Heterogeneous Devices in PV Rich Distribution Systems: Gently Cooperated Voltage Control

K.Anitha[1], G.Punnam Chander[2], E.Shravan[3], Sameena Jilla[4], R.Premalatha[5]

[1, 2, 3, 4] *Assistant Professor, Department of Electrical and Electronics Engineering, Vaagdevi College of Engineering, Warangal, Telangana-506005, India.*

[5] *Associate Professor, Department of Electrical and Electronics Engineering, Vaagdevi College of Engineering, Warangal, Telangana-506005, India.*

Abstract

The voltage control interactions between recently installed PV inverters and previously deployed on-load tap-changer (OLTC) transformers become increasingly important as photovoltaic (PV) adoption increases steadily. Current approaches frequently depend on a decision-making algorithm to fully assume control of both inverters and OLTC in order to achieve coordinated voltage regulation. Consequently, in order to be under the new algorithm's control, OLTC must relinquish its independent tap switching logic and carry out the necessary changes. In this research, a soft coordination framework is suggested with the goalof bridging this gap. In particular, OLTC is permitted to retain its independent control state during tap switching, and the decision-making algorithm will only directly regulate the Var output of inverters. The voltage control problem should first be represented as a memory-based Markov decision process (MDP) in order to realise such soft coordination. Based on this foundation, the current soft actor-critical algorithm (RSAC) is proposed for Index Terms—Coordinated voltage control, distribution systems, OLTC transformer, photovoltaic (PV).

I. Introduction

The increasing integration of photovoltaic (PV) technology into distribution systems presents both opportunities and challenges for reducing carbon emissions. Issues with voltage violations are now the primary barrier to increased PV power integration in distribution networks.

On-load tap-changer (OLTC) transformers installed upstream are usually used as tap switches in conventional distribution systems to regulate system voltage. Line drop compensation (LDC) is one of the most often used OLTC tap control logics. To be more precise, OLT uses an analogue circuit to simulate the distribution line's voltage drop.

Consequently, if the predicted voltage is outside of the permitted range for a longer period of time than the time delay, the OLTC can sense remote voltage variations based on local measurements and adapts its tap position accordingly [1, 2]. There are reports of additional OLTC tap control logics in [3, 4], both with and without remote monitoring. Originally intended to adjust for voltage changes brought on by slow load variations, these commonly used OLTC transformers may still be effective if PV penetration is low[4,5].However, because of significant reverse power flow [6] and unequal PV power distribution [7], OLTC transformers alone would not be able to effectively handle the overvoltage issue as PV penetration increases over time. Fast swings in PV power can also lead to excessive tap operations of OLTC transformers [8], which speeds up the ageing process of the devices.

Voltage control is made more flexible by inverters' ability to seamlessly modify their Var output in real time, as opposed to OLTC transformers' usually poor response times and discontinuous tap operations. Coordination of voltage control techniques for distributed PV inverters has been attempted to be designed. To optimize PV inverters' Var output for real-time voltage control, for instance, a distributed method resilient to communication asynchrony was presented in [9]. For unbalanced distribution systems, an inter-phase coordinated voltage control technique was developed after the Volt-Var interaction between phases was examined in [10]. Inverter clusters' coordinated voltage regulation is also covered by, but not limited to, [11–13].

In most situations, PV inverters are installed in distribution systems whose voltage is currently under the regulation of previously deployed OLTC transformers. Existing research for coordinated voltage control with heterogeneous devices can be mainly divided into two categories:

1. Coordinating according to rules. For heterogeneous devices, a variety of coordination principles have been developed to accomplish coordinated voltage control. For instance, in [14], the permissible voltage range was appropriately split into a number of zones, with OLTC or inverters taking appropriate corrective action based on each zone. PV inverters and a battery energy storage system (BESS) were intended to participate in voltage adjustment temporarily in [15].The moment an OLTC tap operation was initiated, they decreased their involvement in voltage management. This meant that there was no overuse of the BESS or PV inverters. In[16], a coordination plan for BESS and OLTC control was put forth. To initiate the OLTC tap, the weighted average of the estimated voltage across all buses was used as the control signal

2. cooperation based on algorithms. In order to coordinate heterogeneous devices in voltage control, techniques based on both reinforcement learning (RL) and optimization have been developed. on [19], for instance, a distribution system voltage control problem was framed within an RL framework. In this scenario, deep Q-net (DQN) collaboratively dispatched inverters, capacitor banks (CBs), and OLTC on the same time scale. In [20], a multi-timescale co-optimization model was developed as a mixed-integer second-order cone programme, taking into account the varying response speeds of various devices. Network reconfiguration, OLTC, and inverters (battery and PV) were scheduled on a daily, hourly, and 20-minute basis, respectively. Similar to this, in [21], a two-layer cost-effective control technique was used to coordinate OLTC, CBs, PVs, and mobile energy storage systems (MESS) for both cost minimization and voltage management. Refer to [22, 23] for additional information Even while rule-based coordination solutions are useful in regulating voltage, they frequently require empirical design. It is a more versatile method of coordinating heterogeneous devices through RL-based algorithms or optimization. The majority of RL-based or optimisation techniques need direct device control (also known as hard coordination techniques). Therefore, in order to be controlled by a new algorithm, previously installed OLTC transformers that have already been operating efficiently for many years must fully reverse their current operation rules and carry out the necessary upgrades. It is important to note that these OLTC modifications would increase prices and complicate field implementation. Moreover, the local power provider owns the OLTC transformer. To bridge this gap, a learning-based soft coordination control method that fully respects existing operation rules of previously deployed devices is proposed in this paper, with contributions summarized as follows:

1) Soft coordination framework: An innovative control framework that aims to “softly” collaborate inverters with OLTC transformers for system voltage regulation is proposed in Section II.

Different from most hard coordination methods that rely on the direct control of all devices, in our soft coordination framework, only inverters’ Var output is controlled by a well-trained proxy model,and OLTC is allowed to maintain its autonomous control state for tap switching. The rationale behind the soft coordination lies on the fact that OLTC’ existing autonomous control logic is a voltage-dependent control rule. As a result, through coupled system voltage, it is possible to indirectly control OLTC’ tap position as long as the system voltage profile can be elaborately shaped by inverters’ Var compensation.

2) Memory-based Markov decision process: To realizethe soft coordination, the voltage control problem should first be modelled as a Markov decision process (MDP) before the design of the corresponding RL algorithm. In this paper, the OLTC follows its existing “LDC + time delay” control logic for tap switching, namely OLTC adjusts its tap position according to its memory of voltage variations over a past period of time.

As a result, the standard MDP where state transitions only depend on the action and current system states cannot fully capture the characteristics of OLTC’s tap behavior. To cater for this time series-coupled tap switching mechanism of OLTC, the standard MDP is extended to the memory-based MDP in Section III. Correspondingly, the proxy model makes decisions according to the historical trajectory (time series-coupled information) it has observed instead of only current system states.

3) Recurrent soft actor-critic method: The RSAC algorithm, which is described in Section IV, can be used to train the proxy model with episodes that are found in memory-based MDPs. In contrast to conventional actor-critic based algorithms, our proposed RSAC algorithm's actor network (proxy model) is a deep neural network (DNN) equipped with a Gated Recurrent Unit (GRU). This network is specifically designed to process time series-coupled data efficiently and make decisions based on historical trajectory. Consequently, the proxy model can effectively learn the time series-based tap switching features of OLTC using our proposed RSAC algorithm, and a well-trained proxy model is capable of making informed judgements in memory-based MDPs to accomplish the soft coordination..

II. Soft Coordination Framework

A. *Ta p Control Logic of OLTC Transformers*

OLTC transformers play a dominating role in voltage regulation in most distribution systems. Following its own autonomous control rule, an OLTC transformer can adaptively adjust its tapposition to compensate system voltage variations. Fig. 1 demonstrates one of the most popular tap control logics applied on OLTC transformers in the industry.

As shown in Fig. 1 a), an internal model called LDC circuit is used to match the distribution line impedance. Distribution system operators (DSOs) can set R and$X$ values in the LDC circuit through the load-center method or voltage-spread method [1] to adjust the compensation. With the LDC circuit, a downstream voltage level$Vest$can be estimated according to detected voltage $V_0$and current I_0on the secondary side of OLTC as

$Vest$=V0−I0(R+jX), (1)

We will compare this predicted VCO with the voltage control target VCO and the dead band Vd. When $Vest$ leaves its permitted range [Vtg −ΔVdb,Vtg +ΔVdb], the OLTC transformer's timer begins to count. Temporary voltage breaches, as seen in Fig. 1b), will not cause tap switches to activate since the OLTC timer will be terminated once V is back within its permitted range. Simultaneously, the timer reading TDtexceeds the setting of time delay Td. In the event that $Vest$is larger (lower) than Vtg +ΔVdb ($Vtg - \Delta Vtb$), OLTC will step down (up) its tap positionZt. OLTC transformers frequently employ this "LDC + time delay" control algorithm for adaptive tap switching. In this case, as shown in Fig. 2b), OLTC tap operations will occasionally be initiated, which modifies the inverters' operating points and lessens the load of Var compensation on them. The OLTC transformer, which was initially intended to compensate for slow changes in load, must directly face the fast-fluctuating PV power if inverters do not participate in voltage regulation at all. This will result in excessive OLTC tap operations, as shown in Fig. 2c), and make a transformer more susceptible to damage.

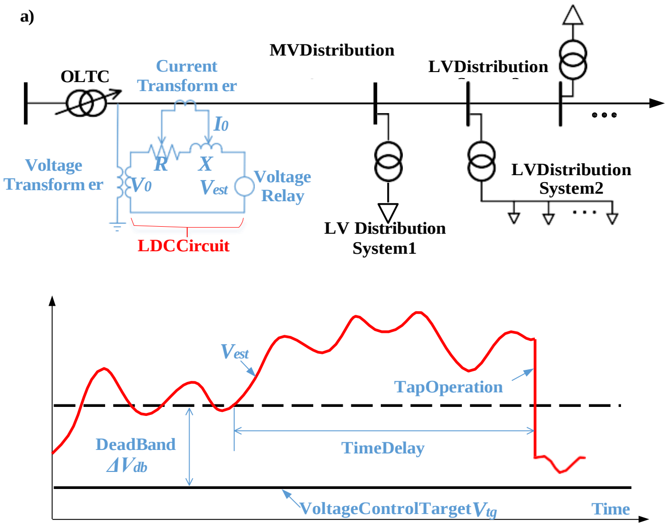


Fig. 1. A typical distribution system with an OLTC transformer for voltage regulation. a) LDC circuit, b) Tap operation

B. *Interaction of Heterogeneous Devices in Voltage Control*

Interactions between inverters and OLTC in system voltage control are inevitable because of linked system voltage, which allows the upstream OLTC transformer to sense the voltage correction from the inverters' Var output. A distributed inverter may instantly modify its Var output, which allows it to accurately shape the voltage profile of the system. As a result, an OLTC that adheres to its "LDC+ time delay" rule is indirectly controlled by the inverters. For instance, inverters will increase their voltage correction to stop the OLTC timer before it reaches the time delay Td in order to prevent a tap switch; On the other hand, inverters must reduce their voltage correction and permit a specific degree of voltage violation hazards to persist longer than the OLTC time delay Td in order to activate a tap switch.

Fig.2 demonstrates distinct changes of OLTC behaviors

using various Var output techniques for inverters. No OLTC tap switch will be activated if inverters are built to maintain a constant point of common coupling (PCC) voltage by Var compensation, as shown in Fig. 2 a). Consequently, only the inverters' Var compensation will be able to control PV power-induced overvoltage for an extended period of time, leaving the inverters susceptible to Var saturation. Significant reactive power flow will also result in further system line loss. Inverters that loosen their control over the system voltage—for instance, by applying Volt-Var droop curves—can cause changes in PV power, which will cause the PCC voltage to fluctuate.

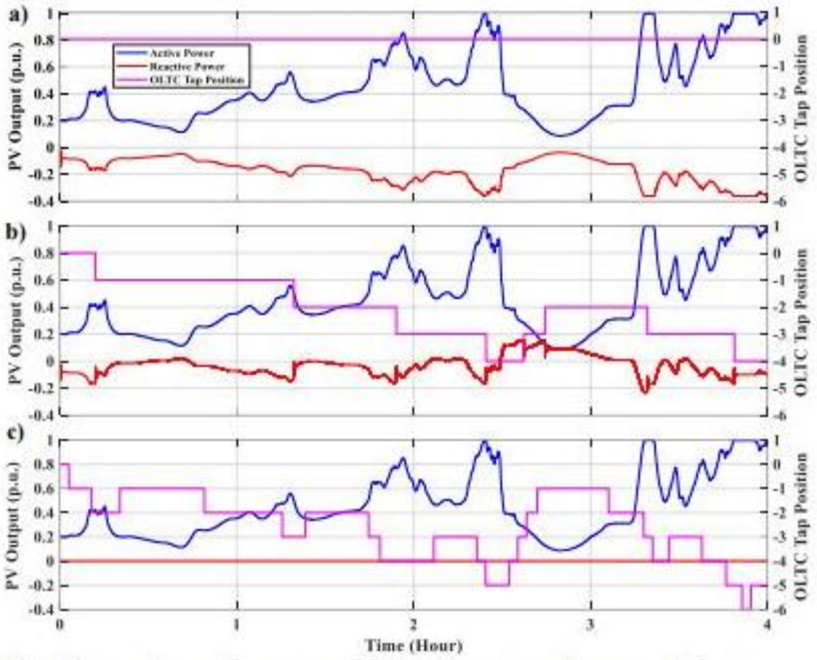


Fig.2. Interactions of inverters and OLTC in system voltage regulation.

C. *Hard Coordination Versus Soft Coordination*

As discussed in Section II-B, OLTC under its autonomous operation state has potential to be indirectly controlled by inverters' Var output. Inspired by this idea, an innovative soft coordination framework is proposed in this paper.The traditional "hard coordination" methods are in opposition to the concept of "soft coordination" as presented in our manuscript. The coordination algorithm must seize control of every device involved in hard coordination. Specifically, an optimization or learning-based algorithm determines the outputs of every device in real time. On the other hand, with soft coordination, a subset of devices is directly controlled by the coordination algorithm, but the remainder devices are permitted to continue operating independently. Particularly, in this work, the trained proxy model only instructs distributed inverters to directly regulate their Var output, whilst the OLTC modifies its tap position in accordance with its own "LDC + time delay" control logic (self-governing status). Notably, even though the coordination approach does not directly regulate the tap position of an OLTC as it does in hard coordination, coordinated voltage regulation, and Realization of Soft Coordination in Voltage Control

D. *Realization of Soft Coordination in Voltage Control*

To realize such a soft coordination mechanism, the voltage control problem is modeled as a memory-based MDP, where the OLTC-equipped distribution system is regarded as the environment in the context of RL, and the proxy model

controlled inverters will output reactive power (action At) to interact with the environment, as shown in Fig. 3. The proxy model is a GRU-equipped DNN, and its inputs in time step t are current system states $St$and previous action $At-1$. According to the historical trajectory Kt,namely a series of(St,$At-1$)input in different time steps

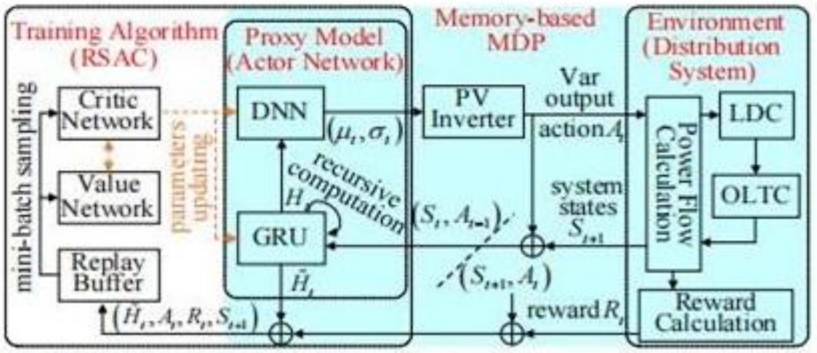


Fig. 3. The schematic figure of our proposed method.

1. Actions

Within our designed soft coordination framework, the Var outputs of inverters are decision variables(actions). Therefore, at time instant$t$, the action$At\in \mathcal{A}$ is the set of the Var output of PV invertersQpv,namely$A$=Qpv,and it supper and
Lower limits are$Qpv$and$-Qpv$,respectively.
Proxy model are action probability distribution$\pi\phi(\cdot|Kt)$,

2.Mark ov States

According to the OLTC operation mechanism introduced in Section II-A, the voltage of all buses Vt, tap position Zt, and the timer reading TDt(which indicates how long the over/under voltage risk has lasted) are necessary Markov states in depicting the OLTC behavior in its voltage control process. In addition, the system voltage is also influenced by active and reactive power flows. Therefore, the complete Markov states
$St\in S$ of the distribution system voltage control are given as:
S=($Pload$,$Qload$,Ppv,V,Z,TD), (2)

3 Rewards

The soft format we designed has two control objectives: a)Eliminate voltage interruptions, b) Reduce line losses. That's all
Note that only excessive OLTC operations should be performed[25] Decay and there is no need to reduce their numberOpen the OLTC faucet. Line loss reduction can be improved Operational efficiency of the distribution system, especially mine
PV-rich systems require additional reactive power Suppression of mutations. Usually additional reaction effects. It causes more losses in the system lines. Like most learning-based algorithms, the penalty framework Added bonus feature to calculate possible systems Power cut Especially in the event of a power cut Among all buses, the reward function is modeled as (3) for quantization System voltage deviation. Otherwise, the reward function is As shown in Equation (4), try to minimize the system line loss. According to this Reward Design The agent model will receive negative rewards if Its action will cause any bus to lose power. more than On the other hand, if the system can get positive rewards The voltage is within the allowed range and the bus is lost This is less than if no action had been taken

$$Rt = M \sum j \in \mathcal{N}[max(Vt\ j - V, 0) + max(V - Vt\ j\ , 0)]\ (3)$$
$$Rt = \lambda(Pt\ loss{,}0 - Pt\ loss),\ (4)$$

MEMORY-BASED MD SYSTEM VOLTAGE CONTROL

Tuple of Memory-based MDP in System Voltage Control

The voltage control problem should be first depicted as MDPs before the further design of the RL algorithm. For examples in[19,22],based on the standard MDP,DQN Algorithms are established for inverters' Volt-Var control[12], the adversarial MDP was proposed for conducting adversarial learning, through which proxy model-controlled inverters are able to provide robust Volt-Var control against the model mismatch. Considering the time series-coupled tap switching mechanism of OLTC, a memory-based MDP with the tuple ($\mathcal{A}$,$\mathcal{S}$,$\mathscr{R}$,$\mathcal{P}$) is used to mathematically model the system State transitions in the coordinated voltage regulation

$$Rt = M \sum j \in \mathcal{N}[max(Vt\ j - V, 0) + max(V - Vt\ j\ , 0)]\ (3)$$
$$Rt = \lambda(Pt\ loss{,}0 - Pt\ loss),\ (4)$$

where $Rt \in \mathscr{R}$ is the reward; $M < 0$ represents the penalty coefficient of voltage violations;$V$and$V$ are upper and lower limits of system voltage, respectively;Vjdenotes the voltage of bus j at time instant t; $\mathcal{N}$ is the set of all buses; $\lambda > 0$ represents the incentive factor; $Ploss$represents the line loss if actionAt is taken at time instant t,and$Ploss$,0corresponds to problem, and $\mathcal{A}$, $\mathcal{S}$,$\mathscr{R}$, and$\mathcal{P}$ represent the sets of all possible actions, Markov states, rewards, and the probability distribution of state transition, respectively

4) Probability Distribution of State Transition

The system state will transition from St toSt+1with a reward Rt after an action At is taken at time instant t. Such a state transition obeys a probability distribution $\rho \in \mathcal{P}$, denoted as below

$(St{+}1\ , Rt\ ){\sim}\rho(\cdot\ |St\ , At\ )$, (5)

which models the impact of stochastic disturbances (e.g., load variations) on system state transitions.

B. Historical Trajectory and Decision Making

Current system states are not sufficient for making appropriate decisions, since OLTC adjusts its tap position depending on its memory of voltage variations over a past period of time, as demonstrated in Section II-A. To address this issue, the historical trajectory $Kt \in \mathcal{K}$, defined as in (6), is used to support proxy model for its decision making, in order to successfully grasp the characteristics of OLTC's tap operation in a time series.

$$Kt = (S0\ , A0\ , S1\ , A1\ , \ldots\ , St{-}1\ , At{-}1\ , St),\ (6)$$

where the Kt is a series of states and actions obtained from inverter-environment interactions up to now, and $Kt$is comprised of $Kt-1$and $(At-1,St)$. It is worth noting, $(At-1,St)$ is the input of the proxy model at time instant t, while the proxy model makes decisions according to the whole historical trajectoryKt. The proxy model is a neural network in nature, which establish es the mapping from the historical trajectory$Kt$to the action probability distribution denoted as $\pi\phi(\cdot|Kt)$. Finally, the action At is obtained through random sampling from the probability distribution, as in (7):

$$At{\sim}\pi\phi(\cdot|Kt),\ \ (7)$$

In the context of RL, $\pi\phi$is called as the action policy of the proxy model, and it will be further introduced in Section IV-A. The action distribution$\pi\phi(\cdot|Kt)$ will converge to its optimal value with very small variances after sufficient training.

C. MDP Considering Entropy

To sufficiently explore the action space and avoid premature convergence, the maximum entropy learning technique is adopted in this paper. Here in, the entropy of the Probability distribution$\pi\phi(\cdot|Kt)$is denoted as $\mathscr{H}(\pi\phi(\cdot|Kt))$, And larger $\mathscr{H}(\pi\phi(\cdot|Kt))$ means a more random selection of action from its distribution $\pi\phi(\cdot\ |Kt)$.

Different from standard deep reinforcement learning (DRL) that aims to maximize the expectation of accumulated rewards

$\mathbb{E}\tau\sim\pi\phi\ \sum N\ n=0\ \gamma\ nRt+n$, in maximum entropy learning, an entropy term [26] is added in the state value function as:

$$V\ \pi\phi(kt) = \tau\sim\ \mathbb{E}\ \pi\phi\ \{\sum N\ n=0\ \gamma\ n\ [Rt+n + \alpha\mathscr{H}(\pi\phi(\cdot\ |Kt+n\ ))]|Kt=kt\ \}\ .\ (8)$$

Herein, $V\ \pi\phi(kt)$ as it is shown in (8) represents the expected value of the accumulated rewards and entropy in the future trajectory if the historical trajectory $Kt = kt$ and policy $\pi\phi$ is applied for future action making; τ represents future trajectories following the action policy $\pi\phi$; N denotes the length of trajectories; $\gamma \in (0,1)$ is the discount factor; $\alpha > 0$ is a temperature parameter used to balance the Exploration and Exploitation during the learning process.

Correspondingly, the state-action value function $Q\ \pi\phi$ is expressed in (9):

$Q\ \pi\phi(kt\ ,\ at) = \tau\sim\ \mathbb{E}\ \pi\phi\ \{\sum N\ n=0\ \gamma\ nRt+n\ +$

$\alpha\sum N\ n=1\ \gamma\ n\mathscr{H}(\pi\phi(\cdot\ |kt+n\ ))\ |Kt=kt\ ,At=at\ \}.\ (9)$

RECURRENTS OF TACTOR- CRITICAL ALGORITHM

Structure of the RSAC Algorithm

To maximize the total rewards that it can obtain, our suggested RSAC approach must be used to train the proxy model enough. The RSAC algorithm is based on an enhanced version of the actor-critic algorithm [26], which can handle time series-coupled data better. Traditional actor-critic based algorithms have been applied in distribution system voltage regulation [12, 27, 28]. The RSAC algorithm consists of an actor network and a critic network, as the name implies. Actions are made by the actor network, and then they are quantitatively assessed using Q values by the critic network. Importantly, the actor network in the RSAC method is the proxy model, which determines the Var output of inverters. For this reason, it will be referred to as Actor Network

As shown in Fig. 3, the actor network (proxy model) is established as a GRU-equipped DNN, which is designed to make proper actions (Var compensation) for inverters according to a series of $(St,At-1)$ it has observed (namely, the historical trajectory Kt). Herein, GRU is a type of recurrent neural networks (RNNs) with a gating mechanism [29] Such

refinement of RNN includes an update gate and a reset gate, which determine what information is allowed through to the output, and it can be trained to retain information over time. With this simulated ability to remember information, GRU is used in this paper to process the time-series coupled variablelength historical trajectory Kt as:

$$Ht=g\phi(St,At-1,Ht-1)=g\phi(\tilde{H}t),\ (10)$$

Where$g\phi r$ denotes the GRU mapping rule, and ϕrrepresents its network parameter.Htrepresents the GRU hidden state at time instant$t$, which will be updated at each time step; $(St,At-1,Ht-1)$is denoted by$\tilde{H}t$for ease of description. Through recursive computations as presented in (10), the variable-length historical trajectory Ktas in (6) is projected into a fixed-dimension$\tilde{H}t$.The updated hidden state$Ht$is then transferred to a DNN a sits input(referringtoFig.3),and its output is the probability distribution of actions. In this paper, possible actions $At \in \mathbb{R}1\times NA$are designed to obey normal distributions, and correspondingly the outputs of DNN are $\mu t \in \mathbb{R}1\times NAand\sigma t \in \mathbb{R}1\times NA$representing the sets of expectations and variances of these distributions, respectively:

$$(\mu t, \sigma t) = f\phi d(Ht), \quad (11)$$

WhereNA is the dimension of actionAt;ϕd means the DNN parameter; $f\phi d$ denotes the DNN mapping rule. So far, the GRU-DNN based actor network has been parameterized as:

$$(\mu t\ , \sigma t\ ) = f\phi d\ [g\phi r\ (Ht)] = \pi\phi(Ht), (12)$$

where $f\phi d\ [g\phi r\ (\cdot)]$ is denoted by $\pi\phi$ with $\phi = [\phi r\ ,\phi d]$, and therefore $\pi\phi$ represents actor network's action policy characterized by ϕ. According to normal distributions given by (12), action At can finally be obtained through random sampling as:

$$At = \mathbb{P}(\mu t + \varepsilon t \odot \sigma t\ ), (13)$$

where εt is a random variable obeying the standard normal distribution $\mathcal{N}(0,1)$, and $\odot$ is the dot product operator. Since
inverters' Var outputs (i.e., action At) have physical boundaries, $\mathbb{P}$ represents the projection from infinity to the interval $[-Qmax\ pv,Qmax\ pv\ ]$.
2) Critic Network The actor network (with policy $\pi\phi$) takes an action At according to known Ht at time instant t, and the value of this
action, namely the future rewards and entropy that the actor network will totally obtain after action At is taken is expressed as the state-action value function $Q\ \pi\phi$ in (9). In this paper, this state-action value function $Q\ \pi\phi$ is approximated by a neural network $Q\theta\ \pi\phi$ with parameter θ as

$$Q\ \pi\phi \approx Q\theta\ \pi\phi\ (Ht\ , At), (14)$$

where $Q\theta\ \pi\phi$ is called critic network in the RSAC algorithm. A well-trained critic network $Q\theta\ \pi\phi$ can properly output Q values to evaluate actions, and a larger Q value means a better action.
3) Value Network
Similarly, the state value function $V\ \pi\phi$ in (8) can be approximated by the value network $V\psi\ \pi\phi$ with parameter ψ as:

$$V\ \pi\phi \approx V\psi\ \pi\phi\ (Ht). (15)$$

4) Replay Buffer

The latest Nb sets of actor network's experiences (called episodes), which are a series of tuples ($Ht\ , At\ , Rt\ , St+1$) obtained from inverter-environment interactions, are stored in the replay buffer $\mathcal{D}$. A mini-batch of episodes will be randomly sampled from the replay buffer and used for each
round of parameter updating.
B. Network Parameter Updating

In our proposed RSAC algorithm, parameters of the actor network (ϕ) and the critic network (θ) are alternately and iteratively updated by strategies of Policy Evaluation and Policy

Improvement, respectively, while the value network (ψ) works as an auxiliary in parameter updating. Through Policy Evaluation strategy, the critic network $Q\theta\ \pi\phi$ with updated θ could have better performance inaction value estimation for a given policy $\pi\phi$. The improved critic network $Q\theta\ \pi\phi$ will further be used to update the current actor network ($\phi\ '$) through Policy Improvement strategy, and the updated actor network (ϕ) will have better performance in action selection. Namely, actions made by updated policy $\pi\phi$ tend to have larger Q values compared with that of the original policy $\pi\phi'$.

1) Critic Network Updating Through Policy Evaluation Following the definition of state value function $V\ \pi\phi$ and state-action value function $Q\ \pi\phi$ as in (8) and (9) respectively, the following Bellman equations can be derived:

$$\begin{aligned}
&Q_\theta^{\pi_\phi}(\widetilde{H}_t, A_t)\\
&= \mathop{\mathbb{E}}_{\tau\sim\pi_\phi}\left\{\textstyle\sum_{n=0}^{N}\gamma^n R_{t+n} + \alpha\sum_{n=1}^{N}\gamma^n \mathcal{H}\left(\pi_\phi(\cdot\,|\widetilde{H}_{t+n})\right)\right\}\\
&= \mathop{\mathbb{E}}_{\tau\sim\pi_\phi}\left\{R_t + \gamma\textstyle\sum_{n=1}^{N}\gamma^{n-1}[R_{t+n} + \alpha\mathcal{H}\left(\pi_\phi(\cdot\,|\widetilde{H}_{t+n})\right)]\right\}\\
&= \mathop{\mathbb{E}}_{(S_{t+1},R_t)\sim\rho}\left\{R_t + \gamma V_\psi^{\pi_\phi}(\widetilde{H}_{t+1})\right\} \qquad (16)
\end{aligned}$$

$$\begin{aligned}
&V_\psi^{\pi_\phi}(\widetilde{H}_t) = \mathop{\mathbb{E}}_{\tau\sim\pi_\phi}\left\{\textstyle\sum_{n=0}^{N}\gamma^n\left[R_{t+n} + \alpha\mathcal{H}\left(\pi_\phi(\cdot\,|\widetilde{H}_{t+n})\right)\right]\right\}\\
&= \mathop{\mathbb{E}}_{\tau\sim\pi_\phi}\left\{\textstyle\sum_{n=0}^{N}\gamma^n R_{t+n} + \alpha\sum_{n=1}^{N}\gamma^n \mathcal{H}\left(\pi_\phi(\cdot\,|\widetilde{H}_{t+n})\right) +\right.\\
&\left.\alpha\mathcal{H}\left(\pi_\phi(\cdot\,|\widetilde{H}_t)\right)\right\}\\
&= \mathop{\mathbb{E}}_{A_t\sim\pi_\phi}\left\{Q_\theta^{\pi_\phi}(\widetilde{H}_t, A_t) - \alpha\log P_{\pi_\phi}(A_t|\widetilde{H}_t)\right\}, \qquad (17)
\end{aligned}$$

where $Q\theta\ \pi\phi\ (Ht\ ,\ At)$ is expressed as a function of $V\psi\ \pi\phi\ (Ht+1)$ in (16), and $V\psi\ \pi\phi\ (Ht)$ can also be transferred to a function of $Q\theta\ \pi$
$\phi\ (Ht\ ,At)$ as in (17). It is worth noting, the variable-length historical trajectory Kt as the input of $V\ \pi\phi$ and $Q\ \pi\phi$ in (8) and (9) has beenprojected into a fixed-dimension Ht in the state value network $V\psi\ \pi\phi$ and critic network $Q\theta\ \pi\phi$. In addition, the entropy
$\mathscr{H}(\pi\phi(\cdot\ |Ht))$ in this paper is defined as:
$\mathscr{H}(\pi\phi(\cdot\ |Ht)) = -\ At\sim\ \mathbb{E}\ \pi\phi\ \{log\ P\pi\phi\ (At\ |Ht)\}$, (18)
where $P\pi\phi\ (At\ |Ht)$ represents the probability density of action At in the probability distribution $\pi\phi(\cdot\ |Ht)$. The Bellman Equations (16) and (17) finally yield two loss functions as follows:

$$J_Q(\theta) = \mathop{\mathbb{E}}_{(\widetilde{H}_t,A_t,R_t,S_{t+1})\sim\mathcal{D}}\left\{\left[Q_\theta^{\pi_\phi}(\widetilde{H}_t, A_t) - R_t - \gamma V_{\bar{\psi}}^{\pi_\phi}(\widetilde{H}_{t+1})\right]^2\right\} \qquad (19)$$

$$\begin{aligned}
&J_V(\psi) = \mathop{\mathbb{E}}_{\widetilde{H}_t\sim\mathcal{D},A_t\sim\pi_\phi}\left\{\left[V_\psi^{\pi_\phi}(\widetilde{H}_t) - Q_\theta^{\pi_\phi}(\widetilde{H}_t, A_t) +\right.\right.\\
&\left.\left.\alpha\log P_r(A_t|\widetilde{H}_t)\right]^2\right\}, \qquad (20)
\end{aligned}$$

where $(\tilde{H}t\ ,At\ ,\ Rt\ ,\ St+1)\sim\mathcal{D}$ means that the tuple $(Ht\ ,At\ ,\ Rt\ ,\ St+1)$ is randomly sampled from the replay buffer $\mathcal{D}$, and so does $Ht\sim\mathcal{D}$. It is worth noting that a target state value network $V\psi\ \pi\phi$ rather than the state value network $V\psi\ \pi\phi$ is used in (19) to improve the stability of the algorithm.

$\theta \leftarrow \theta + \eta \nabla \theta J Q(\theta)$ (21)
$\psi \leftarrow \psi + \eta \nabla \psi J V(\psi)$. (22)

$\psi \leftarrow \beta\psi + (1-\beta)$.
(23)

Actor Network Updating Through Policy Improvement After sufficient training, the critic network $Q\theta\ \pi\phi$ can
properly output Q values for action estimation, where a larger Q value indicates a better action. To improve the performance
of the actor network, one direct idea is to build such a probability distribution, that actions with larger Q values have
larger probability densities. Therefore, based on the current policy $\pi\phi'$ and its critic network $Q\theta\ \pi\phi'$, the target probability
distribution $\pi\phi(\cdot\ |\tilde{H}\ t)$ of the actor network is aimed to be shaped as:

$$\pi_{\phi}(\cdot\ |\widetilde{H}_t) \rightarrow \frac{exp\left\{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,\cdot)/\alpha\right\}}{\int exp\left\{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,A_t)/\alpha\right\} dA_t}, \qquad (24)$$

The Kullback-Leibler (KL) divergence is used to measure the deviation between two distributions. Correspondingly, the
training objective is designed to minimize the KL divergence between $\pi\phi(\cdot\ |Ht)$ and its target distribution as:

$$\phi| \leftarrow \arg\min_{\phi}\ D_{KL}\left\{\pi_{\phi}(\cdot\ |\widetilde{H}_t)\middle\|\frac{exp\left\{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,\cdot)/\alpha\right\}}{\int exp\left\{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,A_t)/\alpha\right\} dA_t}\right\}, \quad (25)$$

where DKL represents the KL divergence between two distributions. Since the value of $\int exp$ $\{Q\theta\ \pi\phi'\ (Ht\ ,\ At)/\alpha\}\ dAt$ in (25) Is independent to the selection of action At , it can be denoted as $F\theta\ \pi\phi\ (Ht)$ for ease of description. According to (25), the loss function is given as :

$$\begin{aligned} J_\pi(\phi) &= D_{KL}\left\{\pi_{\phi}(\cdot\ |\widetilde{H}_t)\middle\|\frac{exp\left\{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,\cdot)/\alpha\right\}}{F_\theta^{\pi_{\phi'}}(\widetilde{H}_t)}\right\} \\ &= \int P_{\pi_\phi}(A_t|\widetilde{H}_t) \log\frac{P_{\pi_\phi}(A_t|\widetilde{H}_t)}{exp\left\{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,A_t)/\alpha\right\}/F_\theta^{\pi_{\phi'}}(\widetilde{H}_t)} dA_t \\ &= \mathop{\mathbb{E}}_{\widetilde{H}_t\sim\mathcal{D},A_t\sim\pi_\phi}\left\{\log P_{\pi_\phi}(A_t|\widetilde{H}_t) + \log F_\theta^{\pi_{\phi'}}(\widetilde{H}_t) - \frac{Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t,A_t)}{\alpha}\right\}. \end{aligned} \qquad (26)$$

By substituting (13) into (26), we can equivalently express the loss function as :

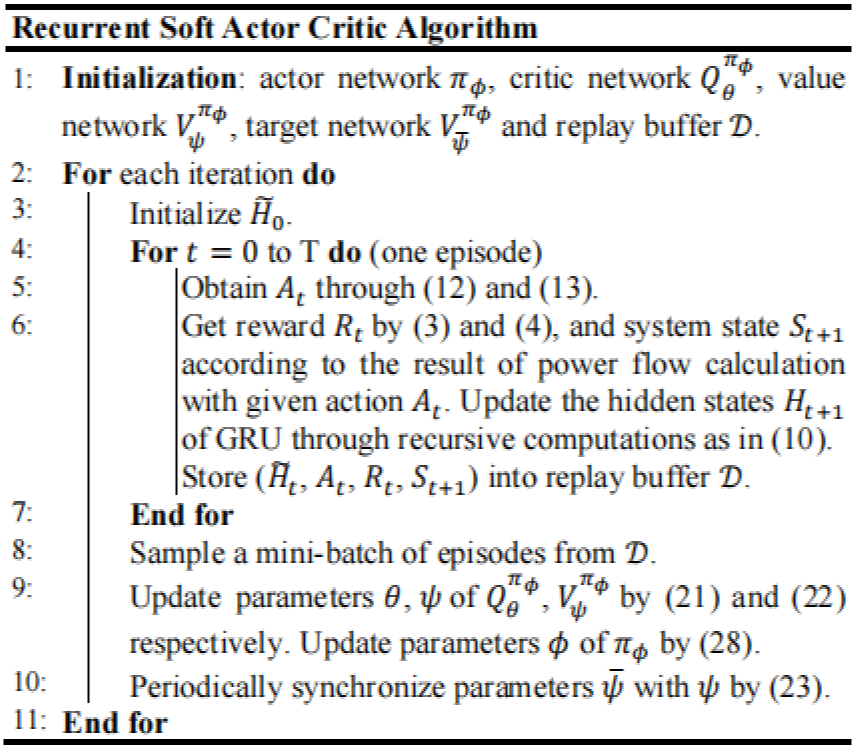
$$\begin{aligned} J_\pi(\phi) = \mathop{\mathbb{E}}_{\widetilde{H}_t\sim\mathcal{D},\varepsilon\sim\mathcal{N}(0,1)} &\Big\{\log P_{\pi_\phi}(\mathbb{P}(\mu_t + \varepsilon_t \odot \sigma_t)|\widetilde{H}_t) + \\ &\log F_\theta^{\pi_{\phi'}}(\widetilde{H}_t) - Q_\theta^{\pi_{\phi'}}(\widetilde{H}_t, \mathbb{P}(\mu_t + \varepsilon_t \odot \sigma_t))/\alpha\Big\}. \end{aligned} \qquad (27)$$

Finally, the parameter of the action network can be updated through the mini-batch stochastic gradient descent
Method :

$$\phi \leftarrow \phi + \eta \nabla_{\phi} J_{\pi}(\phi). \qquad (28)$$

C. Training Procedure

The procedure of network training (i.e., actor network $\pi\phi$, critic network $Q\theta\ \pi\phi$, value network $V\psi\ \pi\phi$ and target network $V\psi\ \pi\phi$) is summarized as below :



```
Recurrent Soft Actor Critic Algorithm
Initialization: actor network $\pi_{\phi}$, critic network $Q_{\theta}^{\pi_{\phi}}$, value network $V_{\psi}^{\pi_{\phi}}$, target network $V_{\bar{\psi}}^{\pi_{\phi}}$ and replay buffer $\mathcal{D}$.
For each iteration do
    Initialize $\widetilde{H}_0$.
    For $t = 0$ to T do (one episode)
        Obtain $A_t$ through (12) and (13).
        Get reward $R_t$ by (3) and (4), and system state $S_{t+1}$ according to the result of power flow calculation with given action $A_t$. Update the hidden states $H_{t+1}$ of GRU through recursive computations as in (10).
        Store $(\widetilde{H}_t, A_t, R_t, S_{t+1})$ into replay buffer $\mathcal{D}$.
    End for
    Sample a mini-batch of episodes from $\mathcal{D}$.
    Update parameters $\theta, \psi$ of $Q_{\theta}^{\pi_{\phi}}, V_{\psi}^{\pi_{\phi}}$ by (21) and (22) respectively. Update parameters $\phi$ of $\pi_{\phi}$ by (28).
    Periodically synchronize parameters $\bar{\psi}$ with $\psi$ by (23).
End for
```

V. CASE STUDIES:

A. IEEE 33-Bus Balanced Distribution System

1. Test System Introduction A modified IEEE 33-bus distribution system with a 16-step OLTC transformer and 7 distributed Pv inverters, as in Fig. 4, is used for case studies. The maximum active and reactive power of PV inverters on different buses are also shown in this figure. Considering the system peak load 3.8MW, the total PV With a maximum installation capacity of 4.6 MW, the modified IEEE 33-bus system is a PV-rich distribution network. The OLTC transformer, as described in Section II-A, uses the control parameters listed in Table I to regulate system voltage in accordance with its LDC rule. Fig. 5 shows the active and reactive load profiles for the entire day. In this paper, the system voltage allowed range is specified at 0.95p.u.~1.05p.u. It is important to note that Bus 15 is the bus with PV inverters that is located the farthest away, making it the most susceptible to overvoltage issues when reverse power flow happens. In order to examine the effectiveness of various approaches for voltage regulation, only the voltage profiles of Bus 15 are shown in this.

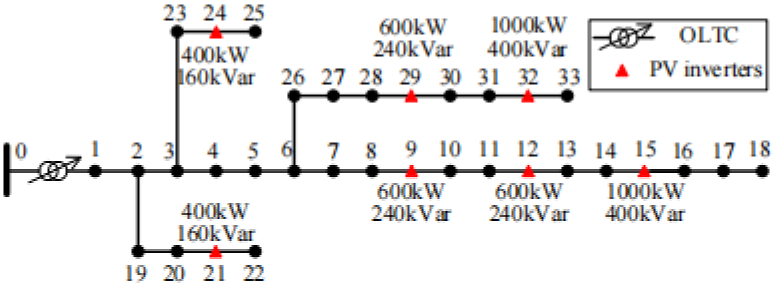


Fig. 4. Topology of the modified IEEE 33-bus system.

TABLE I CONTROL PARAMETERS OF THE OLTC TRANSFORMER

V_{tg}	R	X	ΔV_{db}	T_d	regulator range
1 (p.u.)	0.864 (p.u.)	0.538 (p.u.)	0.008 (p.u.)	180s	±5%

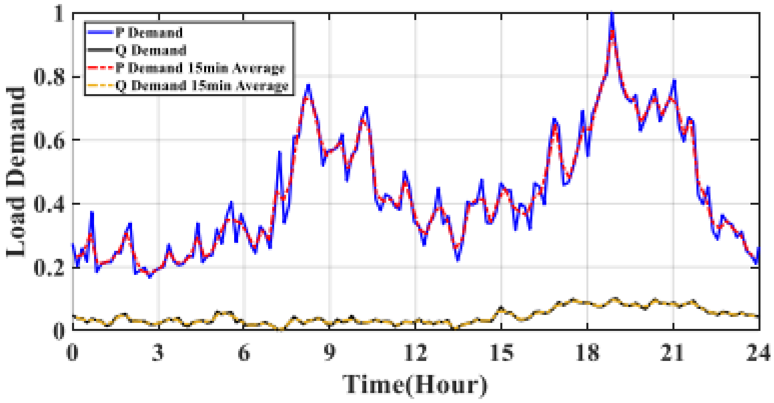


Fig. 5. Active and reactive load profiles with their predictions.

Open DSS is used as the programming environment for the 33-bus distribution system model, and Py Torch [30] is used to build the suggested RSAC algorithm in Python for network training. By using co-simulations between Open DSS and Python, training data is produced. In particular, the environment is altered by the actor network created in Python, changing the power flow outcomes of Open DSS in the process. The rewards and revised system states are then returned to Python. The RSA algorithm's hyper parameters are displayed in Table II..

1) Training Process

Fig. 6 shows the episode average reward value during the training process of successive 1500 episodes. Initially, in the early learning phase, the action policies lead to negative rewards due to limited positive learning experiences and un optimized action policies. These negative rewards illustrate that the actor network (proxy model) is incapable of maintaining the system voltage security and simultaneously reducing system line loss. However, as the training progresses,

the actor network gradually evolves and obtains positive rewards more frequently. A positive reward implies that there

is no voltage violation, and the system line loss is further reduced by taking actions. It is observed that the episode average reward keeps fluctuating, but with an upward trend. The training process converges after about 1000 episodes.

$$\theta \leftarrow \theta + \eta \nabla \theta J Q(\theta) \qquad (21)$$

$$\psi \leftarrow \psi + \eta \nabla \psi J V(\psi). \qquad (22)$$

$$\psi \leftarrow \beta\psi + (1 - \beta) \qquad . (23)$$

Actor Network Updating Through Policy Improvement After sufficient training, the critic network $Q\theta\ \pi\phi$ can properly output Q values for action estimation, where a larger Q value indicates a better action. To improve the performance of the actor network, one direct idea is to build such a probability distribution, that actions with larger Q values have larger probability densities. Therefore, based on the current policy $\pi\phi'$ and its critic network $Q\theta\ \pi\phi'$, the target probability distribution $\pi\phi(\cdot\ |\tilde{H}\ t)$ of the actor network is aimed to be shaped as:
The Kullback-Leibler (KL) divergence is used to measure the deviation between two distributions. Correspondingly, the training objective is designed to minimize the KL divergence between $\pi\phi(\cdot\ |Ht)$ and its target distribution as:
where DKL represents the KL divergence between two distributions. Since the value of $\int exp$ $\{Q\theta\ \pi\phi'\ (Ht\ ,\ At)/\alpha\}\ dAt$ in (25) Is independent to the selection of action At , it can be denoted as $F\theta\ \pi\phi\ (Ht)$ for ease of description. According to (25), the loss function is given as :
By substituting (13) into (26), we can equivalently express the loss function as :
Finally, the parameter of the action network can be updated through the mini-batch stochastic gradient descent.

Method :
C. Training Procedure
The procedure of network training (i.e., actor network $\pi\phi$, critic network $Q\theta\ \pi\phi$, value network $V\psi\ \pi\phi$ and target network $V\psi\ \pi\phi$) is summarized as below :

V. CASE STUDIES:
A. IEEE 33-Bus Balanced Distribution System
1. Test System Introduction A modified IEEE 33-bus distribution system with a 16-step OLTC transformer and 7 distributed Pv inverters, as in Fig. 4, is used for case studies. The maximum active and reactive power of PV inverters on different buses are also shown in this figure. Considering the system peak load 3.8MW, the total PV With a maximum installation capacity of 4.6 MW, the modified IEEE 33-bus system is a PV-rich distribution network. The OLTC transformer, as described in Section II-A, uses the control parameters listed in Table I to regulate system voltage in accordance with its LDC rule. Fig. 5 shows the active and reactive load profiles for the entire day. In this paper, the system voltage allowed range is specified at 0.95p.u.~1.05p.u. It is important to note that Bus 15 is the bus with PV inverters that is located the farthest away, making it the most susceptible to overvoltage issues when reverse power flow happens. In order to examine the effectiveness of various approaches for voltage regulation, only the voltage profiles of Bus 15 are shown in this.

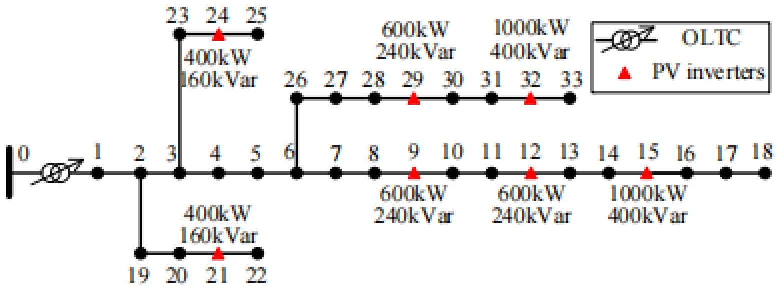


Fig. 4. Topology of the modified IEEE 33-bus system.

TABLE I CONTROL PARAMETERS OF THE OLTC TRANSFORMER

V_{tg}	R	X	ΔV_{db}	T_d	regulator range
1 (p.u.)	0.864 (p.u.)	0.538 (p.u.)	0.008 (p.u.)	180s	±5%

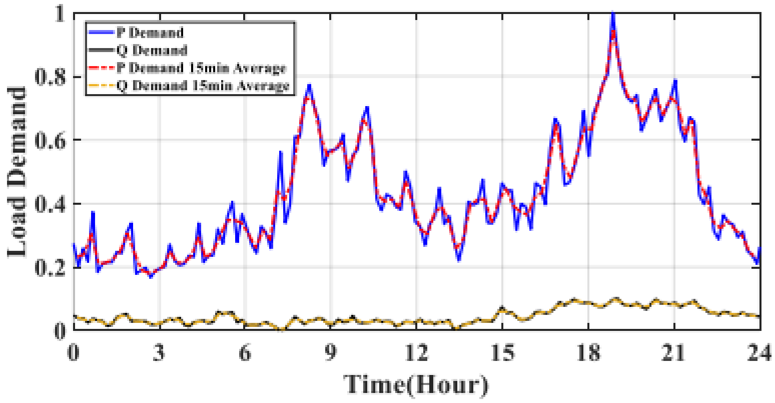


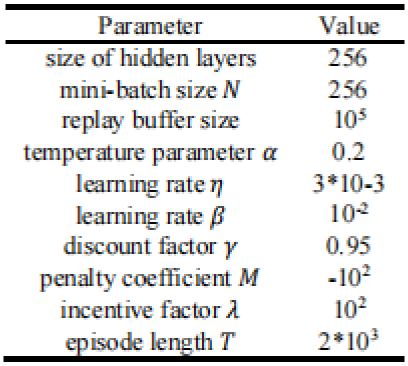
Fig. 5. Active and reactive load profiles with their predictions.

Open DSS is used as the programming environment for the 33-bus distribution system model, and Py Torch [30] is used to build the suggested RSAC algorithm in Python for network training. By using co-simulations between Open DSS and Python, training data is produced. In particular, the environment is altered by the actor network created in Python, changing the power flow outcomes of Open DSS in the process. The rewards and revised system states are then returned to Python. The RSA algorithm's hyper parameters are displayed in Table II..

1) Training Process

Fig. 6 shows the episode average reward value during the training process of successive 1500 episodes. Initially, in the early learning phase, the action policies lead to negative rewards due to limited positive learning experiences and un optimized action policies. These negative rewards illustrate that the actor network (proxy model) is incapable of maintaining the system voltage security and simultaneously reducing system line loss. However, as the training progresses, the actor network gradually evolves and obtains positive rewards more frequently. A positive reward implies that there is no voltage violation, and the system line loss is further reduced by taking actions. It is observed that the episode average reward keeps fluctuating, but with an upward trend. The training process converges after about 1000 episodes.

TABLE II HYPERPARAMETERS OF THE PROPOSED RSAC ALGORITHM



Parameter	Value
size of hidden layers	256
mini-batch size N	256
replay buffer size	10^5
temperature parameter α	0.2
learning rate η	3*10-3
learning rate β	10^{-2}
discount factor γ	0.95
penalty coefficient M	-10^2
incentive factor λ	10^2
episode length T	$2*10^3$

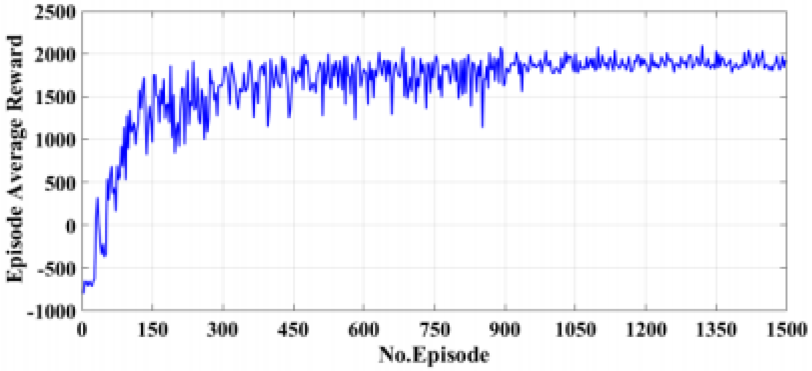


Fig.6. Training process of the RSAC algorithm.

1) Baseline Methods

1) Two typical approaches are implemented as baselines to assess the voltage regulation performance and system line loss of the proposed method. Here are the specifics:

2) 1) Baseline-1: In order to minimize system line loss and control the system voltage within the permitted range, the OLTC transformer and distributed PV inverters are coordinated using a hard coordination mechanism with a two-layer structure. In the first layer, an optimization problem based on the daily PV power forecast is solved to schedule the OLTC tap positions in advance for every 15 minutes. Based on this, a soft actor-critic (SAC) algorithm optimizes the operation points of PV inverters for real-time Volt-Var control in the second layer.

3) Baseline-2: The popular Volt-Var droop curves are employed to control local voltage in photovoltaic inverters, and the OLTC transformer that was previously in use keeps its "LDC + time delay" control logic for tap switching. Under such a method, neither coordinated voltage control nor line loss minimization is specifically designed, and both inverters and OLTC are in their autonomous control stages in Baseline-2. Baseline-2 is not a coordinating method as a result.

2) Strong PV Power Fluctuating Scenario

IntheBaseline-1, the15-minuteaverageofthegroundtruth load and PV power profiles shown in Fig. 5 and Fig. 7 respectively are regarded as the prediction for day-long OLTC tap operation scheduling. If the forecasts come true to 100%, Baseline-1 might attain ideal voltage control performance. On the other hand, swift-moving clouds have the ability to quickly engulf a distribution system in one or two minutes, causing a sudden loss of nearly 70% of the PV power in this area [15]. Because of this, it is exceedingly challenging to forecast the PV power profile over the course of a day, particularly for distribution systems with narrow area ranges. Forecast mistake might seriously impair Baseline-1's ability to regulate voltage. Furthermore, OLTC adhering to its "LDC + time delay" control rule (as in Baseline-2 and our suggested approach) has the ability to adaptively modify its tap position in order to regulate voltage as needed. However, based on its scheduling in the OLTC, the tap location will remain constant every 15 minutes.

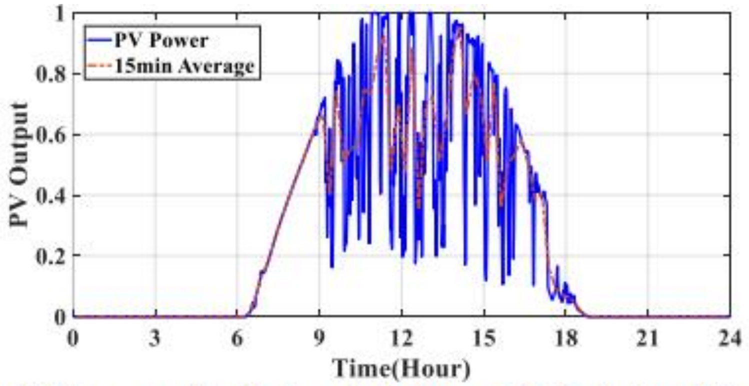
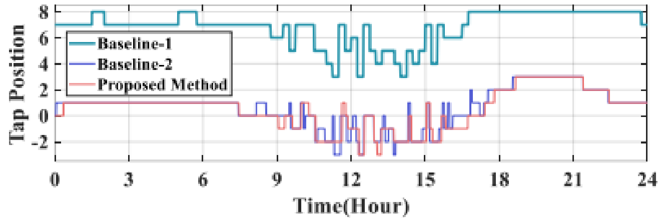
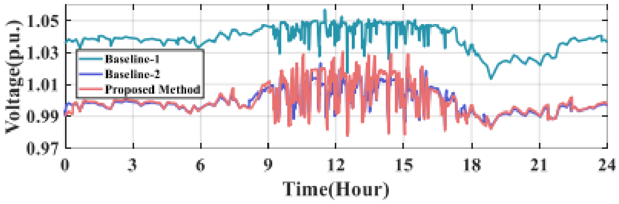


Fig. 7. PV power profile with strong fluctuations and its 15-minute prediction.





IV. CONCLUSION AND FUTURE WORK

This article proposes a novel control structure aimed at coordinating inverters and On-Load Tap Changers (OLTC) to adjust system voltage in a "soft" manner. It allows the previously deployed OLTC to maintain its autonomous operation state for tap switching within this soft coordination framework. Moreover, by adjusting the Var output of inverters, it is possible to reduce system line losses and achieve coordinated voltage control. Importantly, our proposed algorithm demonstrates resilience to variations in the line's X/R ratio, effectively distributing the voltage control responsibilities between OLTC and the Var compensation provided by inverters. Case studies have shown that our method can consistently regulate system voltage without leading to Var saturation in inverters, even under conditions of high PV penetration (exceeding 100%) and X/R ratios below one.

The performance of our proposed method, in terms of voltage regulation and system line loss, is evaluated against two baseline methods through case studies. Baseline-1, a two-layer method, relies on the accuracy of PV power prediction and maintains a constant OLTC tap position within each dispatch interval (e.g., 15 minutes), which limits its flexibility in addressing voltage fluctuations caused by variable PV power. This can result in unnecessary Var compensation, increased system line losses, or even voltage control failures on days with significant PV power variability. Baseline-2, lacking a specialized coordination mechanism, depends solely on local control for both OLTC and inverters, yet it demonstrates consistent performance across various scenarios.

Our proposed technique not only achieves overvoltage reduction and line loss minimization in all cases but also allows OLTC to retain its autonomous control state, enhancing the overall system flexibility. The operational states of the system, as discussed in the paper, remain well within the limits of the test systems' maximum power transmission capabilities. However, in networks that are heavily loaded or experiencing post-disturbance conditions, employing a learning-based algorithm for online voltage collapse risk identification might be viable. In such scenarios, the priority for OLTC and inverters collaboration could shift towards increasing the security margin instead of minimizing line losses, once the risk threshold is breached. This opens avenues for further research into detailed algorithm design.

REFERENCES

[1] T. A. Short, "Electric Power Distribution Equipment and Systems," United States of America: CRC Press, 2006.

[2] W. H. Kersting, "Distribution System Modeling and Analysis," CRC Press, 2002.

[3] C. Long and L. F. Ochoa, “Voltage Control of PV-Rich LV Networks: OLTC-Fitted Transformer and Capacitor Banks,” IEEE Transactions on Power Systems, vol. 31, no. 5, pp. 4016-4025, 2016.

[4] A. T. Procopius and L. F. Ochoa, “Voltage Control in PV-Rich LV Networks Without Remote Monitoring,” IEEE Transactions on Power Systems, vol. 32, no. 2, pp. 1224-1236, 2017.

[5] N. Yorino, Y. Zoka, M. Watanabe, and T. Kurushima, “An Optimal Autonomous Decentralized Control Method for Voltage Control Devices by Using a Multi-Agent System,” IEEE Transactions on Power Systems, vol. 30, no. 5, pp. 2225-2233, 2015.

[6] M. I. Hossain, R. Yan, and T. K. Saha, “Investigation of the Interaction Between Step Voltage Regulators and Large-Scale Photovoltaic Systems Regarding Voltage Regulation and Unbalance,” IET Renewable Power Generation, vol. 10, no. 3, pp. 299-309, 2016.

[7] L. Wang, T. K. Saha, and R. Yan, “Voltage Regulation for Distribution Systems with Uneven PV Integration in Different Feeders,” 2017 IEEE Power & Energy Society General Meeting, 2017.

[8] L. Wang, R. Yan, and T. K. Saha, “Voltage Management for Large Scale PV Integration into Weak Distribution Systems,” IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 4128-4139, Sep. 2018.

[9] J. Li, C. Liu, M. E. Khodayar, M.-H. Wang, Z. Xu, B. Zhou, and C. Li, “Distributed Online VAR Control for Unbalanced Distribution Networks With Photovoltaic Generation,” IEEE Transactions on Smart Grid, vol. 11, no. 6, pp. 4760-4772, 2020.

[10] L. Wang, R. Yan, F. Bai, T. K. Saha, and K. Wang, “A Distributed Inter-Phase Coordination Algorithm for Voltage Control with Unbalanced PV Integration in LV Systems,” IEEE Transactions on Sustainable Energy, vol. 11, no. 4, pp. 2687-2697, Oct. 2020.

[11] Y. Wang, M. H. Syed, E. Guillo-Sansano, Y. Xu, and G. M. Burt, “Inverter-Based Voltage Control of Distribution Networks: A Three-Level Coordinated Method and Power Hardware-in-

the-Loop Validation," IEEE Transactions on Sustainable Energy, vol. 11, no. 4, pp. 2380-2391, 2020.
[12] H. Liu and W. Wu, "Two-Stage Deep Reinforcement Learning for Inverter-Based Volt-VAR Control in Active Distribution Networks," IEEE Transactions on Smart Grid, vol. 12, no. 3, pp. 2037-2047, 2021.
[13] L. Wang, L. Xie, Y. Yang, Y. Zhang, K. Wang, and S.-j. Cheng, "Distributed Online Voltage Control with Fast PV Power Fluctuations and Imperfect Communication," IEEE Transactions on Smart Grid, vol. 14, no. 5, pp. 3681-3695, 2023.
[14] M. R. Jafari, M. Parniani, and M. H. Ravanji, "Decentralized Control of OLTC & PV Inverters for Voltage Regulation in Radial Distribution Networks with High PV Penetration," IEEE Transactions on Power Delivery, 2022.
[15] L. Wang, F. Bai, R. Yan, and K. T. Saha, "Real-Time Coordinated Voltage Control of PV Inverters and Energy Storage for Weak Networks With High PV Penetration," IEEE Transactions on Power Systems, vol. 33, no. 3, pp. 3383-3395, May 2018.
[16] T. Tewari, A. Mohapatra, and S. Anand, "Coordinated Control of OLTC and Energy Storage for Voltage Regulation in Distribution Network With High PV Penetration," IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 262-272, 2021.
[17] M. Chamana, B. H. Chowdhury, and F. Jahanbakhsh, "Distributed Control of Voltage Regulating Devices in the Presence of High PV Penetration to Mitigate Ramp-Rate Issues," IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 1086-1095, 2018.
[18] K. M. Muttaqi, A. D. T. Le, M. Negnevitsky, and G. Ledwich, "A Coordinated Voltage Control Approach for Coordination of OLTC, Voltage Regulator, and DG to Regulate Voltage in a Distribution Feeder," IEEE Transactions on Industry Applications, vol. 51, no. 2, pp. 1239-1248, 2015.
[19] Y. Zhang, X. Wang, J. Wang, and Y. Zhang, "Deep Reinforcement Learning-Based Volt-VAR Optimization in Smart Distribution Systems," IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 361-371, 2021.
[20] R. Zafar and H. R. Pota, "Multi-Timescale Coordinated Control with Optimal Network Reconfiguration using Battery Storage System in Smart Distribution Grids," IEEE Transactions on Sustainable Energy, pp. 1-12, 2023.
[21] X. Sun, J. Qiu, Y. Yi, and Y. Tao, "Cost-Effective Coordinated Voltage Control in Active Distribution Networks With Photovoltaics and Mobile Energy Storage Systems," IEEE Transactions on Sustainable Energy, vol. 13, no. 1, pp. 501-513, 2022.
[22] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis, and J. Sun, "Two-Timescale Voltage Control in Distribution Grids Using Deep Reinforcement Learning," IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2313-2323, 2020.
[23] Y. Huo, P. Li, H. Ji, H. Yu, J. Yan, J. Wu, and C. Wang, "Data-Driven Coordinated Voltage Control Method of Distribution Networks With High DG Penetration," IEEE Transactions on Power Systems, vol. 38, no. 2, pp. 1543-1557, 2023.
[24] Z. Gajić, D. Carlson, and M. Kockott, "Advanced OLTC Control to Counteract Power System Voltage Instability," ABB Power Technologies, Substation Automation, SE-721 59 VÄSTERÅS, SWEDEN.
[25] F. Bai, R. Yan, T. K. Saha, and D. Eghbal, "An Excessive Tap Operation Evaluation Approach for Unbalanced Distribution Networks With High PV Penetration," IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 169-178, 2021.

[26] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," [Online]. Available: https://arxiv.org/pdf/1801.01290.pdf.

[27] D. Cao, J. Zhao, W. Hu, N. Yu, F. Ding, Q. Huang, and Z. Chen, “Deep Reinforcement Learning Enabled Physical-Model-Free Two-Time scale Voltage Control Method for Active Distribution Systems,” IEEE Transactions on Smart Grid, vol. 13, no. 149-165, 2022.

[28] D. Hu, Z. Ye, Y. Gao, Z. Ye, Y. Peng, and N. Yu, “Multi-Agent Deep Reinforcement Learning for Voltage Control With Coordinated Active and Reactive Power Optimization,” IEEE Transactions on Smart Grid, vol. 13, no. 6, pp. 4873-4886, 2022.

[29] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches." [Online]. Available: https://arxiv.org/abs/1409.1259.

[30] "OLC12641 Aerial Catalogue," [Online]. Available: https://www.olex.com.au/eservice/Australia-en_AU/fileLibrary/Download_540225171/Australasia/files/OLC12641_AerialCat.pdf.

[31] "IEEE PES Test Feeder," [Online]. Available: https://cmte.ieee.org/pes-testfeeders/resources/.