A Deep Q-Network based optimized modulation scheme for Dual-Active-Bridge converter to reduce the RMS current

Tang, Yuanhong; Hu, Weihao; Xiao, Jian; Lu, Zhengdong; Liu, Zhou; Chen, Zhe; Blaabjerg, Frede

Published in:
Energy Reports

DOI (link to publication from Publisher):
10.1016/j.egyr.2020.11.055

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Tang, Y., Hu, W., Xiao, J., Lu, Z., Liu, Z., Chen, Z., & Blaabjerg, F. (2020). A Deep Q-Network based optimized modulation scheme for Dual-Active-Bridge converter to reduce the RMS current. *Energy Reports*, *6*(9), 1192-1198. https://doi.org/10.1016/j.egyr.2020.11.055

Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 6 (2020) 1192–1198
www.elsevier.com/locate/egyr

2020 7th International Conference on Power and Energy Systems Engineering (CPESE 2020), 26–29 September 2020, Fukuoka, Japan

A Deep Q-Network based optimized modulation scheme for Dual-Active-Bridge converter to reduce the RMS current

Yuanhong Tang[a], Weihao Hu[a,*], Jian Xiao[b], Zhengdong Lu[a], Zhou Liu[c], Zhe Chen[c], Frede Blaabjerg[c]

[a] *The Power System Wide-area Measurement and Control Sichuan Provincial Key Laboratory, School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[b] *School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[c] *Department of Energy Technology, Aalborg University, Aalborg, Denmark*

Received 23 October 2020; accepted 8 November 2020

Abstract

In order to reduce the conduction losses of the Dual-Active-Bridge (DAB) converter, this paper proposes an optimized modulation scheme based on deep reinforcement learning (DRL). Owing to the Extended-Phase-Shift (EPS) modulation based Deep Q-Network (DQN) algorithm, the optimal phase-shift-angles can be defined, which reduces the root-mean-square (RMS) current tremendously. Moreover, the zero-voltage-switching (ZVS) performance can be guaranteed for the whole operation conditions. A 200 W prototype of the DAB converter is built and tested to prove the effectiveness of the proposed optimized modulation scheme. Experimental results demonstrates that the proposed optimized modulation scheme can obtain lower RMS current and higher operation efficiency in comparison to other three modulations.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Power and Energy Systems Engineering, CPESE, 2020.

Keywords: DAB converter; RMS current; Deep reinforcement learning; Deep Q-Network; ZVS

1. Introduction

In the early 1990s, the Dual-Active-Bridge (DAB) converter was first proposed by Doncker et al. [1]. Due to the benefits of the electrical isolation, high power density, and wide soft switching range, the DAB converter has been widely used in smart grids, uninterruptible power supply and other technical fields.

As the simplest modulation scheme for the DAB converter, the Single-Phase-Shift (SPS) is the most widely employed, which enables easy realization, fast dynamic response and soft switching. However, this method suffers from the narrow soft switching, large circuital current and large root-mean-square (RMS) current, which may cause poor operation efficiency for the DAB converter, especially under light load conditions [2]. Aiming to overcome

* Corresponding author.
E-mail address: whu@uestc.edu.cn (W. Hu).

https://doi.org/10.1016/j.egyr.2020.11.055

Peer-review under responsibility of the scientific committee of the 7th International Conference on Power and Energy Systems Engineering, CPESE, 2020.

the defects existing in the SPS modulation, many improved modulation schemes have been proposed, such as Dual-Phase-Shift (DPS), Extended-Phase-Shift (EPS) and Triple-Phase-Shift (TPS) modulations [3]. By introducing more control variables, these modulations can improve the control flexibility and the operation performance.

Furthermore, based on these multiple control variables modulations, many researchers started to make efforts to improve the efficiency of the DAB converter. An efficiency optimization scheme based on the DPS modulation was proposed to reduce the power losses, but this method suffered from the complicated calculation [4]. And a current stress optimization scheme based on EPS modulation was proposed to reduce the peak current for improving the operation efficiency, but this optimal current stress could lose especially under light load condition [5]. The RMS current and the reactive power can be decreased with the help of the TPS modulation [6,7]. However, the power model is complex under the TPS modulation, which will cause complicated calculation process.

On that basis, it can be found that the efficiency improvement of the DAB converter can be realized by reducing the power losses, the peak current, the RMS current and the reactive power, etc. The power losses existing in the DAB converter mainly include the conduction losses and the switching losses of the power switches, and the copper losses and the core losses of the magnetic components [8]. And, the conduction losses occupies the dominating part of all power losses over the whole operation range, which indicates that reducing RMS current is of a high importance for the DAB converter to improve the operation performance [9].

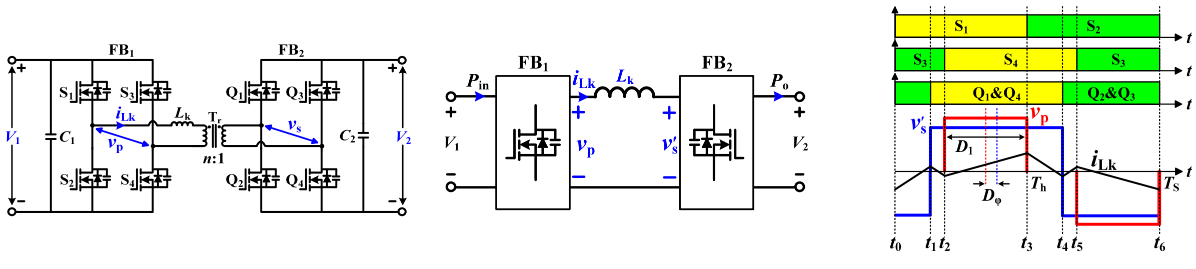
Today, the artificial intelligence (AI) is widely used for solving the optimized control problems. As the typical AI methods, the deep learning (DL) has a strong perceptual ability, while lacks of certain decision-making ability; and the reinforcement learning (RL) has decision-making ability, but it cannot perceive problems [10,11]. Therefore, the deep reinforcement learning (DRL) by combining these two methods can obtain complementary advantages and provide solutions for the cognitive decision-making problem of complex systems [12]. As one of the DRL algorithms, the deep Q-network (DQN) can reduce the computational complexity tremendously, when the states of the environment are too many; and it can provide control strategy in real time [13].

In This paper, a DRL based optimized modulation scheme is proposed to reduce conduction losses of DAB converter. Specifically, the DQN algorithm based on EPS modulation is applied to achieve the minimum RMS current over the whole operation range. During the training of the DQN algorithm, the ZVS constrains are considered to guarantee the soft switching. After the effective implementation of these algorithms, the operation performance can be improved for the DAB converter for whole continuous operation range.

The rest of this paper is organized as follows. In Section 2, the EPS modulation principle and the ZVS constrains will be given firstly. In Section 3, the optimized modulation scheme based on the DQN algorithm will be presented. In Section 4, experimental results will be demonstrated to verify the validity of the proposed optimized modulation scheme. Finally, conclusions will be drawn in Section 5.

2. EPS modulation principle of the DAB converter

The typology of the DAB converter is shown in Fig. 1(a), which consists of one input port, one output port, two full bridge (FB_1, FB_2), one isolation transformer and one series inductor. There are four power switches in each full bridge. L_k denotes the leakage inductor of the transformer and the series inductor, and n represents the turns ratio of the transformer. The AC square-wave voltage v_p and v_s exist at the primary side and secondary side of the transformer, respectively. Furthermore, the corresponding equivalent circuit is illustrated in Fig. 1(b), where P_{in}



(a) (b) (c)

Fig. 1. Dual-Active-Bridge converter. (a) main circuit; (b) equivalent circuit; (c). key waveforms.

is the input power, P_o is the output power, and v_s' is the equivalent value of v_s at the primary side, i.e. $v_s'=n\times v_s$. Fig. 1(c) illustrates the key waveforms of the EPS modulation, where i_{Lk} denotes the current flowing through L_k, D_1 indicates the duty ratio of the v_p and D_φ denotes the phase shift between the center points of v_p and v_s. In the EPS modulation, the duty ratio of the v_s is set as 1, and the dead time between power switches under the same leg is ignored to simplify the analysis.

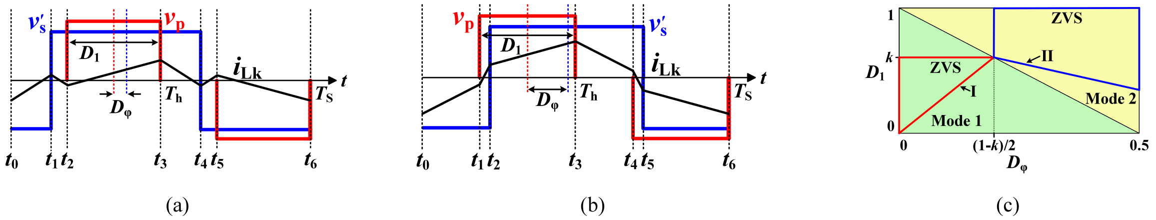


Fig. 2. Working waveforms for different operation modes and corresponding ZVS constraints. (a) *Mode* 1; (b) *Mode* 2; (c). ZVS constraints.

Here we assume that the power is transmitted from the primary side to the secondary side. The voltage conversion ratio k is defined as $k=nV_2/V_1$, where $k<1$ denotes the buck mode operation and $k>1$ denotes the boost mode operation. In this paper, we assume that $k\leq 1$, while the other operation mode $k>1$ can be analyzed similarly. Moreover, the unified output power can be defined as $P_{base} = V_1^2/8L_k f_s$ and the unified current can be defined as $I_{base} = P_{base}/V_1 = V_1/8L_k f_s$, where V_1 represents the input DC voltage, f_s is the switching frequency. Thus, the normalized output power P_{o_u} can be calculated as $P_{o\_u} = P_o/P_{base}$, and the normalized RMS current I_{rms_u} can be calculated as $I_{rms\_u} = I_{rms}/I_{base}$, where P_o denotes the output power and I_{rms} is RMS current flowing through L_k.

In EPS modulation, D_1 is usually limited in [0,1], and D_φ is usually limited in [0, 0.5] to reduce the RMS current I_{rms} [9]. As shown in Fig. 2(a) and (b), two operation modes are contained in the EPS modulation. Moreover, the corresponding ZVS constrains are summarized in Fig. 2(c). As illustrated in Fig. 2(c), the ZVS boundary I can be calculated as $2kD_\varphi/(1\text{-}k) \leqslant D_1$, and the ZVS boundary I can be calculated as $2k(1\text{-}D_\varphi)/(1+k) \leq D_1$. The calculation formulas and corresponding constraints of the EPS are shown in Table 2. The normalized RMS current I_{rms_u} in *Mode* 1 and *Mode* 2 can be calculated by (1) and (2), respectively.

$$I_{rms\_u} = \frac{2\sqrt{3}}{3}\sqrt{kD_1^3 - 2D_1^3 + D_1^2 - 3kD_1 + 12kD_1D_\varphi^2 + k^2} \tag{1}$$

$$I_{rms\_u} = \frac{2\sqrt{3}}{3}\sqrt{-8kD_\varphi^3 + 12kD_\varphi^2 - 6k(1-D_1)^2D_\varphi - 2D_1^3 + 3kD_1^2 + 3D_1^2 - 6kD_1 + k^2 + k} \tag{2}$$

3. Optimized EPS modulation by using the DQN algorithm

3.1. DQN algorithm

The DQN algorithm is adopted in this paper to solve the optimized phase shift angles (D_1, D_φ), according to different operation environments (V_1, V_2, P_o), where V_1 and V_2 represent the input and output DC voltage respectively, and P_o is the output power. Thus the minimum RMS current can be obtained for the DAB converter. Assuming that D_1 is the training variable of the DQN algorithm, the other phase shift angle D_φ can be calculated by the formula of the output power as shown in Table 1. A detailed description of the DQN algorithm is given as follows.

Table 1. Mode constraints, output power and ZVS constrains under the EPS modulation.

Modes	Mode constraints	Normalized output power P_{o_u}	ZVS constraints
1	$0\leq D_\varphi\leq 0.5(1\text{-}D_1)$	$4kD_1D_\varphi$	$2kD_\varphi/(1\text{-}k) \leq D_1\leq k$
2	$0.5(1\text{-}D_1) \leq D_\varphi\leq 0.5+0.5D_1$	$k[1\text{-}(1\text{-}2D_\varphi)^2\text{-}(1\text{-} D_1)^2]$	$2k(1\text{-}D_\varphi)/(1+k) \leq D_1\leq 1$, $0.5(1\text{-}k) \leq D_\varphi\leq 0.5$

• **State s:** The phase shift angle D_1 in the EPS modulation is defined as the state s, i.e. $s = D_1$.

• **Action *a*:** Since Q-learning is used in the DQN algorithm, the state s should be discrete during the training process. The action space is defined as A={δ, -δ}, where δ indicates quantized value of the action a. Thus, the next state will become s'=s+a after action a=δ, while it will become s'=s-a after action a=-δ. In the letter, δ is set as 5×10^{-4}.

• **Reword function *r*:** In order to find the optimized phase shift angles (D_1, D_φ), a reword function is defined as:

$$r(s,a) = \begin{cases} 1, & \Delta I_{rms} < 0, \\ -|\frac{\Delta I_{rms}}{I_{ref}}|, & I_{ref} > \Delta I_{rms} \geq 0, \\ -1, otherwise, & \\ 50, & I_{rms\_c} \leq I_{rms\_\min} \end{cases} \tag{3}$$

where I_{rms} is the RMS current which can be calculated by (1) or (2), I_{ref} is the reference value of I_{rms}, $I_{\mathrm{rms\_min}}$ is the minimum RMS current which will be updated during the training of the DQN algorithm, and $I_{\mathrm{rms\_c}}$ is the RMS current under present state. The increment of the RMS current can be described as $\Delta I_{\mathrm{rms}} = I_{\mathrm{rms\_c}}$- $I_{\mathrm{rms\_p}}$, where $I_{\mathrm{rms\_p}}$ is the RMS current under previous state.

• **The Deep-Neural-Networks of DQN:** In the training process of the DQN, the update of the Q-value should follow the *Bellman* equation:

$$Q^k(s,a) = (r^k + \gamma \max_{a_i' \in A_i} Q^k(s',a')) \tag{4}$$

where $Q^k(s, a)$ represents the Q-value after k times action from state s, γ indicates the discount factor, and $Q^k(s', a')$ represents the Q-value for next state s'.

In the DQN algorithm, the back propagation (BP) neural network is adopted. Thus, the errors of the BP neural network Δ_Q is equal to the difference between both sides of (4). Moreover, an experience replay mechanism is used for the training of DQN algorithm. During the training process, the target value of the BP neural network (y) can be calculated by the right-hand side of (4). The replay memory of experience replay, which is represented by D, includes many batches of (s, a, s', r). Assuming that the capacity of the replay memory is C_{s}, the old samples will be covered by new samples once the sample size is larger than C_{s} during the training process. In each training process, a random mini-batch is sampled from D to train BP neural network.

In this paper, a structure of 1 input (current state s) of the input layer, 2 hidden layers and 1 output (corresponding Q-value) of the output layer are used to batch train the BP neural network. Each hidden layer contains 20 neurons.

• **Action select:** In the DQN algorithm, the ε-greedy principle is used to find the optimal action. The value of $I_{\mathrm{rms\_min}}$ is updated as the minimum RMS current of the last state. After N times training using the ε-greedy principle, minimum value of $I_{\mathrm{rms\_min}}$ will be adopted. And then, the maximum Q-value based action selection will be used to training the agent until the learned strategy get converged. Thus, the action selecting principle can be given by:

$$a_i' = \arg\max_{a_i \in A_i'} Q_i(s_i, a_i). \tag{5}$$

3.2. Training process

Before the training of the DQN algorithm, it is important to set the training parameters and design specifications of the DAB converter. Specifically, the range of the input voltage V_1 is set as 100 V to 140 V, the output voltage V_2 is set as 40 V, the range of the output power P_{o} is chosen as 0 W to 200 W, the switching frequency f_{s} is chosen as 50 kHz, the turns ratio of the transformer is chosen as 1:1, and the value of inductor L_{k} is chosen as 41 μH. The training parameters are illustrated in Table 2, and the corresponding DQN-based agent training diagram is shown in Fig. 3. The training process of the DQN algorithm is shown in Table 3. After the DQN training, the optimal phase shift angles (D_1, D_φ) can be obtained from the trained agent over the whole operation environment conditions (V_1, V_2, P_{o}).

Table 2. Training parameters of the DQN algorithm.

Parameter	Value
Discounting factor (γ)	0.9
State quantity (δ)	5×10^{-4}
Maximum training times (N_T)	10^5
BATCH_SIZE	32
Reference value of I_{rms} (I_{ref})	20
C_s	5000

Table 3. Training progress of the DQN algorithm.

Algorithm: DQN algorithm.

Create action space, sample pool and BP neural network.
Set the N_T, γ, I_{ref}
For each episode do
  Initialize state s and action a randomly
  While (not meet episode end condition) do
    Calculate the I_{rms_c} using (1) or (2)
    Calculate the reward value of last state–action $r(s_i, a_i)$
using (3)
    Store transition (s, a, r, s') in D
    Sample random *BATCH_SIZE* train data form D
    Calculate the expected output of BP neural network (y)
    Train BP neural network
    Select action using ε-greedy police
  End While
End For

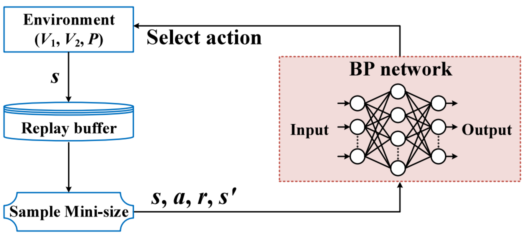


Fig. 3. DQN-based agent training diagram.

4. Experimental results

In order to validate the proposed optimized EPS scheme (O-EPS) based on DQN algorithm, a 200 W experiment hardware prototype has been built, as illustrated in Fig. 4. The brief design specifications of the DAB converter have been depicted in Section 3.2. The detail experimental results will be presented as follows.

Figs. 5 and 6 illustrate the experimental results when $V_1 = 100$ V and $V_2 = 40$ V respectively. As illustrated from Fig. 5(a) and Fig. 6(a), the output voltage V_{out} can be restored quickly and keep stable when P_o is changed from 200 W to 80 W. The corresponding amplified experimental waveforms are shown in Fig. 5(b), Fig. 5(c), Fig. 6(b) and Fig. 6(c), where green circle indicate the ZVS conditions of S_1 and S_2, green dotted circles indicate the ZVS conditions of S_3 and S_4, blue circles indicate the ZVS conditions of Q_1 and Q_2, and blue dotted circles indicate the ZVS conditions of Q_3 and Q_4. According to these figures, the proposed O-PES scheme can realize the ZVS for different load conditions.

Moreover, Fig. 7 depicts the curves of the measured RMS current, and the related efficiency varying with output power between different modulation methods. Fig. 7(a) and Fig. 7(b) depict the measured RMS currents, when $V_1 = 100$ V and $V_1 = 140$ V respectively. According to Fig. 7(a) and Fig. 7(b), the proposed O-EPS can obtain lower RMS currents in comparison to other three modulations. Moreover, Fig. 7(c) and Fig. 7(d) show the curves

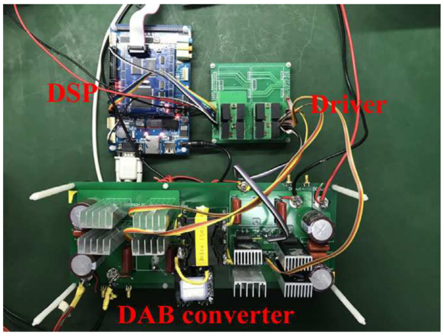
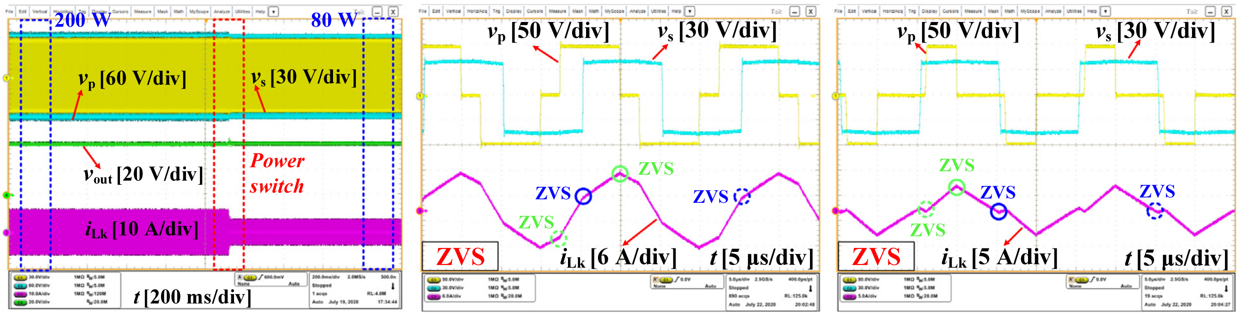
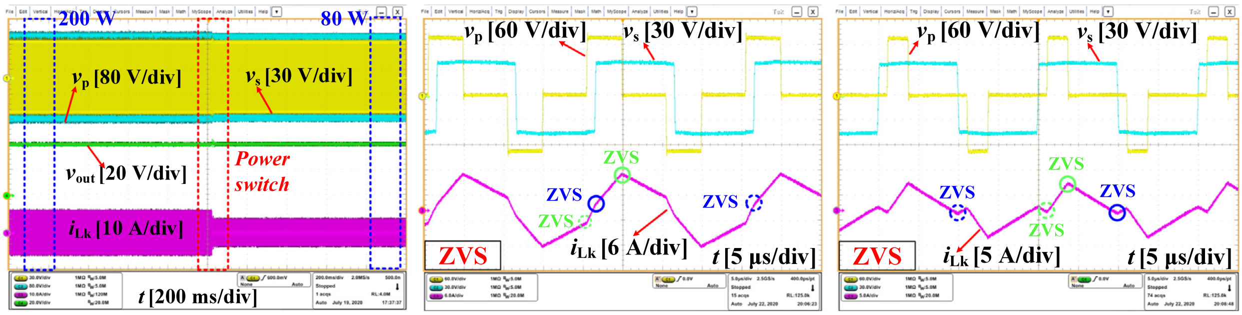


Fig. 4. Experiment hardware prototype.



(a) (b) (c)

Fig. 5. Experiment results for different load conditions with $V_1 = 100$ V and $V_2 = 40$ V. (a) dynamic response when P_o is varied from 200 W to 80 W; (b) amplified experimental waveforms at 200 W. (C) amplified experimental waveforms at 80 W.



(a) (b) (c)

Fig. 6. Experiment results for different load conditions with $V_1 = 140$ V and $V_2 = 40$ V. (a) dynamic response when P_o is varied from 200 W to 80 W; (b) amplified experimental waveforms at 200 W. (C) amplified experimental waveforms at 80 W.

of the measured efficiency with respect to P_o with $V_1 = 100$ V and $V_1 = 140$ V respectively. According to Fig. 7(c) and Fig. 7(d), the proposed O-EPS scheme can obtain higher efficiency than other three modulations.

Based on the experimental results depicted from Figs. 5 to 7, it can be observed that the proposed O-EPS scheme can realize the ZVS for different operation conditions. Furthermore, the RMS current can be reduced and the efficiency can be improved with the help of the proposed O-EPS scheme. Therefore, the proposed O-EPS scheme can be used for the DAB converter under whole operation conditions with excellent performance.

5. Conclusion

In this paper, a deep reinforcement learning based optimized modulation scheme is proposed for the Dual-Active-Bridge (DAB) converter to reduce the root-mean-square (RMS) current. The Deep Q-Network (DQN) algorithm

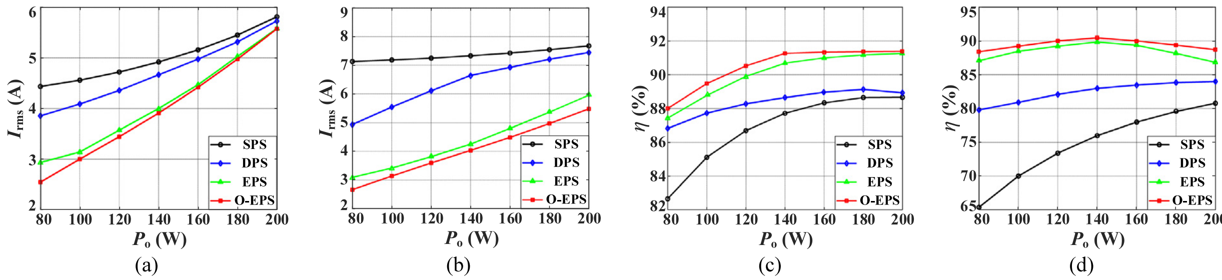


Fig. 7. Measured RMS current and efficiency varied with the output power under different input voltage. (a) Measured RMS current when $V_1 = 100$ V; (b) Measured RMS current when $V_1 = 140$ V; (c). Measured efficiency when $V_1 = 100$ V; (d). Measured efficiency when $V_1 = 140$ V.

based on the Extended-Phase-Shift (EPS) is adopted with the consideration of the zero-voltage-switching (ZVS) constraints during the agent training of the DQN algorithm. Therefore, the trained agent of the DQN algorithm can provide optimal phase shift angles in real time under whole continuous operations. Experimental results indicate that the proposed optimized modulation scheme can reduce the RMS current and improve the efficiency of the DAB converter under different operations. Based on these, the proposed optimized modulation scheme is suitable for smart grids, uninterruptible power supply and other technical fields, when the DAB converter is used.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Sichuan Science and Technology Program, China (Sichuan Distinguished Young Scholars) under Grant 2020JDJQ0037.

References

[1] De Doncker RWAA, Divan DM, Kheraluwala MH. A three-phase soft-switched high-power-density DC/DC converter for high-power applications. IEEE Trans Ind Appl 1991;27(1):63–73.
[2] Xu G, Sha D, Xu Y, et al. Hybrid-bridge-based DAB converter with voltage match control for wide voltage conversion gain application. IEEE Trans Power Electron 2017;33(2):1378–88.
[3] Tang Y, Hu W, Xiao J, et al. Reinforcement learning based efficiency optimization scheme for the DAB DC-DC converter with triple-phase-shift modulation. IEEE Trans Ind Electron 2020. (Early Access).
[4] Zhao B, Song Q, Liu W. Efficiency characterization and optimization of isolated bidirectional DC–DC converter based on dual-phase-shift control for DC distribution application. IEEE Trans Power Electron 2012;28(4):1711–27.
[5] Zhao B, Yu Q, Sun W. Extended-phase-shift control of isolated bidirectional DC–DC converter for power distribution in microgrid. IEEE Trans Power Electron 2011;27(11):4667–80.
[6] Shi H, Wen H, Hu Y, et al. Reactive power minimization in bidirectional DC–DC converters using a unified-phasor-based particle swarm optimization. IEEE Trans Power Electron 2018;33(12):10990–1006.
[7] Tong A, Hang L, Li G, et al. Modeling and analysis of a dual-active-bridge-isolated bidirectional DC/DC converter to minimize RMS current with whole operating range. IEEE Trans Power Electron 2017;33(6):5302–16.
[8] Shih LC, Liu YH, Chiu HJ. A novel hybrid mode control for a phase-shift full-bridge converter featuring high efficiency over a full-load range. IEEE Trans Power Electron 2018;34(3):2794–804.
[9] Liu B, Davari P, Blaabjerg F. An optimized hybrid modulation scheme for reducing conduction losses in dual active bridge converters. IEEE J Emerg Sel Top Power Electron 2020. (Early access).
[10] Sutton RS, Barto AG. Reinforcement learning: An introduction. IEEE Trans Neural Netw 1998;9(5). 1054–1054.
[11] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7.
[12] Arulkumaran K, Deisenroth MP, Brundage M, et al. Deep reinforcement learning: A brief survey. IEEE Signal Process Mag 2017;34(6):26–38.
[13] Lin CC, Deng DJ, Chih YL, et al. Smart manufacturing scheduling with edge computing using multiclass deep Q network. IEEE Trans Ind Inf 2019;15(7):4276–84.