energies

MDPI

Article

Learning-Based Model Predictive Control of DC-DC Buck Converters in DC Microgrids: A Multi-Agent Deep Reinforcement Learning Approach

Hoda Sorouri, Arman Oshnoei *, Mateja Novak, Frede Blaabjerg and Amjad Anvari-Moghaddam

Department of Energy (AAU Energy), Aalborg University, 9220 Aalborg, Denmark; hoso@energy.aau.dk (H.S.); nov@energy.aau.dk (M.N.); fbl@energy.aau.dk (F.B.); aam@energy.aau.dk (A.A.-M.)
* Correspondence: aros@energy.aau.dk

Abstract: This paper proposes a learning-based finite control set model predictive control (FCS-MPC) to improve the performance of DC-DC buck converters interfaced with constant power loads in a DC microgrid (DC-MG). An approach based on deep reinforcement learning (DRL) is presented to address one of the ongoing challenges in FCS-MPC of the converters, i.e., optimal design of the weighting coefficients appearing in the FCS-MPC objective function for each converter. A deep deterministic policy gradient method is employed to learn the optimal weighting coefficient design policy. A Markov decision method formulates the DRL problem. The DRL agent is trained for each converter in the MG, and the weighting coefficients are obtained based on reward computation with the interactions between the MG and agent. The proposed strategy is wholly distributed, wherein agents exchange data with other agents, implying a multi-agent DRL problem. The proposed control scheme offers several advantages, including preventing the dependency of the converter control system on the operating point conditions, plug-and-play capability, and robustness against the MG uncertainties and unknown load dynamics.

Keywords: DC microgrid; finite set model predictive control; dc-dc buck converter; deep reinforcement learning; constant power load

Citation: Sorouri, H.; Oshnoei, A.; Novak, M.; Blaabjerg, F.; Anvari-Moghaddam, A. Learning-Based Model Predictive Control of DC-DC Buck Converters in DC Microgrids: A Multi-Agent Deep Reinforcement Learning Approach. *Energies* **2022**, *15*, 5399. https://doi.org/10.3390/en15155399

Academic Editor: Mario Marchesoni

Received: 31 May 2022
Accepted: 24 July 2022
Published: 26 July 2022

1. Introduction

Microgrids (MGs) are a group of interconnected loads and distributed generations (DGs), and they are usually interfaced to the grid through power converters to reduce pollution and power transmission losses with the flexibility of different installation location. This is an important concept for future distribution systems and will be more utilized in renewable energy integration that is the fastest-growing energy source globally [1,2]. The MGs can be utilized in both islanded and grid-connected operation modes [3]. The use of clean and sustainable energy resources such as photovoltaic systems, batteries, and chargers, has created a lot of interest in DC-MGs [4]. DC-MGs also have several benefits in comparison with their AC counterparts. For example, controlling reactive power or unbalanced electrical signals is not a problem in a DC-MG, while protection is still a challenging task [5]. The critical issue for AC islanded MGs is to ensure voltage and frequency stability when inverters are connected to power sources with lines and loads [3,6].

The power converters' role is like voltage sources interface between loads and different types of sources that are responsible to share the power based on the availability and capability of the energy sources [7]. The most typical interfaces employed in DC-MGs are DC-DC buck and boost converters. Once the converters are tightly controlled, they act as constant power loads (CPLs) [8]. CPLs hold negative impedance, which may induce instability in the DC bus, and consequently, the whole MG may fail [9]. The CPL's impact becomes more critical once MG works in islanded mode due to decreased damping.

Various solutions have been proposed in previous studies to deal with this problem. For example, introducing virtual impedance loops to converter control systems offers promising solutions for increasing the precision of power sharing and damping oscillatory currents in DC-MGs [10]. There are diverse control strategies for current and voltage control of DC-DC converters, including sliding mode control (SMC), fuzzy logic, proportional-integral (PI), model predictive control (MPC), and state-dependent Riccati equations control [11]. Linear controllers are the most straightforward control methods to reach the voltage regulation in DC-MGs [12]. These methods evaluate the network's stability around only one equilibrium point [13,14] supply load power. An integrated CPL raises the degree of nonlinearity in DC-MGs. Thus, traditional linear strategies are questioned and face stability restrictions. A nonlinear PI stabilization controller has been developed in [15] to ensure stability in DC-MGs. This method has the challenge of variable switching frequency as it affects converter efficiency [16]. The authors in [9] have proposed a nonlinear SMC to develop a control rule that guarantees an area larger than local stability while improving large-signal stability. The main drawback of SMC is that it is challenging to impose restrictions or control abstract quantities. To cover these drawbacks, FCS-MPC has been identified as one of the most favorable controllers for power electronic applications due to its capability over real-time solutions to multiple objectives and constraints [17,18]. The performance of FCS-MPC is deeply influenced by the weighting coefficients, the tuning of which is still a challenge to be undertaken. In this regard, Ref. [8] has employed an artificial neural network method in off-line mode for weighting coefficient design in uninterrupted power supply (UPS) system. This method, however, demands a high number of calculations for the adaptation and training process, and also the conducted analyses for identifying the optimal values of weighting coefficients are dependent on operating conditions, which may give rise to a flawed performance of the control system.

Recently, model-free intelligent controllers such as fuzzy logic and neural network have been developed to decrease the sensitivity to modeling inaccuracy. The main characteristic of intelligent controllers is the model-free design that enables them to manage model non-linearity, complexity, and uncertainty in power electronic applications. Nevertheless, these methods are only suitable for a specific time interval as suffering from the lack of the capability to learn online [19]. With the rapid development in machine learning, reinforcement learning (RL)-based techniques have gained significant attention. They have become a vital mechanism in developing intelligent networks. RL approaches have successfully solved complicated problems by integrating them with a deep neural network, called DRL [20]. As a DRL algorithm, Deep Q Network (DQN) is developed to address the limitations of conventional Q networks [21]. The DQN has been utilized in different applications such as Automatic Underwater Vehicles [22], Aerial Robots [23], and quadrotor control [24]. Nevertheless, DQN utilizes discrete steps for estimating the value function, limiting its use for problems with continuous steps. Hence, a deep deterministic policy gradient (DDPG) algorithm is formulated to address this challenge [24,25]. In [26], the DRL is employed as a voltage controller for a DC-DC buck converter. In [27], the application of the DRL method is investigated for optimizing the weighting coefficient for an FCS-MPC controlled inverter in a UPS system. However, those studies address a single-agent RL problem, which may not suit multiple-inverter systems like DC-MGs.

Motivated by the previous discussion, this paper uses the FCS-MPC to improve the voltage regulation of DC-DC buck converters used in a DC-MG. To avoid the dependency of the converter control system on the operating conditions, the weighting coefficients appearing in the FCS-MPC objective function for each converter are regulated in an online fashion via distributed DRL algorithm. The DRL problem is solved by a DDPG algorithm in a critic-actor framework. The DRL agent is trained for each buck converter in the MG, and the weighting coefficients in the FCS-MPC are obtained based on reward computation with the interactions between the MG and agent. Under the proposed strategy, each agent is established at the local converter to reach the optimal purpose simultaneously. The simulation validations under different operational conditions are provided to illustrate

the effectiveness of the proposed control scheme. Table 1 summarizes a taxonomy of existing publications in the area and compares previous studies in this field to highlight the main contributions of this paper.

Table 1. Comparison of the contributions of this paper with the previous studies.

Refs.	Controller	PnP Capability	Robust	Adaptive	Multi DG Units	CPL
[7]	ANN-Backstepping	–	✓	✓	–	✓
[12]	FCS-MPC	–	–	–	–	–
[16]	FCS-MPC	–	–	–	–	–
[18]	FCS-MPC	–	✓	–	–	✓
[20]	ANN-MPC	–	✓	✓	–	–
[26]	DDPGiPI	–	✓	✓	–	✓
[27]	RL-MPC	–	✓	✓	–	–
This paper	DDPG-MPC	✓	✓	✓	✓	✓

ANN: Artificial Neural Network, DDPGiPI: Deep Deterministic Policy Gradient Intelligent Proportional-Integer.

The contribution of this paper can be summarized as follows:

- A learning-based FCS-MPC is proposed to regulate the output voltage of DG units in a DC-MG. A multi-agent DRL-based approach is used to provide an online and adaptive tuning of weighting coefficients of the FCS-MPC.
- Unlike the FCS-MPC with constant coefficients, which are typically designed for a specified operating condition, the proposed approach avoids the dependency of the converter control system on the operating conditions.
- Usually, the control design of the converters follows this presumption that the CPLs are ideal, while in practice, the CPLs are of unknown and/or time-varying character. Hence, the performance of the proposed controller is investigated against the power changes in the non-ideal CPLs.
- One of the critical issues in MGs is DGs' plug-and-play (PnP) operation due to the inherently discontinuous nature of renewable energy sources. To address this issue, the dynamic performance of the proposed controller is examined under the PnP operation of DG units.

2. Model of Microgrid

The diagram of a single-bus DC-MG with multiple DG units and loads is depicted in Figure 1. The buck converter used for each DG is regulated by the duty ratio of an IGBT switching to maintain the output voltage stable. The DGs are connected to a common DC bus through LC filters. The DC-MG is assumed to operate in an islanding model. For DG_i unit, the following differential equations can represent the output voltage and current of the converter:

$$DG_i = \begin{cases} \frac{dV_{oi}}{d_t} = \frac{1}{C_{ti}} I_{ti} - \frac{1}{C_{ti}} I_{L_i} \\ \frac{dI_{ti}}{d_t} = -\frac{R_{ti}}{L_{ti}} I_{ti} - \frac{1}{L_{ti}} V_{oi} + \frac{1}{L_{ti}} V_{ti} \end{cases} \quad (1)$$

where I_{L_i} and I_{t_i} are the currents of load and converter, respectively; V_{ti} is the converter's output voltage; V_{oi} represents the capacitor voltage; and L_{ti} and C_{ti} are the filter parameters. There is an assumption that buck converter dynamics, inherently switching, have been averaged over time. Nevertheless, this is a soft approximation for converters operating at high frequencies. The output voltage and current for DG_i can be described in the state-space form as follows:

$$DG_i : \begin{cases} \dot{x}_{[i]}(t) = A_{ii} x_{[i]}(t) + B_i u_{[i]}(t) + M_i d_{[i]}(t) \\ y_{[i]}(t) = C_i x_{[i]}(t) \end{cases} \quad (2)$$

where $x_{[i]} = [V_{oi}, I_{ti}]^T$ is the state variable; $u_{[i]} = V_{ti}$ is the control input; and $d_{[i]} = I_{L_i}$ is the exogenous input. The corresponding vectors and matrices are as follows:

$$A_{ii} = \begin{bmatrix} 0 & \frac{1}{C_{ti}} \\ -\frac{1}{L_{ti}} & -\frac{R_{ti}}{L_{ti}} \end{bmatrix} B_i = \begin{bmatrix} 0 \\ \frac{1}{L_{ti}} \end{bmatrix} M_i = \begin{bmatrix} -\frac{1}{C_{ti}} \\ 0 \end{bmatrix} C_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3}$$

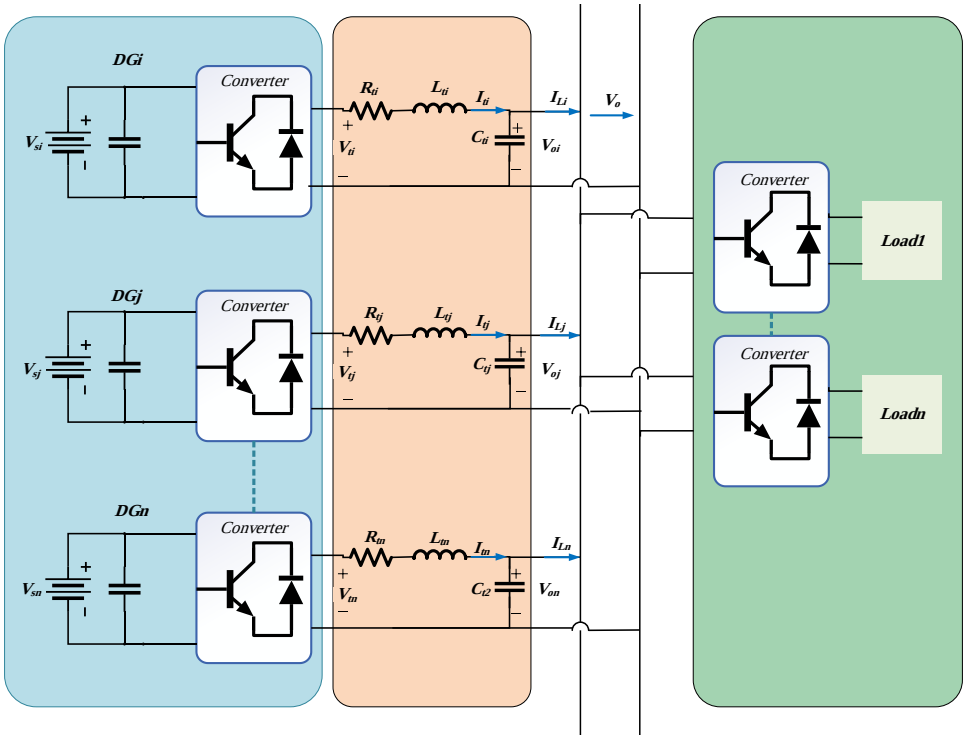

Figure 1. A diagram of a typical DC-MG.

Continuation of (3), the overall model of the MG consisting of three DGs (i.e., DG_i, DG_j and, DG_k) is expressed by:

$$\begin{bmatrix} \dot{x}_{[i]} \\ \dot{x}_{[j]} \\ \dot{x}_{[k]} \end{bmatrix} = A \begin{bmatrix} x_{[i]} \\ x_{[j]} \\ x_{[k]} \end{bmatrix} + \begin{bmatrix} B_i & 0 & 0 \\ 0 & B_j & 0 \\ 0 & 0 & B_k \end{bmatrix} \begin{bmatrix} u_{[i]} \\ u_{[j]} \\ u_{[k]} \end{bmatrix} + \begin{bmatrix} M_i & 0 & 0 \\ 0 & M_j & 0 \\ 0 & 0 & M_k \end{bmatrix} \begin{bmatrix} d_{[i]} \\ d_{[j]} \\ d_{[k]} \end{bmatrix}$$
$$\begin{bmatrix} y_{[i]} \\ y_{[j]} \\ y_{[k]} \end{bmatrix} = \begin{bmatrix} c_{[i]} & 0 & 0 \\ 0 & c_{[j]} & 0 \\ 0 & 0 & c_{[k]} \end{bmatrix} \begin{bmatrix} x_{[i]} \\ x_{[j]} \\ x_{[k]} \end{bmatrix} \tag{4}$$

where

$$A = \begin{bmatrix} A_{ii} & A_{ij} & A_{ik} \\ A_{ji} & A_{jj} & A_{jk} \\ A_{ik} & A_{kj} & A_{kk} \end{bmatrix} \tag{5}$$

It should be mentioned that due to the neglect of line dynamics, the matrices A_{ik}, A_{jk}, A_{ji}, A_{ki}, A_{ij} and A_{kj} are equal to zero.

3. Proposed Controller Design

Figure 2 illustrates the DRL-based FCS-MPC-operated converter. In this approach, a proper control command is obtained based on the prediction from the converter model and a objective function (OF). The Equations (2) and (3) describe the continuous state-space

model of a dc-dc buck converter. To get a discrete representation appropriate for a digital control system, this paper uses the zero-order hold (ZOH) discretization technique to discrete the continuous-time model. The discrete state-space model with a sampling time *Ts* can be described as follows [19]:

$$x(k+1) = A_d x(k) + B_d u(k) + M_d d(k) \tag{6}$$

where

$$A_d = e^{A_{ii} T_s} \tag{7}$$

$$B_d = \int_0^{T_s} e^{A_{ii}\tau} B_i d\tau \tag{8}$$

$$M_d = \int_0^{T_s} e^{A_{ii}\tau} M_i d\tau \tag{9}$$

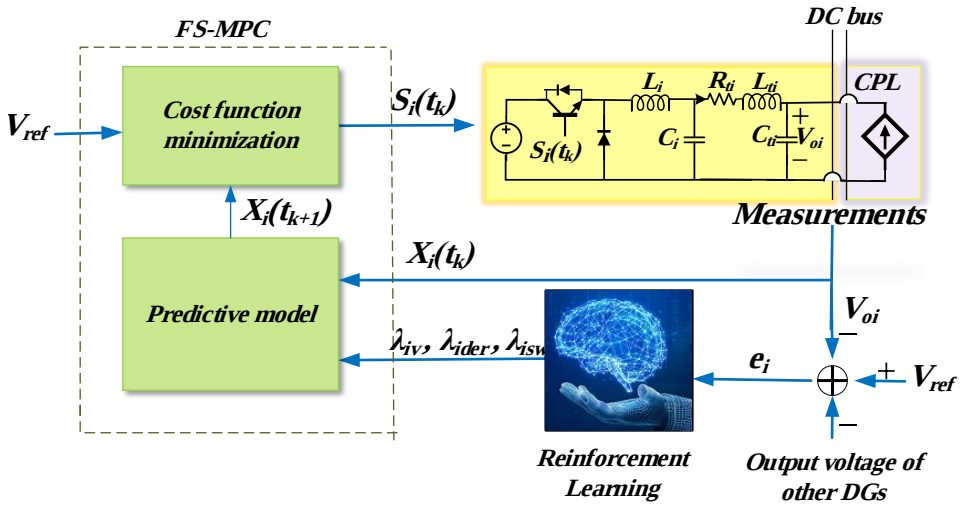

Figure 2. Proposed control scheme for a dc-dc buck converter.

These equations are used in the FCS-MPC prediction step.

System model and OF design are the two main stages of FCS-MPC controller design. The switching signal determines the voltage vector and has two initial states of zero and one. FCS-MPC method is mainly used in digital controls and works based on synchronized switching and sampling instants [28]. The main goal of the control system is to properly adjust the voltage V_{ti} so that the voltage V_{oi} can follow the reference voltage precisely. The fundamental function of FCS-MPC is to predict values of V_{oi} and I_{ti} and apply optimal V_{ti} based on a OF. The OF with minimum value is then executed to the converter. Therefore, determining an appropriate OF is a prominent part of the FSC-MPC approach.

OFs with multi-step prediction horizons have been offered to enhance the steady-state performance of the control system, which are typical for high-power multilevel converters. In contrast, a single-step prediction is typically a better choice in a converter with high switching frequencies. This implies more performance areas by using longer prediction horizons [29]. It should be noted that the implementation of a single-stage horizon requires less computation and is flexible in integrating linear and non-linear control objectives and constraints. This study uses a OF with a single-step prediction horizon. For dc voltage regulation on the converter, the OF of the single-step horizon is expressed as follows:

$$g_{con} = (V_{ref} - V_{oi}(k+1))^2 \tag{10}$$

where V_{ref} is the voltage reference.

Additional current reference term is added to improve the steady state performance:

$$g_c = (I_L - I_{ti}(k+1) + C_t\omega_r V_{oi}(k+1))^2 \tag{11}$$

where I_L is the load current; $V_{oi}(k+1)$ and $I_{ti}(k+1)$ are the predicted voltage and current; and $\omega_r = 2\pi f_r$ is the angular frequency.

The g_{con} and g_c terms are multiplied with weighting coefficients λ_v and λ_{der}. In addition, the current limiting term h_{lim}, and switching penalization term sw, are also defined:

$$h_{lim} = \begin{cases} 0, & if \quad |\bar{i}_t(k)| \leq i_{max} \\ \infty, & if \quad |\bar{i}_t(k)| > i_{max} \end{cases} \tag{12}$$

$$sw = |\Delta S(i)|^2 \tag{13}$$

where $|\Delta S(i)|$ is 1 if switch change happens at instant i and 0 otherwise. The terms expressed in (11)–(13) are added to (10), which eventually produces the modified OF as:

$$g_p = \lambda_v g_{con} + \lambda_{der} g_c + \lambda_{sw} sw + h_{lim} \tag{14}$$

As can be seen, the system performance is highly influenced by the weighting coefficients λ_v,λ_{der}, and λ_{sw}, which should be adjusted optimally. In [19], these weighting coefficients were tuned offline. However, once the operating point varies significantly, securing a good response is a challenging task. In this paper, the DRL approach is used to adjust the weighting coefficients in a online manner and quick way and thereby improve the performance of the converter control system. The design process of the multi-agent DRL-based regulation scheme will be discussed in the following section.

4. Multi-Agent DRL-Based Regulation Scheme

In this paper, the problem of the design of weighting coefficients, i.e., λ_v,λ_{der} and λ_{sw}, is formulated in a multi-agent DRL framework. Each DG has its controller, whose weighting coefficients are determined by the DRL agent in a distributed manner. In the distributed procedure, the DRL agent associated with each DG exchanges data with others. Thus, each DG uses a local reward function. The multi-agent DRL environment is a network model of a DC-MG including three DGs. DRL agents operate together to regulate the output voltage. A schematic of the proposed strategy is shown in Figure 3. This method includes two phases of offline and online learning. In the offline phase, the DRL agents follow a centralized learning process to explore the environment. A reward function is then generated to assess the actions generated by the agents. By updating critic and actor networks, each agent generates the optimal control command (updates the weighting coefficients) to improve the system performance. In the online phase, agents will take action in a distributed framework to determine weighting coefficients. Distributed control is one of the most desired communication-based control techniques that does not need a central controller. Each agent adjusts the FCS-MPC coefficients of its DG based on its observations e_i. The observation e_i for each DG, considering the communication network (communication links transfer measured data of each DG unit), is an error between the average of voltages broadcasted from each DG and the reference voltage, which is expressed by

$$e = V_{MG} - V_{ref} = \frac{\sum_{j=1}^{N} V_{oj}}{N} - V_{ref} \tag{15}$$

where N is the number of DGs in the MG.

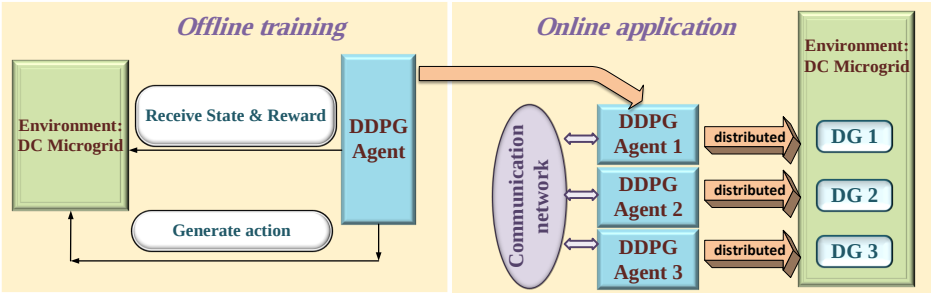

Figure 3. The schematic diagram of the proposed multi-agent DRL-based regulation scheme.

The DRL concept is to find the best policy together with the states and actions while getting maximum rewards through the interaction between the agent and the environment. The DRL problem is described as a Markov decision-making process (MDP). An MDP is defined with 5 parameters (s, A, P, R, γ) so that the s is the state, A is action, $P = s \times A \times s \Rightarrow [0, 1]$ is a state transition probability, where $P = p(s_{t+1} \mid s_t, a_t)$, $R : s \times A \Rightarrow R$ is reward and $\gamma \in [0, 1]$ is the discount factor. At any time step, the DRL agent monitors the state s_t and selects the appropriate action a_t according to the policy $\pi(a_t \mid s_t)$. Then, the agent observes the reward value r_t appropriate to this action and determines the next state (s_t) accordingly. The definition of discount reward with γ discount factor is as follows [26]:

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{16}$$

The goal is to maximize the discount return, that is:

$$J = E_{r_i, s_i \sim En, a_i \sim \pi}[G_1] \tag{17}$$

where En denotes the environment. Actions a_i are determined based on the policy π. In most DRL problems, the action-value function (AVF) expresses the anticipated return G_t after the action a_t is applied to the state s_t, and the AVF in DRL is illustrated as:

$$Q^{\pi}(s_t, a_t) = E_{r_{i \geq t}, s_{i \geq t} \sim En, a_{i>t} \sim \pi}[G_t \mid s_t, a_t] \tag{18}$$

Therefore, the main purpose of DRL is to calculate the AVF $Q^{\pi}(s_t, a_t)$ and find the appropriate policy value π accordingly.

Many RL methods use the Bellman equation to estimate the AVF, which is given by:

$$Q^{\pi}(s_t, a_t) = E_{r_t, s_{t+1} \sim E}\left[r_t(s_t, a_t) + \gamma E_{a_{t+1} \sim \pi}[Q^{\pi}(s_{t+1}, a_{t+1})]\right] \tag{19}$$

In the following, the DDPG algorithm is used to design the DRL agents. Figure 4 shows the execution process of the DDPG, consisting of two networks, the actor and the critic. The actor-network adjusts the weight of θ^{μ} of policy $\mu(s \mid \theta^{\mu})$ based on observation or state to the corresponding action, and the critic network modifies the weights of action function $Q(s, a \mid \theta^{Q})$. Critic coefficients are updated through minimization of the following loss function:

$$L(\theta^{Q}) = E_{(}s, a)[(Q(s_t, a_t \mid \theta^{Q}) - y_t)^2] \tag{20}$$

where

$$y_t = r_t(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_t \mid \theta^{\mu}) \mid \theta^{Q}) \tag{21}$$

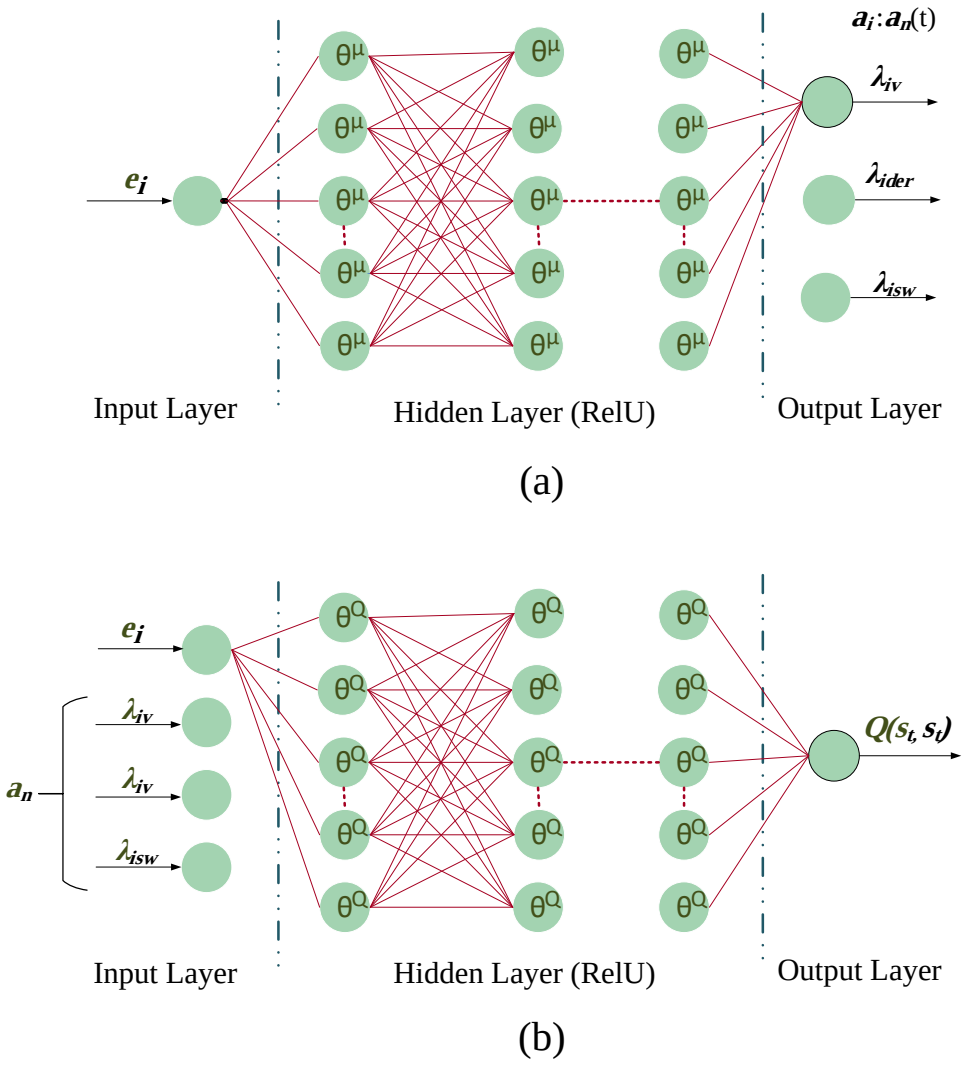

Figure 4. Illustration of the (**a**) actor-network (**b**) critic-network.

Furthermore, the actor coefficients θ^{μ} are updated as:

$$\begin{aligned} \nabla_{\theta^{\mu}} J^{\theta^{\mu}} &\approx E_{s_t \sim \rho^{\beta}}[\nabla_{\theta^{\mu}} Q(s, a \mid \theta^{Q}) \mid_{a=\mu(s|\theta^{\mu})} \nabla_{\theta^{\mu}} \mu(s \mid \theta^{\mu})] \\ &= E_{s_t \sim \rho^{\beta}}[\nabla_{a} Q(s, a \mid \theta^{Q}) \mid_{a=\mu^{\theta}(s)} \nabla_{\theta^{\mu}} \mu(s \mid \theta^{\mu})] \end{aligned} \tag{22}$$

where ρ is a discounted distribution; and β is a specific policy to the current policy π. Also, an exploration noise W has been added to the actor actions (i.e., $a_t = \mu(s_t \mid \theta^{\mu}) + W$) to improve the training process [30]. The design of actor and critic networks in this paper consists of an input layer, an output layer, and three hidden layers, including 80, 80, and 30 neurons between the input and output layers, shown in Figure 4. The input signal to the actor network is a vector state of the e, and its outputs are λ_v, λ_{der} and λ_{sw}. The developed control system aims to minimize the output voltage error in the shortest possible time to stabilize the MG-DC. Hence, the reward function is considered as:

$$r_t = \frac{1}{\mid e \mid} = \frac{1}{V_{MG} - V_{ref}} \tag{23}$$

Based on the reward signal, the weight coefficients of the actor and critic networks are trained in such a way that the error between the reference voltage V_{ref} and the average value of the output voltage of the DGs is minimized. Flowchart of the proposed multi-

agent DRL based design of the weighting coefficients in the FCS-MPC is shown in Figure 5. The DDPG design process for each agent is available in Algorithm 1.

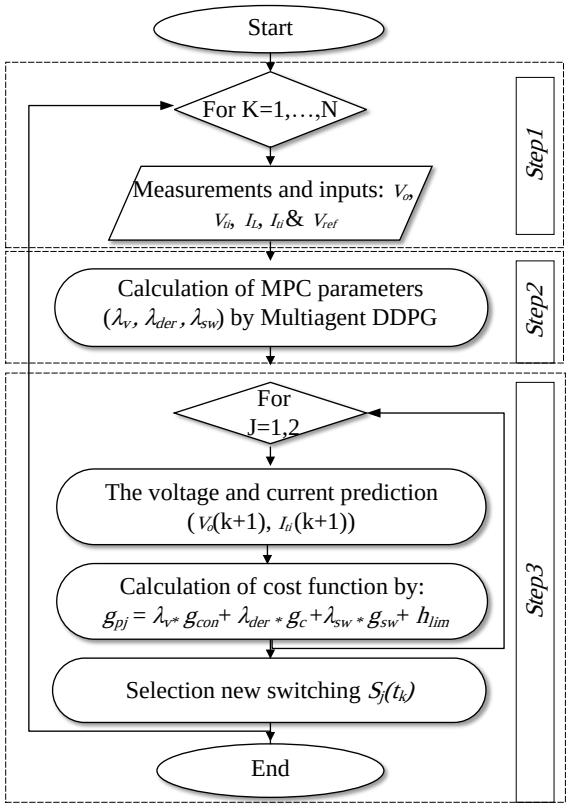

Figure 5. Flowchart of the proposed multi-agent DRL based design of the weighting coefficients in the FCS-MPC.

Algorithm 1 The pseudo-code for the standard DDPG

Randomly initialize critic $Q(s, a \mid \theta^Q)$ and actor $\mu(s \mid \theta^\mu)$ networks with weights θ^Q and θ^μ, respectively.
Initialize Q' and μ' networks based on new weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
for episode = 1 to M **do**
  Start with Ornstein-Uhlenbeck Noise (OU) for exploration and get the initial observation state s_1
  for t = 1 to T **do**
    Select the control actions (λ_v, λ_{der} and λ_{sw}) according to $a_t = \mu(s_t \mid \theta^\mu) + W$.
    Apply action a_t to the environment and observe the e as the next state s_{t+1}.
    Calculate the reward r_t according to Equation (23) by difference between the values of the simulated and observed behavior.
    Save (s_t, a_t, r_t, s_{t+1}) into the replay buffer F.
    Sample random minibatch of m transition from F.
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} \mid \theta^{\mu'}) \mid \theta^{Q'})$
    Update critic with the loss: $L = \frac{1}{m}\Sigma_i (y_i - Q(s_i, a_i \mid \theta^Q))^2$
    Update the actor policy based on the sampled policy gradient: $\nabla_{\theta^\mu} J^{\theta^\mu} \approx \frac{1}{m}\Sigma_i \nabla_a Q(s, a \mid \theta^Q) \mid_{a=\mu^\theta(s)} \nabla_{\theta^\mu} \mu\ (s \mid \theta^\mu)$
    Update the target network:
    $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \theta^{Q'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
  end for
end for

5. Simulation Results

Various case studies are conducted to assess the performance of the proposed control scheme. The DC-MG shown in Figure 1, including three DGs interfaced with dc-dc buck converters, is simulated in MATLAB/Simulink software environment. The specifications of the studied network are available in Table 2, and the design parameters of the DDPG algorithm can be found in Table 3. The proposed scheme is tested in the following scenarios:

- Unknown load dynamics
- Variation of input voltage
- PnP operation
- Variation of reference voltage

Table 2. Parameters of the test system.

Parameters	Values
DG1 parameters	$R_1 = 0, L_1 = 2.5$ mH, $C_1 = 83.3$ µF
DG2 parameters	$R_2 = 0, L_2 = 0.14$ mH, $C_2 = 18.75$ µF
DG3 parameters	$R_3 = 0, L_3 = 0.21$ mH, $C_3 = 12.5$ µF
LC filter on the DG1	$R_{t1} = 0.2, L_{t1} = 450$ µH, $C_{t1} = 220$ µF
LC filter on the DG2	$R_{t2} = 0.2, L_{t2} = 450$ µH, $C_{t2} = 220$ µF
LC filter on the DG1	$R_{t3} = 0.2, L_{t3} = 450$ µH, $C_{t3} = 220$ µF
Input voltage	$V_s = 300$ V
Sampling time	$T_s = 20$ µs
Switching frequency	$f_r = 10$ kHz
CPL	$P_{CPL} = 120$ W
Reference voltage	$V_{ref} = 200$ V

Table 3. Parameters settings of the DDPG.

Parameters	Values
Discount factor, γ	0.9995
Learning rate, λ	0.0001
Mini-batch size	128
Reply buffer size	1,000,000

5.1. Study 1: Unknown Load Dynamics

This study demonstrates the robust performance of the proposed control scheme against sudden load changes. For this purpose, the load power declines to half of its initial value at $t = 0.2$ s, returning to the previous value at $t = 0.3$ s. The deviations in voltage and current of the load are shown in Figures 6 and 7, respectively. As shown in Figure 6, the load voltage remains constant once the load power changes, indicating the proper performance of the proposed scheme. The output voltage of each DG is illustrated in Figure 8. As the figure presents, the overshoot occurrences are minor, denoting the effectiveness of the proposed control scheme. The generated weighting coefficients for each DG by the DRL-based approach under the applied load changes are presented in Figures 9–11. The figures reveal that the DRL-based method regulates the weighting coefficients such that a stable operation of the DC-MG is achieved.

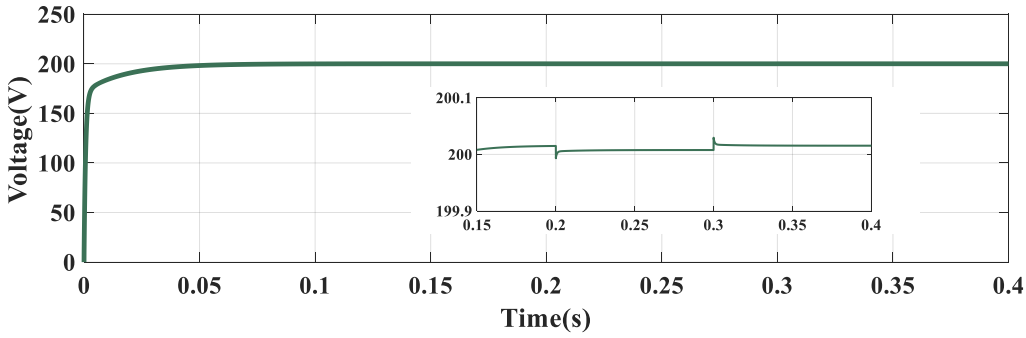

Figure 6. Study 1: load voltage.

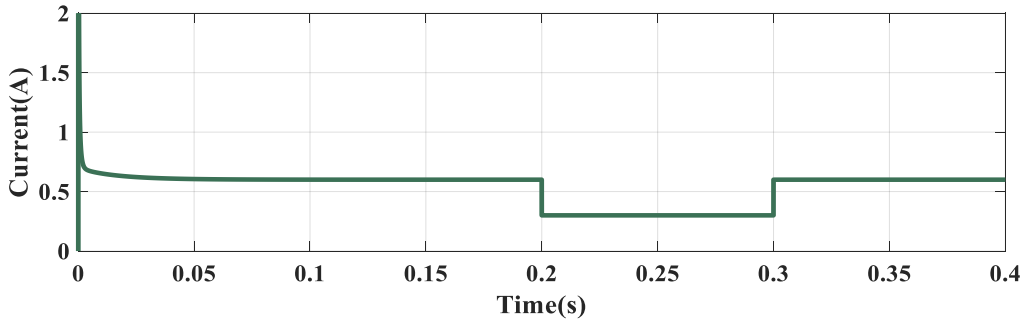

Figure 7. Study 1: load current.

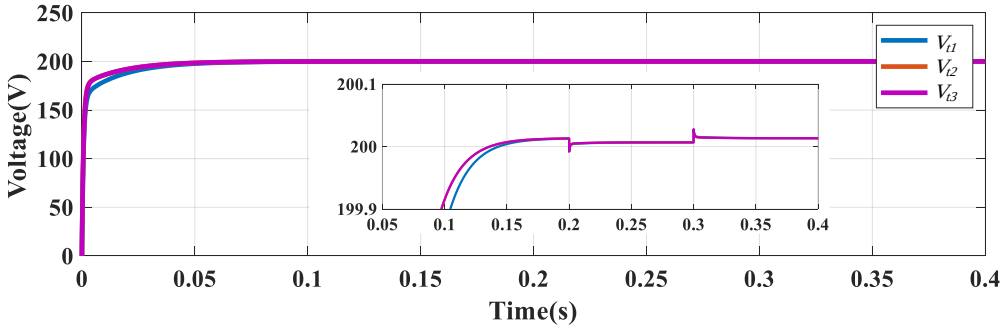

Figure 8. Study 1: output voltage of each DG.

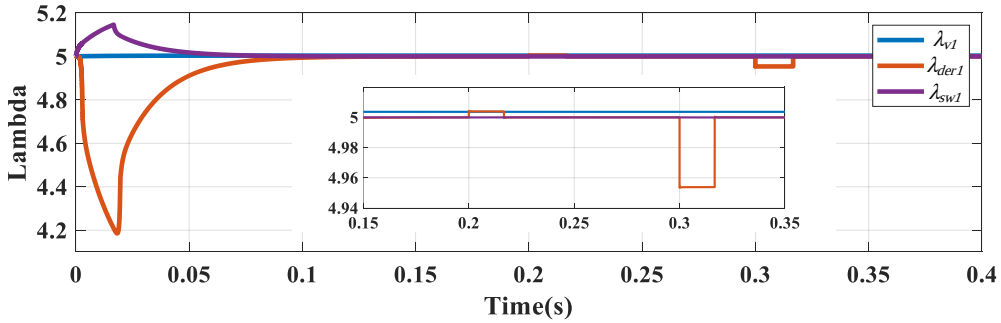

Figure 9. Study 1: generated weighting coefficients for FCS-MPC in DG1.

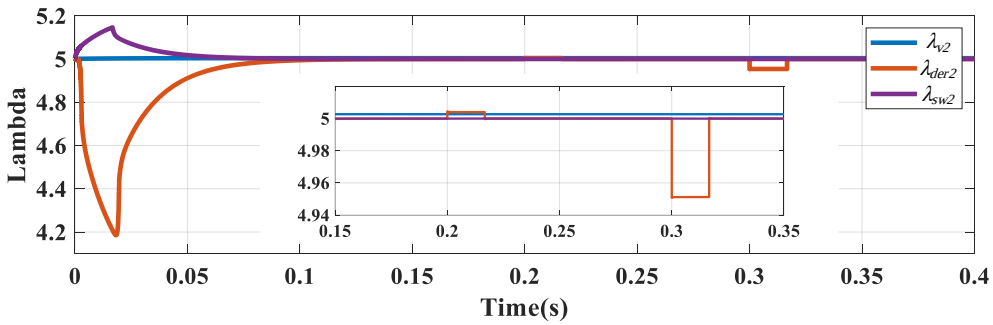

Figure 10. Study 1: generated weighting coefficients for FCS-MPC in DG2.

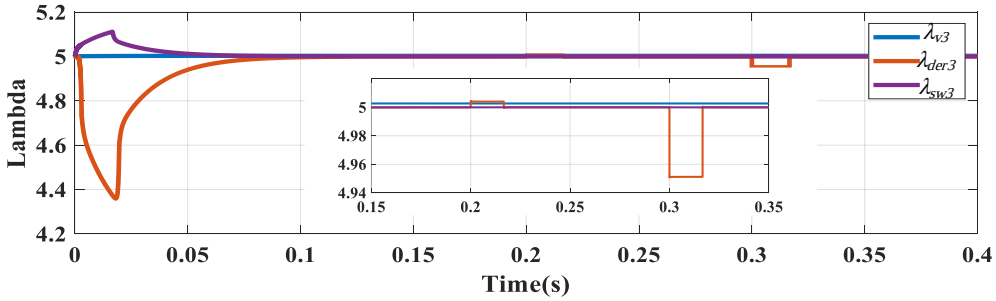

Figure 11. Study 1: generated weighting coefficients for FCS-MPC in DG3.

5.2. Study 2: Input Voltage Variations

This study examines the robustness of the proposed control scheme under the input voltage changes. A step increase of 10 v at $t = 1$ s is assumed in the input dc voltage of the buck converters. The load voltage and current are shown in Figures 12 and 13, respectively. As the figure indicates, the proposed scheme could withdraw the offset induced by input voltage deviation in a short time. The deviations in output voltages of DGs are shown in Figure 14. As shown, by use of the proposed scheme, the voltage fluctuation of DG1 is less than 1%, which is quite satisfactory. Similar argument is correct about the rest of DGs. From the figures it can be concluded that the current has more changes in the return of voltage to the reference value. The generated weighting coefficients by the DRL method are presented in Figures 15–17 in response to the input voltage variation.

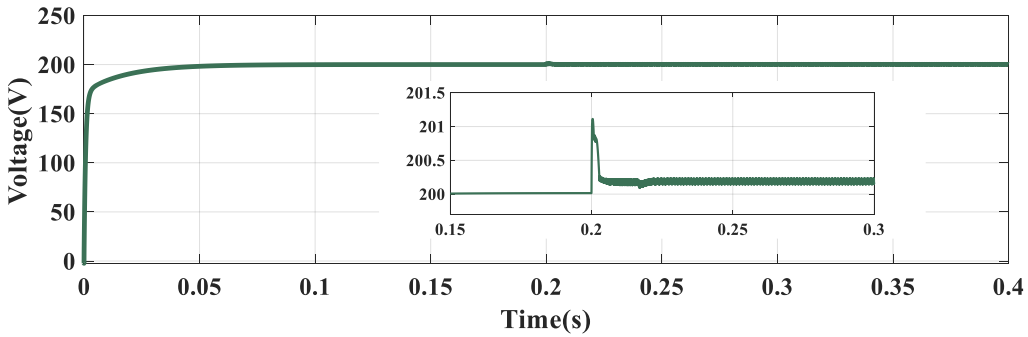

Figure 12. Study 2: load voltage.

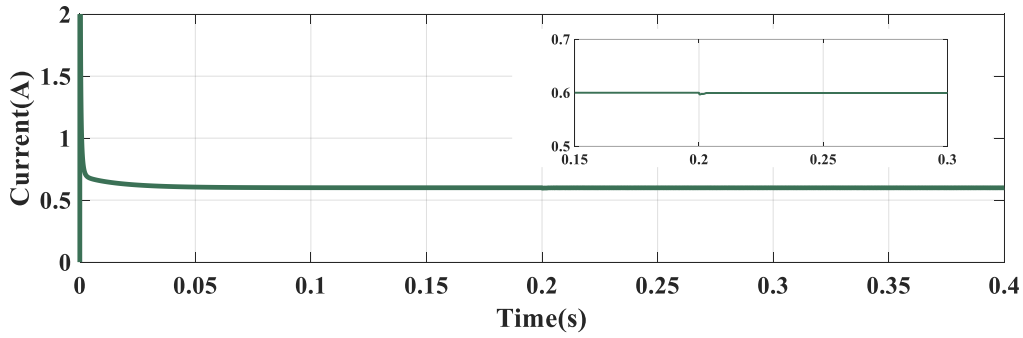

Figure 13. Study 2: Load current.

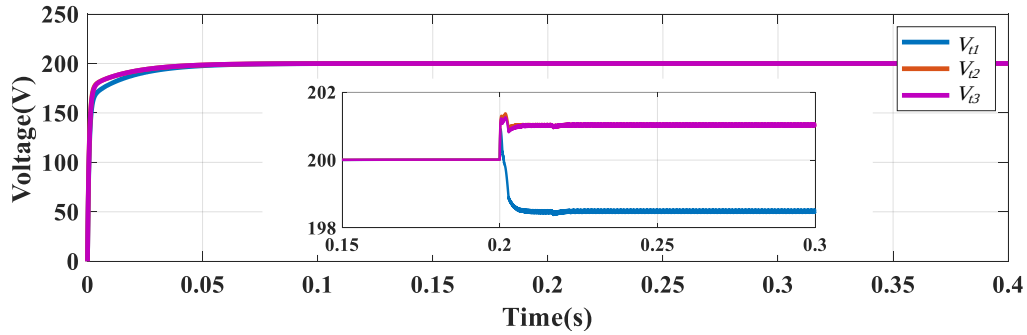

Figure 14. Study 2: output voltage of each DG.

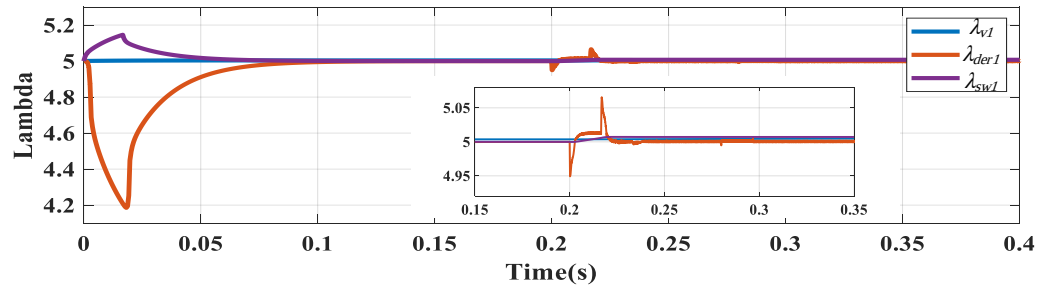

Figure 15. Study 2: generated weighting coefficients for FCS-MPC in DG1.

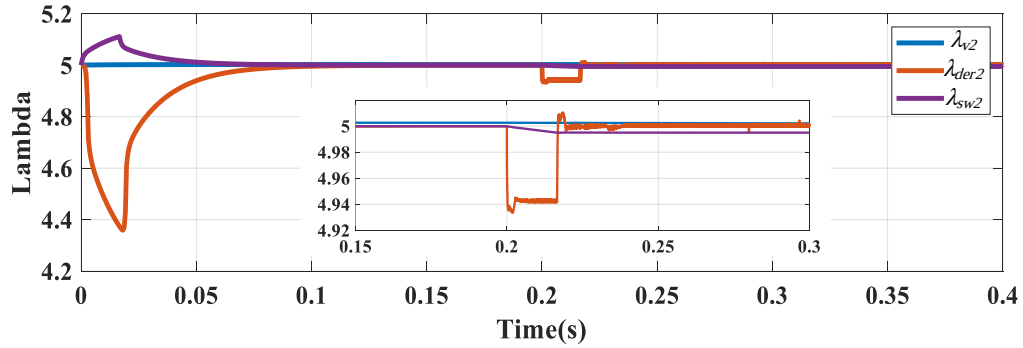

Figure 16. Study 2: generated weighting coefficients for FCS-MPC in DG2.

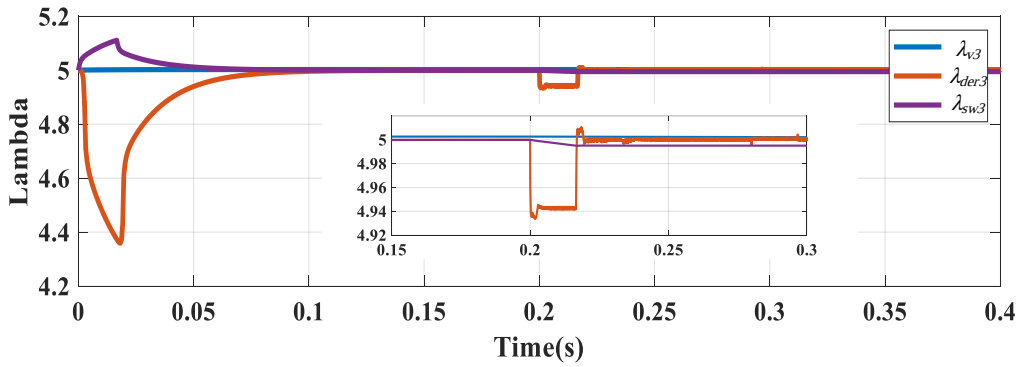

Figure 17. Study 2: generated weighting coefficients for FCS-MPC in DG3.

5.3. Study 3: PnP Operation

Here, the PnP capability of the proposed scheme is examined. For this goal, it is assumed that DG3 is disconnected from the MG at $t = 0.3$ s, and then, is connected at $t = 0.4$ s. The load voltage and current are displayed in Figures 18 and 19. As it can be seen, the voltage variations in stable time are less than 0.02 volt and has achieved the control goals with the quick reaction time. Also, the variation of load current is almost equal to the nominal value of the system. Figure 20 shows the output voltage of each DGU. As seen at t = 0.3, the voltage of DG3 when it is plugged out drops about 0.05 volt, and the voltage of DG1 and DG2 increase by 0.01 volt, which is almost equal to the reference voltage. The advantage is that each unit is controlled separately, which is not possible in centralized controls. As well, the generated weighting coefficients by DRL are illustrated in Figures 21–23. The DRL generates the weighting coefficients online and dynamically during transient state and after disconnection of DG3 from the DC-MG.

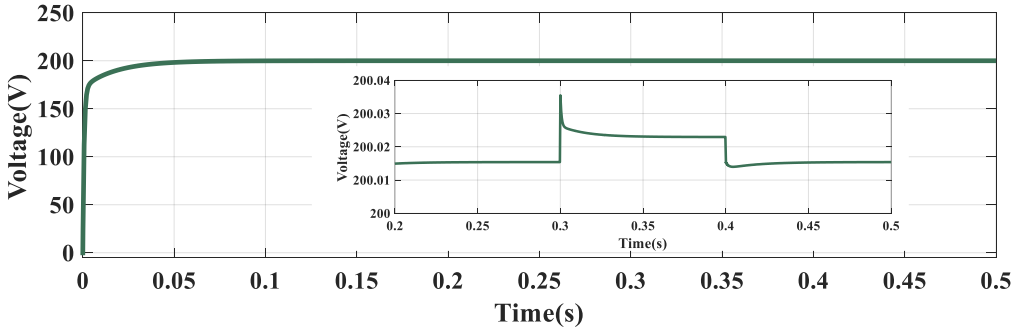

Figure 18. Study 3: load voltage.

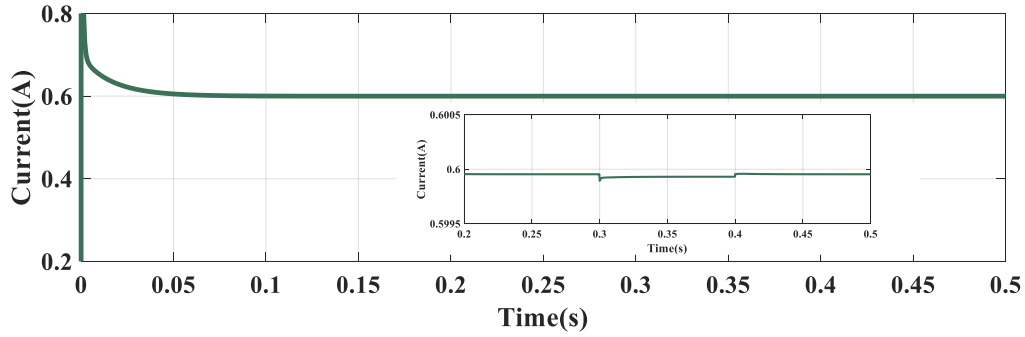

Figure 19. Study 3: load current.

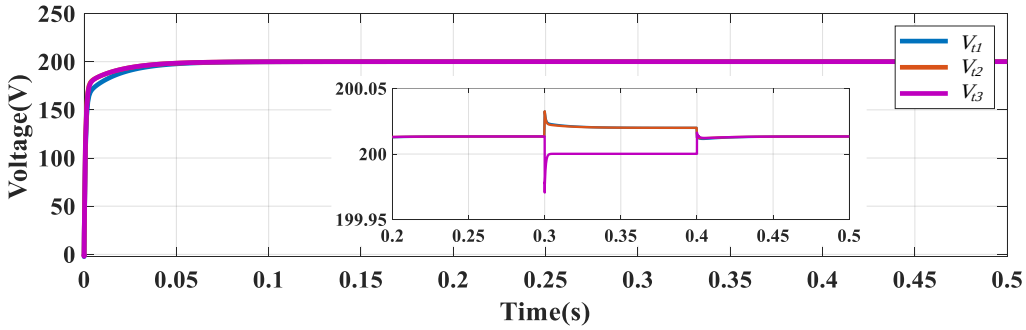

Figure 20. Study 3: output voltage of each DG.

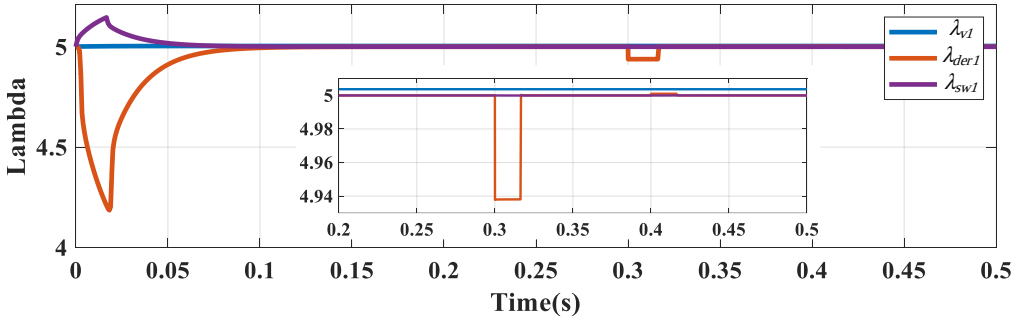

Figure 21. Study 3: generated weighting coefficients for FCS-MPC in DG1.

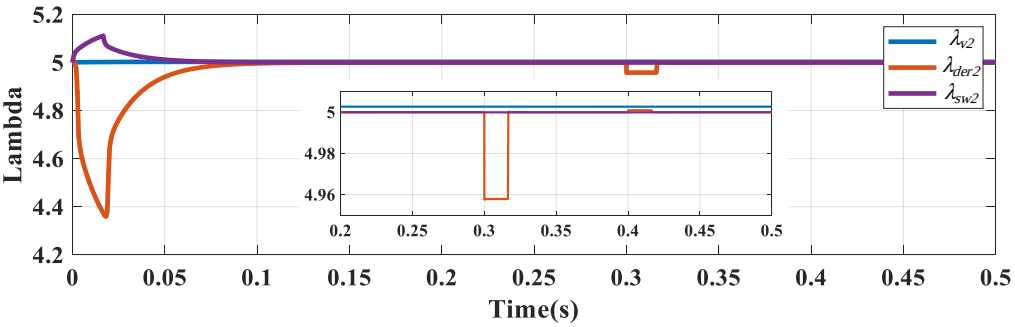

Figure 22. Study 3: generated weighting coefficients for FCS-MPC in DG2.

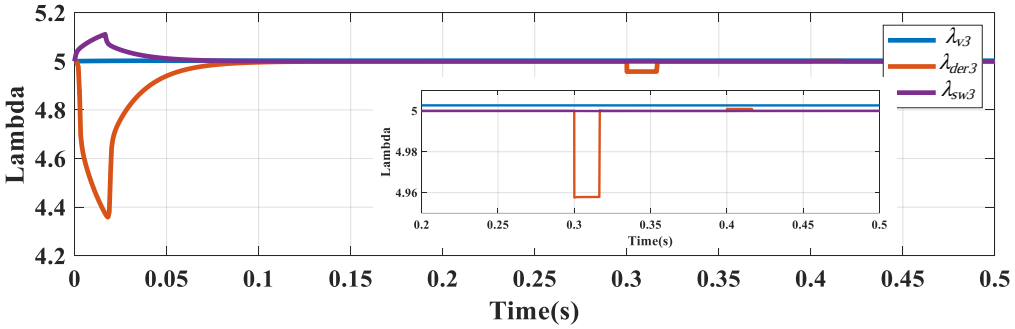

Figure 23. Study 3: generated weighting coefficients for FCS-MPC in DG3.

5.4. Study 4: Variation of Reference Voltage

The voltage reference changes may be required to adjust the current between the DG units or to control the state of charge of batteries embedded in the islanded MG. In this study, the performance of the proposed strategy is evaluated under the variation of reference voltage. For this end, at t = 0.3 s the reference voltage is reduced to 180 volts and at t = 0.6 it returns to the initial value of 200 volts. It means, the reference voltage has changed by 10%. A comparison is also made to examine the effect of weighting coefficients tuned by the multi-agent DRL approach and with those tuned by a trial and error method. Figures 24 and 25 illustrate the voltage and current of the load. As shown, the voltage reaches the reference value in a proper time without any ripple, overshoot, or undershoot. Similar view is correct about the current, where it changes in such a way that it can produce a constant power. Figure 26 shows the output voltage of each DG. The weighting coefficients variation of FCS-MPC in DG1, DG3 and DG3 are available in Figures 27–29. The figures indicate that the DRL-regulation scheme regulates the weighting coefficients in such a way that the least fluctuations are achieved in the dynamic responses. Another study is conducted under reference voltage changes, wherein the weighting coefficients are fixed. The coefficients λ_v,λ_{der} and λ_{sw} are considered equal to 4.9, 4.65, and 5, respectively. The load voltage and current in this case when the reference voltage changes is shown in Figures 30 and 31, respectively. As it can be observed, although the control system with fixed coefficients can keep the voltage constant at 200 volts, when the reference voltage changes, it can not follow the new value correctly, and there is an error of 2.8 volts.

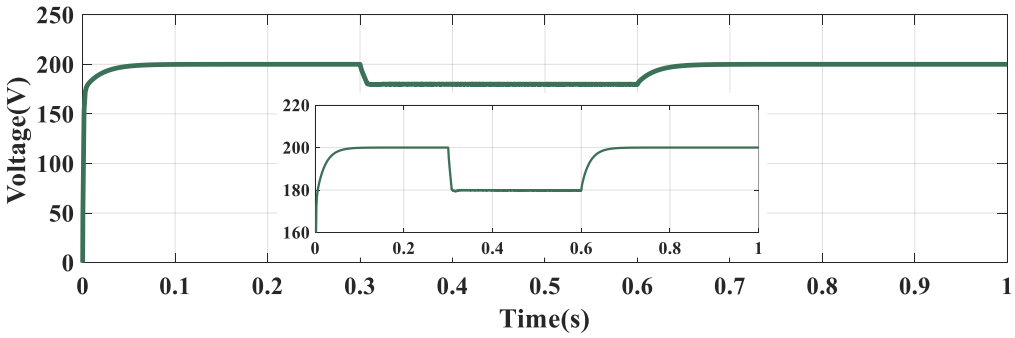

Figure 24. Study 4: load voltage.

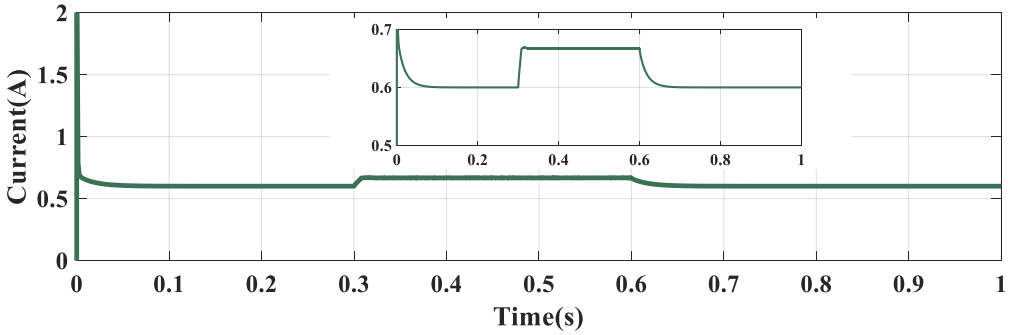

Figure 25. Study 4: load current.

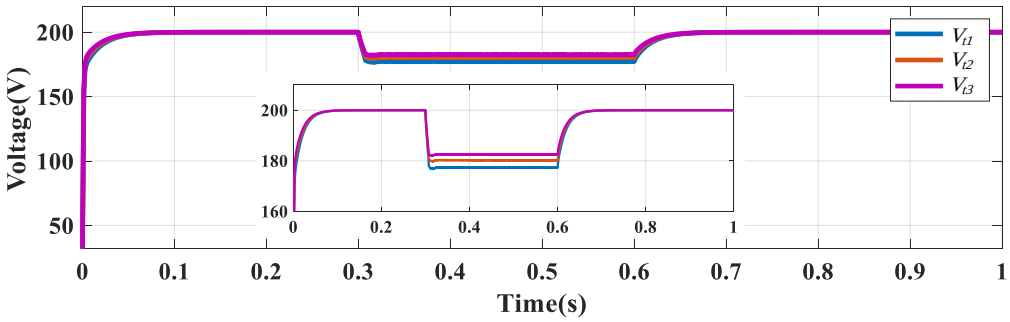

Figure 26. Study 4: output voltage of each DG.

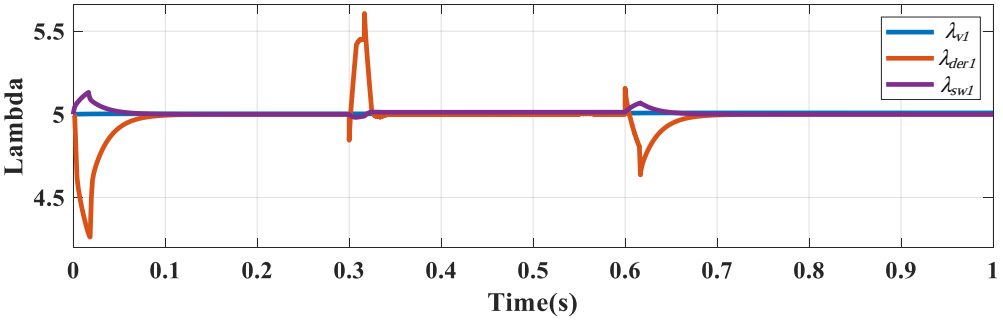

Figure 27. Study 4: generated weighting coefficients for FCS-MPC in DG1.

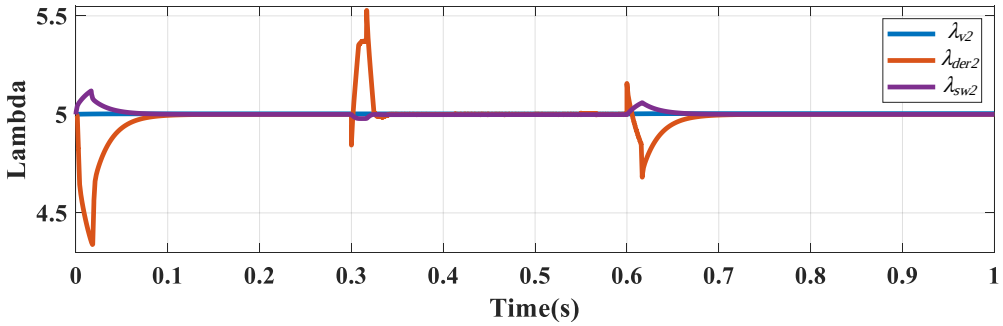

Figure 28. Study 4: generated weighting coefficients for FCS-MPC in DG2.

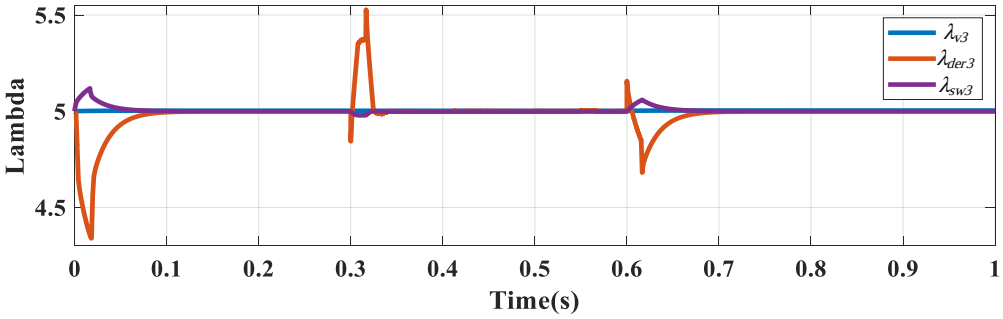

Figure 29. Study 4: generated weighting coefficients for FCS-MPC in DG3.

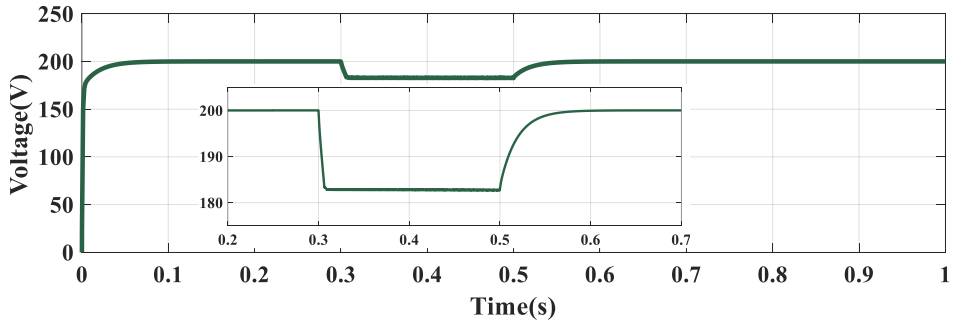

Figure 30. Study 4: load voltage (fixed weighting coefficients).

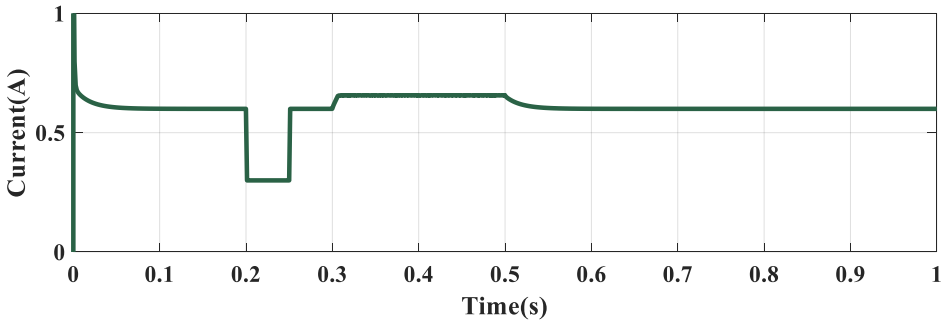

Figure 31. Study 4: load current (fixed weighting coefficients).

6. Conclusions

This study proposed a real-time solution employing the multi-agent DRL algorithm to design the weighting coefficients appearing in the FCS-MPC used for buck converters interfaced with CPLs in a DC-MG. A DDPG method is employed to learn the optimal weighting coefficient design policy. Minimizing the voltage and current divisions of each DG were the main objectives behind the DRL-based FCS-MPC method. The proposed method's key features are the online learning capacity, minimal computational complexity, and no need for prior knowledge of MG dynamics. Finally, the simulation results obtained from a benchmark DC-MG with three DGs demonstrated the effectiveness of the proposed solution with different operational conditions. For example, it was shown that, by use of the proposed scheme, the voltage fluctuation of DG1 under input voltage variations is less than 1%, which is quite satisfactory. The results confirmed that the proposed control scheme: (1) has superior performance in comparison with FCS-MPC with fixed weighting coefficients; (2) indicates a robust performance against the uncertainties such as input and reference voltage variation; (3) deals with the power changes in the non-ideal CPLs; (4) presents the PnP capability; and (5) avoids the dependency of the converter control system on the operating point conditions, thereby supporting a wide range of operating conditions.

Author Contributions: Conceptualization, H.S., A.O., M.N., F.B. and A.A.-M.; methodology, H.S., A.O. and M.N.; software, H.S. and A.O.; validation, M.N., F.B. and A.A.-M.; formal analysis, H.S., A.O. and M.N.; resources, H.S.; data curation, H.S., A.O. and M.N.; writing—original draft preparation, H.S. and A.O.; writing— review and editing, M.N., F.B. and A.A.-M.; supervision, F.B. and A.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Reliable Power Electronic-Based Power Systems (REPEPS) project at the AAU Energy Department, Aalborg University, as a part of the Villum Investigator Program funded by the Villum Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Acronyms

FCS-MPC	Finit Control Set-Model Predictive Control
DC-MG	DC Microgrid
MG	Microgrid
DG	Distributed Generation
CPL	Constant Power Load
SMC	Sliding Mode Control
PI	Proportional-integer
MPC	Model Predictive Control
UPS	Uninterrupted Power Supply
RL	Reinforcement Learning
DRL	Deep Reinforcemnt Learning
DQN	Deep Q Network
DDPG	Deep Deterministic Policy Gradient
PnP	Pluy and Play
OF	Objective Function
ZOH	Zero-Order Hold
MDP	Decision-making process
AVF	Active-Value Function
Variables and Parameters	
I_{L_i}	Load current
I_{t_i}	Converter current
V_{ti}	Converter's output voltage
V_{oi}	Capacitor voltage
C_{ti}, L_{ti}	Filter parameter
x_i	State variable
u_i	Control input
d_i	Exogenous input
V_{ref}	Reference voltage
ω_r	Angular frequency
$V_{oi}(k+1)$	Predicted voltage
$I_{ti}(k+1)$	Predicted current
λ_v	Voltage weighting coefficient
λ_{sw}	Switching weighting coefficient
λ_{der}	Current weighting coefficient
h_{lim}	Current limiting term
sw	Switching penalization
e	Error between the average of voltages broadcasted from each DG and the reference voltage
N	Number of DGs
s	State
A	Action
P	State transition probility
R	Reward
F	Replay buffer
m	Total number of transitions in the replay buffer F
γ	Discount factor
λ	Learning rate
G_t	Anticipated return
J	Discount return
E_n	Environment
ρ	Discounted distribution
β	Specific policy to the current policy π

References

1. Oshnoei, S.; Aghamohammadi, M.; Oshnoei, S.; Oshnoei, A.; Mohammadi-Ivatloo, B. Provision of Frequency Stability of an Islanded Microgrid Using a Novel Virtual Inertia Control and a Fractional Order Cascade Controller. *Energies* **2021**, *14*, 4152. [CrossRef]
2. Oshnoei, S.; Aghamohammadi, M.; Oshnoei, S. A novel fractional order controller based on fuzzy logic for regulating the frequency of an Islanded Microgrid. In Proceedings of the International Power System Conference (PSC), Tehran, Iran, 9–11 December 2019; pp. 320–326.
3. Aguirre, M.; Kouro, S.; Rojas, C.A.; Vazquez, S. Enhanced Switching Frequency Control in FCS-MPC for Power Converters. *IEEE Trans. Ind. Electron.* **2021**, *68*, 2470–2479. [CrossRef]
4. De Bosio, F.; De Souza Ribeiro, L.A.; Freijedo, F.D.; Pastorelli, M.; Guerrero, J.M. Effect of State Feedback Coupling and System Delays on the Transient Performance of Stand-Alone VSI with LC Output Filter. *IEEE Trans. Ind. Electron.* **2016**, *63*, 4909–4918. [CrossRef]
5. Li, D.; Ho, C.N.M. A Module-Based Plug-n-Play DC Microgrid with Fully Decentralized Control for IEEE Empower a Billion Lives Competition. *IEEE Trans. Power Electron.* **2021**, *36*, 1764–1776. [CrossRef]
6. Zhu, X.; Meng, F.; Xie, Z.; Yue, Y. An Inertia and Damping Control Method of DC-DC Converter in DC Microgrids. *IEEE Trans. Energy Convers.* **2020**, *35*, 799–807. [CrossRef]
7. Sorouri, H.; Sedighizadeh, M.; Oshnoei, A.; Khezri, R. An intelligent adaptive control of DC–DC power buck converters. *Int. J. Electr. Power Energy Syst.* **2022**, *141*, 108099. [CrossRef]
8. Kwasinski, A.; Onwuchekwa, C.N.; Member, S. Dynamic Behavior and Stabilization of DC Microgrids With Instantaneous Constant-Power Loads. *IEEE Trans. Power Electron.* **2011**, *26*, 822–834. [CrossRef]
9. Hossain, E.; Perez, R.; Nasiri, A.; Padmanaban, S. A Comprehensive Review on Constant Power Loads Compensation Techniques. *IEEE Access* **2018**, *6*, 33285–33305. [CrossRef]
10. Cespedes, M.; Xing, L.; Sun, J. Constant-power load system stabilization by passive damping. *IEEE Trans. Power Electron.* **2011**, *26*, 1832–1836. [CrossRef]
11. Dragicevic, T.; Vazquez, S.; Wheeler, P. Advanced control methods for power converters in DG systems and microgrids. *IEEE Trans. Ind. Electron.* **2020**, *68*, 5847–5862. [CrossRef]
12. Garcia, C.; Mohammadinodoushan, M.; Yaramasu, V.; Norambuena, M.; Davari, S.A.; Zhang, Z.; Khaburi, D.A.; Rodriguez, J. FCS-MPC based pre-filtering stage for computational efficiency in a flying capacitor converter. *IEEE Access* **2021**, *9*, 111039–111049. [CrossRef]
13. Karamanakos, P.; Geyer, T.; Kennel, R. Computationally efficient optimization algorithms for model predictive control of linear systems with integer inputs. In Proceedings of the 54rd IEEE Conference on Decision and Control, Osaka, Japan, 15–18 December 2015; pp. 3663–3668. [CrossRef]
14. Grainger, B.M.; Zhang, Q.; Reed, G.F.; Mao, Z.H. Modern controller approaches for stabilizing constant power loads within a DC microgrid while considering system delays. In Proceedings of the 2016 IEEE 7th International Symposium on Power Electronics for Distributed Generation Systems (PEDG 2016), Vancouver, BC, Canada, 27–30 June 2016; pp. 3–8. [CrossRef]
15. Tiwari, R.; Ramesh Babu, N.; Arunkrishna, R.; Sanjeevikumar, P. Comparison between PI controller and fuzzy logic-based control strategies for harmonic reduction in grid-integrated wind energy conversion system. *Lect. Notes Electr. Eng.* **2018**, *435*, 297–306. [CrossRef]
16. Yoo, H.J.; Nguyen, T.T.; Kim, H.M. MPC with constant switching frequency for inverter-based distributed generations in microgrid using gradient descent. *Energies* **2019**, *12*, 1156. [CrossRef]
17. Ding, D.; Yeganeh, M.S.; Mijatovic, N.; Wang, G.; Dragicevic, T. Model predictive control on three-phase converter for PMSM drives with a small DC-link capacitor. In Proceedings of the 2021 IEEE International Conference on Predictive Control of Electrical Drives and Power Electronics (PRECEDE), Jinan, China, 18–20 September 2021; pp. 224–228.
18. Sorouri, H.; Sedighizadeh, M. Robust control of DC-DC converter supplying constant power load with Finite-Set Model Predictive Control. In Proceedings of the 12th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC), Tabriz, Iran, 2–4 February 2021; pp. 1–3.
19. Khorsandi, A.; Ashourloo, M.; Mokhtari, H.; Iravani, R. Automatic droop control for a low voltage DC microgrid. *IET Gener. Transm. Distrib.* **2016**, *10*, 41–47. [CrossRef]
20. Dragičević, T.; Novak, M. Weighting Factor Design in Model Predictive Control of Power Electronic Converters: An Artificial Neural Network Approach. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8870–8880. [CrossRef]
21. Khezri, R.; Oshnoei, A.; Oshnoei, S.; Bevrani, H.; Muyeen, S. M. An intelligent coordinator design for GCSC and AGC in a two-area hybrid power system. *Appl. Soft Comput. J.* **2019**, *76*, 491–504. [CrossRef]
22. Oshnoei, A.; Sadeghian,O.; Mohammadi-Ivatloo, B.; Freijedo, F.D. Anvari-Moghaddam, A. Data-driven coordinated control of AVR and PSS in power systems: A deep reinforcement learning method. In Proceedings of the 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Bari, Italy, 7–10 September 2021. [CrossRef]
23. Yan, Z.; Xu, Y. Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search. *IEEE Trans. Power Syst.* **2019**, *34*, 1653–1656. [CrossRef]

24. Zhu, J.; Zhu, J.; Wang, Z.; Guo, S.; Xu, C. Hierarchical Decision and Control for Continuous Multitarget Problem: Policy Evaluation with Action Delay. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 464–473. [CrossRef]
25. Wang, Y.; Sun, J.; He, H.; Sun, C. Deterministic Policy Gradient with Integral Compensator for Robust Quadrotor Control. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 3713–3725. [CrossRef]
26. Gheisarnejad, M.; Farsizadeh, H.; Khooban, M.H. A Novel Nonlinear Deep Reinforcement Learning Controller for DC-DC Power Buck Converters. *IEEE Trans. Ind. Electron.* **2021**, *68*, 6849–6858. [CrossRef]
27. Wan, Y.; Dragičević, T.; Mijatovic, N.; Li, C.; Rodriguez, J. Reinforcement learning based weighting factor design of model predictive control for power electronic converters. In Proceedings of the IEEE International Conference on Predictive Control of Electrical Drives and Power Electronics (PRECEDE), Jinan, China, 18–20 September 2021.
28. Riar, B.S.; Scoltock, J.; Madawala, U.K. Model Predictive Direct Slope Control for Power Converters. *IEEE Trans. Power Electron.* **2017**, *32*, 2278–2289. [CrossRef]
29. Sampedro, C.; Bavle, H.; Rodriguez-Ramos, A.; de la Puente, P.; Campoy, P. Laser-based reactive navigation for multirotor aerial robots using deep reinforcement learning. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1024–1031.
30. Dragičević, T. Dynamic Stabilization of DC Microgrids with Predictive Control of Point-of-Load Converters. *IEEE Trans. Power Electron.* **2018**, *33*, 10872–10884. [CrossRef]