*Article*

# Twin-Delayed Deep Deterministic Policy Gradient Algorithm to Control a Boost Converter in a DC Microgrid

**Rifqi Firmansyah Muktiadji [1,*]** , **Makbul A. M. Ramli [1]** and **Ahmad H. Milyani [1,2]**

1   Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia; mramli@kau.edu.sa (M.A.M.R.); ahmilyani@kau.edu.sa (A.H.M.)
2   Center of Research Excellence in Renewable Energy and Power Systems, K.A.CARE Energy Research and Innovation Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia
*   Correspondence: rmuktiadji@stu.kau.edu.sa; Tel.: +966-562-052-769

**Abstract:** A stable output voltage of a boost converter is vital for the appropriate functioning of connected devices and loads in a DC microgrid. Variations in load demands and source uncertainties can damage equipment and disrupt operations. In this study, a modified twin-delayed deep deterministic policy gradient (TD3) algorithm is proposed to regulate the output voltage of a boost converter in a DC microgrid. TD3 optimizes PI controller gains, which ensure system stability by employing a non-negative, fully connected layer. To achieve optimal gains, multi-deep reinforcement learning agents are trained. The agents utilize the error signal to obtain the desired output voltage. Furthermore, a new reward function used in the TD3 algorithm is introduced. The proposed controller is tested under load variations and input voltage uncertainties. Simulation and experimental results demonstrate that TD3 outperforms PSO, GA, and the conventional PI. TD3 exhibits less steady-state error, reduced overshoots, fast response times, fast recovery times, and a small voltage deviation. These findings confirm TD3's superiority and its potential application in DC microgrid voltage control. It can be used by engineers and researchers to design DC microgrids.

**Keywords:** boost converter; DC microgrid; PI controller; twin-delayed deep deterministic policy gradient

## 1. Introduction

Recently, the integration of distributed generation (DG), comprising renewable energy and energy storage, has emerged as a key strategy to tackle environmental pollution and decrease the dependence on fossil energy sources for traditional power plants [1]. This integration, particularly when DGs are incorporated into power distribution systems, provides several advantages [2], such as improved voltage and power quality and reduced power loss [3]. The aggregation of DGs, energy storage, and loads forms a microgrid, an evolving concept in modern power systems [4]. The microgrid has two forms: AC and DC. Unlike the AC microgrid, the DC microgrid provides a number of advantages, such as high efficiency [5], ease of control [6], versatility [7], and reliability [8,9]. Furthermore, in a DC microgrid, there are no issues with reactive power flow, quality, or frequency [10]. DC microgrids have been extensively used in many applications [11], such as solar arrays, ship electrical systems, aerospace, electric vehicles, data centers, energy storage systems, and telecommunications [12].

However, in DC microgrids, the fluctuating output voltage from DC sources, such as a photovoltaic solar panel, poses a challenge, demanding regulation through converters due to intermittent solar conditions [13,14]. Yet, dealing with converters will lead to complex problems because of nonlinear time variants as well as nonminimum phase characteristics [15]. In addition, the presence of load uncertainty can lead to instability in the system [16] and become more challenging to resolve [17]. To tackle these problems, several industries are currently implementing a proportional integral (PI) controller because of its straightforward configuration, reliability, simplicity of application, good performance,

and active investigation into fine-tuning PI controllers [18]. Over the years, the parameters of PID controllers have been adjusted based on knowledge, trial–error approaches, and traditional tuning techniques such as Cohen–Coon and Ziegler–Nichols [19]. However, the weaknesses of these methods may lead to poor performance of the boost converter's output under certain conditions, such as during wide-range operation [20].

Prior studies investigated different control techniques for boost converters. In a previous study [21], a fuzzy logic-tuned PI controller was proposed as a method for controlling the converter. The simulation results of this method showed that it had better robustness to mitigate disturbances compared to a traditional fuzzy logic controller (FLC). However, FLC's reliance on expertise in constructing membership functions limits its adaptability [22]. In another study [23], Li et al. developed a cascade method for boost converter control that is effective in handling dynamic disturbances. Despite its efficacy, this method's complexity limits its use in industrial applications. Therefore, there is a need to improve PI and PID controllers, which are widely used in the industry, by tuning their parameters with metaheuristic-based techniques such as firefly algorithm (FA) [24], particle swarm optimization (PSO) [25], ant colony optimization (ACO) [26], and genetic algorithm (GA) [27]. However, these techniques are mainly proposed for nonlinear problems.

In the last few years, reinforcement learning (RL) methods have been applied in microgrid systems for a good solution [28,29]. A comprehensive review of research and practice in deep RL is presented in [30]. Fu et al. proposed an improved RL method for optimal secondary control of voltage, and the results confirm the efficacy and superiority of the proposed method [31]. Moreover, an RL agent design aimed at maximizing rewards ensured network voltage stability [32]. However, conventional RL agents faced restrictions due to low-dimensional action space discretization. To overcome this problem, the authors in [33] proposed a deep RL (DRL) approach to address the impact of constant load power under various voltage references. The results revealed that the presented strategy provides a better system dynamic response. However, this technique has the drawback of having a constant gradient signal [34]. To address continuous control problems, the introduction of the deep deterministic policy gradient (DDPG) algorithm has eliminated state–action discretization [35]. However, DDPG updates Q-values with a deep Q-network, which can lead to overestimation of Q-values and result in suboptimal policies.

To address the weaknesses of DDPG, the twin-delayed deep deterministic policy gradient (TD3) updates delayed actors, double critics, and actors. TD3 has been used in many applications. For instance, it improved training through proportional and differential controllers for unmanned aerial vehicle tracking, demonstrating remarkable multilevel success [36]. Furthermore, TD3 enhanced grid-integrated photovoltaics (PV) output by improving regulator performance utilizing an RL agent and verified it through numerical simulations [37]. The authors of [38] presented an RL twin-actor TD3 (TATD3) algorithm by combining twin-actor networks in the current TD3. The proposed method was used to control challenging batch processes due to complicated nonlinear dynamics and unstable operational environments. The results indicated the superiority of the TATD3 controller compared to existing RL algorithms.

This paper proposes a modified TD3 method for optimizing the parameters of the PI controller for boost converters in DC microgrids. Moreover, a new reward function employed in the method is presented. The contributions of this study are summarized as follows:

- A novel reward formula that combines the absolute error between the output voltage and the controller's action is proposed. The reward formula is crucial for effective performance assessment under input voltage uncertainties and load variations.
- The TD3 algorithm is presented using the newly devised reward function to fine-tune the parameters of a PI controller. This optimization approach significantly improves the voltage regulation of boost converters in DC microgrids.

- Simulation and hardware experiments are conducted under various scenarios, including fluctuations in the input voltage and load variation. They are treated equally to assess the controller's robustness and efficiency.

The subsequent sections are arranged as follows: In Section 2, a DC-DC boost converter system configuration in a DC microgrid is discussed. A control design strategy using the TD3 algorithm is explained in Section 3. In Section 4, numerical simulation results are presented, while experimental verifications are presented in Section 5. Finally, the conclusions of this work are summarized in Section 6.

## 2. The DC-DC Boost Converter

The equivalent circuit of a DC-DC boost converter is presented in Figure 1. The converter mathematical model is obtained by utilizing Kirchhoff's current and voltage laws [39]. The boost converter consists of a transistor for switching, which allows it to work in two conditions: open and closed switches [40]. At the closed condition presented in Figure 2, the inductor $L$ conserves power through $V_{in}$. In this state, the current cannot flow into the capacitor and resistor. Under this condition, the converter is expressed using Equations (1) and (2) [39].

$$V_{in} - L\frac{di_L}{dt} = 0 \tag{1}$$

$$C\frac{dV_c}{dt} = -\frac{V_c}{R} \tag{2}$$
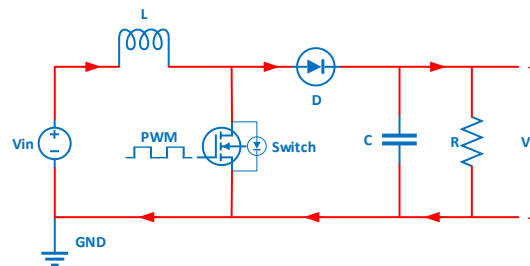


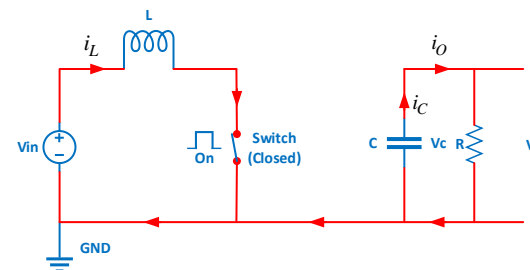**Figure 1.** Circuit Diagram of DC-DC Boost Converter.



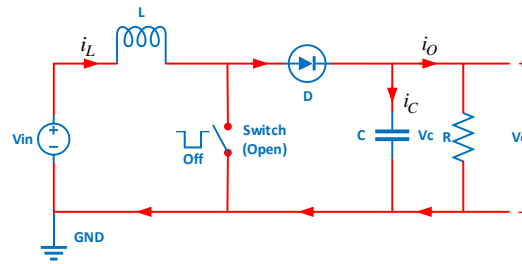**Figure 2.** Circuit diagram of a DC-DC converter for closed switch conditions.

In Equations (1) and (2), the inductor current $i_L$ and capacitor voltage $V_c$ are assumed to be $x_1$ and $x_2$, respectively. The modeling of the converter in the state-space for the closed switch condition is expressed as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \text{ that is } \dot{x} = A_1 x + B_1 u \tag{3}$$

In the open switch mode depicted in Figure 3, the inductor $L$ releases power. The converter formula is formulated using Equations (4) and (5) [39].

$$V_{in} - L\frac{di_L}{dt} - V_c = 0 \tag{4}$$

$$i_L = C\frac{dV_c}{dt} + \frac{V_c}{R} \tag{5}$$



**Figure 3.** Circuit diagram of a DC-DC converter for open switch conditions.

Likewise, by utilizing Equations (4) and (5), the state-space model of the converter in an open switch condition is defined as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \text{ that is } \dot{x} = A_2 x + B_2 u \tag{6}$$

Since the circuit operates in two conditions, the average model is applied [41].

$$\overline{A} = A_1 d_c + A_2 (1 - d_c) \tag{7}$$

$$\overline{B} = B_1 d + B_2 (1 - d_c) \tag{8}$$

$$\dot{x} = \overline{A}x + \overline{B}u \tag{9}$$

To obtain the average model of Equations (3) and (6), Equations (7)–(9) are used. The DC-DC boost converter average model in state-space form is defined in Equation (10) [42].

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{(1-d_c)}{L} \\ \frac{(1-d_c)}{C} & -\frac{\delta}{C} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in}, \ \delta = \frac{1}{R} \tag{10}$$

where $d_c$ represents the duty cycle and $\delta$ is the inverse of resistance.

## 3. A Twin-Delayed Deep Deterministic Policy Gradient (TD3)

RL is a machine learning technique that rewards wanted behaviors and punishes unwanted ones. This technique is a value-based algorithm that is learned through Q-value estimation. DDPG is an RL technique that combines DQN and DPQ. It includes two components: the actor and the critic. The actor is a function that takes the current state as an input and generates a continuously varying action.

In contrast, the critic is defined as a Q-value network that collects the state, acts as an input, and produces the Q. Figure 4 presents a general network of actors and critics. The actor decides which action must be performed, while the critic gives information to the actor to know how great the action is and how it must be tuned. The actor learning is built using a policy gradient approach. The critics assess the action created by the actor by calculating the value function and employing the temporal difference method to update its parameters. In contrast, the actor is updated using the DPG algorithm through $\alpha = \mu(s|\theta_\mu) + N$, where $N$ is defined as a random noise function. For updating the parameters of actor $\theta_\mu$ and critic $\theta_Q$, exponential smoothing is utilized and formulated as [43].

$$\theta_{\mu'} = \tau\theta_\mu + (1-\tau)\theta_{\mu'} \ (actor) \tag{11}$$

$$\theta_{Q'} = \tau\theta_Q + (1-\tau)\theta_{Q'} \ (critic) \tag{12}$$

where $\theta_\mu$ is the parameter update of the actor, $\theta_Q$ defines the parameter update of the critic, and $\tau$ is the smoothing factor. The Bellman equation and the estimation of action value are obtained by employing the critic network [37].

$$Q'(s,a) = E\big[r(s,a) + \gamma Q'(s',a')\big] \tag{13}$$

with a discount factor $\gamma \ll 1$, the TD error utilizes $y = r + \gamma Q'(s',a')$ and minimizes the loss function (LF) for updating the critic's parameters as defined in Equation (14) [44].

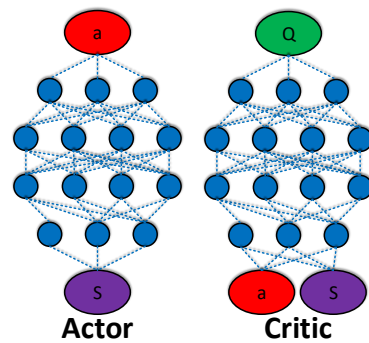$$LF = \frac{1}{M}\sum_{i=1}^{M}(y_i - Q(s_i,a_i))^2 \tag{14}$$



**Figure 4.** Actor and critic networks.

The policy gradient expressed in Equation (15) is employed to maximize the expected discounted reward [35].

$$\nabla_{\theta_\mu} J \approx \frac{1}{M}\sum_{i=1}^{M}\Big[\nabla_a Q(s,a)\big|_{s=s_i,a=\mu(s_i|\theta_\mu)}\nabla_{\theta_\mu}\mu\big(s|\theta_\mu\big)\big|_{s_i}\Big] \tag{15}$$

The TD3 algorithm is an improvement of the DDPG method and is a model-free, online, off-policy RL technique. The computation of TD3 is like the DDPG algorithm. The TD3 algorithm is modified especially to control the output voltage of a DC-DC boost converter in a DC microgrid. This adaptability shows the algorithm's versatility in solving accurate control tasks within complex systems like DC-DC boost converters in a DC microgrid. The value function overestimation influences Q-learning performance. If this overestimation continues during training, the policy update will be negatively influenced. To address this drawback, the double Q-learning and double DQN approaches have been used. These approaches use two networks for separating Q-values and updating the actor's selection. The double Q-learning calculates the next state number, forming double Q-value networks as defined in Equations (16) and (17) [36].

$$y_1 = r + \gamma Q_{\theta'_1}(s',\mu'(s'|\theta_{\mu'})) \tag{16}$$

$$y_2 = r + \gamma Q_{\theta'_2}(s',\mu'(s'|\theta_{\mu'})) \tag{17}$$

The TD error is expressed as follows:

$$y = r + \gamma_{i=1,2}^{min}Q'_i(s',a') \tag{18}$$

where $i$ is the critic index. To prevent overfitting, a smoothed Q-value is necessary. To achieve this, the implementation of clipped normal distribution noise is used, and the result of the updated target is as follows [43]:

$$y = r + \gamma Q_\theta(s',\mu'(s'|\theta_{\mu'}) + \varepsilon) \tag{19}$$

$$\varepsilon \sim clip(N(0,\sigma),-c,c) \tag{20}$$

### 3.1. Constructing an Environment for the Training Agent

The TD3 agent is trained to optimize PI parameters in a DC microgrid environment. The agent collects data on the environment's state at each time step to select appropriate actions. To construct the environment for training, the observation formula in Equation (21) is first constructed, where $e$ is the difference between the output voltage of the boost converter and the voltage reference. This signal is then connected to the RL agent block. Next, a negative reward function is determined for the RL agent, which provides feedback on the system's convergence. The reward function plays a critical role in guiding the agent to take actions that optimize the values.

$$O_b = \begin{bmatrix} \int e\,dt \\ e \end{bmatrix} \tag{21}$$

On the other hand, the agent's output is the control signal formula for the DC microgrid, and its value is determined by a policy that maximizes the reward. In this paper, a new reward function is proposed. The formulation of it is expressed in Equation (22).

$$RF = -|e| - 0.01|u| \tag{22}$$

where $u$ is the control action of the RL agent. The negative sign in this equation produces the maximum reward and minimum error. An environment interface object is also created, and the observation and action dimensions of the environment are extracted. This reward function can determine the PI parameters optimally. Furthermore, as the agent continually updates its parameters, the reward function can generate fast convergence, high performance, and minimal computation.

### 3.2. Constructing the TD3 Agent

After obtaining the observation signal, TD3 determines the next step by using an actor representation. The first step in building this actor is to construct a deep neural network (DNN) that takes the action output and observation inputs. The PI controller is modeled as a neural network (NN) using a combination of one fully connected layer with error-integral observations and error as defined in Equation (23).

$$u = \begin{bmatrix} \int e\,dt & e \end{bmatrix} * \begin{bmatrix} K_i & K_p \end{bmatrix}^T \tag{23}$$

where $K_p$ and $K_i$ define the absolute values of the neural network weights. Two critic cost function descriptions are used by the TD3 agent to estimate the long-term reward provided by actions and observations. A DNN with two inputs—observation and action—and one output is built to generate the critics.

Adjusting the parameters of TD3 includes some factors and methods to optimize its performance in reinforcement learning tasks. The controller sample time is the first parameter adjusted in the proposed method. The effect of this parameter in TD3 influences several aspects, like how often the agent interacts with the environment. A smaller sample time indicates that the agent gains more frequent updates on the environment, potentially leading to quicker learning. However, too many updates can create noise. The second parameter adjusted in the proposed method is the size of the mini batch. This parameter affects the stability of learning. A larger mini-batch size usually presents a more precise estimation. This can lead to more stable updates. The experience buffer length is another parameter that must be adjusted. A longer experience buffer stores a more extensive range of past experiences, which may lead to greater robustness. The next parameter is the targeting policy smoothing model. This parameter can influence the stability and performance of the learning process. The target smooth factor is also to be adjusted. Higher values of this parameter result in introducing more noise into the target actions, whereas

lower values lead to less smoothing. The last parameter is the discount factor. A larger discount factor confirms the importance of long-term rewards. A smaller discount factor selects immediate rewards. Based on the impacts of the parameters of the proposed method, we have adjusted the suitable value for the parameters presented in Table 1 so that the desired output voltage of the converter can be achieved. The detailed architecture of the proposed TD3-based controller for controlling the DC-DC boost converter is presented in Figure 5.

**Table 1.** Parameters of TD3 agent.

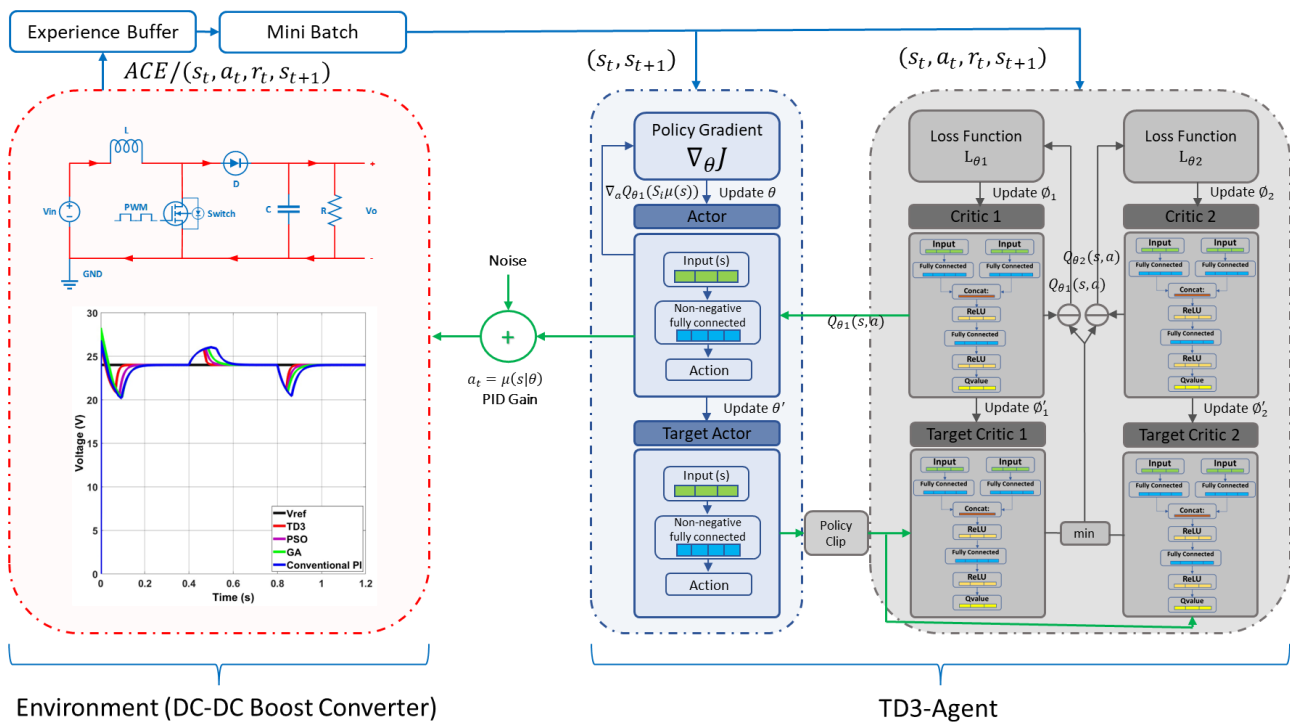| Parameter Type | Symbol | Number |
|---|---|---|
| Sample time of controller | $T_s$ | 0.01 s |
| Size of mini batch | $Z$ | 128 |
| Experience buffer length | $L_E$ | $10^6$ |
| Target policy smoothing model | $\sigma_t$ | 0.1 |
| Target Smooth Factor | $\tau_t$ | 0.005 |
| Discount Factor | $\gamma$ | 0.99 |



**Figure 5.** Proposed TD3-based controller.

The critic network presented in Figure 5 consists of nine layers. Then, a concatenation layer is created to connect all inputs, accompanied by a fully connected layer for each input. As an activation function to optimize the result, this paper employs a rectified linear unit (ReLU) between each layer. For updating the actor and critic network parameters, the Adam optimizer is used. Finally, the agent is created utilizing the critic representation, specific actor representation, and agent options.

### 3.3. Training and Validating the Agent

The training created in this study is run for 1000 episodes, where every episode is 100-time steps. Then, after the agent receives an average accumulative reward bigger than −355 over 100 successive episodes, the training is stopped. The agent in this condition could regulate the boost converter output. Then, the learned agent is validated by simulation. The PI gains of the controller are the absolute weights of actor representation. The learnable

parameters from the actor are extracted to obtain the weights, and the controller gains are obtained. The proposed TD3 method demonstrates the capability to optimize the gains of a PI controller, which is important for accurate voltage control. Utilizing a non-negative, fully connected layer is a remarkable innovation that guarantees the prevention of negative gain parameters, which improves the controller's stability and efficiency.

## 4. Simulation Results

In this paper, the parameters of the DC-DC boost converter are presented in Table 2, while Table 3 presents the gains of the PI controller tuned using TD3, PSO, GA, and the conventional PI. The initial simulation uses a voltage reference $V_{ref}$, 24 V and an input frequency of 100 kHz. The proposed TD3-tuned PI controller is compared with a PSO-tuned PI, a GA-tuned PI, and the conventional PI. For the best performance, the parameters of PSO and GA utilized in this work are shown in Table 4, while for the conventional PI, the Ziegler–Nichols tuning is used. MATLAB R2022a is used to develop the codes of the proposed controller, and the simulation is run on a personal computer with an 11th Gen intel®Core ™ i7-11700T @1.4GHz and 32 GB of RAM. Several system tests are employed to assess the superiority of the TD3, including voltage reference changes, input voltage variation, and load resistance variation. Additionally, the integral time absolute error (ITAE) is applied to confirm the quality of the proposed method. ITAE is the total calculation of the areas under and over the set point signal and the output voltage of the converter multiplied by time.

**Table 2.** The DC-DC boost converter and PI parameters.

| Parameters | Symbol | Value | Units |
|---|---|---|---|
| Input DC Voltage | $V_{in}$ | 12 | V |
| Inductor | $L$ | 100 | µH |
| Capacitor | $C$ | 47 | µF |
| Resistor Load | $R$ | 5 | kΩ |

**Table 3.** The PI parameters of several techniques.

| Parameters | TD3 | PSO | GA | Conventional PI |
|---|---|---|---|---|
| $K_P$ | 10.083 | 13.483 | 15.365 | 11.192 |
| $K_I$ | 1038.414 | 788.421 | 604.9213 | 412.988 |

**Table 4.** The parameters of GA and PSO.

| Methods | Parameters | Value |
|---|---|---|
| PSO | Size of Swarm | 40 |
| | Self-recognition coefficient $c_1$ | 2 |
| | Social coefficient $c_2$ | 2 |
| | Weight $\begin{bmatrix} w_1 & w_2 \end{bmatrix}$ | $\begin{bmatrix} 0.8 & 0.3 \end{bmatrix}$ |
| | Max Velocity | 100 |
| GA | Size of population | 50 |
| | Probability of mutation | 0.04 |
| | Probability of crossover | 0.75 |
| | Scale for mutations | 0.1 |

### 4.1. Signal Response of the Converter for Increasing and Decreasing the Voltage Reference

The proposed controller plays an important role in the system, which is the boost converter system. The controller continuously monitors the output voltage of the converter and compares it to the voltage reference $V_{ref}$. The controller will calculate the error, which is the difference between the voltage reference $V_{ref}$ and the output voltage of the converter. The main task of the controller is to minimize this error as much as possible by generating

a control signal for the switching of the converter so that the desired output voltage of the converter can be achieved. The voltage reference signal is the first system test used in this study. This test is used in the DC microgrid to meet the voltage shift on the load side. The simulation time is set to 1.2 s. At the start of the simulation, the voltage reference is 24 V. Figure 6 shows the signal of four converter output voltages, which are TD3, PSO, GA, and the conventional PI. The output response of TD3 is better than other methods in terms of transient response and overshoot. The TD3 in the initial response has the smallest overshoot, which is 26.3491 V, while the settling time of TD3 is 0.0851 s, which is a faster response than other methods. Then, at $t = 0.4$ s, the voltage reference is increased to 30 V. There is no overshoot for all methods, and the settling time of TD3 is smaller, 0.03 s, compared to others. In another condition, at $t = 0.8$ s, the voltage reference is reduced to 28 V. The response shows that there is no undershoot for all signals, and the smallest settling time is TD3. In this work, the ITAE is applied to check the performance quality of the methods, and the smallest ITAE is TD3, 0.1290. Furthermore, the computational cost of TD3 is smaller, 125 s, than PSO, 213.73 s, and GA, 565.498 s.
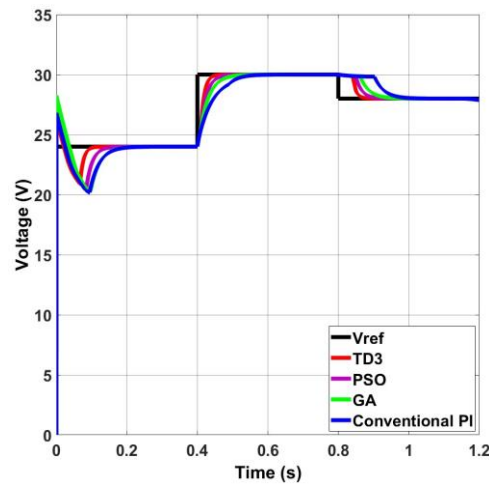


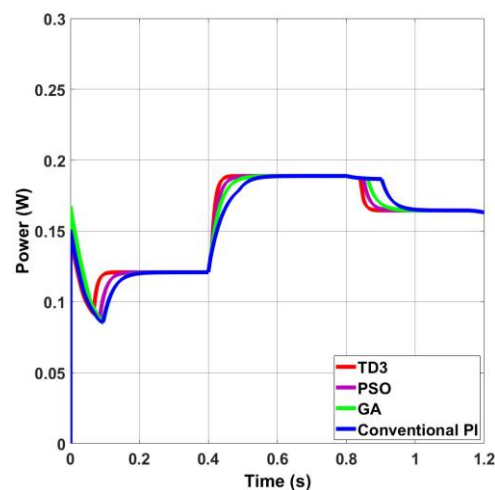**Figure 6.** Converter output voltage when the voltage reference is varied.

The results indicate that the TD3 produces a minimum steady-state error, a slight overshoot, a rapid transient response, and stable conditions in steady-state response. Additionally, the proposed method can be used in the DC microgrid system, even with changes in the voltage reference to meet the shift voltage on the load side. The detailed performances of all methods are shown in Table 5. The inductor current $i_L$ and the output power for voltage reference variations are depicted in Figures 7 and 8, respectively. In the starting condition, the inductor current $i_L$ for TD3 is smaller than in the others. Additionally, when the voltage reference is 24 V, the inductor current is 9.6 mA, while for voltage references of 30 V and 28 V, the inductor current is 15.17 mA and 13.049 mA, respectively. The output power of the converter using TD3 consumes the smallest power, which is 0.14 W in the starting condition, compared to PSO, GA, and the conventional PI, which consume 0.152 W, 0.16 W, and 0.15 W, respectively.

**Table 5.** The performance of all controllers when the voltage reference is varied.

| Controllers | Starting | | Increasing $V_{ref}$(+6 V) | | Decreasing $V_{ref}$(−2 V) | | ITAE |
|---|---|---|---|---|---|---|---|
| | $V_p$ (V) | $t_s$ (s) | $\Delta V$ (V) | $t_s$ (s) | $\Delta V$ (V) | $t_s$ (s) | |
| TD3 | 26.3491 | 0.0851 | 0 | 0.0300 | 0 | 0.0511 | 0.1290 |
| PSO | 27.5740 | 0.1188 | 0 | 0.0446 | 0 | 0.0704 | 0.1721 |
| GA | 28.2389 | 0.1487 | 0 | 0.0616 | 0 | 0.0912 | 0.2361 |
| Conventional PI | 26.8090 | 0.1506 | 0 | 0.0972 | 0 | 0.1343 | 0.3429 |
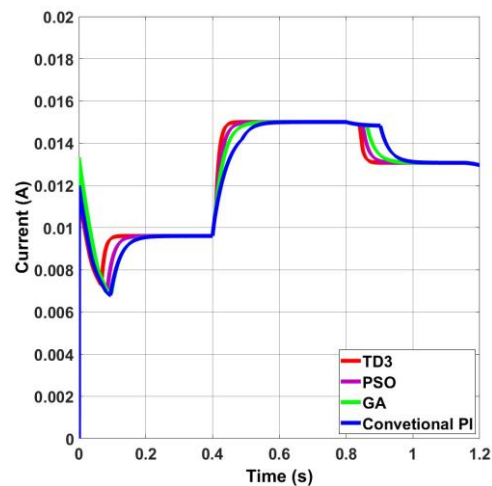
**Figure 7.** Converter inductor current when the voltage reference is varied.



**Figure 8.** Converter output power when the voltage reference is varied.

The combination of input voltage variation and voltage reference change is tested on the system to demonstrate the performance of the boost converter. The desired output voltage is shown in Figure 9, whereas the inductor current and power are shown in Figures 10 and 11, respectively. It is noticeable that when the reference signal is increased and the input voltage is decreased at $t = 0.4$, both the output voltage and the inductor current demonstrate a corresponding increase. Therefore, the characteristics of inductor current are proportional to output voltage instead of inversely proportional. Furthermore, the characteristics of output power in Figure 11 are also proportional to the output voltage instead of being inversely proportional. So, Figures 9–11 reflect the expected characteristics.

*4.2. Signal Response of the Converter for Increasing and Decreasing the Input Voltage $V_{in}$*

The second test employed in this work is changing the input voltage, $V_{in}$. This test is used to verify whether the proposed controller can overcome the variation in the input voltage. The overall simulation time is 1.2 s. The input voltage, $V_{in}$, is 12 V at the beginning of the simulation. The output voltage with the input voltage variation for all methods is shown in Figure 12. From Figure 12, it can be observed that the response output of TD3 is better than the other methods regarding robustness and fast recovery time. At $t = 0.4$ s, when the input voltage $V_{in}$ is increased to 16 V, the voltage deviation of TD3 is smaller, 1.8067 V, compared to other methods. Furthermore, TD3 produces the shortest recovery time, 0.0786 s. For the next condition, at $t = 0.8$ s, the input voltage $V_{in}$ is reduced to 12 V.

The response signals reveal that the voltage deviation of TD3 is smaller, 2.8437 V, than others, as well as that the recovery time of TD3 has the shortest time, which is 0.0583 s.
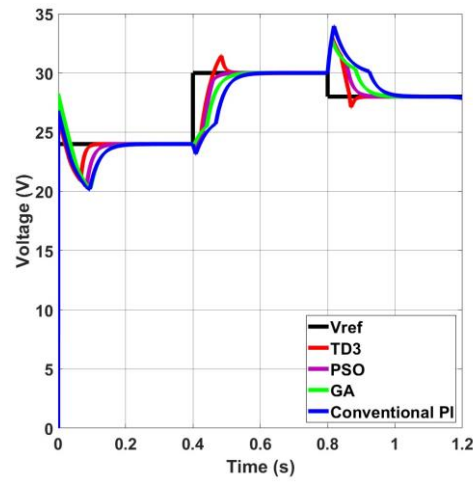
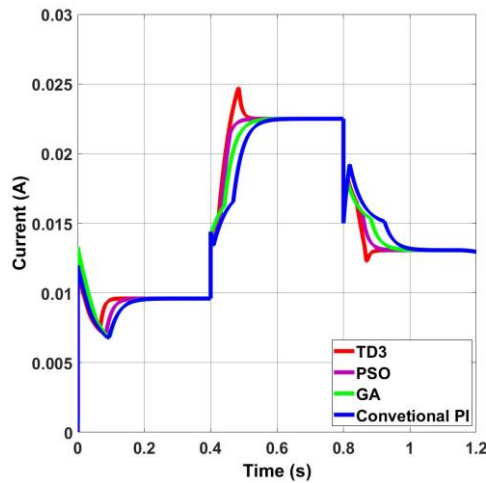**Figure 9.** Converter output voltage when the voltage reference and input voltage are varied.

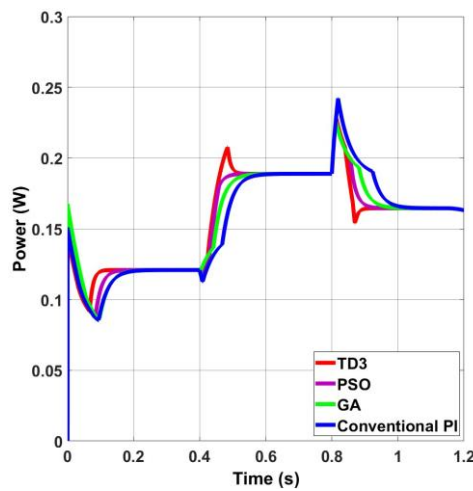**Figure 10.** Current inductor when the voltage reference and input voltage are varied.

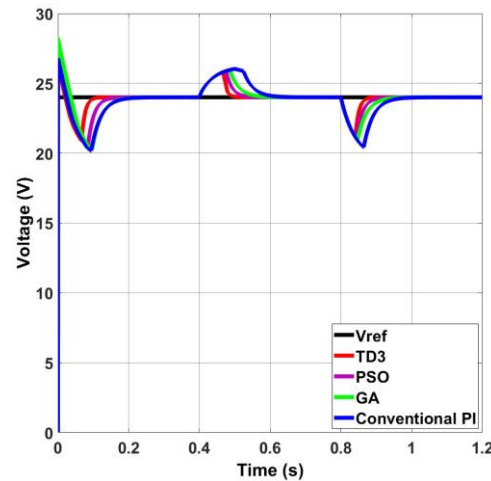**Figure 11.** Output power when the voltage reference and input voltage are varied.

**Figure 12.** The converter output voltage when the input voltage is changed.

ITAE is used to evaluate the overall performance of all methods. The results show that the ITAE of TD3 is 0.1305, which is smaller compared with PSO, GA, and the conventional PI. These results indicate that the proposed TD3 method generates a robust response, faster recovery time in comparison to others, and stable conditions when disturbances occur. Additionally, the proposed controller can be utilized in DC microgrid applications when there is a disturbance, such as input voltage variation. Table 6 presents the comprehensive data for all the methods. Figures 13 and 14 depict the inductor current $i_L$ and output power for input voltage variations, respectively. In initial conditions, the inductor current $i_L$ for TD3 is lower than other methods. Furthermore, when the input voltage is increased to 16 V, the inductor current is 7.19 mA, while when the input voltage is decreased to 12 V, the current inductor is 9.6 mA for steady-state conditions. The output power of the converter consumes 0.115 W for all methods in steady-state conditions.

**Table 6.** The performance of the proposed controller when the input voltage is varied.

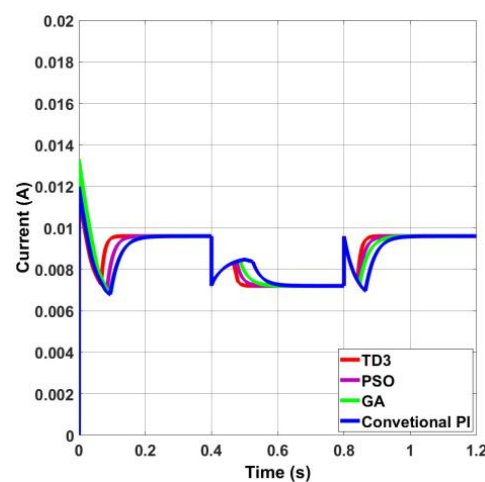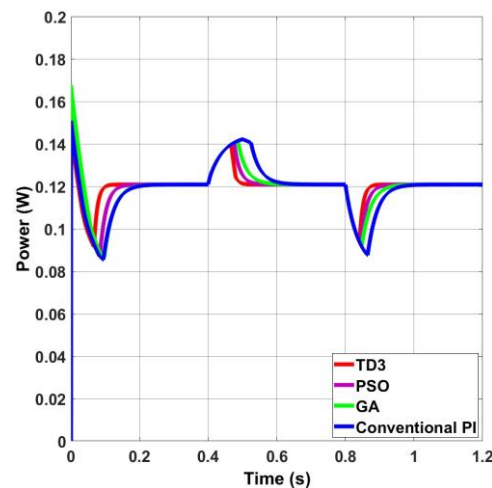| Controllers | Increasing $V_{in}$(4 V) | | Decreasing $V_{in}$(−4 V) | | ITAE |
|---|---|---|---|---|---|
| | $\Delta V$ (V) | $t_{rec}$(s) | $\Delta V$ (V) | $t_{rec}$(s) | |
| TD3 | 1.8067 | 0.0786 | 2.8437 | 0.0583 | 0.1305 |
| PSO | 1.8850 | 0.0986 | 2.9563 | 0.0744 | 0.1792 |
| GA | 1.9587 | 0.1215 | 3.0680 | 0.0945 | 0.2413 |
| Conventional PI | 2.0325 | 0.1611 | 3.5707 | 0.1192 | 0.3514 |



**Figure 13.** Converter inductor current when the input voltage is changed.

**Figure 14.** Converter output power when the input voltage is changed.

*4.3. Signal Response of the Converter for Increasing and Decreasing the Load Resistance R*

The final test signal applied in this study is the variation in load resistance. This system test is employed because the load resistance of the system in the DC microgrid application varies depending on consumer needs, and it is used to determine whether the proposed controller can handle variations in load resistance. The simulation time (t) is set to be 1.2 s. Initially, the load resistance $R$ is 5 k$\Omega$. Then, at $t = 0.4$ s, $R$ is reduced to 1.7 k$\Omega$. The output response signals when the load resistance changes for all methods are presented in Figure 15. TD3 demonstrates a smaller voltage deviation of 1.0682 V compared to PSO at 1.1803 V, GA at 1.2854 V, and the conventional PI at 1.3598 V. Additionally, the recovery time of TD3 is the shortest, which is 0.2152. When the load resistance $R$ is increased to 5 k$\Omega$ at $t = 0.8$ s, TD3 generates a voltage deviation of 3.2942 V, which is smaller than all other tested methods. Furthermore, TD3 has a shorter recovery time of 0.0512 s compared to the others. The ITAE is presented to assess the overall response of all methods. The ITAE of TD3 is 0.1434, which is smaller in comparison with the other methods. Table 7 provides detailed data on the converter signal response. The results show that TD3 has superior performance in terms of robustness, fast recovery, and stability when compared to the other methods.
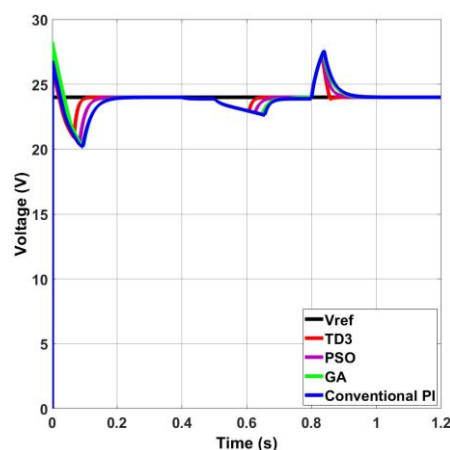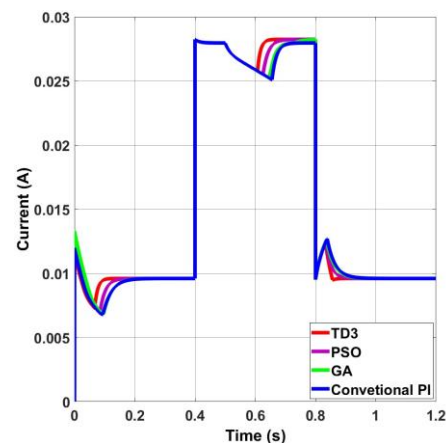


**Figure 15.** The converter output voltage when the load resistor is changed.

In the event of a disturbance, such as a change in load resistance, the proposed controller can be applied to DC microgrid systems. Figures 16 and 17 show the load current and output power of the converter for load variations, respectively. The inductor current $i_L$ reaches 28.24 mA in steady-state conditions when the load resistance decreases to 1.7 k$\Omega$. However, when the load resistance increases to 5 k$\Omega$, the inductor current $i_L$ for all methods

is 9.6 mA under steady-state conditions. The output power produced by all methods is 0.34 W when the load resistance is reduced to 1.7 kΩ, and 0.125 W when it is increased to 5 kΩ.

**Table 7.** The converter performance in terms of resistance change.

| Controllers | Decreasing R to 1.7 kΩ | | Increasing R to 5 kΩ | | ITAE |
|---|---|---|---|---|---|
| | $\Delta V$ (V) | $t_{rec}$ (s) | $\Delta V$ (V) | $t_{rec}$ (s) | |
| TD3 | 1.0682 | 0.2152 | 3.2942 | 0.0512 | 0.1434 |
| PSO | 1.1803 | 0.2412 | 3.3946 | 0.0676 | 0.1935 |
| GA | 1.2854 | 0.2667 | 3.5074 | 0.0871 | 0.2610 |
| Conventional PI | 1.3598 | 0.2738 | 3.6082 | 0.0932 | 0.2860 |



**Figure 16.** Load current of the converter for load resistor variations.
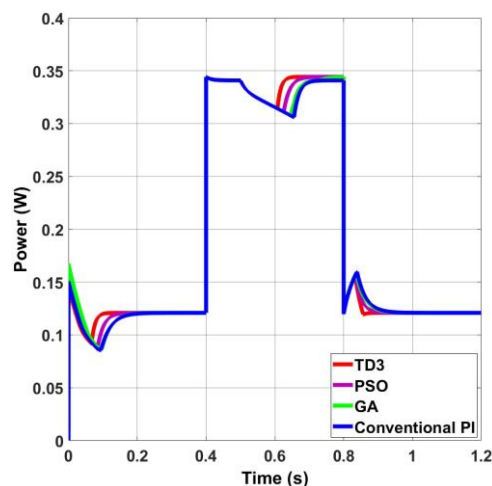


**Figure 17.** The converter output power when the load is changed.

## 5. Experimental Validations

To further verify the effectiveness of TD3, a prototype hardware design, as shown in Figure 18, was constructed. The design included a Siglent SPD3303X-E DC power supply, a Siglent SDL1020X-E DC electronic load, a DC-DC boost converter using the parameters outlined in Table 2, a Siglent SDS1202X-E two-channel digital oscilloscope, a Siglent SDM3045X digital multimeter, an NI Elvis II for data acquisition, and a Dell personal computer for programming the controller. IRFZ44N was used as the switching device and gate drive circuit. The system architecture is shown in Figure 19. In this study, the voltage reference and the frequency used for the initial condition are equal to the simulation, which

is 24 V and 100 kHz, respectively. The proposed controller was evaluated through several system tests, including variations in $V_{ref}$, input voltage, and load resistance.
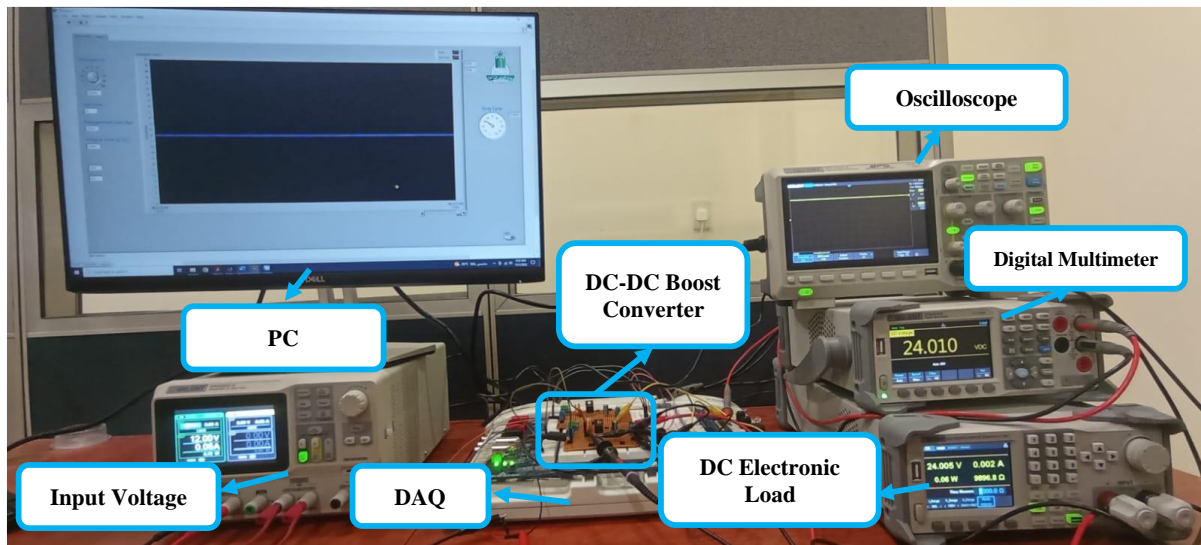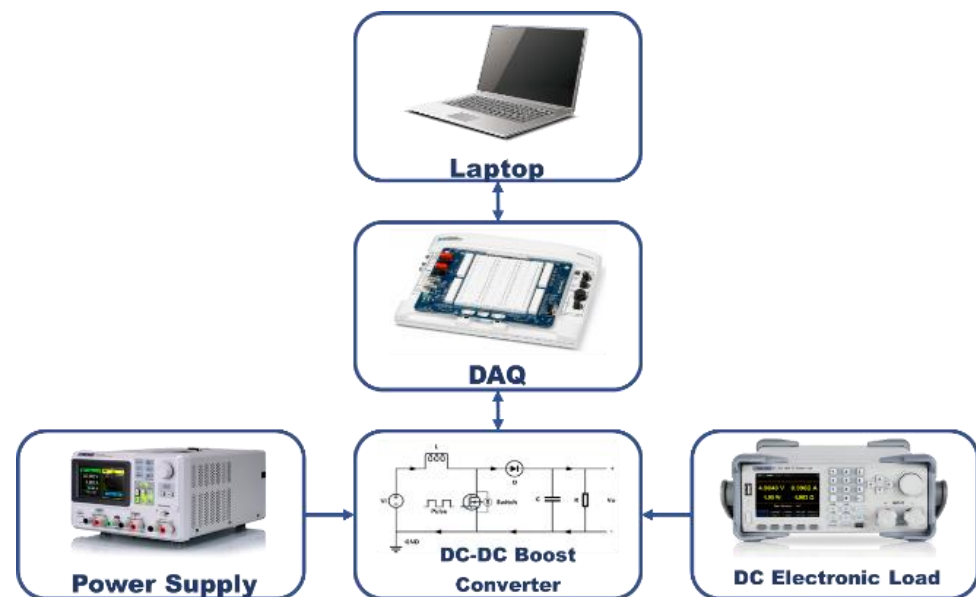


**Figure 18.** Hardware setup.



**Figure 19.** System Architecture.

*5.1. Signal Response of the Converter for Increasing and Decreasing the Voltage Reference in Experimental Validation*

The first system test is the voltage reference change. As previously mentioned, this test is used in a DC microgrid to meet the voltage shift on the load side. In the experimental hardware, the voltage reference $V_{ref}$ is initially raised from 24 V to 26 V. Figure 20 shows the experimental waveform of the converter's output voltage when the voltage reference is changed for all methods. It can be observed that the output voltage of all methods can follow the voltage reference and provide a stable response. The settling time required for the output voltage to match the voltage reference using the TD3 controller as shown in Figure 20a is the shortest, with a time of 0.082 s and without any overshoot, while the settling time for PSO in Figure 20b, GA in Figure 20c, and the conventional PI in Figure 20d is 0.124 s, 0.145 s, and 0.225 s, respectively.
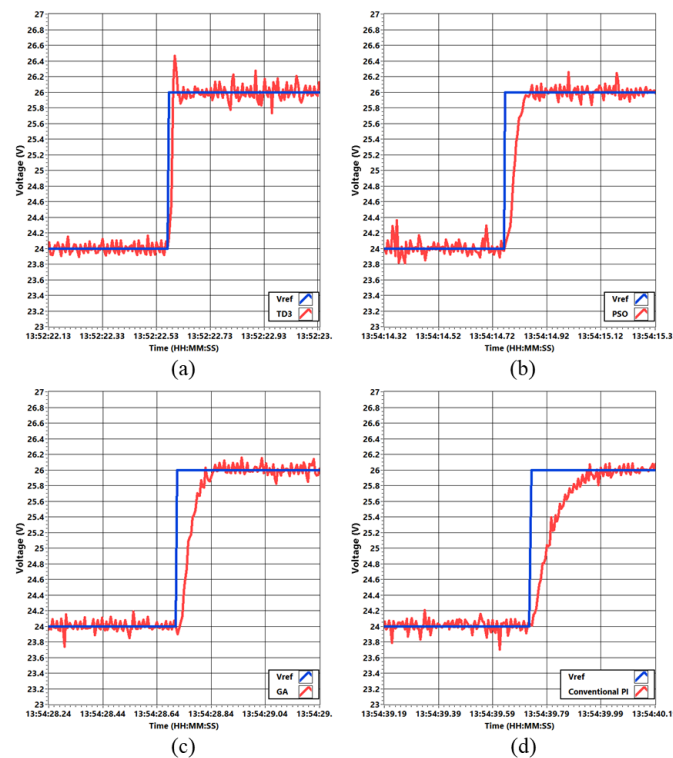
**Figure 20.** The converter output voltage in experimental validation when voltage reference is changed (24 V to 26 V) using (**a**) TD3, (**b**) PSO, (**c**) GA, and (**d**) conventional PI.
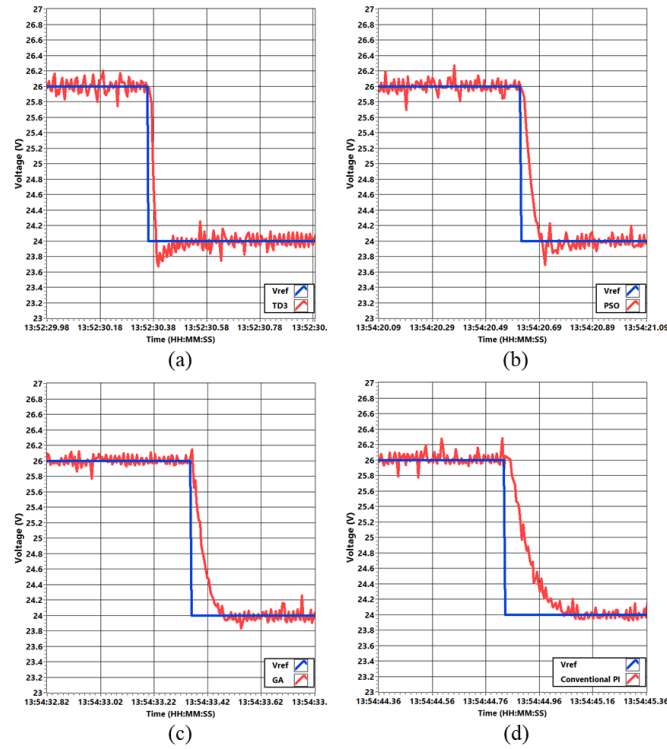
Additionally, in Figure 21, the voltage reference $V_{ref}$ is decreased from 26 V to 24 V. It can be observed that the converter's output voltage can follow the voltage reference for all controllers. TD3 presented in Figure 21a produces a smaller settling time than the others, which is 0.06 s, whereas PSO, GA, and the conventional PI depicted in Figure 21b–d generate a settling time of 0.121 s, 0.126 s, and 0.227 s. From the results, therefore, we see that they generate fewer steady-state errors, a reduced overshoot, a faster transient response, and stable conditions in steady-state responses. Furthermore, the TD3 can be applied in the DC microgrid system even when there are changes in the voltage reference to meet the voltage shift on the load side and provide a stable output voltage in steady-state response.

*5.2. Signal Response of the Converter for Increasing and Decreasing the Input Voltage $V_{in}$ in Experimental Validation*

The following system test for the experimental hardware is the input voltage variation $V_{in}$. As stated earlier, this assessment is employed in the DC microgrid application to ensure that the proposed controller can overcome the variation in input voltage, which constantly varies. Figure 22 depicts the experimental waveform of the converter's output signal for all controllers when the input voltage is decreased from 12 V to 10 V. It can be observed that the output voltage utilizing TD3 presented in Figure 22a can return to the voltage reference with a voltage deviation of 0.561 V and a recovery time of 0.231 s, while PSO in Figure 22b generates a voltage deviation of 0.861 V and a recovery time of 0.311 s. Furthermore, GA in Figure 22c provides a voltage deviation of 1.152 V and a recovery time of 0.375 s, whereas the conventional PI in Figure 22d has a voltage deviation of 1.513 V and a recovery time of 0.524 s. The results state that TD3 has the smallest voltage deviation and recovery time. The experimental waveform of the output response when the $V_{in}$ is increased from 10 V to 12 V is presented in Figure 23 for all techniques. It shows that the output signal employing TD3 depicted in Figure 23a can return to the voltage reference with a small voltage deviation of 0.851 V and a recovery time of 0.175 s, while in Figure 23b, the voltage deviation and recovery time of PSO are 1.253 V and 0.212 s, respectively. In addition, GA shown in Figure 23c and the conventional PI presented in Figure 23d produce a voltage

deviation of 1.641 V and 1.92 V, with recovery times of 0.25 s and 0.31 s, respectively. Thus, the results indicate the TD3 generates a smaller voltage deviation, a robust response, and a faster recovery time compared to other controllers. Furthermore, it can be used in the DC microgrid application when there is a disturbance, such as an input voltage variation.



**Figure 21.** The converter output voltage in experimental validation when voltage reference is changed (26 V to 24 V) using (**a**) TD3, (**b**) PSO, (**c**) GA, and (**d**) conventional PI.
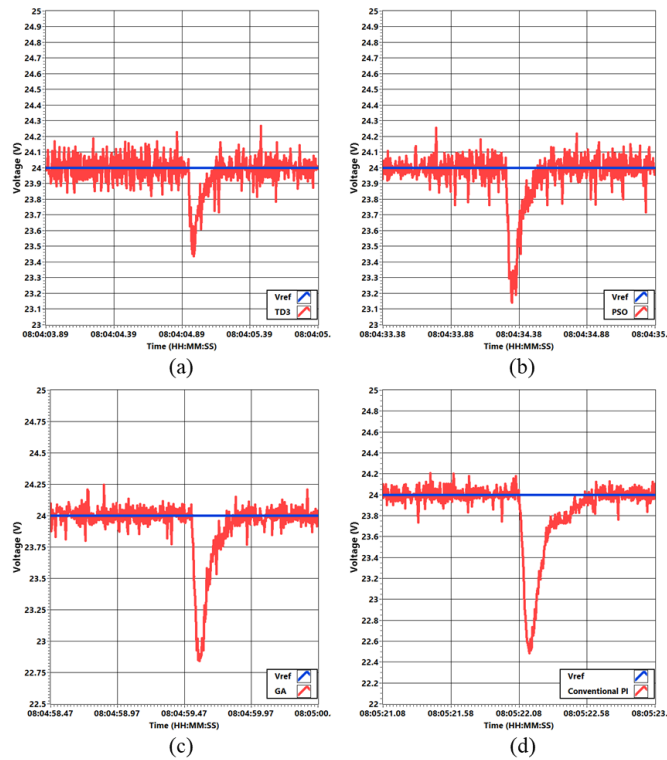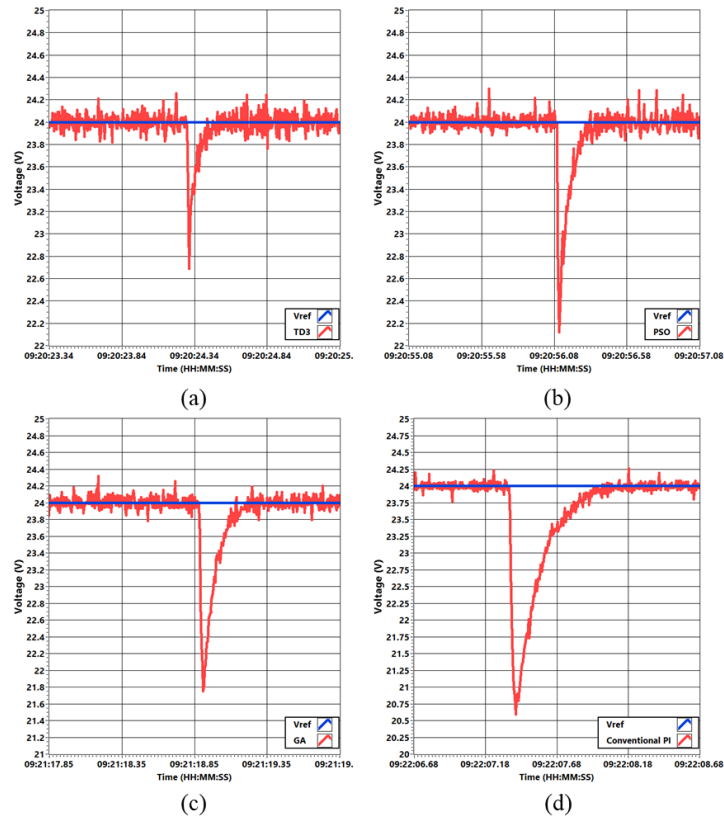


**Figure 22.** The converter output voltage in experimental validation when input voltage is changed (12 V to 10 V) using (**a**) TD3, (**b**) PSO, (**c**) GA, and (**d**) conventional PI.
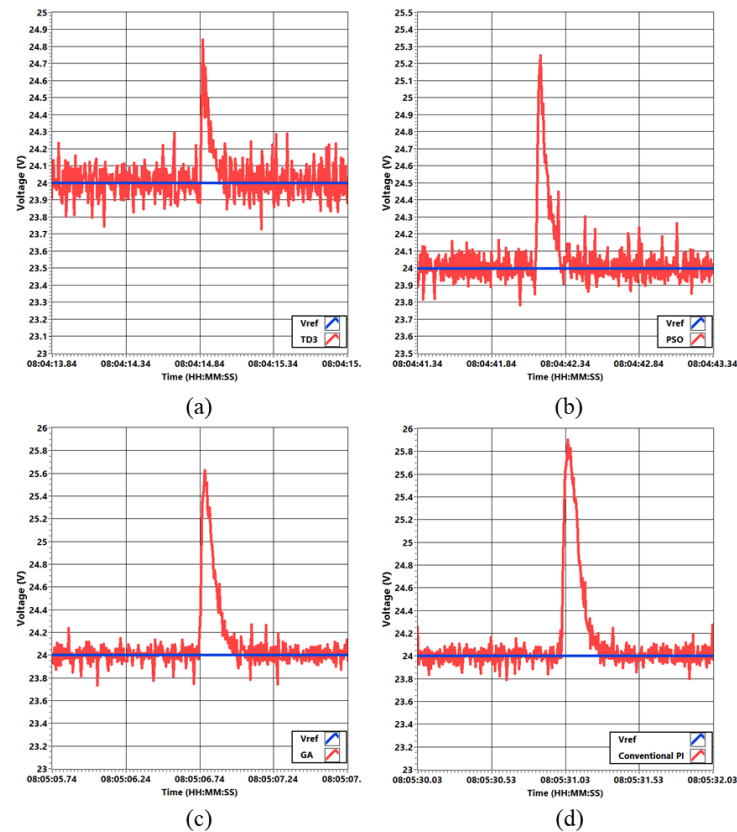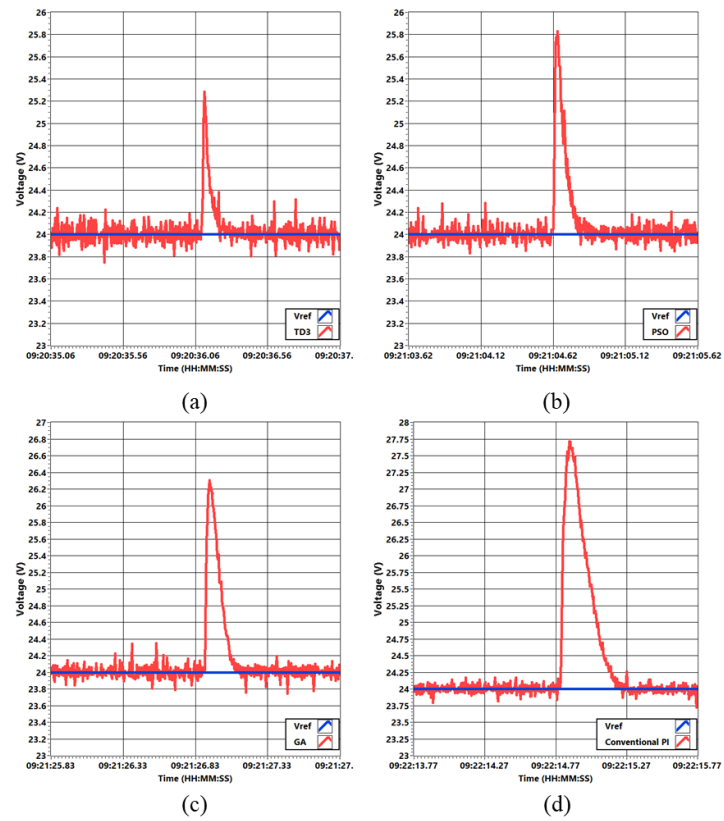
**Figure 23.** The converter output voltage in experimental validation when input voltage is changed (10 V to 12 V) using (**a**) TD3, (**b**) PSO, (**c**) GA, and (**d**) conventional PI.

*5.3. Signal Response of the Converter for Increasing and Decreasing the Load Resistance R in Experimental Validation*

The final system test in the experimental hardware is the load resistance R change. As previously mentioned, due to the load resistance of the system in the DC microgrid application always varying based on consumer needs, this test must be applied to prove whether the proposed controller can overcome the variation in load resistance. Figure 24 presents the experimental waveform of the converter output voltage for all techniques when the load resistance R is decreased from 5 kΩ to 1.7 kΩ. It can be observed that with TD3 shown in Figure 24a, the output signal returns to the voltage reference with a voltage deviation of 1.315 V and a recovery time of 0.175 s, while for PSO presented in Figure 24b, the voltage deviation is 1.851 V and the recovery time is 0.228 s. Moreover, the voltage deviation and recovery time using GA depicted in Figure 24c are 2.223 V and 0.345 s, respectively. While using the conventional PI, as presented in Figure 24d, the voltage deviation and recovery time are 3.441 V and 0.453 s, respectively. Then, the experimental waveform of the converter output voltage when the load resistance R is raised from 1.7 kΩ to 5 kΩ is presented in Figure 25 for all methods. Using TD3, depicted in Figure 25a, shows that the output voltage returns to the voltage reference with a voltage deviation of 1.321 V and a recovery time of 0.151 s. While using PSO presented in Figure 25b, the voltage deviation is 1.802 V, and the recovery time is 0.242 s. Moreover, the voltage deviation and recovery time using GA depicted in Figure 25c are 2.31 V and 0.275 s, respectively, while using the conventional PI presented in Figure 25d are 3.75 V and 0.451 s, respectively. Therefore, the results state that the TD3 has the smallest voltage deviation, is most robust, and has the fastest recovery time. Moreover, TD3 can be applied for DC microgrid applications that have a disturbance, such as a load resistance variation.

**Figure 24.** The converter output voltage in experimental validation when the load resistance is changed (5 kΩ to 1.7 kΩ) using (**a**) TD3, (**b**) PSO, (**c**) GA, and (**d**) conventional PI.



**Figure 25.** The converter output voltage when the load resistance is changed (1.7 kΩ to 5 kΩ) using (**a**) TD3, (**b**) PSO, (**c**) GA, and (**d**) conventional PI.

## 6. Conclusions

This paper presents a modified TD3 algorithm for voltage control in a DC microgrid's boost converter. The proposed method optimizes PI controller parameters using a non-negative fully connected layer, which is crucial to avoid negative gain parameters. Both numerical simulations and experimental prototypes are used to demonstrate the effectiveness and superiority of the proposed method. Tests are conducted to reveal the superior performance of the proposed controller compared to PSO, GA, and the conventional PI methods. The simulation and experimental results show that the TD3 method has smaller steady-state errors, lower overshoots, a faster recovery time, a smaller voltage deviation, and a faster transient response time when compared to other methods such as PSO, GA, and the conventional PI. This study suggests its suitability for DC microgrid applications. For further research, a combination of sliding mode control and RL techniques can be used to enhance the performance of the TD3 algorithm for the optimization of DC microgrids.

## References

1. Zhang, Y.; Yang, H.; Wang, P. The Voltage Stabilizing Control Strategy of Off-Grid Microgrid Cluster Bus Based on Adaptive Genetic Fuzzy Double Closed-Loop Control. *J. Electr. Comput. Eng.* **2021**, *2021*, 5515362. [CrossRef]
2. El-Ela, A.A.A.; El-Sehiemy, R.A.; Abbas, A.S. Optimal Placement and Sizing of Distributed Generation and Capacitor Banks in Distribution Systems Using Water Cycle Algorithm. *IEEE Syst. J.* **2018**, *12*, 3629–3636. [CrossRef]
3. Zishan, F.; Akbari, E.; Montoya, O.D.; Giral-Ramírez, D.A.; Molina-Cabrera, A. Efficient PID Control Design for Frequency Regulation in an Independent Microgrid Based on the Hybrid PSO-GSA Algorithm. *Electronics* **2022**, *11*, 3886. [CrossRef]
4. Bastos, R.F.; Aguiar, C.R.; Balogh, A.; Sütő, Z.; Machado, R.Q. Power-Sharing for Dc Microgrid with Composite Storage Devices and Voltage Restoration without Communication. *Int. J. Electr. Power Energy Syst.* **2022**, *138*, 107928. [CrossRef]
5. Esmaeili, M.; Ahmadi, A.A.; Nateghi, A.; Shafie-khah, M. Robust Power Management System with Generation and Demand Prediction and Critical Loads in DC Microgrid. *J. Clean. Prod.* **2023**, *384*, 135490. [CrossRef]
6. Mahajan, T.; Potdar, M.S. An Improved Strategy for Distributed Generation Control and Power Sharing in Islanded Microgrid. In Proceedings of the 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 5–7 March 2020; pp. 133–136. [CrossRef]
7. Badar, M.; Ahmad, I.; Mir, A.A.; Ahmed, S.; Waqas, A. An Autonomous Hybrid DC Microgrid with ANN-Fuzzy and Adaptive Terminal Sliding Mode Multi-Level Control Structure. *Control Eng. Pract.* **2022**, *121*, 105036. [CrossRef]
8. Sarangi, S.; Sahu, B.K.; Rout, P.K. A Comprehensive Review of Distribution Generation Integrated DC Microgrid Protection: Issues, Strategies, and Future Direction. *Int. J. Energy Res.* **2021**, *45*, 5006–5031. [CrossRef]
9. Ali, S.; Zheng, Z.; Aillerie, M.; Sawicki, J.P.; Péra, M.C.; Hissel, D. A Review of Dc Microgrid Energy Management Systems Dedicated to Residential Applications. *Energies* **2021**, *14*, 4308. [CrossRef]
10. Naik, K.R.; Rajpathak, B.; Mitra, A.; Kolhe, M.L. Adaptive Energy Management Strategy for Sustainable Voltage Control of PV-Hydro-Battery Integrated DC Microgrid. *J. Clean. Prod.* **2021**, *315*, 128102. [CrossRef]
11. Liu, J.; Zhang, W.; Rizzoni, G. Robust Stability Analysis of DC Microgrids With Constant Power Loads. *IEEE Trans. Power Syst.* **2018**, *33*, 851–860. [CrossRef]
12. Aluisio, B.; Dicorato, M.; Ferrini, I.; Forte, G.; Sbrizzai, R.; Trovato, M. Planning and Reliability of DC Microgrid Configurations for Electric Vehicle Supply Infrastructure. *Int. J. Electr. Power Energy Syst.* **2021**, *131*, 107104. [CrossRef]

13. Ait Ayad, I.; Elwarraki, E.; Baghdadi, M. Intelligent Perturb and Observe Based MPPT Approach Using Multilevel DC-DC Converter to Improve PV Production System. *J. Electr. Comput. Eng.* **2021**, *2021*, 6673022. [CrossRef]

14. Liu, X.; Zhang, Y.; Suo, Y.; Song, X.; Zhou, J. Large-Signal Stability Analysis for Islanded DC Microgrids with N+1 Parallel Energy-Storage Converters. *Electronics* **2023**, *12*, 4032. [CrossRef]

15. Al-Baidhani, H.; Kazimierczuk, M.K.; Reatti, A. Nonlinear Modeling and Voltage-Mode Control of DC-DC Boost Converter for CCM. In Proceedings of the IEEE International Symposium on Circuits and Systems, Florence, Italy, 27–30 May 2018; Volume 2018.

16. Alipour, M.; Zarei, J.; Razavi-Far, R.; Saif, M.; Mijatovic, N.; Dragicevic, T. Observer-Based Backstepping Sliding Mode Control Design for Microgrids Feeding a Constant Power Load. *IEEE Trans. Ind. Electron.* **2022**, *70*, 465–473. [CrossRef]

17. Guo, Q.; Bahri, I.; Diallo, D.; Berthelot, E. Model Predictive Control and Linear Control of DC–DC Boost Converter in Low Voltage DC Microgrid: An Experimental Comparative Study. *Control Eng. Pract.* **2023**, *131*, 105387. [CrossRef]

18. Borase, R.P.; Maghade, D.K.; Sondkar, S.Y.; Pawar, S.N. A Review of PID Control, Tuning Methods and Applications. *Int. J. Dyn. Control* **2020**, *9*, 818–827. [CrossRef]

19. Ibrahim, O.; Yahaya, N.Z.; Saad, N. Comparative Studies of PID Controller Tuning Methods on a DC-DC Boost Converter. In Proceedings of the International Conference on Intelligent and Advanced Systems, ICIAS 2016, Kuala Lumpur, Malaysia, 15–17 August 2016.

20. Zehra, S.S.; Dolara, A.; Amjed, M.A.; Mussetta, M. Implementation of Nonlinear Controller to Improve DC Microgrid Stability: A Comparative Analysis of Sliding Mode Control Variants. *Electronics* **2023**, *12*, 4540. [CrossRef]

21. Slamet, S.; Rijanto, E.; Nugroho, A.; Ghani, R.A. A Robust Maximum Power Point Tracking Control for PV Panel Using Adaptive PI Controller Based on Fuzzy Logic. *Telkomnika (Telecommun. Comput. Electron. Control)* **2020**, *18*, 2999–3009. [CrossRef]

22. Hasanien, H.M.; Muyeen, S.M. A Taguchi Approach for Optimum Design of Proportional-Integral Controllers in Cascaded Control Scheme. *IEEE Trans. Power Syst.* **2013**, *28*, 1636–1644. [CrossRef]

23. Li, H.; Liu, X.; Lu, J. Research on Linear Active Disturbance Rejection Control in Dc/Dc Boost Converter. *Electronics* **2019**, *8*, 1249. [CrossRef]

24. Gupta, D.K.; Soni, A.K.; Jha, A.V.; Mishra, S.K.; Appasani, B.; Srinivasulu, A.; Bizon, N.; Thounthong, P. Hybrid Gravitational-Firefly Algorithm-Based Load Frequency Control for Hydrothermal Two-Area System. *Mathematics* **2021**, *9*, 712. [CrossRef]

25. Faisal, S.F.; Beig, A.R.; Thomas, S. Time Domain Particle Swarm Optimization of PI Controllers for Bidirectional VSC HVDC Light System. *Energies* **2020**, *13*, 866. [CrossRef]

26. Wongkhead, S.; Tunyasrirut, S. Implementation of a Dsp- Tms320f28335 Based State Feedback with Optimal Design of Pi Controller for a Speed of Bldc Motor by Ant Colony Optimization. *Prz. Elektrotech.* **2021**, *97*, 9–14. [CrossRef]

27. Belgaid, Y.; Helaimi, M.; Taleb, R.; Youcef, M.B. Optimal Tuning of PI Controller Using Genetic Algorithm for Wind Turbine Application. *Indones. J. Electr. Eng. Comput. Sci.* **2019**, *18*, 167–178. [CrossRef]

28. Darshi, R.; Shamaghdari, S.; Jalali, A.; Arasteh, H. Decentralized Reinforcement Learning Approach for Microgrid Energy Management in Stochastic Environment. *Int. Trans. Electr. Energy Syst.* **2023**, *2023*, 1190103. [CrossRef]

29. Kolodziejczyk, W.; Zoltowska, I.; Cichosz, P. Real-Time Energy Purchase Optimization for a Storage-Integrated Photovoltaic System by Deep Reinforcement Learning. *Control Eng. Pract.* **2021**, *106*, 104598. [CrossRef]

30. Arwa, E.O.; Folly, K.A. Reinforcement Learning Techniques for Optimal Power Control in Grid-Connected Microgrids: A Comprehensive Review. *IEEE Access* **2020**, *8*, 208992–209007. [CrossRef]

31. Fu, Y.; Guo, X.; Mi, Y.; Li, Z.; Yuan, M. Distributed Economic Droop Control for DC Microgrid Based on Reinforcement Learning. *Dianli Zidonghua Shebei/Electric Power Autom. Equip.* **2021**, *41*, 1–7. [CrossRef]

32. Kosaraju, K.C.; Sivaranjani, S.; Suttle, W.; Gupta, V.; Liu, J. Reinforcement Learning Based Distributed Control of Dissipative Networked Systems. *IEEE Trans. Control Netw. Syst.* **2022**, *9*, 856–866. [CrossRef]

33. Hajihosseini, M.; Andalibi, M.; Gheisarnejad, M.; Farsizadeh, H.; Khooban, M.H. DC/DC Power Converter Control-Based Deep Machine Learning Techniques: Real-Time Implementation. *IEEE Trans. Power Electron.* **2020**, *35*, 9971–9977. [CrossRef]

34. Lowe, R.; Wu, Y.; Tamar, A.; Harb, J.; Abbeel, P.; Mordatch, I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 2017.

35. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. In Proceedings of the 4th International Conference on Learning Representations, ICLR, San Juan, Puerto Rico, 2–4 May 2016.

36. Abo Mosali, N.; Shamsudin, S.S.; Alfandi, O.; Omar, R.; Al-Fadhali, N. Twin Delayed Deep Deterministic Policy Gradient-Based Target Tracking for Unmanned Aerial Vehicle with Achievement Rewarding and Multistage Training. *IEEE Access* **2022**, *10*, 23545–23559. [CrossRef]

37. Nicola, M.; Nicola, C.I.; Selișteanu, D. Improvement of the Control of a Grid Connected Photovoltaic System Based on Synergetic and Sliding Mode Controllers Using a Reinforcement Learning Deep Deterministic Policy Gradient Agent. *Energies* **2022**, *15*, 2392. [CrossRef]

38. Joshi, T.; Makker, S.; Kodamana, H.; Kandath, H. Twin Actor Twin Delayed Deep Deterministic Policy Gradient (TATD3) Learning for Batch Process Control. *Comput. Chem. Eng.* **2021**, *155*, 107527. [CrossRef]

39. Muktiadji, R.F.; Ramli, M.A.M.; Bouchekara, H.R.E.H.; Milyani, A.H.; Rawa, M.; Seedahmed, M.M.A.; Budiman, F.N. Control of Boost Converter Using Observer-Based Backstepping Sliding Mode Control for DC Microgrid. *Front. Energy Res.* **2022**, *10*, 8978. [CrossRef]

40. Muktiadji, R.F.; Ramli, M.A.M.; Seedahmed, M.M.A.; Uswarman, R. Endryansyah Power Sharing Control and Voltage Restoration in DC Microgrid Using PI Fuzzy. In Proceedings of the 2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE), Surabaya, Indonesia, 10–11 September 2022; pp. 130–135.

41. Sira-Ramirez, H.; Perez-Moreno, R.A.; Ortega, R.; Garcia-Esteban, M. Passivity-Based Controllers for the Stabilization of DC-to-DC Power Converters. *Automatica* **1997**, *33*, 499–513. [CrossRef]

42. Chincholkar, S.; Jiang, W.; Chan, C.Y.; Rangarajan, S.S. A Simplified Output Feedback Controller for the Dc-dc Boost Power Converter. *Electronics* **2021**, *10*, 493. [CrossRef]

43. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [CrossRef]

44. Dankwa, S.; Zheng, W. Twin-Delayed DDPG: A Deep Reinforcement Learning Technique to Model a Continuous Movement of an Intelligent Robot Agent. In Proceedings of the 3rd International Conference on Vision, Image and Signal Processing, Vancouver, BC, Canada, 26–28 August 2019.