**DOI**：10.13234/j.issn.2095-2805.2025.4.255 **中图分类号**：TM619 **文献标志码**：A

# 基于改进 PPO 的 HCSY-MG 并网系统分布式混合储能充放电优化控制

李锦键(中国电源学会学生会员)，王兴贵，丁颖杰(中国电源学会学生会员)
(兰州理工大学电气工程与信息工程学院，兰州 730050)

**摘要**：为平抑微源半桥变流器串联星型结构微电网 HCSY-MG(half-bridge converter series Y-connection micro-grids)并网系统中微源出力的波动，保证各相直流侧电压之和相等，与并网电流三相平衡，提出 1 种基于改进近端策略优化 PPO(proximal policy optimization)的分布式混合储能系统 HESS(hybrid energy storage system)充、放电优化控制策略。在考虑 HCSY-MG 系统并网电流与分布式 HESS 特性的条件下，确定影响并网电流的主要系统变量，以及 HESS 接入系统的最佳拓扑结构。然后结合串联系统的特点，将分布式 HESS 的充、放电问题转换为深度强化学习的 Markov 决策过程。同时针对 PPO 算法中熵损失权重难以确定的问题，提出 1 种改进的 PPO 算法，兼顾智能体的收敛性和探索性。最后以某新能源发电基地的典型运行数据为算例，验证所提控制策略的可行性和有效性。

**关键词**：串联微电网；分布式混合储能系统；近端策略优化；充放电功率；深度强化学习

# Optimal Control of Distributed Hybrid Energy Storage Charge-discharge for HCSY-MG Grid-connected System Based on Improved PPO

LI Jinjian, *Student Member, CPSS*, WANG Xinggui, DING Yingjie, *Student Member, CPSS*
(College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

**Abstract:** To mitigate the fluctuations in the output from micro-sources within a half-bridge converter series Y-connection microgrid (HCSY-MG) grid-connected system and ensure both the balance of the sum of DC-side voltages in each phase and the balance of three-phase grid-connected current, a distributed hybrid energy storage system (HESS) charge-discharge optimal control strategy based on improved proximal policy optimization (PPO) is put forward. With the consideration of grid-connected current in the HCSY-MG system and the characteristics of the distributed HESS, the key system variables affecting the grid-connected current are identified, and the optimal topology for integrating the HESS into the system is determined. Subsequently, by combining the characteristics of the series-connected system, the charge-discharge problem of the distributed HESS is transformed into a Markov decision process within the framework of deep reinforcement learning. Meanwhile, to address the challenge of determining the entropy loss weight in the PPO method, an improved PPO method is proposed, which balances the agent's convergence and exploration capabilities. Finally, the typical operational data of a renewable energy generation site is used as a case study, and the feasibility and effectiveness of the proposed control strategy is validated.

**Keywords:** Series-connected microgrid; distributed hybrid energy storage system (HESS); proximal policy optimization (PPO); charge-discharge power; deep reinforcement learning

**收稿日期**：2024-08-14；**修回日期**：2024-09-18；**录用日期**：2024-10-07；**网络首发日期**：2024-10-25
**基金项目**：国家自然科学基金资助项目(51967011)
This work is supported by National Natural Science Foundation of China under the grant 51967011

在传统交流微电网中，可再生能源发电单元与负载多通过电力电子变流器并联在交流母线上，该拓扑结构导致传统交流微电网中存在诸如环流和谐波抑制困难，控制复杂等问题。针对交流微电网

目前存在的问题，文献[1]提出 1 种半桥变流器串联星型结构微电网 HCSY-MG(half-bridge converter series Y-connection microgrid)系统，该系统凭借串联特性，从拓扑结构上解决了传统交流微电网中存在的问题，同时使系统控制更为简单。HCSY-MG 系统为补偿可再生能源发电固有的波动性，会在各发电单元 GM(generator module)直流侧并联混合储能系统 HESS(hybrid energy storage system)，因为 HESS 分布在各 GM 直流侧，所以该种储能方式也被称为分布式储能。与集中式储能相比，分布式储能可以根据不同微源的特性，更有针对性地采取相应的控制策略，并且在协调控制时能达到更优的效果。与传统 H 桥链式结构组成的微电网相比，HCSY-MG 系统在微源数量一定的条件下，所使用的开关器件数量减半，可显著降低微电网系统的投资成本，同时半桥变流器的控制也更加简单、灵活[1]。但 HCSY-MG 系统仍旧存在诸如可再生能源波动造成输出电压、电流中含有部分直流和 50 Hz 低频脉动分量，以及分布式 HESS 投资成本会不可避免上升等问题。为解决可再生能源波动性造成的上述问题，研究人员广泛采用优化 HESS 充、放电功率的方法。例如，文献[2]针对光伏微电网中的混合储能系统，提出 1 种模型预测控制策略，提高了直流母线电压的稳定性和动态响应能力。但仅有部分关于串联型微电网中分布式 HESS 单体控制策略的研究，结合串联系统特点的 HESS 充、放电功率协调优化控制鲜有研究。基于此，本文将研究 1 种 HCSY-MG 并网系统中分布式 HESS 的充、放电控制策略，在充分利用系统特点和优势的基础上，降低可再生能源发电波动性对系统的不利影响。

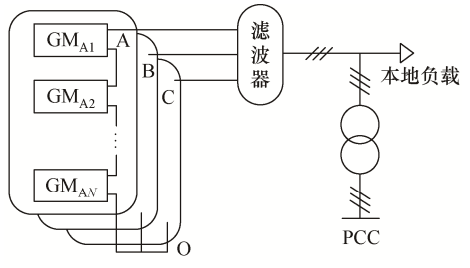
微电网中储能系统的充、放电优化控制可以描述为考虑多种因素变化时的充、放电功率或状态的时序安排，即序贯决策问题[3]。为解决序贯决策问题，目前广泛采用的方法包括经典优化方法、基于规划的方法、启发式优化方法和深度强化学习 DRL (deep reinforcement learning)方法等[4]。其中，基于数据驱动的DRL 方法能够利用微电网大量数据，根据不同运行要求和优化目标，在实时决策的同时给出相应的控制方案和优化策略，已经成为当下的研究热点[5]。针对不同控制目标，目前已有DRL 中的多种算法被应用至微电网储能装置的调度与控制。文献[6]考虑了可再生能源出力的未来不确定性，利用双重深度 $Q$ 学习算法对微电网中储能装置进行调度，成功降低了微电网的运行成本；文献[7]在考虑多种其他因素的基础上，通过双重深度 $Q$ 学习算法降低微电网运行成本的同时，与启发式方法对比，说明了 DRL 方法的可行性和优越性；文献[8]利用深度确定性策略梯度方法对微电网中的储能进行控制，实现了运行利润的最大化。除考虑微电网运行的经济性以外，文献[9-10]分别从最小化网损和储能系统功率损耗 2 个方面出发，利用近端策略优化 PPO(proximal policy optimization)方法和 $Q$ 学习算法对微电网中的能量调度进行优化；文献[11]提出了 1 种两步超前的储能调度DRL 方法，提高了储能和可再生能源的利用率；文献[12]利用柔性策略评价算法控制微电网中储能的充、放电，使微电网中的电压偏移最小。在 DRL 方法中，基于策略梯度的算法通过构建策略网络，避免了对动作的离散化，使相应算法能够应用至连续动作空间的环境中[13]。针对传统策略梯度算法对步长选择较为敏感的问题，文献[14]为提高算法稳定性和对超参数的鲁棒性，提出了置信域策略优化算法，避免了策略参数发生大范围变化，但该算法的计算效率较低；文献[15]在置信域策略优化算法的基础上提出了 1 种基于剪切函数的 PPO 算法，通过多个训练回合对参数进行小批量更新，在提高算法计算效率的同时，保留了置信域策略优化算法的优点。

本文在确定适用于 HCSY-MG 并网系统 HESS 拓扑结构的基础上，充分考虑串联型微电网的特性，以各相直流侧电压之和偏差最小为目标，提出 1 种基于改进 PPO 算法的 HESS 充、放电功率优化控制策略。同时根据训练过程中损失函数的变化，提出 1 种改进的 PPO 算法，使熵损失函数中的权重因子自适应更新，解决了熵损失权重难以确定的问题。最后以某新能源发电基地的典型运行数据为算例，对所提方法的可行性和有效性进行了验证。

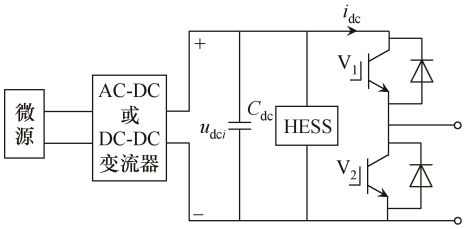
# 1　HCSY-MG 并网系统

## 1.1 HCSY-MG 系统结构

HCSY-MG 系统的拓扑结构如图 1 所示。在该系统中，每相均由 $N$ 个 GM 串联组成，并以星型连接的方式组成三相逆变环节，再经滤波器、变压器等装置连接于交流母线。




(a) HCSY-MG并网系统




(b) 半桥变流器

**图 1　HCSY-MG 系统拓扑结构**

**Fig. 1 Topology of HCSY-MG system**

由图 1(b)可见，每个 GM 由微源、AC-DC 或 DC-DC 变流器、稳压电容 $C_{dc}$、HESS 和半桥变流器组成。为便于分析，将每个 GM 等效为受控电压源，则可得到 HCSY-MG 并网系统的等效电路如图 2 所示。图中：$R_e$ 为线路等效电阻；$i_x$ 为系统输出电流；$u_{xN}$ 为等效单相输出电压；$i_{Lx}$ 为负载电流；$i_{ex}$ 为并网电流；$e_x$ 为电网电压。其中，$x$=A，B，C。

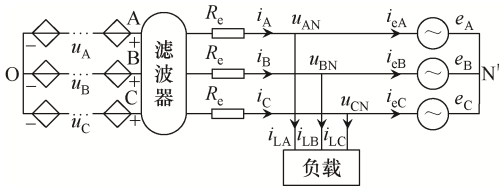



**图 2　HCSY-MG 并网系统等效电路**

**Fig. 2 Equivalent circuit of HCSY-MG grid-connected system**

## 1.2 HCSY-MG 系统并网电流模型

由图 2 可得 HCSY-MG 等效单相并网系统的基尔霍夫电压定律 KVL(Kirchhoff's voltage law)方程为

$$u_{x\mathrm{N}} = L_{\mathrm{f}}\frac{\mathrm{d}i_x}{\mathrm{d}t} + R_{\mathrm{e}}i_x + e_x \tag{1}$$

式中，$L_{\mathrm{f}}$ 为滤波电感。

进一步，令 $i_{ex} = i_x - i_{Lx}$，可得到三相并网电流的矩阵表达式为

$$L_{\mathrm{f}}\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} i_{\mathrm{eA}} \\ i_{\mathrm{eB}} \\ i_{\mathrm{eC}} \end{bmatrix} + \begin{bmatrix} R_{\mathrm{e}} & 0 & 0 \\ 0 & R_{\mathrm{e}} & 0 \\ 0 & 0 & R_{\mathrm{e}} \end{bmatrix}\begin{bmatrix} i_{\mathrm{eA}} \\ i_{\mathrm{eB}} \\ i_{\mathrm{eC}} \end{bmatrix} = \begin{bmatrix} u_{\mathrm{AN}} - e_{\mathrm{AN'}} \\ u_{\mathrm{BN}} - e_{\mathrm{BN'}} \\ u_{\mathrm{CN}} - e_{\mathrm{CN'}} \end{bmatrix} \tag{2}$$

式中，$e_{\mathrm{AN'}}$、$e_{\mathrm{BN'}}$、$e_{\mathrm{CN'}}$分别为电网 A、B、C 相的相电压；$u_{\mathrm{AN}}$、$u_{\mathrm{BN}}$、$u_{\mathrm{CN}}$的表达式[16]分别为

$$u_{\mathrm{AN}} = \frac{2u_{\mathrm{AO}} - u_{\mathrm{BO}} - u_{\mathrm{CO}}}{3} = \frac{1}{3}\left[\left(u_{\mathrm{dcA}} - \frac{1}{2}u_{\mathrm{dcB}} - \frac{1}{2}u_{\mathrm{dcC}}\right) + \frac{\sqrt{3}}{4}M(u_{\mathrm{dcB}} - u_{\mathrm{dcC}})\cos(\omega_0 t) + \frac{M}{4}(4u_{\mathrm{dcA}} + u_{\mathrm{dcB}} + u_{\mathrm{dcC}})\sin(\omega_0 t)\right] \tag{3}$$

$$u_{\mathrm{BN}} = \frac{2u_{\mathrm{BO}} - u_{\mathrm{AO}} - u_{\mathrm{CO}}}{3} = \frac{1}{3}\left[\left(u_{\mathrm{dcB}} - \frac{1}{2}u_{\mathrm{dcA}} - \frac{1}{2}u_{\mathrm{dcC}}\right) - \frac{\sqrt{3}}{4}M(2u_{\mathrm{dcB}} + u_{\mathrm{dcC}})\cos(\omega_0 t) - \frac{M}{4}(2u_{\mathrm{dcA}} + 2u_{\mathrm{dcB}} - u_{\mathrm{dcC}})\sin(\omega_0 t)\right] \tag{4}$$

$$u_{\mathrm{CN}} = \frac{2u_{\mathrm{CO}} - u_{\mathrm{AO}} - u_{\mathrm{BO}}}{3} = \frac{1}{3}\left[\left(u_{\mathrm{dcC}} - \frac{1}{2}u_{\mathrm{dcA}} - \frac{1}{2}u_{\mathrm{dcB}}\right) + \frac{\sqrt{3}}{4}M(u_{\mathrm{dcB}} + 2u_{\mathrm{dcC}})\cos(\omega_0 t) - \frac{M}{4}(2u_{\mathrm{dcA}} - u_{\mathrm{dcB}} + 2u_{\mathrm{dcC}})\sin(\omega_0 t)\right] \tag{5}$$

式中：$u_{xO}$为三相逆变环节输出电压；$u_{dcx}$为逆变侧 $x$ 相 GM 直流侧电压之和，$u_{dcx}=\sum_{i=1}^{N}u_{dcx}(i)$；$M$ 为调制比；$\omega_0$ 为调制波角频率。

由式(2)所示的 HCSY-MG 系统并网电流数学模型可以看出，若各相 GM 直流侧电压之和存在偏差导致不相等时，会造成系统输出电压中产生直流分量和基频偏差量，进而使并网电流中同样存在直流分量和基频偏差量。因此在可再生能源发电存在波动的条件下，通过对 HCSY-MG 系统中分布式 HESS 的充、放电功率进行优化控制，可以保证系统各相 GM 直流侧电压之和保持一致，从而降低并网电流中的直流分量和基频偏差量。

根据以上对 HCSY-MG 系统并网电流数学模型的分析，为降低可再生能源发电波动性对并网电流的影响，在结合系统串联特性的基础上，对系统中分布式 HESS 的充、放电功率进行优化控制十分必要。

## 2 含有 HESS 的 GM 内部功率特性分析

目前，HESS 接入微电网直流侧的拓扑结构主要有 4 种：蓄电池和超级电容均通过双向 DC-DC 变流器接入 GM 直流侧，仅蓄电池通过双向 DC-DC 变流器接入，仅超级电容通过双向 DC-DC 变流器接入，蓄电池和超级电容均直接接入。考虑到超级电容和蓄电池的响应时间，以及功率开关器件和控制系统的动作时间，若超级电容通过双向 DC-DC 变流器接入 GM 直流侧，不仅会增大系统的控制成本，还可能影响 HESS 对 GM 高频功率波动的抑制效果。蓄电池一般用于抑制 GM 的低频功率波动，为保证抑制效果，必须通过双向 DC-DC 变流器将蓄电池接入 HESS。因为超级电容的本质为电容器，所以本文将 GM 内部的稳压电容直接替换为 HESS 中的超级电容，同时承担稳定 GM 直流母线电压和平抑高频功率波动的任务。

综上，本文所采用的 HESS 拓扑结构及其内部功率流动情况如图 3 所示。图中：$P_M$ 为微源输出功率；$P_C$ 为超级电容充、放电功率；$P_B$ 为蓄电池充、放电功率；$P_G$ 为 GM 输出功率；$u_{dc}$ 为 GM 直流侧电压；$i_{dc}$ 为输出电流；$C_u$ 为超级电容；$r_B$ 为蓄电池等效电阻；$u_B$ 为蓄电池端电压；$L_B$ 为双向 DC-DC 变流器的储能电感。

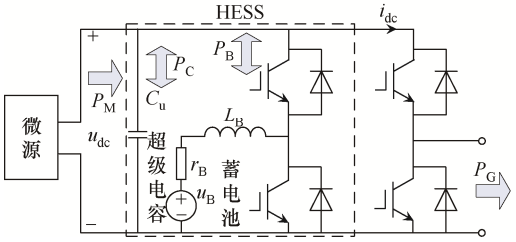



**图 3 GM 内部功率流动情况**

**Fig. 3 Power flow within GM**

当 GM 正常工作时，其内部功率满足

$$P_M-P_C-P_B=P_G \tag{6}$$

$$P_C=u_{dc}C_u\frac{du_{dc}}{dt} \tag{7}$$

由式(7)可以看出，当 $u_{dc}$ 保持恒定时，$P_C$=0。此时，GM 内部功率满足

$$P_B=P_M-P_G=\Delta P_{GM} \tag{8}$$

式中，$\Delta P_{GM}$ 为微源输出功率与 GM 输出功率的差值。

由以上 GM 内部功率流动情况可以看出，当 GM 内部功率流动保持平衡时，GM 直流母线的电压也将保持稳定，因此需要利用 HESS 平抑微源输出功率的波动，以保证 GM 直流侧电压的稳定。结合对 GM 内部功率流动的分析，以及本文所采用的 HESS 拓扑结构，使用 DRL 方法中的改进 PPO 算法对 HCSY-MG 并网系统中各 GM 直流侧的 HESS 充、放电功率进行优化控制，以保证系统各相直流侧电压之和稳定，降低可再生能源波动性对并网电流的不利影响。

## 3 基于改进 PPO 的 HESS 充放电功率优化

### 3.1 强化学习问题建模

DRL 一般可以描述为马尔科夫决策过程 MDP (Markov decision process)。传统的 MDP 包含有 4

要素，由($S, A, R, \pi$)组成的四元数组给出，其中，$S$ 为环境状态集合，$A$ 为智能体采取的动作集合，$R$ 为智能体采取某种动作的奖惩函数，$\pi$ 为在某种状态下选择动作的策略[17]。结合 HCSY-MG 并网系统的特点，以上各部分可以进行如下描述。

(1)环境状态集合 $S$。环境状态集合由智能体需要感知的环境状态变量组成，即环境需要给智能体提供的信息，包括：每个微源的实际出力 $P_{xi}$、每相 GM 直流侧电压之和 $u_{\mathrm{dc}x}$、每个 GM 直流侧 HESS 中蓄电池的荷电状态 $\mathrm{SOC}_{xi}$、三相并网电流的总谐波畸变率 $\mathrm{THD}_x$。故环境状态集合 $S$ 可以表示为

$$S=[P_{xi}(t),u_{\mathrm{dc}x}(t),\mathrm{SOC}_{xi}(t),\mathrm{THD}_x] \tag{9}$$

式中：$x$ 为三相，$x$=A, B, C；$i$=1, 2, $\cdots$, $N$。

(2)动作空间 $A$。动作空间是智能体在获得环境的状态信息后，可能采取动作的集合。因为本文的目的是 HCSY-MG 并网系统中分布式 HESS 充、放电功率的优化控制，所以动作空间定义为各 GM 中蓄电池的充、放电功率，即

$$A=[P_{\mathrm{B},xi}(t)] \tag{10}$$

式中，$P_{\mathrm{B},xi}(t)$为蓄电池的充、放电功率。

(3)奖惩函数 $R$。奖惩函数是环境对当前动作的即时回报，直接影响着智能体对动作的选择。本文所选择的奖惩函数由正向奖励和反向惩罚 2 个部分组成，当各相直流侧电压之和的偏差在允许范围内时，给予奖励；当 HCSY-MG 并网系统的状态越限时，给予惩罚。故本文选择的奖惩函数 $R$ 可以表示为

$$\begin{aligned}R=&100\{\Delta u_{\mathrm{dc}x}\leqslant\Delta u_{\mathrm{dc}}^{\max}\}-10\{\Delta u_{\mathrm{dc}x}>\Delta u_{\mathrm{dc}}^{\max}\}-\\&10\{P_{\mathrm{B},xi}(t)<P_{\mathrm{B},xi}^{\min},P_{\mathrm{B},xi}(t)>P_{\mathrm{B},xi}^{\max}\}-\\&10\{\mathrm{SOC}_{\mathrm{B},xi}(t)<\mathrm{SOC}_{\mathrm{B},xi}^{\min},\mathrm{SOC}_{\mathrm{B},xi}(t)>\\&\mathrm{SOC}_{\mathrm{B},xi}^{\max}\}-10\{\mathrm{change}\}-\\&100\{\mathrm{THD}_x\geqslant 5\%\}\end{aligned} \tag{11}$$

式中：$\Delta u_{\mathrm{dc}}^{\max}$ 为各相直流侧电压之和偏差的最大允许范围；$P_{\mathrm{B},xi}^{\min}$ 与 $P_{\mathrm{B},xi}^{\max}$ 分别为蓄电池的最小与最大充、放电功率；$\mathrm{SOC}_{\mathrm{B},xi}^{\min}$ 与 $\mathrm{SOC}_{\mathrm{B},xi}^{\max}$ 分别为蓄电池 SOC 的最小与最大值。change 表示更换充、放电的 GM，以提高储能装置的使用寿命，避免频繁启停，故此时给予 1 个较小的惩罚系数；当并网电流 $\mathrm{THD}_x\geqslant 5\%$ 时，将不满足电气和电子工程师协会 IEEE(Institute of Electrical and Electronics Engineers)关于并网电流 THD 的要求，故此时给予 1 个较大的惩罚系数。

(4)动作选择策略 $\pi$。动作选择策略是指智能体在当前环境条件下选择动作的策略，本文动作选择策略由 PPO 算法给出。

### 3.2 改进的近端策略优化算法

根据 3.1 节中关于动作选择策略 $\pi$ 的叙述，本文选用 PPO 算法求解动作选择策略。PPO 算法具有 3 个网络，分别为策略网络、策略网络副本、价值网络，各网络的作用[18]是：①策略网络(Actor 网络)的输入为环境状态集合 $S$，输出为动作空间 $A$，该网络可根据经验回放池中的数据，多次更新网络参数 $\theta$；②策略网络副本(Actor 网络副本)与策略网络相同，负责与环境交互，将获得的小批量数据($S, A, R$)存入经验回放池，策略网络副本的参数 $\theta_{\mathrm{old}}$ 定期从 $\theta$ 复制；③价值网络(Critic 网络)的输入为环境状态集合 $S$ 和动作空间 $A$，输出为价值函数，该网络可根据当前环境状态，评估状态价值，估计优势函数。

PPO 算法的目标函数 $L_t(\theta)$[15]可以表示为

$$L_t(\theta)=\hat{E}_t\left[L_t^{\mathrm{CLIP}}(\theta)-c_1L_t^{\mathrm{VF}}(\theta)+\omega S[\pi_\theta](s_t)\right] \tag{12}$$

式中：$\hat{E}_t$ 为有限批样本经验平均值的期望；$L_t^{\mathrm{CLIP}}(\theta)=\hat{E}_t\left\{\min\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t|s_t)}\hat{A}_t,\mathrm{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t|s_t)},1-\varepsilon,1+\varepsilon\right)\hat{A}_t\right]\right\}$，其中 $a_t$ 为采取的动作，$\pi_\theta$ 为策略网络给出的随机策略，$\pi_{\theta_{\mathrm{old}}}$ 为策略网络副本给出的随机策略，$\varepsilon$ 为剪切系数，$\hat{A}_t=\delta_t+(\gamma\lambda)\delta_{t+1}+\cdots+(\gamma\lambda)^{T-t+1}\delta_{T-1}$，$\delta_t=r_t+\gamma V(s_{t+1})-V(s_t)$，$\gamma$ 为折扣系数，$\lambda$ 为通用优势估计器的参数；$c_1$ 为平方误差系数；$L_t^{\mathrm{VF}}(\theta)$为平方误差损失，$L_t^{\mathrm{VF}}(\theta)=[V(s_t)-G(t)]^2$，其中 $V(s_t)$为价值函数的估计值，$G(t)=r_{t+1}+\gamma r_{t+2}+\cdots+\gamma^{T-t}r_{T+1}+\gamma^{T-t+1}\cdot V(S_{T+1})$，$r_{t+1}$ 为时间步 $t$+1 的即时奖励；$\omega$为熵损失系数；$S[\pi_\theta]$为熵损失，$S[\pi_\theta]=\frac{1}{2}\sum_{k=1}^{c}\ln(2\pi\cdot e\cdot$

$\sigma_{k,i}^2$)；$s_t$为时间步 $t$ 时的环境状态。

式(12)中，熵损失系数$\omega$作为一个关键参数，影响着智能体对于动作的更新。当$\omega$过高时，智能体动作的不确定性会增加，从而需进行更多探索，帮助智能体更快跳出局部最优；当$\omega$过低时，智能体动作的不确定性会减弱，易陷入局部最优，导致最终不能收敛至最优结果。因此为解决熵损失系数难以确定的问题，本文将利用自适应思想，使熵损失系数在训练过程中自适应更新，以在寻优过程中兼顾收敛性和探索性。

在改进策略中，通过比较当前训练回合中目标函数与已执行训练回合平均目标函数的大小，决定熵损失系数在下一轮训练回合中的大小，即

$$\omega=\begin{cases}\omega_{\max} & \text{other}\\ \omega_{\min}+\dfrac{[L_t(\theta)-L_{\min}(\theta)](\omega_{\max}-\omega_{\min})}{L_{\text{ave}}(\theta)-L_{\min}(\theta)} & L_t(\theta)<L_{\text{ave}}(\theta)\end{cases} \tag{13}$$

式中：$\omega_{\min}$与$\omega_{\max}$分别为熵损失系数的最小与最大值；$L_{\min}(\theta)$为已执行训练回合中目标函数的最小值；$L_{\text{ave}}(\theta)$为已执行训练回合中目标函数的平均值。

根据上述改进 PPO 算法，可以得到该算法的流程如图 4 所示。

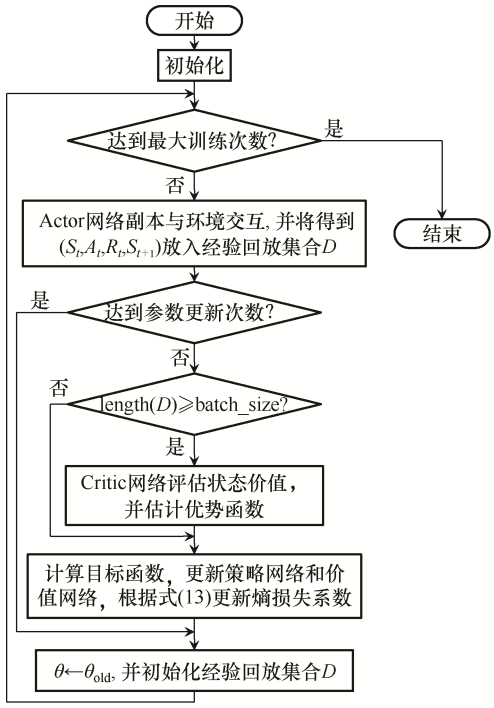



**图 4 改进 PPO 算法流程**

**Fig. 4 The flow of the improved PPO algorithm**

# 4 算例分析

## 4.1 算例基本参数与数据

算例所使用的数据源于某新能源发电基地的典型运行数据，其他主要参数见表 1。

**表 1 算例中主要参数**

**Tab. 1 Main parameters in example**

| 类别 | 参数 | 数值 |
|---|---|---|
| 环境参数 | 串联 GM 数量 $N$ | 4 |
| | GM 直流侧电压 $u_{\text{dc}}$/V | 160 |
| | 蓄电池 SOC 最小值 $\text{SOC}_{\min}$/% | 20 |
| | 蓄电池 SOC 最大值 $\text{SOC}_{\max}$/% | 80 |
| | 蓄电池充、放电功率最大值 $P_{\text{B}}^{\max}$/kW | 1 |
| 改进 PPO 参数 | 平方误差系数 $c_1$ | 0.05 |
| | 熵损失系数最大值$\omega_{\max}$ | 0.80 |
| | 熵损失系数最小值$\omega_{\min}$ | 0.01 |
| | 训练次数 $n$ | 800 |
| | 每轮更新频率 $k$ | 10 |
| | 剪切系数$\varepsilon$ | 0.2 |
| | 经验回放集合容量 length($D$) | 1 000 |
| | Actor 网络学习率 | 0.000 1 |
| | Critic 网络学习率 | 0.000 5 |

Actor 网络与 Actor 网络副本的结构相同，且输入为环境状态集合 $S$，因此输入层的神经元数量为 30；输出为相应的动作 $a_t$，故输出层的神经元数量为 1；隐含层设置为 2 层，每层分别由 60 和 128 个神经元组成，隐含层激活函数选择为 ReLu。Critic 网络的输入则为状态和动作，所以输入层和输出层的神经元数量分别为 31 与 1，隐含层及其激活函数与 Actor 网络设置相同。本文所采用的改进 PPO 算法的神经网络结构如图 5 所示，图中，$\Delta P_{xi}$为微源实际出力与额定值之间的差值。

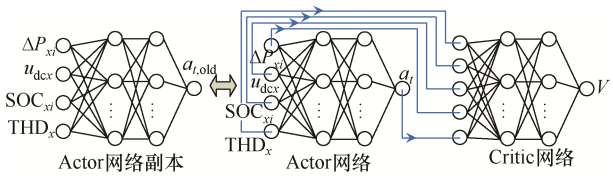



**图 5 神经网络结构**

**Fig. 5 Architecture of neural network**

## 4.2 训练过程

训练过程中，设置每相微源的组成为：风电、

风电、光伏、光伏，并以 A 相为例进行算例分析。为突出 HCSY-MG 系统串联的特点，假设 A 相中第 1 个风电微源和第 3 个光伏微源的出力波动曲线如图 6 所示。

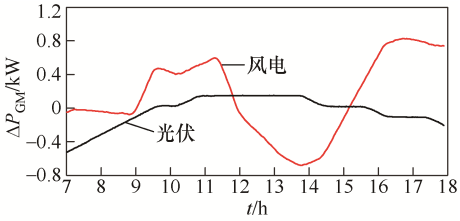



**图 6　风电和光伏微源的功率波动曲线**

**Fig. 6 Power fluctuation curves for wind and PV micro-sources**

在该种工况下，A 相中的微源出力出现波动，GM 直流侧电压之和也将出现波动；而 B、C 相中的微源出力稳定，GM 直流侧电压之和稳定在 640 V。同时设置相应 GM 中蓄电池的 SOC 初始值分别为 60%、60%、40%、40%；风电和光伏 GM 中蓄电池的容量分别为 3 和 2 kW·h[19]。在训练过程中，采用传统 PPO 算法与本文所提改进 PPO 算法的平均奖励函数曲线如图 7 所示。

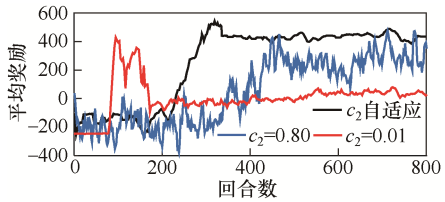



**图 7　训练过程中的奖励函数曲线**

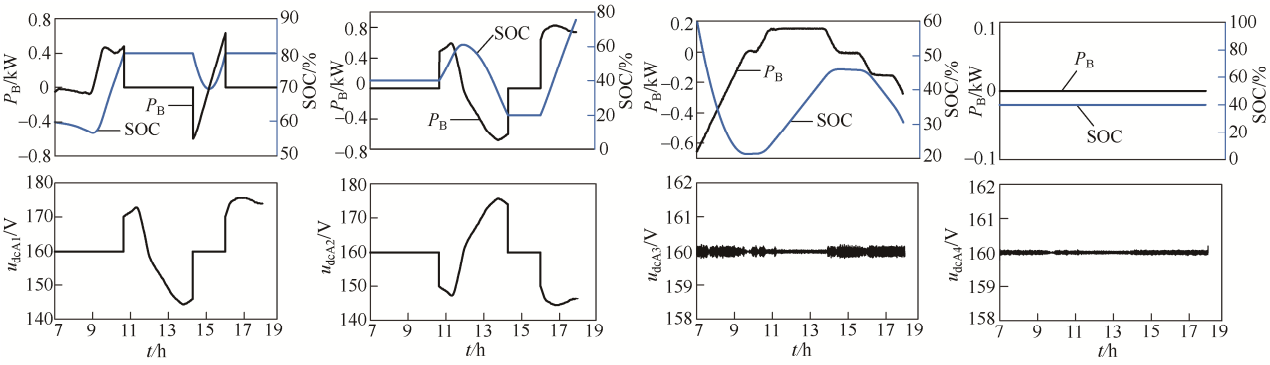
**Fig. 7 Reward function curves during training**

由图 7 可见：当$\omega$=0.01 时，尽管在训练的初始阶段可以获得奖励较大的动作，但会陷入局部最优，且由于智能体探索能力不足，在陷入局部最优时不能快速跳出，导致达到最大训练回合数时无法学得奖励最大的动作；当$\omega$=0.80 时，平均奖励曲线波动剧烈，说明此时智能体的搜索能力较强，可以充分探索动作空间，但在训练后期存在不易收敛的特点；当$\omega$按照本文所提自适应策略进行更新时，在训练初始阶段可以较快学得奖励较大的动作，并且在达到最大训练回合数时，收敛效果较好。以上关于训练过程中奖励函数曲线的分析与 3.2 节关于$\omega$的理论分析基本一致，证明了本文所提熵损失系数自适应更新策略可以在保证 PPO 算法收敛性的同时兼顾探索性。

### 4.3　测试结果

在所设置的环境条件下，利用具有奖励函数值较高的动作与环境进行交互测试，得到 A 相 4 个 GM 中蓄电池的充放电功率 $P_B$ 曲线、对应的 SOC 曲线和直流侧电压曲线如图 8 所示。

由图 8 可见，A 相中各 GM 的蓄电池充、放电功率均未超过最大值 1 kW，同时对应的 SOC 均在[20%, 80%]的规定区间内，说明智能体得出的最优动作满足相应约束条件。在约 10.5 h 时，1#风电 GM 的蓄电池充电至 SOC=80%，因为要继续维持吸收能量的状态，所以此时 A 相中 SOC 最低，剩余容量最多的 2#风电 GM 的蓄电池开始充电；在



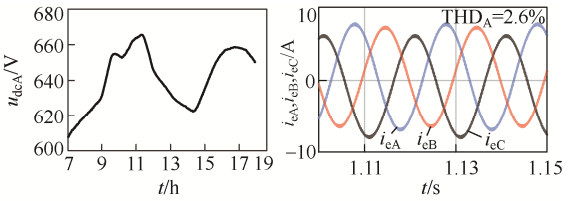

(a) 1#风电GM　(b) 2#风电GM　(c) 1#光伏GM　(d) 2#光伏GM

**图 8　HESS 中蓄电池的充、放电功率及其 SOC 曲线**

**Fig. 8 Charge-discharge power of batteries in HESS and their SOC curves**

约 14.5 h 时，2#风电 GM 中的蓄电池放电至 SOC=20%，因为要继续维持释放能量的状态，所以此时 A 相中 SOC 最高，储存能量最多的 1#风电 GM 中的蓄电池开始放电。1#光伏 GM 直流侧功率发生波动时，由于波动范围较小，且式(11)中更换充、放电模块的惩罚项，所以在满足相应约束条件的情况下，该模块的功率波动由该模块自身的蓄电池进行平抑。同时，结合 A 相中各 GM 的直流侧电压曲线可以看出，本文所提 HESS 充、放电优化控制可以较好地结合 HCSY-MG 并网系统的特点，协调控制各 GM 中配备的 HESS 系统。在 1#风电 GM 中蓄电池 SOC 超过允许值时，结合 HCSY-MG 并网系统的特点，在本文所提 HESS 充、放电优化控制方法下，利用 2#风电 GM 中的蓄电池平抑可再生能源波动对系统直流侧电压之和及并网电流的影响。若不结合 HCSY-MG 并网系统的特点，单独利用各 GM 中配备的 HESS 对其可再生能源出力波动进行平抑，本文所设定的蓄电池容量无法满足平抑需求。在此条件下，若要利用 HESS 对可再生能源出力波动进行平抑，以减轻对系统直流侧电压之和及并网电流的影响，则蓄电池的容量势必要增加，从而会增加相应的投资成本。因此本文在考虑 HCSY-MG 并网系统特点的条件下，提出的 HESS 充、放电功率优化控制方法可以使 HESS 的容量保持在较低的范围，从而降低 HESS 的投资成本。

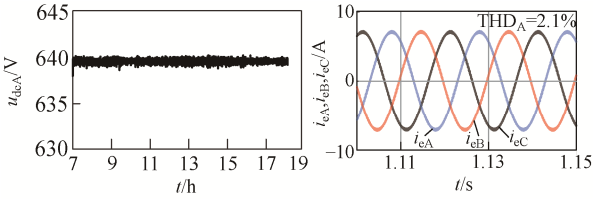
A 相各 GM 直流侧 HESS 中的蓄电池直接接入(即不对其充、放电功率进行控制)时，以及采用本文所提充、放电功率优化控制策略条件下，HCSY-MG 并网系统的 A 相 GM 直流侧电压之和的曲线，以及三相并网电流的曲线分别如图 9 和图 10 所示。




(a) 直流侧电压之和 (b) 三相并网电流

**图 9 蓄电池直接接入 GM 直流侧时的系统特征**

**Fig. 9 System characteristics when battery is directly connected to DC-side of GM**




(a) 直流侧电压之和 (b) 三相并网电流

**图 10 蓄电池采用本文所提控制策略时的系统特征**

**Fig. 10 System characteristics when proposed control strategy is applied to battery**

由图 9(a)可见，HESS 中蓄电池直接接入 GM 直流侧的方法不能完全平抑可再生能源出力的波动性，A 相直流侧电压之和仍存在波动。当 $u_{dcA}$=610 V 时，三相并网电流的曲线如图 9(b)所示，此时三相并网电流中分别含有−1.0、0.5 和 0.3 A 的直流分量，且 A 相并网电流的 THD=2.6%。由图 10(a)可见，当 HESS 中蓄电池的充、放电功率采用本文所提控制策略时，A 相直流侧电压之和仅在一个小范围内波动。相同时间点的三相并网电流曲线如图 10(b)所示，此时三相并网电流中的直流分量分别为−0.020、0.010 和 0.004 A，均下降了约 98%，且 A 相并网电流的 THD 降低至 2.1%。该结果证明本文所提 HESS 充、放电功率控制策略能够有效降低可再生能源波动对 HCSY-MG 系统直流侧电压之和及并网电流的不利影响。

以上测试结果证明了本文所提基于改进 PPO 算法的 HCSY-MG 并网系统分布式 HESS 充、放电优化控制策略是可行的、有效的，具有一定的经济价值。

## 5 结论

针对 HCSY-MG 并网系统中微源出力波动造成各相 GM 直流侧电压之和不相等，进一步对并网电流造成不利影响的问题，本文在确定影响 HCSY-MG 系统并网电流的主要变量，以及适用于该微电网 HESS 拓扑结构的基础上，提出 1 种基于改进 PPO 算法的分布式 HESS 充、放电功率优化控制策略。同时，为降低熵损失系数设置对智能体动作选择的影响，提出 1 种熵损失自适应更新的改

进 PPO 算法。本文所提 HCSY-MG 并网系统 HESS 充、放电控制方法能够充分发挥系统的特点和优势，可以在满足约束的条件下得到最优的充、放电功率，同时，所提改进 PPO 算法能够使智能体在选择动作时兼顾收敛性和探索性。

本文尽管成功将深度强化学习理论引入串联型微电网的控制中，但仅考虑了拓扑结构的串联特性。为充分发挥串联型微电网的特点和优点，后续还需进一步研究多智能体技术在串联型微电网控制中的应用，例如利用储能提升系统稳定性、功率协调控制等方面。

**参考文献：**


[1] Wang Xinggui, Ding Yingjie, Li Jinjian, et al. Characteristics analysis of micro-source half-bridge converter series Y-connection based microgrid systems [J]. Journal of Power Electronics, 2023, 23(10): 1483-1495.

[2] 海涛, 刘昱麟, 王钧, 等. 基于模型预测控制的光伏混合储能系统控制策略[J/OL]. 电源学报, 1-12. (2024-04-08)[2024-09-18]. https://link.cnki.net/urlid/12.1420.TM.20240407.1045.002.

Hai Tao, Liu Yulin, Wang Jun, et al. Control strategy of photovoltaic hybrid energy storage system based on model predictive control [J/OL]. Journal of Power Supply, 1-12.(2024-04-08)[2024-09-18]. https://link.cnki.net/urlid/12.1420.TM.20240407.1045.002. (in Chinese)

[3] 胡维昊, 曹迪, 黄琦, 等. 深度强化学习在配电网优化运行中的应用[J]. 电力系统自动化, 2023, 47(14): 174-191.

Hu Weihao, Cao Di, Huang Qi, et al. Application of deep reinforcement learning in optimal operation of distribution network [J]. Automation of Electric Power Systems, 2023, 47(14): 174-191. (in Chinese)

[4] 张自东, 邱才明, 张东霞, 等. 基于深度强化学习的微电网复合储能协调控制方法 [J]. 电网技术, 2019, 43(6): 1914-1921.

Zhang Zidong, Qiu Caiming, Zhang Dongxia, et al. A coordinated control method for hybrid energy storage system in microgrid based on deep reinforcement learning [J]. Power System Technology, 2019, 43(6): 1914-1921. (in Chinese)

[5] 李鹏, 钟瀚明, 马红伟, 等. 基于深度强化学习的有源配电网多时间尺度源荷储协同优化调控[J/OL]. 电工技术学报，1-18. (2024-07-19)[2024-09-18]. https://doi.org/10.19595/j.cnki.1000-6753.tces.240316.

Li Peng, Zhong Hanming, Ma Hongwei, et al. Multi-timescale optimal dispatch of source-load-storage coordination in active distribution network based on deep reinforcement learning [J/OL]. Transactions of China Electrotechnical Society, 1-18. (2024-07-19)[2024-09-18]. https://doi.org/10.19595/j.cnki.1000-6753.tces.240316. (in Chinese)

[6] Bui V H, Hussain A, Kim H M. Double deep *Q*- learning-based distributed operation of battery energy storage system considering uncertainties [J]. IEEE Transactions on Smart Grid, 2020, 11(1): 457-469.

[7] 刘俊峰, 陈剑龙, 王晓生, 等. 基于深度强化学习的微能源网能量管理与优化策略研究[J]. 电网技术, 2020, 44(10): 3794-3803.

Liu Junfeng, Chen Jianlong, Wang Xiaosheng, et al. Energy management and optimization of multi-energy grid based on deep reinforcement learning [J]. Power System Technology, 2020, 44(10): 3794-3803. (in Chinese)

[8] Ding Tao, Zeng Ziyu, Bai Jiawen, et al. Optimal electric vehicle charging strategy with Markov decision process and reinforcement learning technique [J]. IEEE Transactions on Industry Applications, 2020, 56(5): 5811-5823.

[9] Cao Di, Hu Weihao, Xu Xiao, et al. Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices [J]. Journal of Modern Power Systems and Clean Energy, 2021, 9(5): 1101-1110.

[10] Qiu Xin, Nguyen T A, Crow M L. Heterogeneous energy storage optimization for microgrids [J]. IEEE Transactions on Smart Grid, 2016, 7(3): 1453-1461.

[11] Kuznetsova E, Li Yanfu, Ruiz C, et al. Reinforcement learning for microgrid energy management [J]. Energy, 2013, 59: 133-146.

[12] Wang Shengyi, Du Liang, Fan Xiaoyuan, et al. Deep reinforcement scheduling of energy storage systems for

real-time voltage regulation in unbalanced LV networks with high PV penetration [J]. IEEE Transactions on Sustainable Energy, 2021, 12(4): 2342-2352.

[13] Williams R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning [J]. Machine Learning, 1992, 8(3): 229-256.

[14] Schulman J, Levine S, Moritz P, et al. Trust region policy optimization [EB/OL]. (2015-02-19)[2024-07-21]. https://arxiv.org/abs/1502.05477v5.

[15] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms [EB/OL]. (2017-07-20)[2024-07-21]. https://arxiv.org/abs/1707.06347v2.

[16] Wang Xinggui, Li Jinjian, Guo Qun, et al. Parameter design of half-bridge converter series Y-connection microgrid grid-connected filter based on improved PSO-LSSVM [J]. International Transactions on Electrical Energy Systems, 2023: 9534004.

[17] 邹伟，鬲玲，刘昱杓. 强化学习[M]. 北京：清华大学出版社, 2020.

Zou Wei, Ge Ling, Liu Yubiao. Reinforcement Learning [M]. Beijing: Tsinghua University Press, 2020. (in Chinese)

[18] 秦湖程，黄炎焱，陈天德，等. 基于 PPO 算法的集群多目标火力规划方法[J]. 系统工程与电子技术，2024(11): 3764-3773.

Qin Hucheng, Huang Yanyan, Chen Tiande, et al. Cluster multi-target fire planning method based on PPO algorithm [J]. Systems Engineering and Electronics, 2024(11): 3764-3773. (in Chinese)

[19] 刘畅，卓建坤，赵东明，等. 利用储能系统实现可再生能源微电网灵活安全运行的研究综述[J]. 中国电机工程学报, 2020, 40(1): 1-18.

Liu Chang, Zhuo Jiankun, Zhao Dongming, et al. A review on the utilization of energy storage system for the flexible and safe operation of renewable energy microgrids [J]. Proceedings of the CSEE, 2020, 40(1): 1-18. (in Chinese)



李锦键

**作者简介：**

李锦键(1995— )，男，中国电源学会学生会员，博士研究生。研究方向：可再生能源发电系统与控制，微电网技术。E-mail：lijinjian0326@163.com。

王兴贵(1963— )，男，通信作者，硕士，教授。研究方向：可再生能源发电系统与控制，微电网技术，电力电子与电力传动。E-mail：wangxg8201@163. com。

丁颖杰(1992— )，女，中国电源学会学生会员，博士研究生。研究方向：可再生能源发电系统与控制，微电网技术。E-mail：dingyj0820@126.com。

(责任编辑：耿越)