Adaptive Volt-Var Control with PV inverter and Solid State Transformer Interface Mobile Energy Storage Systems

RAZA, Muhammad Bilal, NOUMAN, Muhammad, NADEEM, Muhammad Faisal, SAJJAD, Intisar Ali and AKMAL, Muhammad <http://orcid.org/0000-0002-3498-4146>

Adaptive Volt-Var Control with PV inverter and Solid State Transformer Interface Mobile Energy Storage Systems

Muhammad Bilal Raza
Electrical Engineering Department
University of Engineering & Technology
Taxila, Pakistan
22MS-EE-7@students.uettaxila.edu.pk

Muhammad Nouman
Electrical Engineering Department
University of Engineering & Technology
Taxila, Pakistan
engr.nouman1214@gmail.com

Muhammad Faisal Nadeem
Electrical Engineering Department
University of Engineering & Technology
Taxila, Pakistan
faisal.nadeem@uettaxila.edu.pk

Intisar Ali Sajjad
Electrical Engineering Department
University of Engineering & Technology
Taxila, Pakistan
intisar.ali@uettaxila.edu.pk

Muhammad Akmal
Electrical Engineering Department
Sheffield Hallam University
S1 1WB Sheffield, United Kingdom
m.Akmal@shu.ac.uk

***Abstract*—The growing integration of distributed energy resources (DERs) such as PV systems and mobile energy storage systems (MESSs) is transforming the conventional distribution model into the active distribution networks (ADNs). However, this transition introduces a highly dynamic and non-linear nature into the system, posing significant challenges to voltage stability and reactive power management. Conventional volt-var control (VVC) approaches incorporating slow-responding devices lack the capability to respond against dynamic environmental conditions, limiting their potential to mitigate voltage violation rate (VVR) and network power losses. To address these issues, this paper proposes an intelligent VVC strategy leveraging PV inverters and solid-state transformer-enabled mobile energy storage systems (SST-MESS). The SST-MESS is capable of providing bidirectional power flow along with reactive power compensation, alongside PV inverters to improve real-time system efficiency. To optimize the proposed multi-objective framework, a novel two-critic reinforcement learning scheme has been implemented. The conceptualized strategy has been tested on the IEEE-69 bus system; a comparative analysis with the baseline VVC approach indicates the effectiveness of the presented strategy, resulting in a 15.645% and 16.015% improvement for two-critic deep deterministic policy gradient (TC-DDPG) and two-critic soft actor-critic (TC-SAC) compared to the baseline scenario, respectively, showing the effectiveness of the presented strategy. Hence, contributing to the optimal integration of DERs in ADNs.**

***Index Terms*—Deep reinforcement learning, Mobile energy storage system, PV inverter, solid-state transformer, volt-var control**

I. Introduction

Motivated by the goal to recarbonize electricity generation, the advancement of power electronics has facilitated the widespread integration of decentralized generation (DG) and mobile energy storage systems (MESSs), shifting conventional distribution systems into active distribution networks (ADNs) [1]. However, the bidirectional power flow from DG and MESS raises concerns about voltage regulation issues. Although reactive power compensation helps stabilize voltage fluctuations, it might also contribute to high power losses in the system [2]. Therefore, a well-organized volt-var control (VVC) strategy is crucial for ensuring a reliable and efficient operation of ADNs.

For instance, in [3] a multi-time scale VVC coordination has been introduced utilizing PV systems, electric vehicles (EVs), and on-load tap changer (OLTC) voltage setpoints to regulate voltage deviations and minimize power losses. Similarly, [4] presented a double-time scale optimization framework based on robust model predictive control (RMPC). The coordination process involved multiple voltage regulation devices. In the slow-time scale control, OLTC, step-voltage regulators (SVR), and capacitor banks (CBs) are optimized to reduce voltage deviations. Based on this, fast-time scale control (FTC) coordinated active and reactive power outputs of DGs to deal with the fast voltage fluctuations while considering the intermittent nature of DG outputs. In [5], an optimization framework has been modeled based on the non-dominated sorting dung beetle optimizer (NSDBO) and model predictive control (MPC) to design a scheduling plan for PV systems, EVs, and energy storage elements. Furthermore, a joint optimization scheme for dynamic network reconfiguration (DNR) and MESS has been introduced and reformulated as a mixed-integer second-order cone programming (MISOCP) model. The proposed strategy leverages the penalty alternating direction method (PADM) being implemented on the IEEE 33-bus network integrated with the transportation network to achieve enhanced voltage stability, reduced distribution losses, and improved renewable integration [6].

After reviewing the above literature study, we find that all of them can be classified as model-based schemes that rely on complete model information which is inherently uncertain, making accurate modeling of ADN even more challenging

979-8-3315-6520-6/25/$31.00 ©2025 IEEE

[7]. To address the limitation of model-based control [8], data-driven deep reinforcement learning (DRL) schemes have been introduced [9], finding extensive application in areas like games [10], robotics control [11], and energy management [12]. In [13], data-driven modeling based on centralized training and decentralized execution has been nominated for voltage regulation in PV-rich networks. In [14], a data-driven day-ahead VVC scheduling framework has been modeled to incorporate flexible energy resources, particularly MESS, to deal with voltage deviations in a dynamic environment. A two-stage hybrid energy dispatch framework has been modeled for energy storage systems (ESS) to help smooth out the intermittency nature of DGs and respond to real-time disturbances [15]. Meanwhile, [16] introduced a joint scheduling approach that integrates MESS with VVC, taking into account voltage and state of charge (SOC) constraints. To improve VVC performance, modeling strategies based on two-time scale [17], three-time scale [18] have been devised to mitigate voltage deviations and power losses. In [19], a graph-convolutional network (GCN) to determine optimal voltage control actions based on network topology. Moreover, [20] proposed a human-in-the-loop soft actor-critic (HL-SAC) for voltage variations in unbalanced three-phase distribution systems.

In the above study, a key exception is persistent suboptimal errors resulting from inaccuracies in the critic network for the evaluation of the state-action value function [21], which caused underestimation bias [22]. Liu et al. [23] presented a two-critic deep reinforcement learning (TC-DRL) approach to address the approximation complexity of the single-critic learning strategy, by isolating the two tasks for each critic, leading to a faster convergence rate.

Conventional VVC approaching based on capacitor bank [3], OLTCs [4], and static var compensator [23] often suffers from slow responding nature and limited capability to adapt to dynamic voltage fluctuations caused by the integration of DGs and MESS in ADNs. The optimal solution lies in using solid-state transformer (SST) and PV inverter. The SST is capable of providing fast, flexible, and localized reactive power compensation based on real-time charging and discharging patterns of MESS, hence ensuring system efficiency.

The key research contributions are given as follows:

- An innovative VVC approach has been presented to handle intermittent nature of DGs by leveraging the concept of SST-enabled MESS, able to handle bidirectional power flow and dynamic reactive power support along with smart PV inverter, based on energy demand, generation, and charging/discharging behavior of MESS.
- State-of-the art TC-DRL algorithms has been utilized name two-critic deep deterministic policy gradient (TC-DDPG) and two-critic soft actor critic (TC-SAC) to model the control framework based on multi-objective nature of VVC problem.
- A comparative analysis has been presented by taking into account conventional VVC as a baseline scenario, along with that single-critic scheme also has been implemented to showcase the effectiveness of TC-DDPG and TC-SAC in mitigation of voltage violation rate (VVR), along with minimization of network losses.

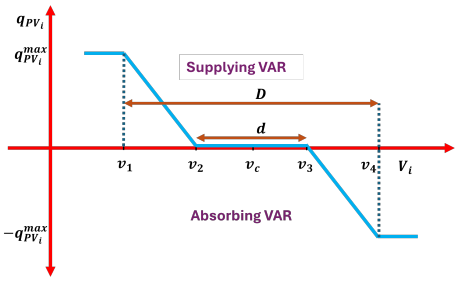

Fig. 1. PV inverter Volt-Var Contorl Setting (VVCS) curve

II. Mathematical Modeling

A. Smart inverter

The smart inverter handles voltage deviations by leveraging the volt-var control setting (VVCS) shown in Fig 1 [24]. The inverter dynamically regulates reactive power output based on terminal voltage fluctuations. It provides reactive power (behaving as a capacitor) when the voltage decreases below the threshold point v_2 and absorbs reactive power (operating as an inductor) when the voltage violates the limit v_3. In VVCS, v_c is considered as a reference point, the voltage width D, and the deadband d. In the VVCS configuration, the curve points (v_1, v_2, v_3, and v_4) are modeled as $v_1 = v_c - \frac{D}{2}$, $v_2 = v_c - \frac{d}{2}$, $v_3 = v_c + \frac{d}{2}$ and $v_4 = v_c + \frac{D}{2}$.

$$Q_{PVi} = g(V_i) = \begin{cases} Q_{\text{PVi}}^{max}, & V_i \le v_1, \\ \frac{v_2 - V_i}{v_2 - v_1} Q_{\text{PVi}}^{max}, & v_1 \le V_i \le v_2, \\ 0, & v_2 \le V_i \le v_3, \\ \frac{v_3 - V_i}{v_4 - v_3} Q_{\text{PVi}}^{max}, & v_3 \le V_i \le v_4, \\ -Q_{\text{PVi}}^{max}, & V_i \ge v_4. \end{cases} \quad (1)$$

Here, V_i denotes local bus voltage, while Q_{PV} presents the inverter's reactive power given as follows:

$$Q_{PV_i} \le \sqrt{S_{PV_i}^2 - P_{PV_i}^2}. \quad (2)$$

where, S_{PV_i} and P_{PV_i} present the apparent power and active power, respectively.

B. Solid State transformer-enabled Mobile energy storage system

The SST behaves as a smart transformer, capable of dynamic voltage regulation to suit MESS requirements while allowing bidirectional power flow. Fig 2 presents the SST-enabled MESS interfaced with the distribution system using a modular active rectifier. It produces high-voltage DC (HVDC) and gives it to a high-frequency DC-to-DC converter. This step comprises a bidirectional converter known as a dual-active bridge (DAB), that allows the transformation of HVDC into

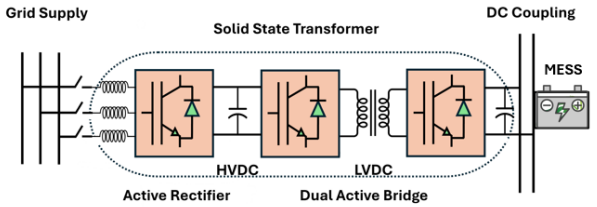

Fig. 2. Solid state transformer-enabled mobile energy storage system

low-voltage DC (LVDC) while providing isolation. The output DC link then makes interference with MESS [25].

1) Active rectifier: In this module, the relationship between the input current I_α, I_β, along with active and reactive power P, Q can be modeled as follows:

$$\begin{bmatrix} P(t) \\ Q(t) \end{bmatrix} = \omega \begin{bmatrix} \phi_\alpha & -\phi_\beta \\ \phi_\alpha & \phi_\beta \end{bmatrix} \begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} \tag{3}$$

where ω presents angular frequency of grid and ϕ_α, ϕ_β models the virtual source fluxes. Moreover, using the forward Euler scheme, we have formulated the following relationship.

$$P^{k+1} = P^k + \omega T_s \left(\frac{d\phi_\alpha}{dt} i_\beta - \frac{d\phi_\beta}{dt} i_\alpha + \frac{di_\beta}{dt} \phi_\alpha - \frac{di_\alpha}{dt} \phi_\beta \right) \tag{4}$$

$$Q^{k+1} = Q^k + \omega T_s \left(\frac{d\phi_\alpha}{dt} i_\alpha + \frac{d\phi_\beta}{dt} i_\beta + \frac{di_\alpha}{dt} \phi_\alpha + \frac{di_\beta}{dt} \phi_\beta \right) \tag{5}$$

where T_s indicate sampling period, and $t = k$ and $t = k + 1$ represent present and future system state. Moreover, relationship between the input voltage U_α, U_β and the output voltage V_{C_α},V_{C_β} can be modeled as follows:

$$\frac{\partial}{\partial t} \begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} = \frac{1}{L} \left(\begin{bmatrix} U_\alpha \\ U_\beta \end{bmatrix} - R \begin{bmatrix} I_\alpha \\ I_\beta \end{bmatrix} - \begin{bmatrix} V_{C\alpha} \\ V_{C\beta} \end{bmatrix} \right) \tag{6}$$

here, R and L denote resistance and inductance of the active rectifier, respectively.

2) Dual active bridge: DAB works as a power converter, capable of providing bidirectional power flow using phase-shift control. The power transferred can be modeled as follows:

$$P_O = \frac{V_i V_O}{2nL_r f_{sw_{dab}}} \varphi(1 - \varphi) \tag{7}$$

where, φ represents the phase angle of the primary side with respect to the secondary side. Meanwhile, L_r denotes auxiliary inductance, $f_{sw_{dab}}$ presents switching frequency, V_i and V_O correspond to input and output voltages, respectively.

III. Problem Formulation

The Markov decision process has been systemized to model VVC problem formulation in ADNs. In the learning framework, the DRL agent observes a state $s \in \mathbf{S}$, performs an action $a \in \mathcal{A}$ based on its optimal policy $\pi : \mathcal{S} \to \mathcal{A}$, obtains a reward $r \in \mathcal{R}$, and transition towards next state s' with probability $P(s'|s, a)$.

1) State: The state presents condition of the DRL environment and can be defined as $s = [p^T, q^T, v^T, q_g^T]^T$, where p, q, and v is the vector presentation of active power, reactive power and voltage profile, respectively. Meanwhile, q_g presents reactive power generation of controllable devices such as smart inverters and SST-enabled MESS.

2) Action: The action determined system response after state observation i.e. $a = q_g'$. In current scenario, the action is reactive power adjustment from PV inverter and SST-enabled MESS, constrained by boundary limits $|q_g| \leq \sqrt{s_g^2 - \overline{p_g^2}}$ and $\underline{q_g} \leq q_g \leq \overline{q_g}$, respectively.

3) Reward: The reward is modeled based on the next state s', comprising negative power losses r_p and VVR r_v. The reward structure is modeled as follows:

$$r_p = -\sum_{i=0}^{N} p_i \tag{8}$$

$$r_v = -\sum_{i=0}^{N} \left[\max(v_i^{'} - \bar{v}_i, 0) + \max(\underline{v_i} - v_i^{'}, 0) \right] \tag{9}$$

where, N represents the number of buses. The total reward structure r is modeled as sum of r_p and $c_v r_v$, where c_v is a penalty coefficient.

A. Two reward structure modeling

The reward function of VVC comprises active power losses r_p and VVR r_v with distinct mathematical structures. Based on branch $i \to j$, the flow of active p_{ij} and reactive power q_{ij} can be modeled as follows:

$$\begin{aligned} \sum_{i:i\to j} p_{ij} + p_j &= \sum_{k:j\to k} p_{jk} \\ \sum_{i:i\to j} q_{ij} + q_j &= \sum_{k:j\to k} q_{jk} \\ p_{l_{ij}} &= r_{ij}(p_{ij}^2 + q_{ij}^2) \\ v_j &= v_i - (r_{ij} p_{ij} + x_{ij} q_{ij}) \end{aligned} \tag{10}$$

here, the resistance and reactance parameters are denoted with r_{ij} and x_{ij}, respectively.

$$p_j = \begin{cases} p_0, & \forall j = 0 \\ p_{gj} - p_{dj}, & \forall j \in N_g \\ p_{cj} - p_{dj}, & \forall j \in N_c \\ -p_{dj}, & \forall j \in N \setminus N_g \end{cases} \tag{11}$$

$$q_j = \begin{cases} q_0, & \forall j = 0 \\ q_{gj} - q_{dj}, & \forall j \in N_g \\ q_{cj} - q_{dj}, & \forall j \in N_c \\ -q_{dj}, & \forall j \in N \setminus (N_g \cup N_c) \end{cases} \tag{12}$$

where N, N_d, N_g, and N_c indicate the total number of buses, load buses, and buses with PV inverters and SST-enabled MESS, respectively. Furthermore, loading at bus j is p_{dj}/q_{dj}, while PV inverters produce p_{gj}/q_{gj}, and SST-enabled MESS gives p_{cj}/q_{cj}.

IV. Two-critic reinforcement learning framework

The TC-DRL learning framework has been presented in Fig. 3. The system operates based on centralized training and centralized execution, allowing a centralized agent to engage with the environment and optimize its policy based on feedback signals from two critic networks.

The Q-value function of the critic networks that deal with the minimization of power losses, $Q_p(s,a)$, and for the mitigation of voltage violation $Q_v(s,a)$ can be modeled as follows:

$$Q_p^{\pi}(s,a) = \mathbb{E}_{a\sim\pi}\left[\sum_{t=0}^{\infty}\gamma^t r_{p,t} \mid s_0 = s, a_0 = a\right] \tag{13}$$

$$Q_v^{\pi}(s,a) = \mathbb{E}_{a\sim\pi}\left[\sum_{t=0}^{\infty}\gamma^t r_{v,t} \mid s_0 = s, a_0 = a\right] \tag{14}$$

For every model interaction, the neural networks $Q_{\phi_p(s,a)}$ and $Q_{\phi_v(s,a)}$ are used for the evaluation of the critic network based on the mean squared error function, expressed as follows:

$$\begin{aligned} L_{Q_p}(\phi_p) &= \frac{1}{|B|}\sum_{(s,a,r_p)\in B}\left(Q_{\phi_p}(s,a) - r_p\right)^2 \\ L_{Q_v}(\phi_v) &= \frac{1}{|B|}\sum_{(s,a,r_v)\in B}\left(Q_{\phi_v}(s,a) - r_v\right)^2 \end{aligned} \tag{15}$$

Moreover, the actor network $\pi_\theta(s)$ has been trained to optimize the loss function, given as,

$$L_\pi(\theta) = \frac{1}{|B|}\sum_{s\in B}\left[Q_{\phi_p}(s,\pi_\theta(s)) + c_v Q_{\phi_v}(s,\pi_\theta(s))\right] \tag{16}$$

In TC-DDPG, to enhance the deterministic policy, an independent exploration noise is incorporated during the training process.

$$a = \text{clip}(\pi_\theta(s) + \xi, a, \bar{a}), \tag{17}$$

here ξ denotes exploration noise produced based on a Gaussian process $\xi \sim \mathcal{N}(0,\sigma)$. The function clipping ensures that values stay within limits $[a,\bar{a}]$ by compensating them to closest boundary limits using the $\min(\max(\pi_\theta(s)+\xi, a), \bar{a})$.

The TC-SAC learning model can be attained by minor modifications in the TC-DDPG learning architecture. The deterministic policy is replaced with a stochastic policy parameterization.

$$\pi_\theta(\cdot|s) = \tanh(\mu_\theta(s) + \sigma_\theta(s)\odot\xi), \quad \xi\sim\mathcal{N}(0,I) \tag{18}$$

where μ_θ and σ_θ are the approximation parameters for mean and standard deviation. In addition, the actor loss function $L_\pi(\theta)$ is substituted with the entropy regularization.

$$\begin{aligned} L_\pi(\theta) = \frac{1}{|B|}\sum_{s\in B, a\sim\pi_\theta(\cdot|s)} &\left(Q_{p_\phi}(s,a) + c_v Q_{v_\phi}(s,a)\right) \\ &- \alpha\log\pi_\theta(\cdot|s) \end{aligned} \tag{19}$$

The learning rate α is updated based on the following equation.

$$\alpha \leftarrow \alpha + \lambda_\alpha \nabla_\phi L(\alpha) \tag{20}$$

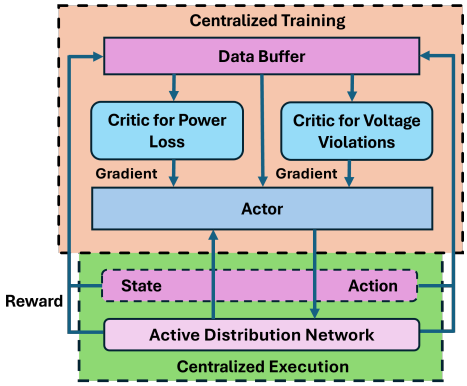

Fig. 3. Two critic learning framework

Both learning frameworks, TC-DDPG and TC-SAC, bring down the variance of the policy gradient by employing the critic value function as a baseline.

V. Simulations and Results

To validate the competence of the proposed VVC scheme, numerical simulations are performed on Baran and Wu's IEEE-69 bus distribution system based on the TC-DRL approach [26]. The system comprised 4 PV inverters with a rating of 2 MVar reactive power and 1.5 MW active power, integrated to buses 5, 22, 44, and 63 respectively, and 1 SST-enabled MESS of 2 MVar reactive and 1.6 MW active power rating, connected to bus 13 of the system. The load and generation profile were adjusted based on daily variations by incorporating 96 data points from [27], while the charging and discharging patterns of MESS were extracted and modified from [28]. Moreover, an additional 20% uniformly distributed noise was incorporated to account for uncertainties. The boundary limit for voltage was considered in the range of 0.95 to 1.05 per unit. Matpower [29] was used to source network data and later introduced into Pandapower [30]. The DRL algorithms were executed using Python 3.9.18, with balanced load flow analysis using Pandapower 3.0.0. Moreover, the simulation procedure for DRL algorithms was based on PyTorch 2.5.1. The total simulation time per training episode on average comprises of 30 seconds based on specified hardware design comprising 12$^{\text{th}}$ Gen Intel(R) Core(TM) i7-12700 2.10 GHz processor with 32 GB RAM. The time primarily stems from iterative interactions between the agent and the power flow solver i.e. MATPOWER, during environment transitions as shown in the figure 4. While current training times may not be suitable for online or real-time retraining, it is important to note that once the model is trained offline, the inference time for deploying a trained agent is less than one second, making it viable for real-time operation with control intervals ranging from 5 to 15 minutes.

A. Discussion and result analysis

The training process was spread over 300 days, utilizing parametric settings from [23]. To ensure consistency and

TABLE I
COMPARATIVE ANALYSIS OF DIFFERENT VVC APPROACHES AND CONTROL STRATEGIES

Algorithm	Reward		P_{loss} /(MW)		VVR/(p.u.)	
	69-bus	69-bus [23]	69-bus	69-bus [23]	69-bus	69-bus [23]
Deterministic policy						
DDPG	-3.727	-4.092	3.7269	4.006	8.639e-5	1.717e-3
TC-DDPG	-2.855	-3.440	2.8554	3.385	7.089e-7	1.099e-3
Stochastic policy						
SAC	-3.067	-3.625	3.0671	3.548	1.791e-6	1.527e-3
TC-SAC	-2.860	-3.455	2.8605	3.406	1.459e-5	9.751e-4

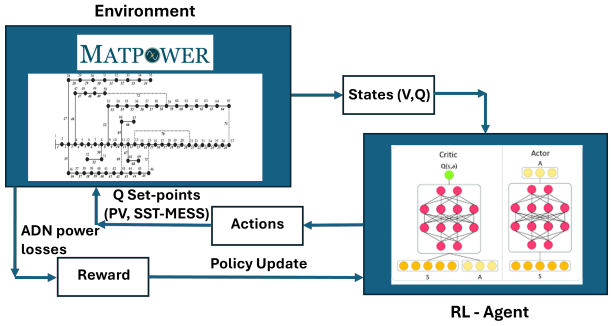

Fig. 4. Schematic representation of interaction between agent and ADN environment

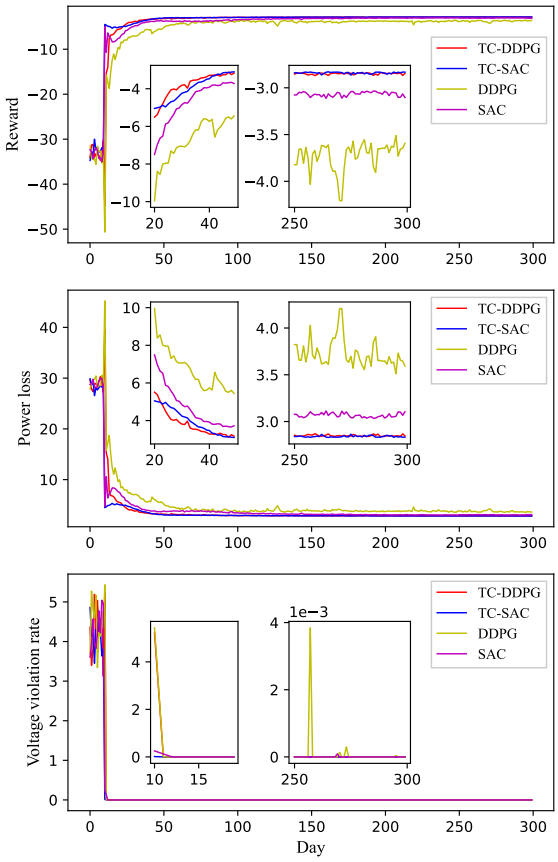

Fig. 5. IEEE-69 Bus System test results for DDPG, TC-DDPG, SAC and TC-SAC: (a) Reward (b) Power loss (c) Voltage violation rate

diminish randomness, all DRL approaches were tested on three random seeds, and the average outputs were considered.

1) Deterministic policy learning: In the deterministic policy, two state-of-the-art learning frameworks, i.e. DDPG and TC-DDPG, have been presented to model the VVC scheme.

2) Stochastic policy learning: Meanwhile, in the stochastic policy framework, we use SAC and its two-critic derivative TC-SAC that has been proposed. Fig. 5 shows the training performance of TC-DRL learning schemes, illustrating the aggregated reward, network losses, and VVR. Initially, TC-DRL approach, i.e. TC-DDPG and TC-SAC, achieves improved convergence rates as compared to the single-critic counterpart, i.e. DDPG and SAC, as depicted from the learning trajectories noted between days 20 and 50 in Fig. 5. Second, TC-DDPG and TC-SAC attained higher rewards compared to DDPG and SAC, as highlighted from the learning curve from days 250 to 300 in Fig. 5. As listed in Table I, the proposed VVC approach incorporating PV-inverter and SST-enabled MESS via TC-DRL outperforms the conventional VVC scheme [23]. The presented approach improved the performance of network power loss considerably and voltage violation marginally. It can be observed that VVR quickly converged to zero and slightly remained around zero. Therefore, in the total reward function, voltage violation constitutes a small proportion while the dominance was mainly illustrated in minimizing network power losses. The small variations may be due to conflicting VVC objectives and numerical disturbances in neural network training.

A comparative analysis has been carried out based on key evaluation parameters including: (i) Power loss P_{loss}, indicating total network power loss measured in megawatts (MW); (ii) VVR, highlighting voltage deviations from prescribed limits in per unit (p.u.); (iii) Reward, showcasing the cumulative improvement in voltage regulation and network loss minimization. The performance analysis based on the IEEE-69 bus system illustrates that the two-critic scheme in combination with the proposed VVC approach outperformed the baseline approach [23] and their single-critic counterpart. TC-DDPG (-2.855) received the highest reward surpassing TC-SAC (-2.860), SAC (-3.067), and DDPG (-3.727), and baseline VVC approach (-3.440) [23], showcasing improved control efficiency and system stability. Moreover, TC-DDPG leads to lowest power loss, i.e., 2.8554 MW, which is 23.338% improved over DDPG and 15.645% less than the baseline VVC approach [23]. Similarly, TC-SAC obtained 2.860 MW, which is 6.736% better than SAC and 16.015% improved compared to baseline scenario [23]. In P_{loss} reduction compared to baseline VVC scheme, TC-SAC outperforms other due to its entropy regulation but for the presented approach TC-DDPG have more optimized results i.e. 2.8554 MW compared to TC-SAC (2.8605 MW). Meanwhile, ability of SST-enabled MESS to compensate for reactive power in coordination with smart PV inverter leads to improved voltage regulation compared to

baseline approach [23], while TC-DDPG outperforms other algorithms for the presented scheme.

VI. Conclusion

This study introduced a novel VVC strategy for ADNs using smart PV inverters and SST-enabled MESS, driven by TC-DRL control framework. Unlike traditional VVC control strategies based on slow-responding devices like capacitor bank, OLTCs, and SVCs. This research showcased the fast-reacting nature of SST-enabled MESS capable of providing bi-directional power flow along with reactive power compensation addressing VVR and mitigation of power losses in real-time dynamics. A comparative analysis with baseline VVC control approach (i.e. based on PV inverter and SVCs) indicate the superior performance of TC-SAC by attaining 16.015% improvement over the baseline VVC approach. Meanwhile, in the proposed approach TC-DDPG outperformed other by accomplishment a reward of -2.855 compared to TC-SAC (-2.860), SAC (-3.067), and DDPG (-3.727). In future, we plan to enhance the VVC strategy by incorporating transient cost analysis for unbalanced distribution networks. We intercepted that proposed DRL algorithms will performed efficiently in these cases by simplifying the learning process and distinguishing the impact of multiple control tasks.

References

[1] G. Gutiérrez-Alcaraz, D. Espín-Sarzosa, V. Hinojosa, F. Valencia-Arroyave, P. Mendoza-Araya, and A. Montenegro, "Microgrids and clusters of microgrids," in *Towards Future Smart Power Systems with High Penetration of Renewables*. Elsevier, 2025, pp. 151–169.

[2] J. Qiao, Y. Mi, J. Shen, D. Xia, D. Li, and P. Wang, "Active and reactive power coordination optimization for active distribution network considering mobile energy storage system and dynamic network reconfiguration," *Electric Power Systems Research*, vol. 238, p. 111080, 2025.

[3] M. R. Dar and S. Ganguly, "Voltage regulation and loss minimization of active distribution networks with uncertainties using chance-constrained model predictive control," *IEEE Transactions on Power Systems*, 2024.

[4] G. Song, Q. Wu, W. Jiao, and L. Lu, "Distributed coordinated control for voltage regulation in active distribution networks based on robust model predictive control," *International Journal of Electrical Power & Energy Systems*, vol. 166, p. 110529, 2025.

[5] J. Zhang, J. Wang, J. Yan, and P. Cheng, "Research on multi-time scale volt/var optimization in active distribution networks based on nsdbo and mpc approach," *Electric Power Systems Research*, vol. 238, p. 111141, 2025.

[6] J. Qiao, Y. Mi, J. Shen, D. Xia, D. Li, and P. Wang, "Active and reactive power coordination optimization for active distribution network considering mobile energy storage system and dynamic network reconfiguration," *Electric Power Systems Research*, vol. 238, p. 111080, 2025.

[7] Q. Hou, N. Dai, and Y. Huang, "Voltage regulation enhanced hierarchical coordinated volt/var and volt/watt control for active distribution networks with soft open points," *IEEE Transactions on Sustainable Energy*, 2024.

[8] J. Xu, Y. Mu, W. Cheng, G. Gong, J. Huang, T. Zhang, Q. Hu, K. Chen, and A. Y. Wu, "A distributed robust voltage control framework for active distribution systems," *IEEE Transactions on Industry Applications*, pp. 1–11, 2025.

[9] N. Sutton, "Data-driven spiking neural network models that refine theories of hippocampal function," Ph.D. dissertation, George Mason University, 2024.

[10] G. C. Barros e Sá and C. A. G. Madeira, "Deep reinforcement learning in real-time strategy games: a systematic literature review," *Applied Intelligence*, vol. 55, no. 3, p. 243, 2025.

[11] C. Tang, B. Abbatematteo, J. Hu, R. Chandra, R. Martín-Martín, and P. Stone, "Deep reinforcement learning for robotics: A survey of real-world successes," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 8, 2024.

[12] Y. Li, Y. Ding, S. He, F. Hu, J. Duan, G. Wen, H. Geng, Z. Wu, H. B. Gooi, Y. Zhao *et al.*, "Artificial intelligence-based methods for renewable power system operation," *Nature Reviews Electrical Engineering*, vol. 1, no. 3, pp. 163–179, 2024.

[13] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, Z. Chen, and F. Blaabjerg, "Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of pvs," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4137–4150, 2021.

[14] Y. Mi, C. Lu, C. Li, J. Qiao, J. Shen, and P. Wang, "Data-driven volt-var coordinated scheduling with mobile energy storage system for active distribution network," *IEEE Transactions on Sustainable Energy*, 2024.

[15] Y. Su, J. Teh, S. Fang, Z. Dai, and P. Wang, "Two-stage optimal dispatch framework of active distribution networks with hybrid energy storage systems via deep reinforcement learning and real-time feedback dispatch," *Journal of Energy Storage*, vol. 108, p. 115169, 2025.

[16] S. Jeon, H. T. Nguyen, and D.-H. Choi, "Safety-integrated online deep reinforcement learning for mobile energy storage system scheduling and volt/var control in power distribution networks," *IEEE Access*, vol. 11, pp. 34 440–34 455, 2023.

[17] R. Hossain, M. Gautam, J. Thapa, H. Livani, and M. Benidris, "Deep reinforcement learning assisted co-optimization of volt-var grid service in distribution networks," *Sustainable Energy, Grids and Networks*, vol. 35, p. 101086, 2023.

[18] M. Liu, H.-D. Chiang, and T. Li, "Three-timescale hierarchical multi-objective volt/var control in active distribution network with inverters droop control," *IEEE Access*, 2024.

[19] R. Hossain, M. Gautam, M. MansourLakouraj, H. Livani, and M. Benidris, "Topology-aware reinforcement learning for voltage control: Centralized and decentralized strategies," *IEEE Transactions on Industry Applications*, 2025.

[20] X. Sun, Z. Xu, J. Qiu, H. Liu, H. Wu, and Y. Tao, "Optimal volt/var control for unbalanced distribution networks with human-in-the-loop deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 15, no. 3, pp. 2639–2651, 2023.

[21] H. Kumar, A. Koppel, and A. Ribeiro, "On the sample complexity of actor-critic method for reinforcement learning with function approximation," *Machine Learning*, vol. 112, no. 7, pp. 2433–2467, 2023.

[22] C. Shen, S. Zhu, S. Han, X. Gong, and S. Lü, "Guided deterministic policy optimization with gradient-free policy parameters information," *Expert Systems with Applications*, vol. 231, p. 120693, 2023.

[23] Q. Liu, Y. Guo, L. Deng, H. Liu, D. Li, H. Sun, and W. Huang, "Two-critic deep reinforcement learning for inverter-based volt-var control in active distribution networks," *IEEE Transactions on Sustainable Energy*, 2024.

[24] K. R. Babu and D. K. Khatod, "Smart inverter-based distributed volt/var control for voltage violation mitigation of unbalanced distribution networks," *IEEE Transactions on Power Delivery*, 2024.

[25] Z. Javid, U. Karaagac, I. Kocar, and W. Holderbaum, "Solid-state transformer modelling in power flow calculation," *Energy Reports*, vol. 9, pp. 448–453, 2023.

[26] D. Das, "Optimal placement of capacitors in radial distribution system using a fuzzy-ga method," *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 6-7, pp. 361–367, 2008.

[27] 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, 2020.

[28] Z. Ziya, "Hybrid energy storage system dataset," https://www.kaggle.com/datasets/ziya07/hybrid-energy-storage-dataset, 2023.

[29] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2010.

[30] R. Bolgaryn, E. Prade, G. Banerjee, S. Drauz-Mauel, D. Lohmeier, P. Lytaev, F. Marten, S. Meinecke, M. Vogt, Y. Xiang *et al.*, "Further developments in pandapower and pandapipes," in *2024 Open Source Modelling and Simulation of Energy Systems (OSMSES)*. IEEE, 2024, pp. 1–8.