ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

A Stability Enhancement Control Method for Constitutive Networked Energy Storage Systems Incorporating Kalman Filtering and Reinforcement Learning

Zhibo Li[1], *Feng Peng*[2], *Chunyan Yi*[2] *and Shike Wang*[3,*]

[1] Rundian Energy Science and Technology Co., Ltd, Zhengzhou, Henan, China, 450000
[2] China Resources Power Holdings Co.,Ltd, Shenzhen, Guangdong, China, 518000
[3] China Resources Power Technology Research Insitute Co., Ltd., Dongguan, Guangdong, China, 513808.

SUMMARY: *The limited inertial support and frequency regulation capability of present power networks have become key technical issues that require further improvement. This paper investigates state-of-charge assessment and coordinated regulation for energy-storage converters operating in grid-forming mode. During actual operation, different battery units may show unequal charge levels, and this imbalance can reduce the stabilizing contribution of the whole storage device. To describe the dynamic characteristics of the battery, a two-order resistor-capacitor equivalent model is constructed. The related model coefficients are obtained through recursive least-square estimation with a forgetting mechanism. Based on the identified model, an EKF estimator is developed to predict the charge state of the battery pack. In addition, according to the PCS dynamic equations and regulation features, a coordinated control method is established for the front-end three-phase PWM converter. Its parameters are further tuned by the DDPG-based reinforcement learning algorithm. The experimental results indicate that the proposed EKF estimation approach can reduce SOC prediction deviations. Tests on the IEEE 39-node New England system also confirm that the designed control method is practical and effective. Overall, the combination of battery charge-state estimation and PCS regulation enhances the operational stability of the grid-supporting storage converter system.*

KEYWORDS: *SOC estimation; PCS; EKF; DDPG; energy storage system*

1 Introduction

With the rapid construction of renewable-energy-based power systems, energy storage technologies have become essential for supporting renewable power integration and maintaining stable system operation due to their fast response characteristics and flexible deployment [1-3]. By the end of 2024, China's accumulated installed capacity of new energy storage had reached 48.18 GW, and this figure is expected to rise to 313.9 GW in 2030. The control mode adopted by the energy storage converter serves as the key interface between the battery unit and the power grid, exerting a direct influence on the dynamic response and operational stability of the entire system [4]. According to different synchronization mechanisms, energy storage converter control can generally be divided into two categories: grid-following control and grid-forming control. Among them, grid-forming energy storage systems (GFESSs) have attracted growing attention as an important technical solution for

*harmonynext2025@163.com
https://doi.org/10.65102/is2026492
Published: 30 April, 2026

enhancing the stability of modern power systems [5, 6].

Unlike the grid-following mode, the grid-forming mode regulates the energy storage system to remain synchronized with the utility grid by adjusting active power output. The automatic grid-forming control can determine the amplitude and frequency of the AC-side voltage, while operating as a voltage-source unit to provide voltage and frequency assistance for the grid [7, 8]. Owing to these advantages, grid-forming energy storage systems (GFESSs) are gradually regarded as one of the key approaches for strengthening the stability of contemporary power systems. Many countries have issued guidelines and technical standards that support the deployment and development of grid-forming energy storage technology. For example, one study [9]proposed a constructive grid-type control structure under a phase-locked loop to a solar power energy storage system, which enhances the performance of synchronization and limits the fault currents of battery storage devices when there is a grid-frequency fluctuation. Literature [10] introduced a constructive grid-forming control design based on phase-locked loop for solar power energy storage systems, contributing to improving the synchronization ability and fault current limitation of battery energy storage facilities when grid frequencies deviate. Literature [11] developed a coordinated control and scheduling strategy for GFESS technology, which allows it to provide diverse services including frequency-limited reserve and voltage regulation while maintaining support for stochastic generation, elimination, and photovoltaic generation. Literature [12] developed a system strength enhancement design for GFESS technology to facilitate grid frequency control. Literature [13] developed a new speed-power feedback controlled GF technique called “motor-generator pair”, which has spontaneous inertia and higher response speed for primary frequency regulation, and simulations show that the transient stability of the motor-generator pair is better than that of the conventional GF inverter. Literature [14] determines the stability control region of the proportional-integral based auxiliary controller through Kharitonov theory, which is combined with a Bayesian optimization algorithm to enhance the robust optimal frequency response of the GFESS to support grid frequency regulation. The combination of grid-forming wind turbines with GFESS allows them to act together to coordinate their control, so that the system can be able to resist wind-power deficits and transient load restoration when restoring the power system under power deficit conditions. Literature [15] uses both grid-forming wind turbines and GFESS to provide coordinated control whereby the system will be able to survive every shortage in wind-power generation and transient loading recovery thus guaranteeing stability in power-system restoration when there is a power deficit. Literature [16] suggests that feedback linearization is one way of enhancing power decoupling control of GFESS in wind-power generation systems. This approach improves the accuracy of power regulation by changing the voltage vector reference in grid forming control. GFESS is a voltage source and it demonstrates excellent stability at weak-grid operation conditions but is susceptible to synchronization instability at strong-grid conditions [17]. Moreover, the incorporation of battery energy storage systems directly into medium and high voltage networks increases the complexity of grid-forming control stability due to a reduced electrical distance between a grid and a lower grid impedance, as well as the growing amount of renewable energy being integrated into a network via power electronics converters [18].

Kalman filtering (KF) has been a popular method of estimating the state of a dynamic system with noisy measurement sequences. The paper presented in [19] created an observer based on the Kalman filtering design, which estimated energy states of the capacitor and made changes to the critical control parameters such as damping and inertia of the GF inverter, depending on the results of the estimation. This technique minimized power and voltage overshoot, increased response rate and enhanced voltage stability. Literature [20], included the

switching control model of the GF inverter into the KF filter, enabling more smooth control switching prior to synchronization.

Reinforcement learning (RL), which is one of the primary branches of ML, discovers how to act in the best way by interacting with the environment around it. In previous work Literature [21] , a deep deterministic policy gradient algorithm based on deep RL was presented to control GF inverters and achieve the best possible grid voltage and frequency settings with minimal variation in voltage and frequency in the inverter. Literature [22] investigated a successful control approach of distributing power and regulating voltage in the DC GFESS using deep RL and current-sharing mechanisms of parallel converter blocks to enhance the stability of GFESS with many constant-power loads. Literature [23] developed a pre-training and fine-tuning meta-RL technique that has the ability to react quickly, whereby control parameters of GF inverters in GFESS are changed to restore the frequency of the grid at various disturbances and operating conditions. However, the existing studies on control methods for grid-forming energy storage still appear inadequate, and the stability risks of energy storage systems under low grid-impedance operating scenarios have not been fully resolved [24].

The stability in operation of a grid-forming storage system is closely related to the condition of SOC of every battery cell and the control abilities of the energy storage converter. In this research, the combination of both Kalman filtering and reinforcement learning has been used to determine the state of charge of batteries and enhance the performance of PCS control. To begin with, a two-order RC equivalent model is chosen as the fundamental battery representation and its parameters are determined using a recursive least-square algorithm with a forgetting factor. Then these parameters are used in an Extended Kalman Filter (EKF) to compute the battery charge state. Following that, deep reinforcement learning is used to optimize the control parameters associated with the PCS. Namely, the PI gains in the voltage outer loop are changed to the PCS when the charging takes place, whereas the damping coefficient and the inertia parameter in the power loop are dynamically adjusted when the discharging occurs. Such a design offers some groundwork of enhancing coordinated control performance of the grid-forming energy storage system.

2 Battery state estimation based on Kalman filtering

Battery energy storage systems are able to store electricity and offer bidirectional energy exchange, and they become one of the most efficient technologies to integrate renewable power into the modern grids. Besides enhancing the use-efficiency of power system equipment, the BESSs could help in the regulation of frequency at times of peak load, supply reactive power support, and improve the stability, economics, and flexibility of grid operation. The performance of the storage system may be enhanced even more when several sub-BESS units work together. This is why the precise determination of the battery state of charge (SOC) has become a critical requirement of the reliable energy management system.

SOC reflects the remaining charge level of a battery under a given charge–discharge condition. It is usually expressed as the proportion of available charge relative to the rated full-charge capacity. A higher SOC indicates that more usable energy remains in the battery, while a lower SOC suggests that the battery is closer to depletion. In practical operation, precise SOC estimation helps prevent overcharging and overdischarging, supports balanced control among battery modules, and provides essential state information for subsequent Kalman-filter-based estimation and energy storage control.

2.1 Equivalent Modeling and Parameter Identification of Batteries

2.1.1 Battery model

The equivalent circuit model that is most often applied in engineering practice includes one resistor and two RC networks. It is on this arrangement that both the GNL equivalent circuit model and the Thevenin equivalent circuit model are synthesized, and the resulting second-order RC equivalent circuit model is shown in Fig. 1. In such a model, the capacitor is modeled as an idealized voltage source, which is an idealized version of the open-circuit voltage of the battery. Open-circuit voltage has some mapping relationship with the state of charge, i.e., at different states of charge, the open-circuit voltage changes. Battery internal resistance is the parameter, which describes ohmic losses that occur during charging and discharging. The first RC network, including a resistor and a capacitor in parallel, is employed to explain the diffusion process inside the battery, namely, the slow and gradual stabilization of the discharge voltage at a relatively large time constant, where the voltage across this network is the first polarization component. The second RC network is in turn another resistor and capacitor in parallel and used to describe the redistribution of charges in the electric double-layer region of the cell - a phenomenon wherein the discharge voltage changes rapidly within a relatively small time constant, the voltage across this network being the second polarization component.

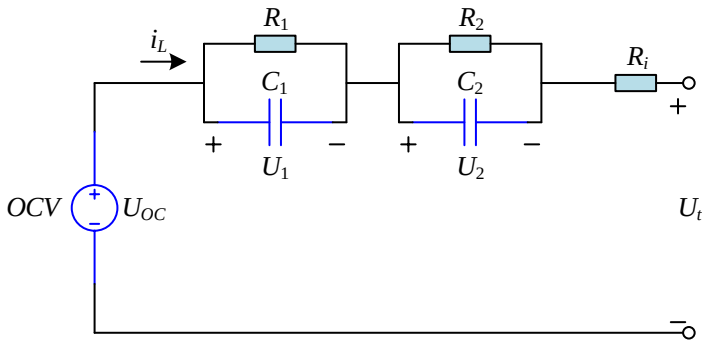

Figure 1: Second order RC equivalent circuit model

2.1.2 Parameter identification

(1) Recursive least squares and better algorithms

Least squares estimation is one of the most widely used algorithms in the field of parameter estimation. Its simplicity of implementation and understanding is evident, and it is able to estimate statistically significant parameters without the need to use any previous observation data or even have any statistical background information. The method is based on a linear system model that describes the connection between system input and system output through a transfer function. Using this transfer function representation, an equivalent differential equation can be obtained, and by setting the unknown system coefficients as the parameters to be estimated, the equation can be rewritten as a concise linear regression model. The process of expanding the system output to higher dimensions and arranging the expressions in matrix form allows constructing a generalized objective function. The principle of least squares states that the best parameter estimates are those which minimize this objective function.

In practical applications, a single computation is insufficient to bring estimated values sufficiently close to the true parameters. For this reason, the recursive least squares (RLS) algorithm is employed for iterative parameter estimation. In each recursive step, the current estimate is updated by incorporating the predicted observed value at the next moment, the actual system output, and a computed gain factor, together yielding a corrected parameter estimate.

The initial estimate can be set to any reasonable value, while the initial gain matrix is typically chosen as a large-valued unit matrix.

However, since the memory length of RLS is theoretically infinite, the progressive accumulation of historical data causes the corrective influence of newer observations to diminish as the number of recursive steps grows. When the total volume of accumulated data becomes sufficiently large, newly acquired measurements can no longer exert a meaningful corrective effect on the estimates, leading to the problem of data saturation. This manifests as an increase in estimation error and a deterioration in parameter identification performance, and the issue is particularly severe in time-varying systems. To overcome this limitation, a forgetting factor is introduced into the RLS framework, giving rise to the Forgetting Factor Recursive Least Squares (FFRLS) method. FFRLS adopts a modified performance metric that discounts the contribution of older data in proportion to the forgetting factor: a smaller forgetting factor enables faster tracking of parameter variations but may also introduce oscillatory behavior, whereas setting the forgetting factor to one reduces the algorithm to the standard recursive least squares method.

(2) Identification of the parameters through the enhanced recursive least squares approach

The battery model is transformed into a form that can be identified to calculate the parameters of the second-order RC equivalent circuit model. Two time constants are introduced, which relate to the respective RC networks, and the model equation is rewritten in a more condensed form. Using these time constants, and a collection of intermediate substitution variables, the model is discretized based on the sampling period, and then after algebraic manipulation a discrete-time linear regression form is derived. It enables the expression to be directly used in the least squares identification setting, wherein a set of composite intermediate parameters is considered as the direct identification objects. Based on these identified composite parameters, the five individual components of the circuit model, namely: the internal resistance, two resistances of the RC branches, and two capacitors of the RC branches, are restored analytically using closed-form algebraic relations, which provide explicit formulas of all the parameters of the circuit model.

The entire process of the online identification of the second-order RC equivalent circuit model parameters is given below. Initially, the system is initiated by defining suitable starting conditions, a constant sampling interval and initial values of the parameters to be estimated through least squares iteration. Secondly, the state of charge of the battery at its present time is calculated with the help of the Anticlockwise Integration Method. Thirdly the voltage across the open circuit is measured by matching the voltage across the open circuit to state of charge with the help of a prior calibrated curve. Fourthly the composite identification parameters are recalculated recursively based on the enhanced FFRLS algorithm. Fifthly, the circuit model parameters of individual circuits are obtained based on the new identification outcomes by the derived analytical expressions. Lastly, the cycle of steps two to five is repeated to achieve the continuous online identification of all equivalent circuit model parameters.

2.2 EKF-based battery SOC estimation

A well-defined battery model and constant updating of the model parameters are also a significant factor that is needed to predict the SOC with certainty. To obtain a good SOC estimation, one has to first ensure that the battery model and the process of identifying its parameters are correctly defined. This section uses the Kalman filter to estimate the SOC of energy storage batteries. The suggested estimation structure is based on the chosen second-order RC equivalent circuit model and the FFRLS-based parameter update algorithm presented in Section 2.1.

2.2.1 Kalman Filter Algorithm

The conventional Kalman filter works on discrete linear systems, which consist of a state equation and an observation equation. In such a representation, the system state variables at every time step are transferred via a state transition matrix and driven by an input control matrix operating on the system inputs, and the process noise is additive. The system output (the observation signal) is produced by a observation matrix and an output control matrix that are used on the state and input respectively and the additive observation noise. The process noise and observation noise are both represented as independent Gaussian white noise processes with zero mean and their covariance matrices are defined as distinct. The initial state of the system is also considered a random variable, with mean vector specifying the predicted initial state and covariance matrix measuring the initial uncertainty. These are all taken to have no correlation among process noise covariance, observation noise covariance, and initial state covariance.

The central concept of Kalman filtering is to update the estimates of the system state variables over time and minimize the mean-square estimation error. The method, using a series of recursive update equations, attempts to find an optimal estimate of the state vectors that reduces the variance between predictions and real measurements as much as possible, based on the available information about the system inputs and measured outputs at each time step. The Kalman filter calculation goes through the next six steps.

First, the predicted state vector of the current time step is calculated based on the last state estimation, utilizing the state transition equation. Secondly, the predicted state error covariance matrix of the current moment is propagated using the optimally filtered covariance of the last step using the state transition and process noise covariance. Thirdly, the Kalman optimal filter gain is calculated based on the predicted covariance, the observation matrix, and the observation noise covariance; this gain controls the relative importance of the prediction over the received measurement. Fourthly, the predicted output variable is constructed by multiplying the observation matrix by the predicted state. Fifthly, the state estimate is refreshed in accordance with the difference between the measured output and the predicted output scaled by the Kalman gain. Sixthly, the state error covariance matrix is refreshed to embody the knowledge obtained by the measurement. The first and second step represent the time update (prediction), and the third to sixth step represents the measurement update (correction).

2.2.2 EKF algorithm

Practically speaking, most physical systems, such as battery electrochemistry, have substantial nonlinearity which precludes using the traditional linear Kalman filter directly. To remedy this, the Extended Kalman Filter (EKF) linearizes the nonlinear system before performing the Kalman filtering steps, and hence is more generally applicable. The basic idea behind EKF is to linearize the nonlinear state transition and measurement functions along a Taylor series expansion around the present state estimate, keeping the first order term and ignoring the higher order terms. It provides an approximation to a locally linear system. A precondition to this linearization is that the two nonlinear functions should be differentiable at all sampling points. The process noise and observation noise can also be viewed as independent Gaussian white noise sequences with their own covariance matrices, in accordance with the linear Kalman filter structure.

Linearization involves replacing the constant state transfer matrix and observation matrix of the standard KF with Jacobian matrices, i.e. matrices of first-order partial derivatives, which are evaluated at the current state estimate of the state transition function and measurement function respectively. The resulting linearised state space equations are subsequently used to perform all the remaining Kalman filtering operations. The EKF routine is similar to the one of the conventional filter: the system is initialized with an initial state mean and error covariance,

time update equations of both the state vector and state error covariance, computation of the Kalman gain based on the Jacobian of the measurement function, state estimate measurement update with the difference between observed and expected measurements, and finally an error covariance measurement update. The main difference with respect to the linear case is that at any given time step, the prediction and the Jacobian matrices are recomputed based on the latest state estimate, thus allowing the local nonlinearity of the system to be taken into consideration.

2.2.3 Implementation of the EKF algorithm in the SOC

When applying EKF on SOC estimation based on the second-order RC equivalent circuit model, the system input is the battery current and the system output is the battery terminal voltage. The state variables are set such that they consist of the battery SOC and the two polarization voltages of the corresponding RC branches. The state variables themselves, the coulomb-counting dynamics of SOC and the first-order RC charging and discharging dynamics of each polarization voltage - are subject to continuous-time state equations which are then discretized with respect to the selected sampling period. Discretized state space matrices are composed of two time constants of the RC networks, both of which are the products of corresponding resistances and capacitances, as well as the sampling period. The observation equation connects the terminal voltage to the open-circuit voltage, the voltage drop due to the internal resistance and to the two polarization voltages.

In order to bring about consistency between the model parameters and the current instantaneous state of the battery, live-time estimation of the model parameters is used in the process of state estimation. The adopted estimation approach is a hybrid estimation algorithm consisting of FFRLS and EKF. The FFRLS algorithm continuously adapts the equivalent circuit model parameters whereas the EKF simultaneously measures the battery SOC based on the most recent model parameters. The sequence of actions is as follows. The initialisation of the parameters and the coefficients of the battery model and the corresponding system state space takes place initially. Secondly, the EKF recursion is used to compute the SOC at the present moment. Thirdly, the pre-calibrated fitting curve that relates OCV to SOC is used to obtain the open-circuit voltage at the current moment. Fourthly, the difference between terminal voltage and open-circuit voltage is considered as an input variable to be provided to FFRLS, and composite identification parameters are updated with an enhanced recursive least squares algorithm. Fifthly, the circuit model parameters are restored based on such identification outcomes based on the analytical inverse relations, followed by the return to the first step, and one cycle of online model parameter update is completed. This feedback loop between the model identification based on FFRLS and the SOC estimation based on EKF makes it possible to adaptively monitor the time-varying properties of the battery during the operation.

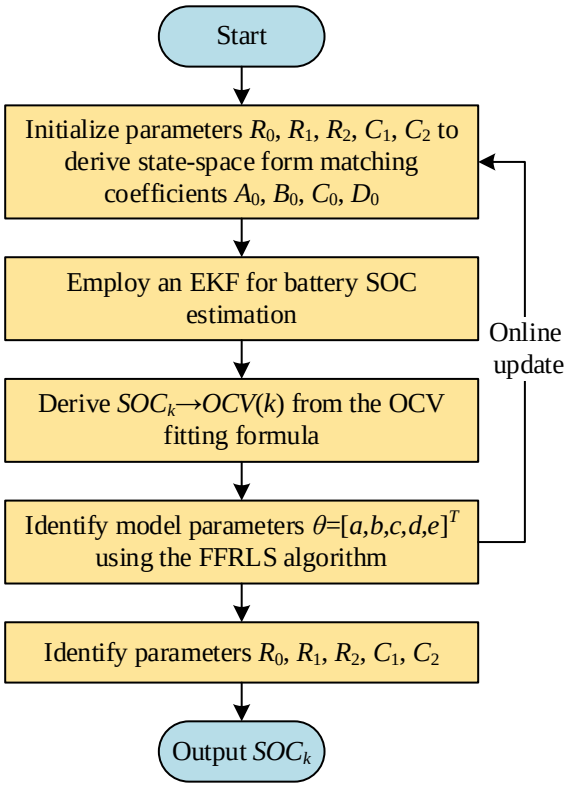

Figure 2: FFRLS+EKF combined estimation process

3 Mathematical model of PCS and its control characteristics

With regard to technology implementation in the grid control field, it implies that such technology may be used to produce an output voltage supply with a certain supporting capacity across the grid connection point to the converter impedance. In fact, this is a power synchronization process, which offers equal inertia support and system strength, complete access to the new energy sources to the network, under both weak and strong grid conditions and grid connected as well as islanded modes. Overall, when power deviations occur on the network, the grid controlled energy storage and synchronous generators will utilize the same power-sharing model and help to stabilize the system frequency and voltage.

Power conversion system (PCS) is the central element of the power electronic converters that play the crucial role of interfaces between the battery energy storage unit and alternating current networks. Due to this role, the PCS plays a significant bridging role in regulating the energy flow between the battery storage system and the external grid. The PCS as a key component of the battery energy storage system has a variety of tasks performed by default. As to the management of charging and discharging, it offers uninterrupted support in the operation of batteries in the grid-connected and islanded modes allowing switching between them without any disruption. The PCS control strategy in relation to power regulation allows the energy storage system to work in all four quadrants of the power plane so that active and reactive power output may be controlled flexibly and independently with the balance between active and reactive power flows within the system maintained.

The current chapter discusses the modeling and control properties of the PCS that will be used to develop a theoretical foundation to the design of the simulation platform of the grid-forming energy storage system which is to be covered in the next chapter.

3.1 Mathematical modeling of PCS

Figure 3 depicts the structural arrangement of the PCS. The main PCS circuit in each modular battery energy storage substation is based on the traditional three-phase bridge voltage source converter topology, where insulated gate bipolar transistors are used as the switching elements because of their advantageous high-power switching properties. On the direct current side, the PWM voltage source converter is connected directly with the battery substation and on the alternating current side there is a series reactor which will help reduce harmonics currents and filter the output waveform. The three-phase symmetrical phase voltages at the common bus point are defined separately per phase in the circuit, and the associated three-phase symmetrical output voltages are determined per phase in the circuit on the alternating current side of the voltage source converter. The switching states of the upper and lower bridge arms in each phase are defined in a complementary binary fashion such that a switch between the upper and lower bridge arms means that the upper bridge arm switch conducts and the lower bridge arm switch opens, and a switch between the lower and upper bridge arm switches implies the opposite conditions.

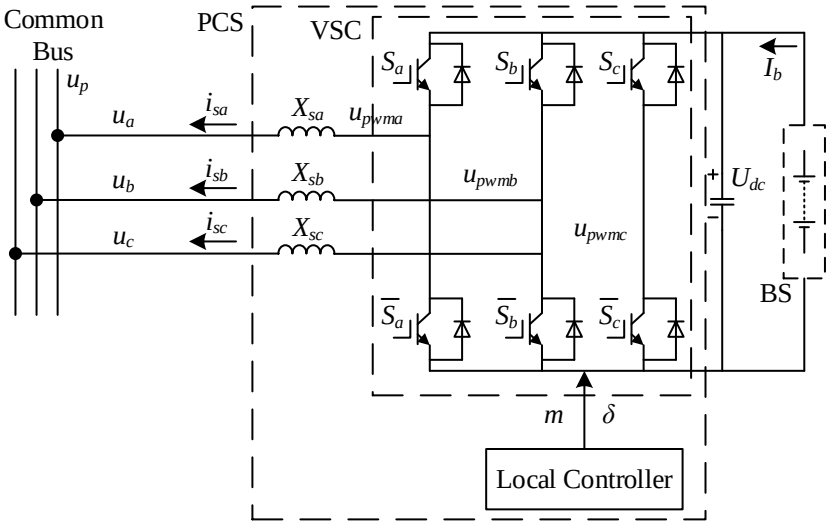

Figure 3: Structure diagram of PCS

PCS allows the converter to cover all four quadrants of the power plane using two coordinated control variables: the phase angle difference between the alternating current side output voltage of the converter and the common bus voltage, and the modulation ratio of the voltage source converter. The joint adjustment of these two variables allows the PCS to provide bi-directional controlled active and reactive powers to the system and act as a generalized impedance converter, and more broadly, as an improved version of a static synchronous compensator. Assuming that the three-phase system is symmetrical and that all power switching devices are considered ideal without losses on conduction or switching times, the voltage relationship ruling the alternating current side of the voltage source converter can be deduced based on the application of Kirchhoff voltage law applied to the reactor branch, which equates the common bus voltage to the combined voltages of the reactor voltage drop and converter output voltage.

$$u_p + u_{x_i} = u_{pwm} \tag{1}$$

where u_{x_i} is the voltage at both ends of the reactor, and its value if the resistance in the input circuit is not considered for the time being:

$$u_{x_i} = L\frac{di_i}{dt} \tag{2}$$

where L is the filter inductance.

For the VSC AC side voltage u_{pwm} can be obtained according to the definition of the switching function:

$$u_{pwm} = u_{dc}\left(S_k - \frac{1}{3}\sum_a^c S_k\right) \tag{3}$$

where u_{dc} is the DC side voltage, and S_k is the switching state of VSC, which is taken as "0" or "1".

The formula (4) and formula (5) were substituted into the formula (2), after finishing the three-phase voltage equation for the AC side of the VSC:

$$u_a = u_{pwma} - L\frac{di_{sa}}{dt} = u_{dc}\left(S_a - \frac{1}{3}(S_a + S_b + S_c)\right) - L\frac{di_{sa}}{dt} \tag{4}$$

$$u_b = u_{pwmb} - L\frac{di_{sb}}{dt} = u_{dc}\left(S_a - \frac{1}{3}(S_a + S_b + S_c)\right) - L\frac{di_{sb}}{dt} \tag{5}$$

$$u_c = u_{pwmc} - L\frac{di_{sc}}{dt} = u_{dc}\left(S_a - \frac{1}{3}(S_a + S_b + S_c)\right) - L\frac{di_{sc}}{dt} \tag{6}$$

3.2 Control Characterization of PCS

To simplify the analysis of PCS control characteristics, this section takes its single-phase operation as an example.

Under steady-state conditions, when the harmonic component of the common-bus voltage is neglected, the following expression can be obtained according to Kirchhoff's voltage law:

$$\dot{U}_p = j\omega X_s \dot{I}_z + \dot{U}_{pwm} \tag{7}$$

and the active power and reactive power exchanged between the PCS and the common bus are respectively:

$$P = \left(\frac{U_{pwm}U_p}{X_s}\cos\delta - \frac{U_p^2}{X_s}\right)\cos\theta + \frac{U_{pwm}U_p}{X_s}\sin\delta\sin\theta \tag{8}$$

$$Q = \left(\frac{U_{pwm}U_p}{X_s}\cos\delta - \frac{U_p^2}{X_s}\right)\sin\theta - \frac{U_{pwm}U_p}{X_s}\sin\delta\cos\theta \tag{9}$$

where δ is the phase angle difference between the voltages and θ is the output impedance angle.

If the output impedance is pure reactance, i.e., $\theta = 90^{\circ}$, it can be obtained from Eqs. (8) and (9):

$$P = \frac{U_p U_{pwm} \sin\delta}{X_s} \tag{10}$$

$$Q = \frac{U_p \left(U_{pwm} \cos\delta - U_p\right)}{X_s} \tag{11}$$

Also, according to the PWM modulation principle can be obtained:

$$U_{pwm} = mU_{dc} \tag{12}$$

where U_{dc} is the PCS dc bus voltage.

If let:

$$r_{ic} = \frac{U_{pwm}}{U_p} \tag{13}$$

This can be obtained by substituting equation (13) into (10) and (11), respectively:

$$P = U_p^2 \frac{r_{ic}}{X_s} \sin\delta \tag{14}$$

$$Q = U_p^2 \frac{r_{ic} \cos\delta - 1}{X_s} \tag{15}$$

From Eq. (14) and Eq. (15), it can be seen that the PCS can be equated to a shunt impedance device connected to the common bus, and its corresponding equivalent impedance is:

$$R_{eq} = \frac{X_s}{k_1} \tag{16}$$

$$X_{eq} = \frac{X_s}{1 - k_2} \tag{17}$$

where $k_1 = r_{ic} \sin\delta$, and $k_2 = r_{ic} \cos\delta$.

Thereby, and also from Eqs. (16) and (17), the magnitude of the equivalent impedance of the PCS connected to the common bus is determined by r_{ic} and δ: for $\delta = 0^{\circ}$, the PCS transmits only reactive power and is equivalent to a purely capacitive reactance when $r_{ic} < 1$, whereas when $r_{ic} > 1$, the PCS is equivalent to pure inductive reactance. For $\delta \neq 0^{\circ}$, the PCS transmits only active power regardless of the value of r_{ic}, and when $\delta < 0^{\circ}$, the PCS is equivalent to a resistive load that absorbs active power, whereas when $\delta > 0^{\circ}$, the PCS is

equivalent to a power source that emits active power outward. In particular, when $\delta = 0^{\circ}$ and $r_{ic} = 1$, the PCS does not transmit any active and reactive power.

To sum up, both the size of the converter alternating current side output voltage and the angle between this voltage and the common bus voltage are factors that define the size and direction of active and reactive power supplied by the PCS to the grid. Moreover, the modulation ratio of the voltage source converter directly determines the amplitude of the converter output voltage. Therefore, independent control of the modulation ratio and the phase angle difference achieves total decoupled control of the active and reactive current component of the PCS output. Exactly such a mechanism of two degrees of freedom allows the PCS to work in a flexible manner in all four quadrants of the power plane with the ability to inject and absorb power in any of the active or reactive dimensions depending on the requirements of the grid conditions.

4 Research on control parameter optimization based on deep reinforcement learning

4.1 Exploring Deep Reinforcement Learning Algorithms

4.1.1 Enhanced learning

Reinforcement learning (RL), also known as augmented learning, is a type of problem and its solution methods in which the learning process occurs over time due to interaction with the environment. Reinforcement learning is one of the key elements of intelligent control, and it is different than supervised learning because it does not need the description of a right policy as supervisory data, nor does it try to discover the local optimum. It rather concentrates on balancing exploration of unknown areas and decisions made using known data. The reinforcement learning has two major parts; the agent and the environment.

4.1.2 Deep Q-networks

The Q-learning technique is grounded in temporal difference learning and belongs to the category of value-based reinforcement learning methods. It constructs either a linear or non-linear approximation of the value function, refines this approximation iteratively, and ultimately improves the agent's decision-making policy. The central quantity in Q-learning is the optimal state-action value function, which represents the maximum expected cumulative discounted reward obtainable from a given state by taking a given action and subsequently following the optimal policy. This optimal value function is defined recursively: the value of a state-action pair equals the immediate reward received plus the discounted maximum value achievable from the resulting next state, with the expectation taken over all possible transitions.

Although Q-learning is a well-established and widely applied algorithm in reinforcement learning, it is inherently limited to environments with discrete and finite state and action spaces, as it can only maintain and update value estimates for previously encountered states. To overcome this limitation and extend applicability to continuous state and action spaces, a parameterized approximation function is introduced to represent the value function. Rather than storing an explicit table of values, this approach uses a function parameterized by a set of learnable weights to approximate the true optimal value function, a technique commonly referred to as value function approximation. The approximating function takes the state and action as vector inputs and produces an estimated value as output, with the model parameters being updated through gradient-based optimization to minimize the approximation error.

The deep Q-network extends this approximation framework by employing a deep convolutional neural network as the function approximator to fit the state-action value function, leveraging the network's capacity to extract high-dimensional features from raw inputs. Two key innovations are further introduced to stabilize and improve the training process:

(1) Sample pool: a dataset consisting of samples recently collected by the intelligentsia, which is used to train the network, solving the correlation and non-stationary distribution problems.

(2) The target network uses an update algorithm that operates on a slower pace so that the Q-value can be delivered independently from the existing Q-network. This separation makes training more stable and efficient. In addition, the TD component will stay within the range of [−1, 1].

4.1.3 Deterministic Policy Gradient Algorithm

Policy Gradient (PG) is a gradient-based reinforcement learning method. Parameterize the policy π as π_θ, input the environment state s to select the action a, and then optimize the parameter θ by gradient ascent to gradually get the optimal policy. As in equation (18):

$$\pi_\theta(s,a) = \mathbb{P}[a \mid s,\theta] \tag{18}$$

Define $J(\theta)$ as the overall performance metric. The strategy gradient algorithm allows the objective function $J(\theta)$ to rise along the gradient to a local maximum: $\Delta\theta = \alpha\nabla_\theta J(\theta)$, and the strategy gradient is: $\nabla_\theta J(\theta) = E_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(s,a) Q^{\pi_\theta}(s,a)\right]$

The PG learning process requires integrating over the entire action state space when calculating the strategy gradient at each step:

$$\nabla_\theta = \int_S \int_A \rho(s)\pi_\theta(a \mid s) Q^{\pi}(s,a) da ds \tag{19}$$

PG learning has good convergence, and can improve in a good direction little by little during the learning process, which increases the probability of positive incentives, but later on, there will be a situation in which the optimal value function has been optimal value function continues to be small oscillations and does not converge.

Stochastic Policy Gradient (SPG) approach calculates the gradient of expected cumulative reward with respect to policy parameters by integration over the whole state and action space. In particular, the policy gradient is calculated as the expected value of the product of the gradient of the log-probability of choosing a specific action in accordance with the current policy and the optimal state-action value of that action, where the expectation is taken across states and actions sampled according to the stationary state distribution and stochastic policy respectively. This formulation demonstrates that the direction of the gradient ascent is dependent on the strength of preference of each action expressed by the policy multiplied by the importance of that action in the given state, which steers the policy parameters to choices that produce greater cumulative returns.

The Deterministic Policy Gradient (DPG) approach is a generalization of this framework to the situation when the policy is deterministic instead of stochastic, with each given state being mapped straight into a particular action without sampling. On such a setup, the policy gradient is calculated by integrating on the state space only, considering the stationary state distribution caused by the deterministic policy. The gradient of the performance objective with regard to the policy parameters is written as the expectation, across states experienced under the present

policy, of the product of the gradient of the deterministic action selection with respect to the policy parameters and the gradient of the state-action value function with regard to the action, assessed at the action recommended by the deterministic policy to that state. This chain-rule decomposition means that the DPG is especially well adapted to continuous action spaces because it directly passes value gradient information backward through the policy without the need to integrate over an action distribution.

That is, the behavior at each step is directly obtained as a deterministic value through the function μ: $a_t = \mu(s_t \mid \theta^{\mu})$, the μ in the function is the optimal behavioral strategy, and it is no longer a stochastic strategy that needs to be sampled. The value function of the deterministic strategy is related to the deterministic strategy, because DPG is a deterministic strategy, and noise can be added to the deterministic action as a stochastic strategy. The DPG algorithm is then combined with the Actor-Critic framework and a Q-learning function learning method is applied to obtain a deterministic optimal behavioral policy function after several training sessions.

Actor-Critic algorithm is one of the significant reinforcement learning methods. It is part of the temporal-difference framework and integrates a value-function-based strategy with a policy-function-based strategy.

At each update, the Actor picks an action a based on the state s of the environment and the policy $\pi_{\theta}(s,a)$. An action leads to a transition into state s', accompanied by an instantaneous reward r issued by the environment. The Critic tunes its evaluation function using the reward received from the environment together with the previous evaluation, which ensures that its evaluations become more aligned with the true reward received from the environment. Next, the Actor updates the policy π_{θ} based on the Critic's evaluation for better decisions in the future. At first, during the training process, the action choice by the Actor is probabilistic, and so is the evaluation by the Critic. Over time, however, the evaluation of the Critic becomes more precise, and the action decision-making of the Actor becomes better.

4.1.4 Deep deterministic policy gradient algorithm

The Deep Deterministic Policy Gradient (DDPG) is a deep reinforcement learning algorithm that can be seen as a control-based approach based on data. In this approach, the mathematical model of the system is discovered with the help of input-output data, and the optimal control of the system is performed in accordance with the predetermined reward function.

Conversely, DDPG does not have the ability to use the Q-learning approach as it is hard to apply the greedy strategy in continuous action space. The limitation is eliminated by considering an Actor-Critic structure with deterministic action policies. In comparison to other approaches based solely on value functions, the Actor-Critic deep policy gradient approach employed in this paper has a higher convergence rate.

DDPG uses the deterministic policy gradient architecture as part of its Actor-Critic structure. The Actor network, which has a single group of learnable weights, indicates a deterministic policy which maps states to actions directly. The Critic network, which has another set of weights, estimates the state-action value function. The goal of the Actor is to maximize the total expected cumulative reward discount. This objective function is maximized through stochastic gradient descent wherein the gradient with respect to the Actor parameters is calculated through the chain rule process (the gradient of the value function with respect to the selected action is multiplied by the gradient of the deterministic policy with respect to the Actor parameters), and the expectation is taken over the states drawn using the replay buffer. The Critic network is trained the same way as in DQN, where the mean-squared temporal difference error between the estimated Q-value and a target value, which is represented by the immediate reward and the

discounted Q-value of the subsequent state, estimated with the target policy and target value networks, is minimized.

DDPG addresses the problem of premature convergence of the policy to local optima in continuous action space by introducing time-correlated noise process to the action selection, which is beneficial to explore more effectively with inertia. On the whole, the DDPG algorithm consists of an action network that approximates the optimal policy and a critic network that approximates the optimal value function, where experience replay, target network freezing, and adaptive noise adjustment are used to achieve training stability. To begin with, the application of DDPG to system control increases the rate at which the convergence of the neural network occurs and enhances real-time control performance over the fuzzy neural network. Secondly, the control method has significant strong dynamic response to the external disturbance. Lastly, regardless of changes in internal parameters, this control method remains able to provide fast dynamic response and minimal steady state error.

4.2 Control parameter optimization based on DDPG algorithm

4.2.1 DDPG Adaptive Controller Control Framework

The energy storage converter has the deep reinforcement learning approach applied on it. When it comes to battery charging, through this approach, the PI control parameters of the outer loop of DC bus voltage are adjusted in order to provide adaptive control of disturbances. Thus, when the disturbances happen in the DC domain, the energy storage converter can stabilize the DC bus voltage by regulating the PI parameters automatically. During the discharging process of batteries, the three-phase PWM converter works as an inverter. Deep reinforcement learning is also used to adaptively control the virtual rotational inertia J and the damping factor D of the power loop in the inversion process. The control framework of the adaptive controller based on the DDPG algorithm is shown in Fig. 4. Where a_t is the deep reinforcement learning output action, when the battery is in charging mode, the action a_t is the K_p, K_i parameter of the outer loop of the DC bus voltage. When the battery is working in discharge mode, the action a_t is the virtual moment of inertia J, damping coefficient D of the power ring.

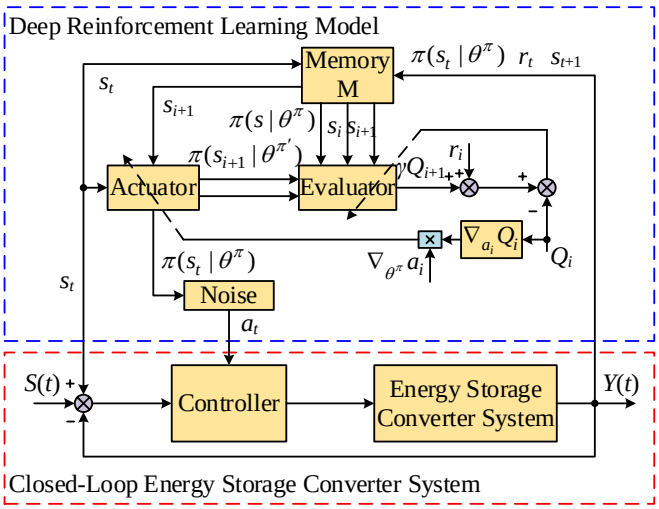

Figure 4: DDPG adaptive controller control framework

4.2.2 Optimization of Voltage Response Characteristic Parameters Based on DDPG Algorithm

When the battery is charging, the three-phase PWM converter that is working in front mostly acts as a rectifier. The main purpose behind it is to provide a steady DC-bus voltage supply to the subsequent DC-DC converter stage. In this instance, deep reinforcement learning is applied to conduct adaptive tuning of PI parameters on the outer DC-bus voltage loop control system. Optimal steady-state and dynamic properties of DC-bus voltage are obtained through adaptive PI parameters tuning which is obtained by the use of DC-bus input voltage and its error. To be more precise, the deep reinforcement learning method gives online PI tuning of the outer DC-bus voltage loop with inputs of input voltage and voltage error. This approach is illustrated in Figure 5.

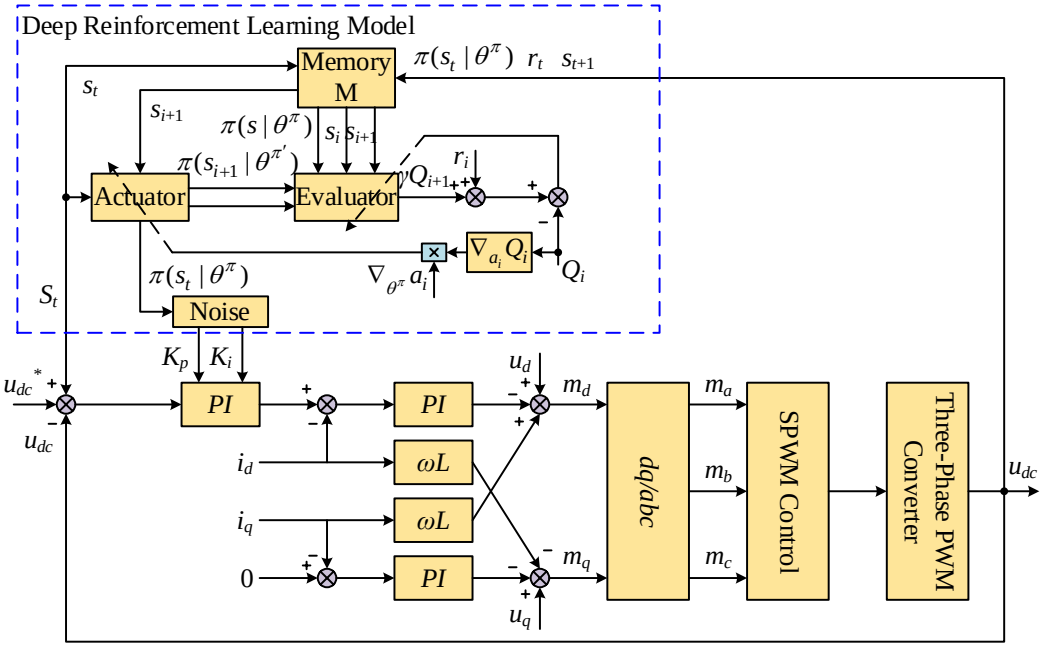

Figure 5: The battery charging process DDPG adaptive controller

The deep reinforcement learning model outputs K_p and K_i parameters to the DC bus voltage outer loop of the front-stage three-phase PWM converter in the energy storage converter system, and measures the DC bus u_{dc} and DC bus voltage deviation Δu_{dc} output under the control of this PI parameter to measure that the Actor module in the deep reinforcement learning model gives K_p, K_i parameters are good or bad, and then adjust the realistic strategy network and realistic value network. Then, a new set of K_p, K_i parameters is output based on the current DC bus voltage deviation Δu_{dc}.

In order to ensure that the designed DDPG adaptive controller converges faster and has better results, the designed reward function is shown in Eq. (20):

$$r_t = -1 + e^{-\frac{10(u_{dc}-u_{dc}^*)^2}{\delta^2}} \tag{20}$$

where δ is the formal parameter of the Gaussian function.

4.2.3 Optimization of frequency response characteristic parameters based on DDPG algorithm

The front-stage three-phase PWM converter switches over to inverter operation when the battery energy storage converter is running in discharge mode under the control scheme of the VSG. Synthesis of three-phase AC output voltage with respect to DC-bus voltage will keep the voltage and frequency stable during weak-grid conditions. The frequency response of the converter system tends to attenuate and oscillate depending on the variation of active power and its dynamic properties are influenced by two critical VSG parameters namely, the virtual inertia constant and the damping coefficient. Deep reinforcement learning is integrated into the VSG control loop to allow adaptive online adjustment of these two parameters in changing grid conditions.

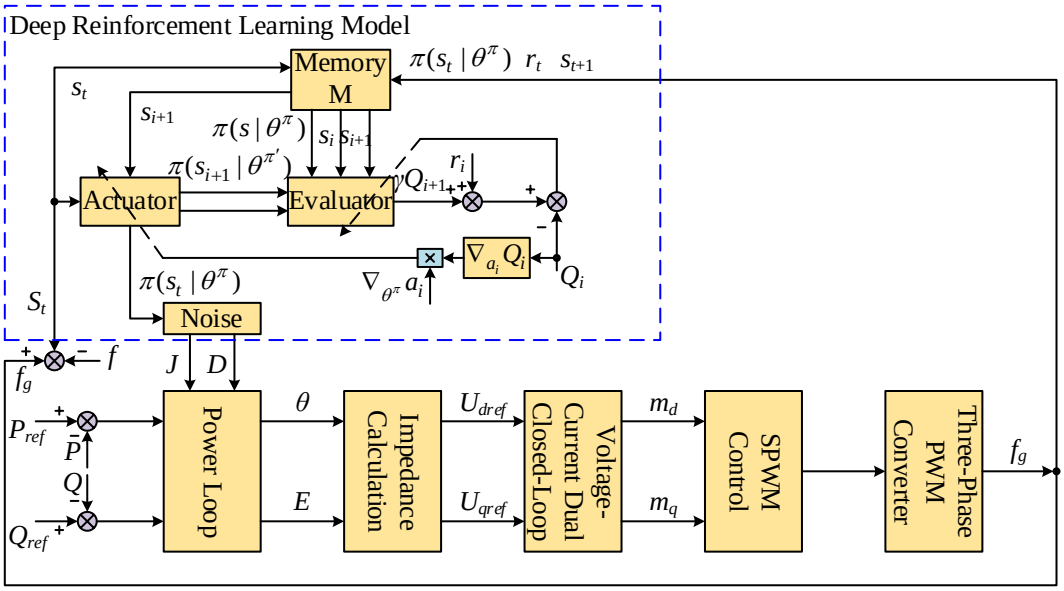

Figure 6: The battery discharging process DDPG adaptive controller

The DDPG algorithm uses the power-loop parameters J and D of the front-stage three-phase PWM converter as its output actions in frequency-response optimization. The performance of the output action is measured through the sampling of the frequency f of the parameter group corresponding to it. Subsequently, the Actor and Critic networks of the DDPG controller are trained based on the reward function.

The reward function is designed to suppress frequency fluctuation when the active output power of the energy storage converter system changes. On one hand, PCS output frequency should maintain good dynamic performance. On the other hand, excessively large or small values of J and D in the steady state should be avoided, because they may damage PCS stability. The reward function is therefore given in Eq. (21):

$$r_t = -1 + e^{-\frac{10(f_g - f)^2}{\sigma^2}} \tag{21}$$

Based on the fact that a very high value of J destabilizes the system dynamics and a very low value of J can lead to visible oscillatory behaviour in frequency responses as the system output power fluctuates, then both situations are not favorable to the weak-grid stability. Hence, J and D are adaptively tuned by deep reinforcement learning alone in case of a significant power

deviation. With the stabilization of the power, J and D revert to their nominal values.

5 Experimental analysis

5.1 Experimental analysis of battery state estimation

The chosen experimental object in this paper is a 3.5 V, 12 Ah lithium iron phosphate battery. The two tests to be used in validations are a constant current pulse discharge test and a dynamic operating condition test. In order to test the feasibility of the battery-state estimation scheme, this paper will compare the EKF algorithm proposed in this paper with the ampere-time integration method.

5.1.1 Constant Current Pulse Experiment

The ambient temperature is set to the standard value of 25.20 C and the fully charged lithium-ion battery is discharged at a steady rate of 3A. Once the experiment has been done, the battery is left to rest during 6.5 h. Within this resting interval, the recorded open circuit voltage is 3.5 V. Using the open-circuit voltage approach, the real SOC is determined as 53.25 percent. The Figure 7 shows the performance comparison of both EKF and ATI algorithms in estimating the SOC. One can see the almost overlapping estimation curves of the two algorithms at the beginning. Nevertheless, the discrepancy between the two starts to emerge over time. The end SOC derived using the ampere-time integration approach is 50.62 and the error is 2.63. Conversely, the EKF algorithm produces a final SOC of 52.79 and the error is only 0.46. The implication of this finding is that EKF results in more accurate estimates of the SOC.

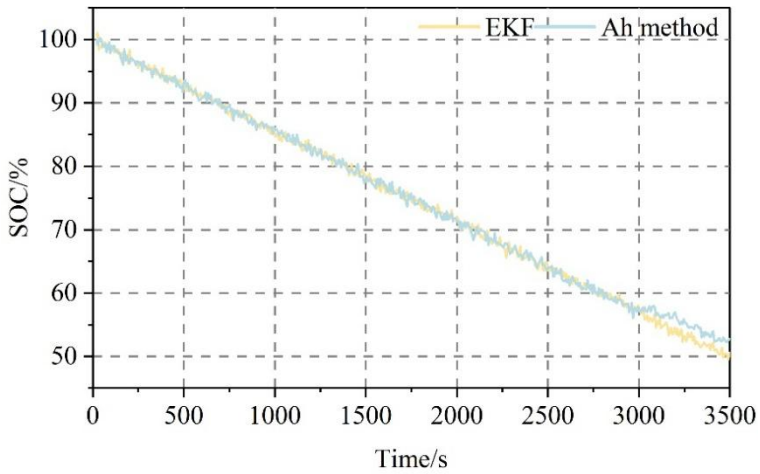

Figure 7: Constant current pulse experiment results

5.1.2 Experimental conditions

The experiment is also conducted in a thermostat at 25.2degC. In Fig. 8, the measured current and voltage curves are shown and in Fig. 8(a), the current curve and in Fig. 8(b), the voltage curve is shown. The battery current ranges between -25 A and 40 A and the voltage between 2.8 Volts and 3.5 Volts.

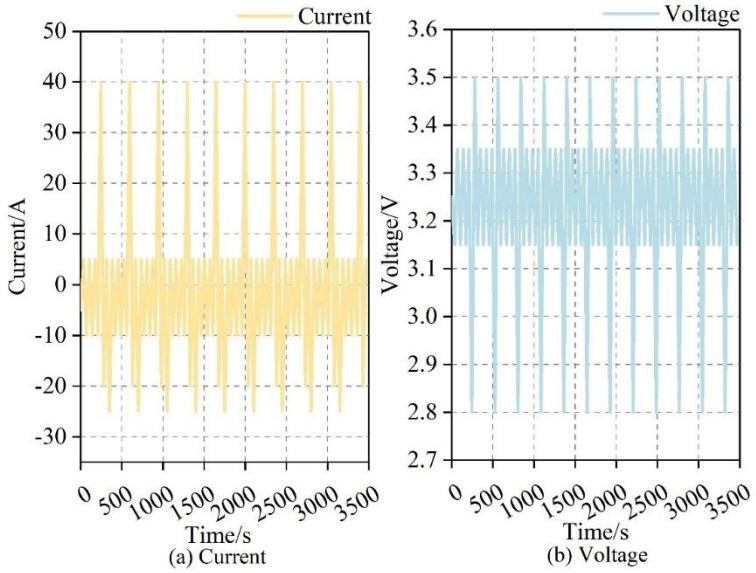

Figure 8: Current and voltage curves

Figure 9 is a comparative analysis of the EKF algorithm and the ampere-time integration method. After the experiment, the battery was left to achieve a completely rested condition and the open-circuit voltage was recorded as 3.3V. With the open-circuit voltage lookup approach being used as a reference, the real SOC was found to be 29.81 percent. The ampere-time integration method gave a SOC of 22.99 percent, which has a significant error of 6.82 percent. Conversely, the EKF algorithm estimated the SOC at 29.21% with a margin of error of just 0.6% in similar circumstances, which shows much better estimation precision.

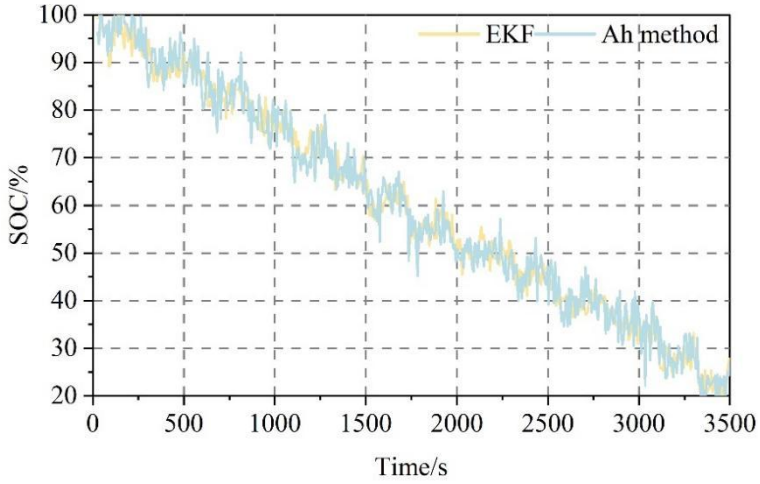

Figure 9: Operating condition experiment results

Based on the aforementioned two experiments, the EKF algorithm is more effective than the ampere-time integration approach. The ampere-time integration approach has increasing error over time, which makes it unsuitable to be used in long term estimates. On the other hand, EKF indicates high correction strength and the deviation to the actual value is less than 1%.

5.2 Experimental analysis of parameter optimization control

5.2.1 Simulation Case Construction

To model a power system using low-inertia properties post-massive integration of new energy, this section will use the modified New England 39-node arithmetic as a validation platform. The structure of the platform is depicted in Fig. 10. On this platform, generating units G1-G10 are placed at low-inertia synchronous machines and nodes 27-29 are chosen as nodes that are connected to energy storage devices that can supply virtual inertia. It is assumed that each energy storage converter has both grid-forming and grid-following control modes. The distribution of inertia across nodes is optimized, and the resulting equivalent inertia coefficients are used as simulation parameters. By time-domain simulation, the properties of synchronous generator nodes, including frequency and active power at various instances following disturbance to the system are confirmed.

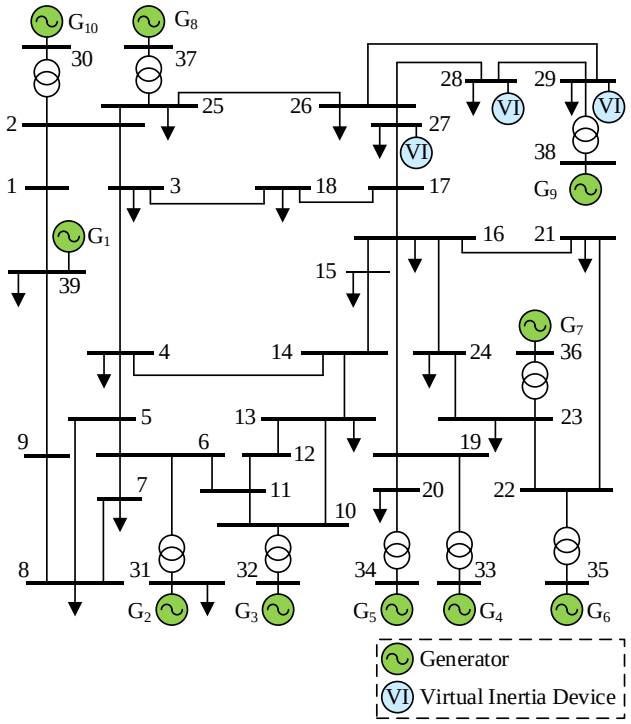

Figure 10: A modified example of 39 nodes in New England

In order to model a system with low-inertia power and large-scale access to new energy, the inertia of equivalent post-generators is set to the values given in Table 1. The inertia values of the 10 generators are between 0.8 and 1.25. The damping coefficients of the post-equivalent generators and accessed inertial devices are configured to $22\,p.u.$.

Table 1: The equivalent inertia of the generator node in the calculation case

Dynamo	Inertia value	Dynamo	Inertia value
1	1	6	1.08
2	1.25	7	1.11
3	1.1	8	0.8
4	1.05	9	0.96
5	0.85	10	0.98

The disturbances in the system like a rapid reduction in renewable energy generation or a

rapid increase in load can be modelled over time using a step disturbance of net active load. It is possible to treat a sudden load drop along with the associated frequency rise in the same manner. In the simulation example, the net active power of node 27 is increased by 57 MW at t=1.2s and the corresponding node disturbance intensity is shown in Fig. 11. Raw disturbance intensities of most of the nodes are about 0.25.

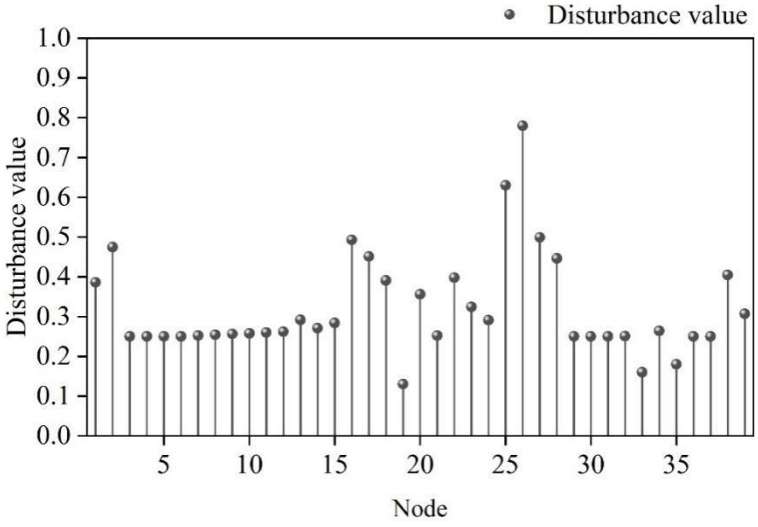

Figure 11: The original disturbance strength of each node in the calculation case

5.2.2 Example simulation results

Revised 39-node New England time-domain simulation case is created in MATLAB/Simulink. Validation of the suggested method of parameter optimization configuration using the DDPG algorithm involves these two steps:

(1) The proposed algorithm is written as an optimization program in MATLAB, and the results of energy storage inertia configuration using the following network type and the structural network type control are calculated according to the needs of the example validation.

(2) According to the results of inertia configuration under different controls, time-domain simulation is carried out in Simulink to obtain the simulation curves of key variables, such as frequency and energy storage active output.

In order to compare and analyze the characteristics and differences of the virtual inertia provided by the grid-constructing type and grid-following type energy storage devices, the corresponding energy storage systems are set up as grid-constructing type and grid-following type control at the nodes accessing the virtual inertia devices, and then the corresponding optimal inertia configuration scheme is determined according to step 1. The total installed inertia values of the network-following type and network-constructing type are 4.955 and 4.625, respectively.In the validation step 1, it can be concluded intuitively that the overall virtual inertia is smaller when the network-constructing type is used compared with the inertia realization using the network-following type control, in other words, in terms of the total configured inertia, the network-constructing type control of the energy storage system is more advantageous.

By plotting the eigenvalue spectra of the state matrices under the grid-configuration device and the grid-following control, it is possible to demonstrate the different impacts of the two different inertia realizations and their corresponding configuration results on the system stability, and a comparison of the eigenspectra of the system matrices for the grid-following/grid-configuration inertia configurations is shown in Fig. 12. The narrower conical mouth of the state matrix eigen-spectrum corresponding to the inertia realization of the

constructed-network type control indicates that: when providing additional inertia support, the constructed-network type control realization has a better damping ratio compared to the following-network type realization, which can shorten the transient process time of the system and benefit the system stability.

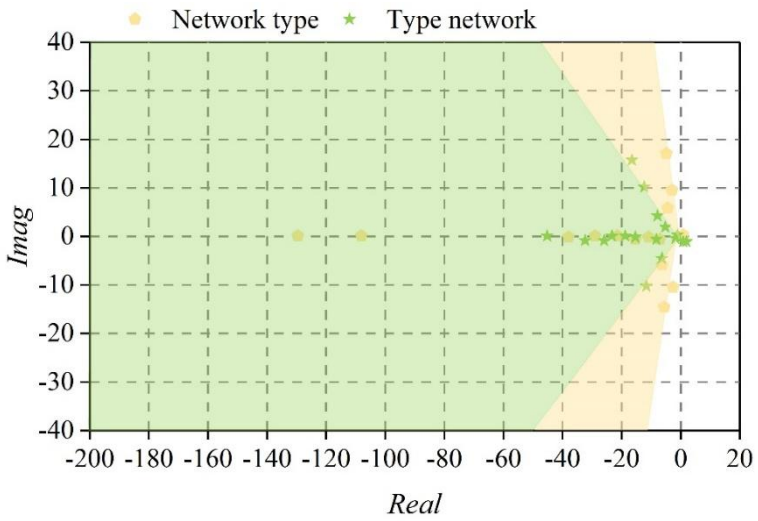

Figure 12: The system matrix feature spectrum is compared

Based on the outcomes of energy-storage virtual inertia configuration in Step 1, the time-domain simulation in Step 2 is performed. A positive step disturbance of 57 MW is imposed on node 27 at t=1.1s which provides the frequency curves of all generator nodes. Taking into account the effect of generator inertia values and equivalent disturbances, it is shown that only the time-domain curves of nodes 36-38, i.e. those corresponding to generators G7-G9, are presented to simplify the graph lines and make them more legible.

In the frequency domain, the results of the simulated grid-following and grid-forming inertia configurations are presented in Figure 13. Figure 13(a), 13(b) and 13(c) depict the simulation outcomes of the initial low-inertia system, the grid-following inertia system and the grid-forming inertia system respectively. It is shown by the results that the extra virtual inertia device may enhance the transient behavior of the system. Further, in terms of frequency deviation, the frequency nadir with grid-forming inertia control is smaller compared to the frequency nadir with grid-following inertia control. Hence, the frequency response is better.

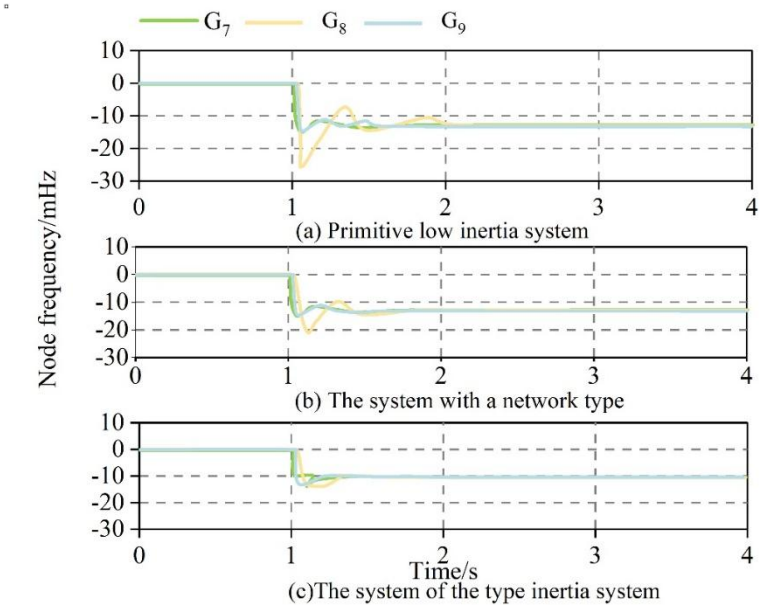

Figure 13: Frequency time domain simulation result

In order to make additional comparisons regarding the effects of various virtual inertia implementation (VI1 to VI3) on the energy storage system, this paper analyzes the active-power output profiles of energy storage devices with both grid-following and grid-forming control as well as the associated generator active-power responses at different operating conditions.

Figure 14 shows the results of active-power time-domain simulation. The subplots (a) and (b) represent the active-power output of the energy storage device under grid-following and grid-forming control respectively. It is worth noting that the maximum active power during grid-forming inertia was found to be about half that observed under grid-following control. It means that the implementation of grid-forming inertia puts significantly more load on the converter capability especially in terms of transient power management of the energy storage system.

The respective impact on generator active-power characteristics with each inertia implementation is illustrated in Figure 15. The subplots (a) and (b) represent the grid-following and grid-forming simulation outcomes respectively. Even though the maximum power of response of the virtual-inertia energy storage system is still quite small in grid-following mode, the resultant frequency-response load is passed across to the generator, which leads to increased active-power stress on it. Inversely, the inertia support offered by the energy storage system under grid-forming control reduces generator stress associated with system transient conditions more efficiently.

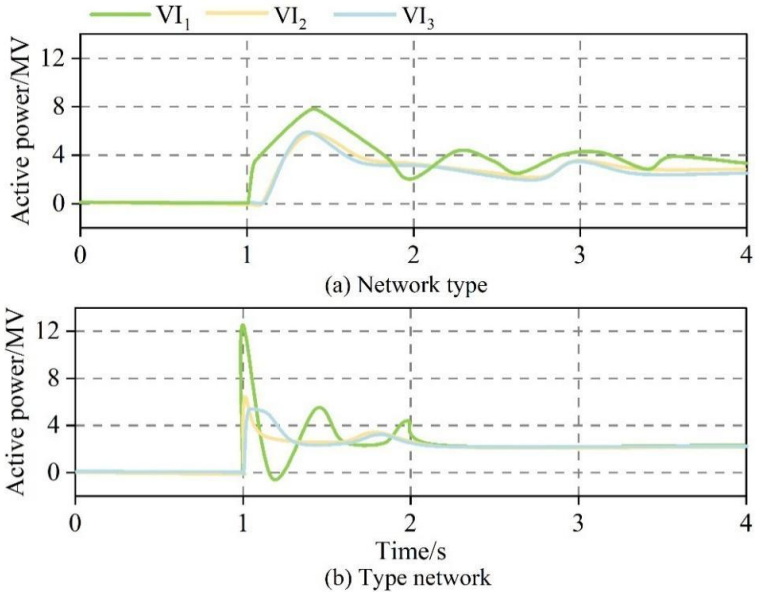

Figure 14: The simulation results of the power domain simulation

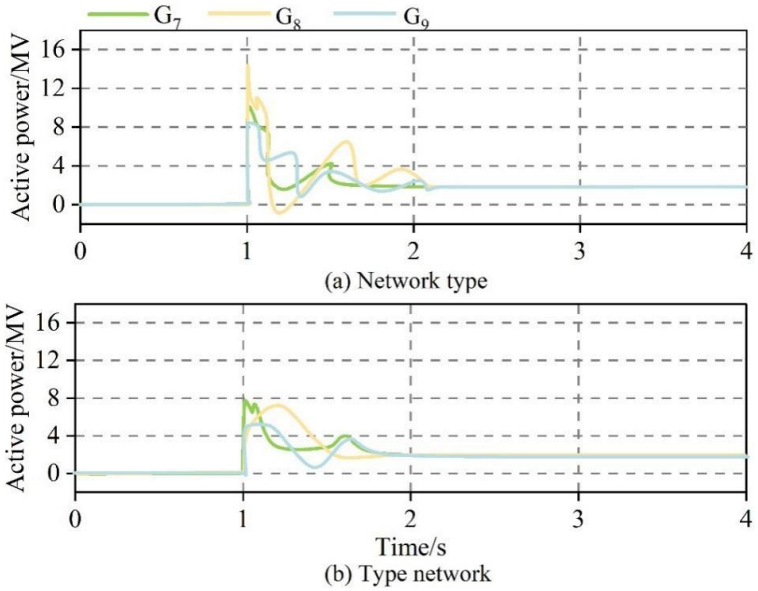

Figure 15: The effect of the power curve

In summary, the strategy proposed in this paper for optimizing the control variables of the three-phase PWM converter at the PCS front stage based on the DDPG algorithm can be more effectively simulated on the energy storage system constructed around the grid-forming mode than on the grid-following one. The proposed strategy can also contribute to improving system stability.

6 Conclusion

The present paper offers a comprehensive framework that integrates SOC estimation and optimal PCS control to improve the stability of grid-forming energy storage systems, where the possibility of the two parts has been confirmed by experimental analysis.

The findings on constant current-pulse discharge and dynamic operating condition

experiments prove that the Extended Kalman Filter is much better compared to other methods in battery state-of-charge estimation. The estimation error is small and the difference between the actual and estimated values is always less than one percentage point, which meets the practical engineering demands.

In order to verify the suggested control strategy, a modified New England 39-node system model is developed to perform the time-domain simulation and comparative study. The findings show that grid-forming inertia realization has better transient performance than grid-following control and can give practical recommendations to grid planners when evaluating the inertia contribution of energy storage devices to the grid during disturbances.

Since SOC estimation and PCS optimal control are coupled by nature, they complement each other to create a mutually enhancing structure that contributes significantly to the stability of grid-forming energy storage systems as a whole.

Funding

This research was supported by the 2023 Key R&D Program of China Resources Power, Research and Demonstration Application Project of Active Support Technology for Grid-Forming Energy Storage (CRP-R&D-LX-2023-025).

About the Author

Li Zhibo, Senior Engineer, holds a Master's degree in Engineering from Wuhan University of Water Resources and Electric Power. He has been engaged in research on relay protection testing in power systems, as well as supervision of thermal power and new energy technologies for a long time. He has won two provincial and ministerial level scientific and technological progress awards, six national patents, published 10 papers, and participated in the compilation of two industry standards.

Feng Peng, 1976.05, male, graduated from Huazhong University of Science and Technology,the General Manager of the East China Region of CR Power. He directed the development of the Hubei Intelligent thermal Power Plant, led the technical verification of the China Resources Xinjiang grid-forming energy storage project, and oversaw the implementation of the integrated operations management system, and presided over the technical and economic demonstrations of major technical transformations in thermal power and new energy businesses.

Chunyan Yi ,1989.02, female, graduated from Huazhong University of Science and Technology ,Senior Technical Management Manager and a member of the Electrical Committee of CR Power. She participated in the review of over 60 thermal power and new energy infrastructure and technical renovation projects, is responsible for the technical demonstration and test management of the Xinjiang grid-forming project, and led the formulation of multiple corporate energy storage technical standards.

Shike Wang received the B.S. and Ph.D. degrees in electrical engineering from Xi’an Jiao Tong University (XJTU), Xi’an, China, in 2014 and 2020, respectively. She then joined China Resources Power Technology Research Institute Co., Ltd. as a R&D engineer in energy storage and renewable energy generation. Her research interests include power grid and source coordination, and small-signal stability of power electronics systems. Dr. Wang received the First Prize Paper Award in IEEE Transactions on Power Electronics in 2021, Science and Technology Progress Special Grade Award of China Power Supply Society in 2023. She is selected to Young Elite Scientists Sponsorship Program by China Association for Science and

Technology.

References

[1] Padmanabhan, N., Ahmed, M., & Bhattacharya, K. (2019). Battery energy storage systems in energy and reserve markets. IEEE Transactions on Power Systems, 35(1), 215-226.

[2] Nadeem, F., Hussain, S. S., Tiwari, P. K., Goswami, A. K., & Ustun, T. S. (2018). Comparative review of energy storage systems, their roles, and impacts on future power systems. IEEE access, 7, 4555-4585.

[3] Datta, U., Kalam, A., & Shi, J. (2021). A review of key functionalities of battery energy storage system in renewable energy integrated power systems. Energy Storage, 3(5), e224.

[4] Dehghani-Sanij, A. R., Tharumalingam, E., Dusseault, M. B., & Fraser, R. (2019). Study of energy storage systems and environmental challenges of batteries. Renewable and Sustainable Energy Reviews, 104, 192-208.

[5] Rana, M. J., & Abido, M. A. (2017). Energy management in DC microgrid with energy storage and model predictive controlled AC–DC converter. IET Generation, Transmission & Distribution, 11(15), 3694-3702.

[6] Zhao, F., Wang, X., Zhou, Z., Harnefors, L., Svensson, J. R., Kocewiak, Ł. H., & Gryning, M. P. S. (2021). Control interaction modeling and analysis of grid-forming battery energy storage system for offshore wind power plant. IEEE Transactions on Power Systems, 37(1), 497-507.

[7] Chen, J., Liu, M., Guo, R., Zhao, N., Milano, F., & O'Donnell, T. (2021). Co-ordinated grid forming control of AC-side-connected energy storage systems for converter-interfaced generation. International Journal of Electrical Power & Energy Systems, 133, 107201.

[8] Li, G., Zhang, Y., Shi, Y., Wang, Z., & Zhou, B. (2025). Distributed coordinated control strategy for grid-forming-type hybrid energy storage systems. Sustainability, 17(4), 1436.

[9] Quan, X., Yu, R., Zhao, X., Lei, Y., Chen, T., Li, C., & Huang, A. Q. (2019). Photovoltaic synchronous generator: Architecture and control strategy for a grid-forming PV energy system. IEEE Journal of Emerging and Selected Topics in Power Electronics, 8(2), 936-948.

[10] Cai, Y., Yu, L., Wu, M., Lv, S., Fu, Z., Tong, W., ... & Shi, S. (2024). Grid-Forming Control for Solar Generation System with Battery Energy Storage. Energies, 17(15), 3642.

[11] Gerini, F., Zuo, Y., Gupta, R., Zecchino, A., Yuan, Z., Vagnoni, E., ... & Paolone, M. (2022). Optimal grid-forming control of battery energy storage systems providing multiple services: Modeling and experimental validation. Electric power systems research, 212, 108567.

[12] Liu, Y., Chen, Y., Xin, H., Tu, J., Zhang, L., Song, M., & Zhu, J. (2024). System strength

constrained grid-forming energy storage planning in renewable power systems. IEEE Transactions on Sustainable Energy.

[13] Li, C., Huang, Y., Deng, H., Zhang, X., & Zhao, H. (2022). A novel grid-forming technology for transient stability enhancement of power system with high penetration of renewable energy. International Journal of Electrical Power & Energy Systems, 143, 108402.

[14] Rehimi, S., Bevrani, H., Tarimoradi, H., Urabe, C. T., Kato, T., & Kato, T. (2024). Robust Optimal Frequency Response Enhancement Using Energy Storage-Based Grid-Forming Converters. Energies, 17(19), 4948.

[15] Zhang, Y., Xie, Y., Cai, S., Wu, Q., Zhu, H., & Xiang, Z. (2025). Coordinated control of grid-forming wind turbines and grid-forming energy storage systems for power system restoration. IEEE Transactions on Sustainable Energy.

[16] Liu, R., Wang, Z., Wang, Y., Shan, Y., Wang, W., & Wu, J. (2025). Power decoupling control for grid-forming battery energy storage system in wind farm. IEEE Transactions on Power Delivery.

[17] Tayyebi, A., Groß, D., Anta, A., Kupzog, F., & Dörfler, F. (2020). Frequency stability of synchronous machines and grid-forming power converters. IEEE Journal of Emerging and Selected Topics in Power Electronics, 8(2), 1004-1018.

[18] Mirmohammad, M., & Azad, S. P. (2024). Control and stability of grid-forming inverters: A comprehensive review. Energies, 17(13), 3186.

[19] Liu, X., Chen, Y., & Fu, Y. (2025). An Adaptive Grid-Forming Control Strategy Based on Capacitor Energy State Estimation. Batteries, 11(9), 337.

[20] Cheng, Y., Wu, W., Chung, H. S. H., Blaabjerg, F., Orabi, M., & Yuan, L. (2024, May). Transfer Control Method Between Grid-Connected and Off-grid Modes Based on Kalman Filter for Grid-Forming Inverter. In 2024 IEEE 10th International Power Electronics and Motion Control Conference (IPEMC2024-ECCE Asia) (pp. 902-906). IEEE.

[21] Balouji, E., Bäckström, K., & McKelvey, T. (2023, October). Deep reinforcement learning based grid-forming inverter. In 2023 IEEE Industry Applications Society Annual Meeting (IAS) (pp. 1-9). IEEE.

[22] Ji, T., Lin, P., Zhu, M., Meng, Q., Zhang, C., & Zhu, J. (2024, August). Deep Reinforcement Learning Based Robust Control Strategy for Power Sharing and Voltage Regulation in DC Grid-Forming Energy Storage System. In 2024 4th Power System and Green Energy Conference (PSGEC) (pp. 1327-1331). IEEE.

[23] Zhang, Z., Xie, Y., & Huang, R. (2025, May). Pre-training and Fine-Tuning Based Meta-Reinforcement-Learning for Grid-Forming-Based Energy Storage System Operation to Recover Power Grid Frequency. In 2025 IEEE 3rd International Conference on Power Science and Technology (ICPST) (pp. 1679-1683). IEEE.

[24] Yameen, M. Z., Lu, Z., El-Sousy, F. F., Younis, W., Zardari, B. A., & Junejo, A. K. (2025). Improving frequency stability in grid-forming inverters with adaptive model predictive

control and novel COA-jDE optimized reinforcement learning. Scientific Reports, 15(1), 16540.