relationship between retail prices and profits through dynamic interactions with customers (CU). Furthermore, deep reinforcement learning adapts to constantly evolving market conditions by continuously integrating both historical and real-time data. This capability not only enhances the effectiveness of demand response mechanisms but also contributes to the development of optimal dynamic pricing strategies. The dynamic pricing problem can be modeled as a discrete finite horizon Markov decision process (MDP) due to uncertainty relationship between wholesale price and load demand. As it is known that MDP has four main elements, which can be defined as follows:

1) State Set: The power demand of customers is defined as a state set, which can be described as $S = \{S_1, S_2, \cdots, S_t, \cdots, S_T\}$, where $S_t = \{(L_{t,n}, l_{t,n}) | n = 1, 2, \ldots, N_c\}$.

2) Action Set: The action set mainly consists of a retail price and a wholesale price, which can be described as $A = \{A_1, A_2, \cdots, A_t, \cdots, A_T\}$, where $A_t = \{(\rho_{t,n}^{retail}, \rho_t^g) | n = 1, 2, \cdots, N_c\}$.

3) Reward Set: The current reward of state S_t and action A_t can be described as $R(S_t, A_t)$. Here, constraint limits are taken as negative parts due to their negative effect on optima approximation, it can be described as $-Vio_{con}$, where Vio_{con} represents the extent of constraint violation, the positive part of reward function is described as E, hence reward functions can be defined as $E-\xi_R Vio_{con}$, where ξ_R denotes the discount factor. The constraint violation Vio_{con} can be described as:

$$\begin{aligned} Vio_{con} = & \sum_{t=1}^{T}\sum_{n=1}^{N_c} \max\{\rho_t^g - \rho_{t,n}^{retail}, 0\} \\ & + \sum_{t=1}^{T}\sum_{n=1}^{N_c} \max\{D_{min} - (L_{t,n}^{con} - l_{t,n}^{con}), 0\} \\ & + \sum_{t=1}^{T}\sum_{n=1}^{N_c} \max\{(L_{t,n}^{con} - l_{t,n}^{con}) - D_{max}, 0\} \\ & + \sum_{t=1}^{T}\sum_{n=1}^{N_c} \max\{\kappa_{min}\rho_{t,min}^g - \rho_{t,n}^{retail}, 0\} \\ & + \sum_{t=1}^{T}\sum_{n=1}^{N_c} \max\{\rho_{t,n}^{retail} - \kappa_{max}\rho_{t,max}^g, 0\} \\ & + \sum_{t=1}^{T} |\rho_{g,t} - G(l_t, l_t - \hat{P}_t)| \\ & + \sum_{t=1}^{T}\sum_{n=1}^{N_c} [\sum_{t'\in\Xi_t^+} \max(\Delta_{t',t,n} - \Delta_{max,t,n}, 0) \\ & + \sum_{t'\in\Xi_t^-} \max(\Delta_{t,t',n} - \Delta_{t,max,n}, 0)] \\ & + |\sum_{t\in T}\sum_{t'\in\Xi_t^+} \Delta_{t',t,n} - \sum_{t\in T}\sum_{t'\in\Xi_t^-} \Delta_{t,t',n}| \end{aligned} \tag{23}$$

4) State-Value Function: The state-value function mainly evaluates the results after a sequence of state and action, which means that it must be a function of state S and action A. Combined with Bellman theory, the state-value function $Q(S, A)$ can be described as;

$$Q(S, A) = \max_A [R(S, A) + \xi_Q \max_{A'} Q(S', A')] \tag{24}$$

where S' and A' represent the state and action vector at the next stage, ξ_Q denotes the discount factor of the state-value function. In the DDPG algorithm, target critic network and target actor network are utilized to enhance the learning efficiency, and the transition is sampled from $Environment$ to be stored in $ReplayBuffer$, the algorithm structure is shown in Fig.2.

After actor-network and critic-network training on weights θ^μ and θ^Q, its generated action can be described as $argmax_A Q(S, A)$. During the training process, two target networks are copied to deduce target value $y_i = R(S^{(i)}, A^{(i)}) + \xi_Q Q(S^{(i)}, A^{(i)})$, where $S^{(i)}$ and $A^{(i)}$ represent the ith sample of state vector and actor vector. Then, the training loss function can be described as:

$$L = \frac{1}{N_{sam}} \sum_{i=1}^{N_{sam}} (Q(S^{(i)'}, A^{(i)'}) - y_i)^2 \tag{25}$$

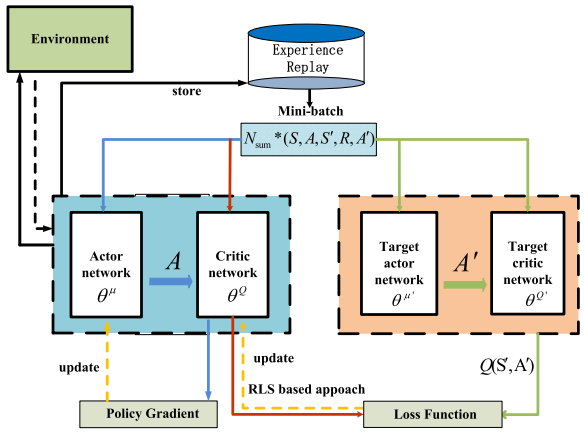


Fig. 2. The algorithm structure of RLS based DDPG.

where N_{sam} denotes the number of samples, $S^{(i)'}$ and $A^{(i)'}$ represent the ith sample of the state vector and actor vector at the next stage. Those traditional DDPG algorithms adopt off-line policy training mechanism, the policy gradient requires to restart when new samples comes for training, which lacks of flexibility to deal with dynamic pricing profile in the demand response model. Hence, a RLS approximation approach is utilized to improve the DDPG algorithm. With consideration of the recursive model for sample learning, the loss function can be rewritten as follows:

$$L = \sum_{i=k-p+1}^{k} (Q(S^{(i)'}, A^{(i)'}) - y_i)^2 \tag{26}$$

where k denotes the current training index, p is the recursive training length. For avoiding an over-fitting issue, a regularization norm must also be taken into consideration. Moreover, state and act vector is a time-related Markov decision process, recent samples have more influence on current training efficiency, and a forgetting factor should also be considered. Then the loss function can be updated as:

$$L = \sum_{i=k-p+1}^{k} \gamma_Q^{k-i} (Q(S^{(i)'}, A^{(i)'}) - y_i)^2 + \lambda_Q \sum_{j=1}^{N_Q} \theta_Q^{(j)2} \tag{27}$$

where $0 < \gamma_Q < 1$ denotes the forgetting factor, λ_Q represents the discount factor. With consideration of neural network learning of a critic network, the state-value function $Q(S^{(i)'}, A^{(i)'})$ can be described as $\sum_{j=1}^{N_Q} \theta_Q^{(j)} \phi_j(A^{(i)'})$, where N_Q is the number of hidden neurons in a critic network, $\theta_Q^{(j)}$ represents the jth component of the critic network weight, $\phi_j(\cdot)$ denotes the network function. The compact version of the loss function can be expressed as follows:

$$L(\mathbf{\Theta}) = (\boldsymbol{Y} - \mathbf{\Phi}^T\mathbf{\Theta})^T \mathbf{\Gamma} (\boldsymbol{Y} - \mathbf{\Phi}^T\mathbf{\Theta}) + \lambda_Q ||\mathbf{\Theta}||^2 \tag{28}$$

where $\mathbf{\Theta}$ represents vector $[\theta_Q^{(1)}, \theta_Q^{(2)}, \cdots, \theta_Q^{(N_Q)}]^T$, $\boldsymbol{Y} = [y_{k-p+1}, y_{k-p+2}, \cdots, y_k]^T$, $\mathbf{\Phi}$ is the $N_Q \times p$ matrix with element $\phi_j(A^{(i)'})$, and matrix $\mathbf{\Gamma}$ can be described as:

$$\begin{pmatrix} \gamma_Q^{p-1} & 0 & 0 & 0 \\ 0 & \gamma_Q^{p-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{29}$$

The optimal solution of the weight vector $\boldsymbol{\Theta}$ can be approximated as follows:

$$\boldsymbol{\Theta}_{m+1} = (\lambda_Q \boldsymbol{I} + \boldsymbol{R}_{m+1})^{-1} \boldsymbol{U}_{m+1} \tag{30}$$

where $\boldsymbol{I}$ represents the identity matrix, $\boldsymbol{R}_{m+1}$ and $\boldsymbol{U}_{m+1}$ denote two exponentially-weighted covariance matrices, which can be deduced as:

$$\begin{cases} \boldsymbol{R}_{m+1} = \gamma_Q \boldsymbol{R}_m + \boldsymbol{\Phi}^T \boldsymbol{\Phi} \\ \boldsymbol{U}_{m+1} = \gamma_Q \boldsymbol{U}_m + \boldsymbol{\Phi}^T \boldsymbol{Y} \end{cases} \tag{31}$$

According to the above procedures, the optimal weight of actor network and critic network can be obtained, then the weight $\theta^{\mu'}$ and $\theta^{Q'}$ of the target actor network and critic network can be updated as follows:

$$\begin{cases} \theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'} \\ \theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'} \end{cases} \tag{32}$$

where $0 < \tau < 1$ denotes the control parameter. After the above procedures, those obtained target network weights can guide learning at the next round.

B. Decomposition Based Probabilistic Multi-Population Evolutionary Algorithm for Multiple Objective Optimization

After dynamic optimization of the price-incentive demand response, the optimal system load at each period can be deduced, and the remaining issue is to manage power output on the power generation side. Due to the stochastic characteristics and multiple-objective requirement of power system, a probabilistic multi-objective evolutionary algorithm is improved to optimize its energy management problem. Obviously, the optimal operation of power system is a multi-objective optimization problem. In this paper, a decomposition based probabilistic PBI algorithm is developed with an adaptive particle swarm optimization technique. The decision variable of the sth scenario can be noted as $x^s = [x_1^s, \cdots, x_b^s, \cdots, x_{Bus}^s]$, and each component x_b^s can be described as:

$$\begin{pmatrix} P_{c,1,1}^{b,s} & \cdots & P_{c,N_b^c,1}^{b,s} & P_{e,1,1}^{b,s} & \cdots & P_{e,N_b^e,1}^{b,s} & \theta_{b,1}^s \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{c,1,T}^{b,s} & \cdots & P_{c,N_b^c,T}^{b,s} & P_{e,1,T}^{b,s} & \cdots & P_{e,N_b^e,T}^{b,s} & \theta_{b,T}^s \end{pmatrix} \tag{33}$$

Since those generated scenarios can increase computational complexity during the optimization process, a scenario reduction technique is also utilized to deal with this problem. The main idea of this scenario reduction is to screen out those similar scenarios with covariance analysis, which can retain the efficiency with less scenarios. For a given scenario x^s, calculate the covariance $Cov_{ss'}$ between scenario $s \in N_s$ and another arbitrary scenario $s' \in N_s$, then the covariance set $Cov_s = \{Cov_{ss'} | s' \in N_s \& s' \neq s\}$, retains one scenario and delete all other similar scenarios from Cov_s. After the above scenario reduction technique on every scenario, then it can obtain streamlined scenarios. Combined with scenario based probabilistic characteristics, the optimal operation model of a hybrid energy system can be divided into N_D probabilistic subproblems, which can be described as:

$$\begin{cases} \min g^{pbi}(x|\lambda^i, z^*) = \sum_{s=1}^{N_s} Pr(s)(d_1^{i(s)} + \zeta d_2^{i(s)}) \\ d_1^{i(s)} = (F(x^s) - z^*)^T \lambda^i / ||\lambda^i|| \\ d_2^{i(s)} = ||F(x^s) - z^* - (d_1^{i(s)}/||\lambda^i||)\lambda^i|| \\ x^s \in \Omega, s = 1, 2, \cdots, N_s \end{cases} \tag{34}$$

where λ^i represents the weight vector of the ith subproblem, z^* denotes the utopia vector, $d_1^{i(s)}$ is the distance between the projection point and z^*, $d_2^{i(s)}$ is the distance between initial point and projection point, $F(x^s)$ denotes the objective function vector, which can be described as $[f_1(x^s), f_2(x^s), \cdots, f_M(x^s)]$, M denotes the number of objective functions, Ω is the feasible domain. Then, each subproblem can be treated as a single objective optimization problem, and a random drift particle swarm optimization algorithm is utilized to optimize it. Here, the velocity and position of each particle can be updated as follows [25]:

$$\begin{cases} v_{i,j}^k = \alpha |C_j^k - x_{i,j}^{k-1}| \delta_{i,j}^k + \beta (E_{i,j}^k - x_{i,j}^{k-1}) \\ x_{i,j}^k = x_{i,j}^{k-1} + v_{i,j}^k, \quad j = 1, 2, \cdots, d \end{cases} \tag{35}$$

where $v_{i,j}^k$ and $x_{i,j}^k$ represent the velocity and position of the ith particle in the jth dimension at the kth step, α and β denote the drift coefficients, d is the number of the element dimension, $\delta_{i,j}^k$ is the random number of the ith particle in the jth component at the kth step, which is generated in a standard normal distribution. C_j^k is the mean of the best position of the jth dimension at the kth step, and $E_{i,j}^k$ is the local focus position of the ith particle in the jth dimension at the kth step, they can be calculated as follows:

$$\begin{cases} C_j^k = \dfrac{\sum_{i=1}^{N_p} y_{i,j}^{k-1}}{N_p} \\ E_{i,j}^k = \kappa_{i,j}^k y_{i,j}^{k-1} + (1 - \kappa_{i,j}^k) y_{i*,j}^{k-1} \end{cases} \tag{36}$$

where N_p represents the number of particles, $y_{i,j}^{k-1}$ denotes the jth dimension of the ith particle at the $k-1$th step, $y_{i*,j}^{k-1}$ is the global best position at the $k-1$th step, $\kappa_{i,j}^k$ denotes the random drift parameter, which can be calculated as follows:

$$\kappa_{i,j}^k = \frac{c_1 r_{1,i,j}^k}{c_1 r_{1,i,j}^k + c_2 r_{2,i,j}^k} \tag{37}$$

where c_1 and c_2 represent acceleration coefficients of seeking $pbest$ and $gbest$ solutions, $r_{1,i,j}^k$ and $r_{2,i,j}^k$ denote two random generated number in (0, 1]. According to the above iterations, each subproblem can generate an optimal solution with a certain weight vector, and all those optimal solutions can form Pareto fronts of a multi-objective optimization problem. In addition, all constraint handling techniques can be found in the literature [26].

C. The Flowchart of Proposed Optimization Strategy for Smart Energy Management

In the process of dynamic pricing based on learning algorithms, a neural network is utilized to approximate the function $G(\cdot)$ that models the relationship between power generation, load demand, and electricity prices. The neural network, which

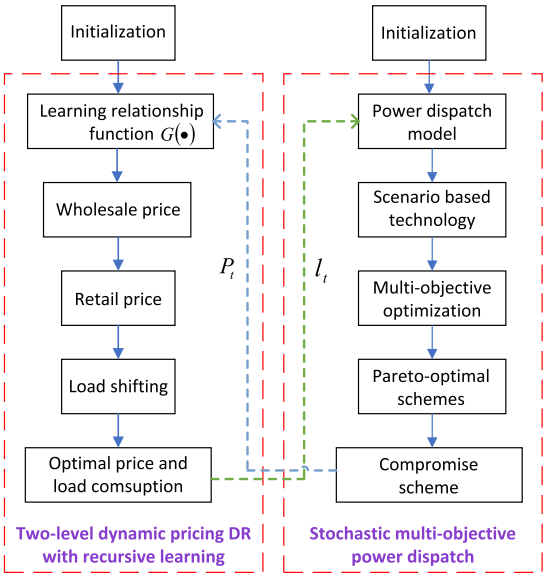


Fig. 3. The flowchart of proposed two-level optimization strategy.

uses power generation and load demand as input variables, seeks to predict electricity prices by training on historical data. Through iterative training, the model incrementally refines its approximation of the objective function $G(\cdot)$, thereby improving its ability to predict wholesale electricity prices under specific conditions of power generation and load demand. Subsequently, reinforcement learning is applied to determine the optimal retail electricity price that motivates customers to shift their load to adjust their energy usage in response to fluctuating prices. This step is crucial for formulating an effective pricing strategy that promotes the desired pattern of electricity consumption. Demand forecasting is crucial in enabling the supply side to plan and optimize power dispatch effectively. In the subsequent phase of scenario-based power dispatch, scenario analysis techniques are employed to envision various potential future scenarios in energy supply, considering the variability and unpredictability of renewable energy sources. A multi-objective optimization approach is then implemented to balance economic costs, emission rates, and voltage stability. The most efficient dispatch strategy is obtained through the analysis of Pareto-optimal solutions. This strategy not only addresses economic and environmental considerations but also ensures voltage stability. The results inform the demand side, facilitating further refinement of the pricing model and dynamic pricing strategies. The flowchart of proposed two-level optimization strategy has been presented in Fig.3.

V. Case Study

To verify the efficiency of the proposed optimization strategy, this case presents a modified IEEE 112-bus test system, where total system load is conducted by grid operator, service provider, and customers. The results analysis of this case mainly consists of dynamic pricing DR and multi-objective power dispatch. The details of dynamic pricing DR can be found in the literature [11], [27], the test system is studied at a peak load 35 p.u. (base power 100KVA). On the power supply side, the IEEE 112-bus system is modified with consideration

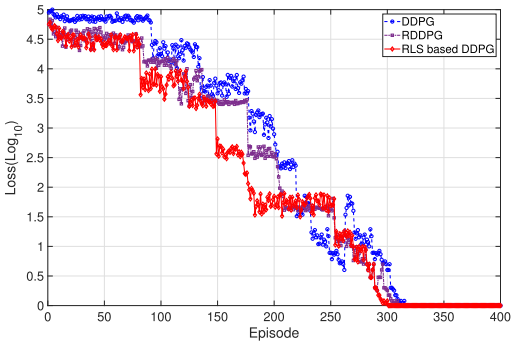


Fig. 4. The training process of DDPG, RDDPG and RLS based DDPG.

of three objectives, which include economic cost, emission rate and SVSI issue, and related details can be seen in the literature [23], [24], [27]. The entire time period is set to 24 timeslots, that is, 24 hours a day. Considering the complexity of the system simulation, the load demand curve of the CU is obtained according to SDG&E, Wholesale electricity prices are based on online data provided by ComEd on June 22, 2017, which can be seen in literature [11]. Some learning parameters can be set as follows: The mini-batch size is set as 20, the forgetting factor γ_Q is set as 0.8, the discount factor λ_Q is set as 0.65, the number of hidden neurons in critic network is set as 10. In the upper-layer model optimization, it mainly presents the wholesale price and retail price results to control load demand and load consumption to maximize the total benefit. In the lower-layer model optimization, four test cases of multi-objective optimization results are presented as: 1) Case 1: Optimization of economic cost and emission rate; 2) Case 2: Optimization of economic cost and SVSI; 3) Case 3: Optimization of emission rate and SVSI; 4) Case 4: Optimization of economic cost, emission rate and SVSI.

A. Upper-Layer Model Optimization With Improved DDPG Approach

Since the relationship between electricity price and system load is unknown, an improved DRL approach is utilized to learn its relationship function to seek minimum economic cost or maximum profit. The learning process of the proposed DRL method is presented in Fig.4, where it can be seen that the proposed DRL converges with nearly 300 samples. In comparison to DDPG and rule based DDPG (RDDPG) in literature [28], the proposed RLS based DDPG performs better for dealing with the training task. The details about relationship between electricity price and total system load can also be found in literature [28]. The optimal price and controllable load are shown in Fig.5, it can be seen that the load peak appears mainly at the 17-20th period, the retail price is always higher than wholesale price, and the peak value appears mainly at nearly load-peak period.

After the price incentive strategy, the system load before adjustment and after adjustment are presented in Fig.6, system load after adjustment has a more flat demand curve in comparison to that before adjustment, which can be better satisfied by power assignment from the power generation side. Moreover, the comparison of the obtained results with DDPG and RDDPG are also provided in Table.I, weight parameter w

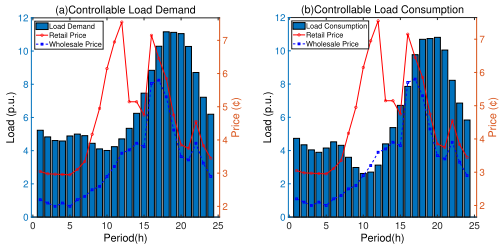


Fig. 5. The obtained load process and retail price by dynamic pricing DR.

TABLE I
THE COMPARISON OF DRL RESULTS ON DYNAMIC PRICING DR MODEL

Index	RLS based DDPG	RDDPG	DDPG
Total benefit ($)	3198	2993	2875
Service benefit ($)	9976	9875	9813
Service cost($)	57033	58115	58367
Peak-valley deviation (p.u.)	12	12	14
Computational time (s)	58	58	61

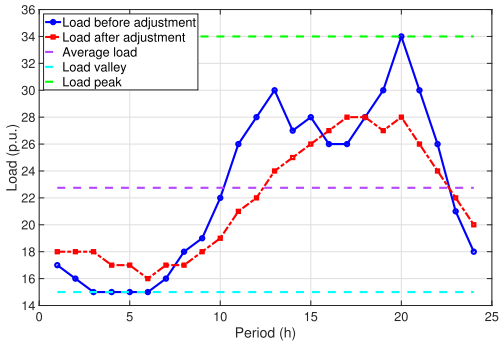


Fig. 6. The load before adjustment and load after adjustment.

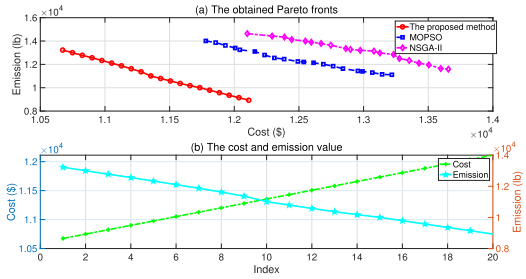


Fig. 7. The obtained schemes in case 1.

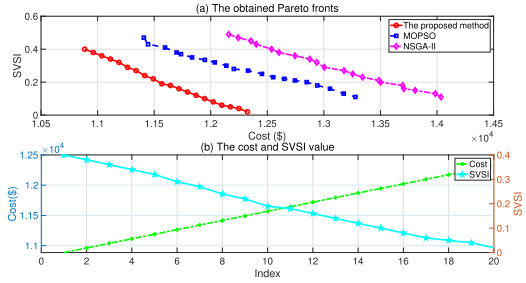


Fig. 8. The obtained schemes in case 2.

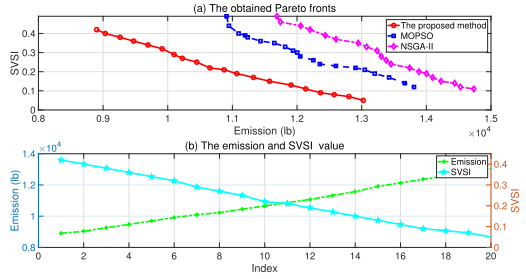


Fig. 9. The obtained schemes in case 3.

is set as 0.9, and total benefit can be calculated by service benefit and service cost, and peak-valley deviation is well controlled within 12MW, which is less than that of DDPG and RDDPG. Combined with the computational time results, it can be found that the proposed method can obtain larger benefit as well as less economic cost within less computational time.

B. Lower-Layer Model Optimization With Probabilistic PBI Based Multi-Objective Optimization Approach

After load control of the dynamic pricing DR model, the remaining task is to assign power output to different generators to satisfy system load on the demand side under the stochastic environment. Stochastic scenarios are generated to simulate the operation process, the PDF of each scenario can be calculated with the PDF of wind power in literature [24]. To optimize economic cost, emission rate and SVSI simultaneously, a PBI framework based multi-objective particle swarm optimization is utilized to obtain the optimal Pareto fronts. Here, the test case consists of four test cases, and it generates 20 Pareto optimal schemes for each test case. (1) Case 1: The economic cost and emission rate are taken as two objectives, those obtained Pareto optimal solutions are shown in Fig.7 (a) and the economic cost and emission rate value are presented in Fig.6 (b). It can be seen that the proposed method can obtain Pareto fronts with both better convergence and diversity distribution in comparison to NSGA-II and MOPSO in literature [29]. Those obtained 20 Pareto optimal values by proposed method are listed in Fig 7 (b), it also reveals that economic cost and emission rate are two conflicting objectives in this optimization model. (2) Case 2: The obtained Pareto optimal solutions of economic cost and SVSI are presented in Fig.8, where the proposed method can obtain Pareto optimal fronts with better convergence and diversity distribution than other alternatives when economic cost and SVSI conflict with each other. (3) Case 3: The obtained results of optimizing emission rate and SVSI are presented in Fig.9, it can be seen that these two objectives conflict with each other, and the proposed method can obtain better Pareto fronts. (4) Case 4: The economic cost, emission rate and SVSI are optimized simultaneously, those obtained Pareto optimal solutions are presented in Fig.10. According to Fig.10, it can be seen that the obtained Pareto front by the proposed method converge better than that of the MOPSO algorithm and NSGA-II, and it also has better diversity distribution than other alternatives. In addition, it can be found that these three objectives are contradictory in the optimization model according to parallel axis plot results in Fig.11.

For further analysis on the obtained optimal scheme, scheme (10) is taken as the compromise scheme (which is labeled in Fig.10), its comparison results with other alternatives are presented in Table.II, where it can see that the proposed

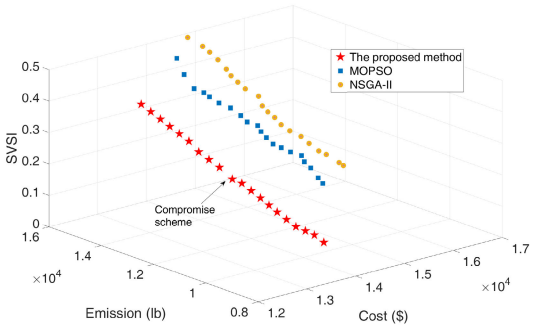


Fig. 10. The obtained Pareto fronts in case 4.

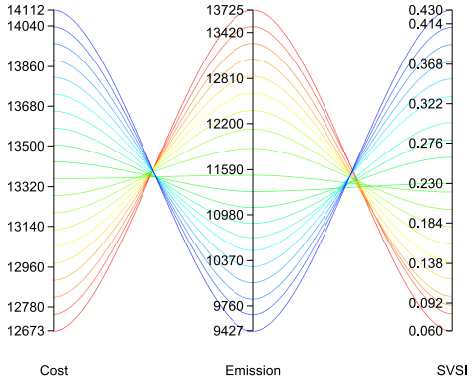


Fig. 11. The parallel axis plot of 3 objective values.

TABLE II
THE COMPARISON OF MULTI-OBJECTIVE OPTIMIZATION RESULT AND EFFICIENCY

Index	Improved PBI method	MOPSO	NSGA-II
Economic Cost ($)	13354.6	12442.7	12812.9
Emission rate (lb)	11512	12277	13299
SVSI	0.23	0.28	0.29
Ave. VD (p.u.)	0.031	0.33	0.34
Computational time (s)	59	65	71

multi-objective optimization can obtain less economic cost and emission rate at a safer security level, the average voltage deviation (VD) is 0.031, which is more stable than other two alternatives. The above result analysis reveals that the proposed method can be a viable alternative for energy management of power system with dynamic pricing DR in a stochastic environment.

VI. CONCLUSION

According to methodology analysis, some merits can be concluded as follows: (1) The proposed RLS based DDPG can accelerate on-line learning efficiency with a recursive least square approach and forgetting factors, which can deal with two-level dynamic pricing DR well. (2) The developed multi-objective optimization algorithm under a probabilistic PBI framework can optimize different objectives simultaneously, as well as reduce computational complexity in a stochastic environment. While smart energy management requires further dynamic coordination between load demand and power generation, the future work is to focus on the power-load dynamic coordination strategy.

REFERENCES

[1] K. Ojand and H. Dagdougui, "Q-learning-based model predictive control for energy management in residential aggregator," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 70–81, Jan. 2022.

[2] J. Ruan et al., "Graph deep learning-based retail dynamic pricing for demand response," *IEEE Trans. Smart Grid*, vol. 14, no. 6, pp. 4385–4397, Nov. 2023.

[3] P. U. Herath et al., "Computational intelligence-based demand response management in a microgrid," *IEEE Trans. Ind. Appl.*, vol. 55, no. 1, pp. 732–740, Jan. 2019.

[4] Y. Li, M. Han, Z. Yang, and G. Li, "Coordinating flexible demand response and renewable uncertainties for scheduling of community integrated energy systems with an electric vehicle charging station: A bi-level approach," *IEEE Trans. Sustain. Energy*, vol. 12, no. 4, pp. 2321–2331, Oct. 2021.

[5] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.

[6] L. Wen, K. Zhou, W. Feng, and S. Yang, "Demand side management in smart grid: A dynamic-price-based demand response model," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 1439–1451, 2024.

[7] R. Schumacher et al., "Self-sustainable dynamic tariff for real time pricing-based demand response: A Brazilian case study," *IEEE Access*, vol. 9, pp. 141013–141022, 2021.

[8] R. Lu, R. Bai, Z. Luo, J. Jiang, M. Sun, and H.-T. Zhang, "Deep reinforcement learning-based demand response for smart facilities energy management," *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8554–8565, Aug. 2022.

[9] H. Li, Z. Wan, and H. He, "Real-time residential demand response," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4144–4154, Sep. 2020.

[10] X. Kou et al., "Model-based and data-driven HVAC control strategies for residential demand response," *IEEE Open Access J. Power Energy*, vol. 8, pp. 186–197, 2021.

[11] R. Lu, S. H. Hong, and X. Zhang, "A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Appl. Energy*, vol. 220, pp. 220–230, Jun. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918304112

[12] A. Karapetyan et al., "A competitive scheduling algorithm for online demand response in islanded microgrids," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3430–3440, Jul. 2021.

[13] A. Montazerolghaem and M. H. Yaghmaee, "Demand response application as a service: An SDN-based management framework," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 1952–1966, May 2022.

[14] M. Z. Oskouei, S. Zeinal-Kheiri, B. Mohammadi-Ivatloo, M. Abapour, and H. Mehrjerdi, "Optimal scheduling of demand response aggregators in industrial parks based on load disaggregation algorithm," *IEEE Syst. J.*, vol. 16, no. 1, pp. 945–953, Mar. 2022.

[15] H. Zhang, D. Yue, C. Dou, K. Li, and X. Xie, "Event-triggered multiagent optimization for two-layered model of hybrid energy system with price bidding-based demand response," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 2068–2079, Apr. 2021.

[16] N. Aguiar, A. Dubey, and V. Gupta, "Pricing demand-side flexibility with noisy consumers: Mean-variance trade-offs," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1151–1161, Mar. 2023.

[17] W. Zhong, K. Xie, Y. Liu, C. Yang, S. Xie, and Y. Zhang, "Distributed demand response for multienergy residential communities with incomplete information," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 547–557, Jan. 2021.

[18] D. Zhang, H. Zhu, H. Zhang, H. H. Goh, H. Liu, and T. Wu, "Multi-objective optimization for smart integrated energy system considering demand responses and dynamic prices," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1100–1112, Mar. 2022.

[19] R. Carli, G. Cavone, T. Pippia, B. De Schutter, and M. Dotoli, "Robust optimal control for demand side management of multi-carrier microgrids," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 1338–1351, Jul. 2022.

[20] S. Wang, H. Gangammanavar, S. D. Eksioglu, and S. J. Mason, "Stochastic optimization for energy management in power systems with multiple microgrids," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 1068–1079, Jan. 2019.

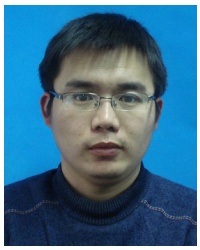
[21] S. Bahrami, Y. C. Chen, and V. W. S. Wong, "Deep reinforcement learning for demand response in distribution networks," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1496–1506, Mar. 2021.

[22] H. Nourianfar and H. Abdi, "A new technique for investigating wind power prediction error in the multi-objective environmental economics problem," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1379–1387, Mar. 2023.
[23] Y. Li, P. Wang, H. B. Gooi, J. Ye, and L. Wu, "Multi-objective optimal dispatch of microgrid under uncertainties via interval optimization," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2046–2058, Mar. 2019.
[24] H. Zhang, D. Yue, W. Yue, K. Li, and M. Yin, "MOEA/D-based probabilistic PBI approach for risk-based optimal operation of hybrid energy system with intermittent power uncertainty," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 4, pp. 2080–2090, Apr. 2021.
[25] W. T. Elsayed, Y. G. Hegazy, M. S. El-Bages, and F. M. Bendary, "Improved random drift particle swarm optimization with self-adaptive mechanism for solving the power economic dispatch problem," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1017–1026, Jun. 2017.
[26] H. Zhang, D. Yue, X. Xie, S. Hu, and S. Weng, "Multi-elite guide hybrid differential evolution with simulated annealing technique for dynamic economic emission dispatch," *Appl. Soft Comput.*, vol. 34, pp. 312–323, Sep. 2015.
[27] S. Chen, "Smart energy management system for microgrid planning and operation," Ph.D. dissertation, Nanyang Technol. Univ., Singapore, 2012.
[28] H. Zhang, D. Yue, C. Dou, and G. P. Hancke, "A three-stage optimal operation strategy of interconnected microgrids with rule-based deep deterministic policy gradient algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1773–1784, Feb. 2024.
[29] Y. Cao, Y. Zhang, H. Zhang, X. Shi, and V. Terzija, "Probabilistic optimal PV capacity planning for wind farm expansion based on NASA data," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1291–1300, Jul. 2017.

Huifeng Zhang (Senior Member, IEEE) received the Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China, in 2013. From 2014 to 2016, he was a Post-Doctoral Fellow with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. From 2017 to 2018, he was granted a Visiting Research Fellow by China Scholarship Council to study with Queens University Belfast and the University of Leeds, U.K. He is currently an Associate Professor with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications. His current research interests include electrical power management, optimal operation of power systems, distributed optimization, and multi-objective optimization.

Jiapeng Huang is currently pursuing the M.Sc. degree in electronic and information engineering with Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include optimization of distribution networks and applications of machine learning in power systems.

Dong Yue (Fellow, IEEE) received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1995. He is currently a Professor and the Dean of the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China, and also a Changjiang Professor with the Department of Control Science and Engineering. His current research interests include the analysis and synthesis of networked control systems, multiagent systems, optimal control of power systems, and the Internet of Things. He is an Associate Editor of the IEEE Control Systems Society Conference Editorial Board and *International Journal of Systems Science*.



Xiangpeng Xie (Senior Member, IEEE) received the B.S. and Ph.D. degrees in engineering from Northeastern University, Shenyang, China, in 2004 and 2010, respectively. From 2012 to 2014, he was a Post-Doctoral Fellow with the Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, China. He is currently a Professor with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include fuzzy modeling and control synthesis, state estimations, optimization in process industries, and intelligent optimization algorithms.

Zhijun Zhang (Member, IEEE) received the Ph.D. degree in information acquisition and control from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2022. He is currently a Research Fellow with Nanyang Technological University, Singapore. His current research interests include optimal operation of power systems, energy storage systems, and electric vehicles.

Gerhard P. Hancke (Life Fellow, IEEE) received the B.Sc. and B.Eng. degrees and the M.Eng. degree in electronic engineering from the University of Stellenbosch, South Africa, in 1970 and 1973, respectively, and the Ph.D. degree from the University of Pretoria, South Africa, in 1983. He is a Professor with the University of Pretoria, South Africa, and recognized internationally as a pioneer and leading scholar in industrial wireless sensor networks research. He initiated and co-edited the first Special Section on Industrial Wireless Sensor Networks in the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS in 2009 and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS in 2013. He has been serving as an Associate Editor and Guest Editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE ACCESS, and previously the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS. Currently, he is a Co-Editor-in-Chief for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.