Reinforcement Learning Based Efficiency Optimization Scheme for the DAB DC-DC Converter with Triple-Phase-Shift Modulation

Tang, Yuanhong; Hu, Weihao; Xiao, Jian; Chen, Zhangyong; Huang, Qi; Chen, Zhe; Blaabjerg, Frede

Published in:
I E E E Transactions on Industrial Electronics

DOI (link to publication from Publisher):
10.1109/TIE.2020.3007113

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

Reinforcement Learning Based Efficiency Optimization Scheme for the DAB DC-DC Converter with Triple-Phase-Shift Modulation

Yuanhong Tang, *Student Member, IEEE*, Weihao Hu, *Senior Member, IEEE*, Jian Xiao, *Student Member, IEEE*, Zhangyong Chen, *Member, IEEE*, Qi Huang, *Senior Member, IEEE*, Zhe Chen, *Fellow, IEEE* and Frede Blaabjerg, *Fellow, IEEE*

***Abstract*- Aim to improve the power efficiency of the dual-active-bridge (DAB) DC-DC converter, an efficiency optimization scheme with triple-phase-shift (TPS) modulation using reinforcement learning (RL) is proposed in this paper. More specifically, the Q-learning algorithm, as a typical algorithm of the RL, is applied to train an agent offline to obtain an optimized modulation strategy, and then the trained agent provide control decision online in real-time manner for the DAB DC-DC converter according to the current operation environment. The main objective is to obtain the optimal phase-shift angles for the DAB DC-DC converter, which can achieve the maximum power efficiency by reducing the power losses. Moreover, all possible operation modes of the TPS modulation are considered during the offline training process of the Q-learning algorithm. Thus, the cumbersome process for selecting the optimal operation mode in the conventional schemes can be circumvented successfully. Based on these merits, the proposed efficiency optimization scheme using the RL can realize the excellent performances for the whole load conditions and voltage conversion ratios. Finally, a 1.2 KW prototyped is built, and the simulation and the experimental results demonstrate that the power efficiency can be improved by using the optimization scheme based on the RL.[1]**

Index Terms- DAB DC-DC converter, Power efficiency, Optimization, Reinforcement Learning (RL), Q-learning.

This work was supported by the Sichuan Distinguished Young Scholars (20JCQN0213). (Corresponding author: Weihao Hu)

Yuanhong Tang, Weihao Hu, Qi Huang are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: yhtang@std.uestc.edu.cn; whu@uestc.edu.cn; hwong@uestc.edu.cn)

Jian Xiao is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: xiaojian_student@163.com)

Zhangyong Chen is with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: zhangyongch@uestc.edu.cn)

Zhe Chen and Frede Blaabjerg are with the Department of Energy Technology, Aalborg University, Aalborg, Denmark (e-mail: zch@et.aau.dk, fbl@et.aau.dk).

I. INTRODUCTION

THE dual-active-bridge (DAB) DC-DC converter, which contains a high frequency power transformer and two H-bridges was firstly proposed in early 1990s [1]. As one of the most popular bidirectional topologies, the DAB DC-DC converter is widely used in electric vehicles (EVs), smart grids and renewable energy systems [2]-[4], etc.

Due to the advantages of easy control, high dynamic performance and soft switching, the single-phase-shift (SPS) is the most widely used control strategies in the DAB DC-DC converter [5]-[7]. However, this control scheme suffers from the low efficiency over the wide operation range, which have motivated the improvement of the phase-shift control strategy. As the typical improved modulation scheme from SPS, the triple-phase-shift (TPS) modulation was proposed to extend the zero-voltage-switching (ZVS) range and decrease the overall power losses [8]-[17]. Remarkably, SPS, extended-phase-shift (EPS) and dual-phase-shift (DPS) modulations can be deemed as a special TPS modulation. However, the calculation process of the TPS is complicated.

To further improve the power efficiency of the DAB DC-DC converter, a growing number of researches have made tremendous efforts in power efficiency optimization strategies. The power-loss-model-based optimization method can achieve the optimal efficiency control by establish an accurate power loss model. However, this optimization method suffers from complicated calculation, especially under the complex operation conditions, such as the varied load conditions and voltage conversion ratios. In order to address this problems, there are many advanced iterative methods were proposed, like Lagrange multiplier method (LLM) [9], the genetic algorithm (GA) [12] and Newtons's method [18]. However, these iterative methods suffer from the time consuming, and highly dependent on the model knowledge and the initial guess setting.

Recently, the fast-growing artificial intelligence (AI) technology have changed the traditional control strategy from the past few decades. For example, an optimized TPS control using neural network which can be used to reduce the reactive power for the DAB DC-DC converter was proposed in [19]. However, this neural network utilization is limited by the labor-intensive, time consuming and over-fitting problems.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

With the rapid development of the reinforcement learning (RL), it was widely applied in the optimization of the control problems [20], [21]. The basic learning mechanism of RL shows that it can accept and deal with incomplete and uncertain information in dynamic environment without the model of the environment, produce the best strategy, choose the best action, and thus maximize the impact of its dynamic environment [22], [23]. As one of the most used representative RL methods, Q-learning algorithm has attracted increasing attention recent years. The off-policy strategy which separates the deferral policy from the learning policy is adopted in Q-learning algorithm. Moreover, the ε-greedy policy and the Bellman optimal equations are used to update the action selection [24], [25]. Compared to other RL algorithms, the Q-learning algorithm has simple Q-functions, which can be used online when agents interact with the dynamic environment [26], [27].

In this paper, an efficiency optimization scheme with TPS modulation using RL is proposed. The main objective is to obtain the optimal phase-shift angles for the DAB DC-DC converter, which can achieve the maximum power efficiency by reducing the power losses. The contributions of this paper are shown as follows:

1) An optimized modulation strategy can be obtained by using the Q-learning algorithm, which can improve the power efficiency under whole load conditions and voltage conversion ratios for the DAB DC-DC converter.
2) All possible operation modes of the TPS modulation are considered during the offline training process of the Q-learning algorithm, thus the cumbersome process for selecting the optimal operation mode in the conventional schemes can be circumvented successfully.

Hence, the proposed efficiency optimization scheme benefit from the outstanding performance under the whole operation circumstances.

This paper is further organized as follows. The TPS modulation principle and the detailed power loss analysis of the DAB DC-DC converter are given in section II. The proposed efficiency optimization scheme based on Q-learning algorithm are developed in section III. In section IV, the performance evaluations and comparisons of the Q-learning algorithm by using the Matlab simulation are presented. In section V, the experimental details and results of a 1.2 kW prototype using the Q-learning algorithm are analyzed to prove the correctness of the theory analysis. Some conclusions are summarized in section VI.

II. TPS Modulation And Loss Analysis Of The DAB DC-DC Converter

A. Operation Principle of the TPS Modulation

Fig.1 shows the main circuit of the DAB DC-DC converter. From Fig. 1, the typical structure of the converter is consist by two symmetrical H-bridges and one magnetic tank, where each H-bridge contains of four power switches, and the magnetic tank contains of a power transformer T_r and an external series inductor. Moreover, L_k indicates the value of external series inductor and the leakage inductor of T_r. Moreover, v_{AB} denotes the voltage between the primary side of T_r, v_{CD} denotes the voltage between the secondary side of T_r and i_{Lk} denotes the primary current following through L_K.

The key waveforms of the DAB DC-DC converter by using TPS modulation are depicted in Fig. 2, where v'_{CD} indicates the equivalent voltage between the secondary side of T_r and $v'_{CD}=n\times v_{CD}$. As shown in Fig.2, three phase-shift angles (D_1, D_2, D_3) are contained, where D_1 denotes the phase-shift angle between S_1 and S_4, D_2 denotes the phase-shift angle between Q_1 and Q_4, and D_3 denotes the phase-shift angle between S_1 and Q_1. More specifically, the TPS modulation could be considered as the SPS modulation if $D_1=D_2=1$, the TPS modulation could be considered as the EPS modulation if $D_1=1$ or $D_2=1$, and the TPS modulation could be considered as the DPS modulation if $D_1=D_2\neq1$. Based on different combinations of D_1, D_2 and D_3 under full, partial and no

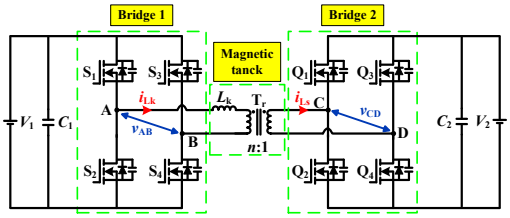


Fig. 1. Main circuit of the DAB DC-DC converter.

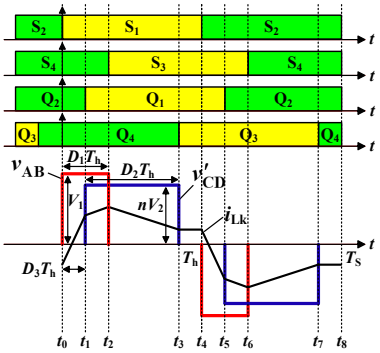


Fig. 2. Key waveforms of the DAB DC-DC converter.

TABLE I
MODE OPERATIONAL CONSTRAINTS AND THE POWER RANGE

Modes	Constraints	Power range (Pu)
1 and 1′	$D_1\geq D_2$, $0\leq D_3\leq(D_1-D_2)$	-0.5~0.5
2 and 2′	$D_2\geq D_1$, $(1+D_1-D_2)\leq D_3\leq 1$	-0.5~0.5
3 and 3′	$D_2\leq(1-D_1)$, $D_1\leq D_3\leq(1-D_2)$	-0.5~0.5
4 and 4′	$D_1\leq D_3\leq 1$, $(1-D_3)\leq D_2\leq(1-D_3+D_1)$	-0.67~0.67
5 and 5′	$(D_1-D_3)\leq D_2\leq(1-D_3)$, $0\leq D_3\leq D_1$	-0.67~0.67
6 and 6′	$(1-D_2)\leq D_1$, $(1-D_2)\leq D_3\leq D_1$	-1~1

TABLE II
THE CLASSIFICATION OF THE POWER LOSSES

Losses	Classification	Value
Power switches losses (P_S)	Conduction losses (P_{C_S})	$R_{DSon}\cdot I^2_{Drms}$
	Switching losses (P_{Sw_S})	$(E_{onM}+E_{offM})\cdot f_{sw}$
	Gate driver losses (P_{Gat_S})	$Q_g\cdot V_{gs}\cdot f_s$
Magnetic losses (P_M)	Copper losses (P_{Cop_M})	$I^2_{Tr\_rms}\cdot[R_{Tr\_pri}+n^2\cdot R_{Tr\_sec}]$
	Core losses (P_{Cor_M})	$k\cdot f_s^{\alpha}\cdot B_{Tr}^{\beta}\cdot V_e$

overlaps, six operation modes and six corresponding complemented modes can be obtained, which are summarized in Table I [28]. The mode operational constraints can be summarized in Table I. Moreover, the transmitted maximum power $P_{o\max}$ can be described as

$$P_{o\max} = \frac{nV_1V_2}{8f_sL_k} \tag{1}$$

The normalized transmitted power can be defined as P_{pu}.

$$P_{pu} = P_o / P_{o\max} \tag{2}$$

where P_o is the transmitted power of the converter. Thus, the power range and the corresponding operations modes of the TPS modulation can be summarized in Table I [28]. For a specific transmitted power, several different operation modes can be used to meet the power requirement.

B. Loss Analysis

The main objective of this paper is to improve the power efficiency by reducing the power losses, thus qualifying the power losses is indispensable. Normally, three kinds of power losses are contained in the DAB DC-DC converter, which namely power switch losses P_S, magnetic losses P_M and unknown losses P_U [29]. More specifically, the power switch losses P_S can be divided into the conduction losses P_{C_S}, the switching losses P_{Sw_S}, and the gate driver losses P_{Gat_S}. The magnetic power losses P_M can be divided into the copper losses P_{Cop_M} and the core losses P_{Cor_M} of the power transformer and the extra series inductor. Furthermore, the unknown losses P_U mainly contains the temperature dependent copper loss and conduction loss relevant to the magnetic devices and the power switch modules respectively, a slightly increased copper loss of Litz wires in the magnetic devices due to skin and proximity effects, and the ohmic losses caused by the DC-link capacitors.

Due to the unknown losses P_U takes a small part of all power losses P_{A_Loss}, the unknown losses P_U is ignored in the power losses analysis to simplify the theoretical analysis. Therefore, the specific loss calculation formula can be calculated and summarized in Table II [15], [30].

Thus, all power losses P_{A_Loss} can be expressed as

$$P_{A\_Loss} = \sum_{i=1}^{8} P_{S_i} + P_{M\_Tr} + P_{M\_L_k} \tag{3}$$

Where the first item indicates the power switch losses in the eight power switches (S_1~S_4, Q_1~Q_4), P_{M_Tr} is the magnetic losses in the power transformer and P_{M_Lk} is the magnetic losses in the extra series inductor. The power efficiency can be improved by reducing the all power losses P_{A_Loss}.

In this paper, the RL is used to find the optimal phase-shift angles (D_1, D_2 and D_3) based on the TPS modulation to obtain the maximum power efficiency, where the training parameters (D_1, D_2 and D_3) are chosen in the twelve operation modes and subject to the modal constraints, which are illustrated in Table I. The detailed training process of the RL will be given in section III.

III. Efficiency Optimization Scheme By Using The Q-Learning Algorithm

A. Q-learning Algorithm

In the Q-learning algorithm, the learned experiences are recorded in a Q-value table, where the optimal action strategy can be obtained according to this Q-value table. More specifically, the Q-value table is made up of the transfer probability for different states, which indicates the behavior with the highest Q-value will be selected directly based on the maximum value behavior selection [27]. In this paper, the Q-learning algorithm is used to solve the optimal control variables (D_1, D_2 and D_3) of the DAB DC-DC converter in a quickly way to obtain the minimum power losses for the whole operation range.

1) State space S: In the DAB DC-DC converter, the reference input quantity consist of the input voltage V_1, output voltage V_2, and transmitted power P_o. For the special reference input quantities V_1, V_2, and P, the power losses P_{A_Loss} are determined by current phase-shift angles D_1, D_2, and D_3. The main objective of this paper is to obtain the optimal phase-shift angles with the minimum power losses by using the Q-learning algorithm. Thus, the state space S can be defined as

$$S = [D_1, D_2, D_3] \tag{4}$$

where the value of D_1, D_2, and D_3 ranges from 0 to 1.

2) Action Space A: The change of the state s is determined by the current action a. According to the current states s, the optimal new state can be obtained by policy π. Due to the current sate s is determined by D_1, D_2, and D_3, the next state s' can be obtained by changing the value of the D_1, D_2, and D_3. Moreover, the value of the D_1, D_2, and D_3 should be changed continuously under the special constraints of Table I, thus the value of the state s should be quantified according to the sensitivity between the transmitted power and phase-shift angles. The variable space C_{Di} is defined as

$$C_{Di} = [0, \pm 1] \times \delta \tag{5}$$

where δ is the quantity value of the state s. The increment ΔD of D_1, D_2, and D_3 should satisfy the constrain as

$$\Delta D_i \in C_{Di} \tag{6}$$

Thus, the action space A can be defined as

$$A = \{C_{D1}, C_{D2}, C_{D3}\} \tag{7}$$

According to (7), the action space A contains 27 options. Thus, the state s will update according to action a.

$$s' = s + a \tag{8}$$

For example, the value of the next state s' will not be changed if the action $a = \{0, 0, 0\}$ is adopted at the state $s = [D_1, D_2, D_3]$, while the next state will become $s' = [D_1+\delta, D_2-\delta, D_3]$ if the action $a = \{\delta, -\delta, 0\}$ is adopted at the state $s = [D_1, D_2, D_3]$.

3) Reward Function r(s, a): Due to the nonlinear equality constraints $P'_o=P_o$ is hard to directionally use in the Q-learning algorithm, a power error function ΔP should is defined as

$$\Delta P(D_1, D_2, D_3) = (P'_o - P_o)^2 \tag{9}$$

where P'_o is the transmitted power during the training process, and P_o is the expected transmitted power. In order to obtained the minimum power error and the minimum power losses, an objective function $F(D_1, D_2, D_3)$ is defined as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

$$F(D_1, D_2, D_3) = P_{A\_Loss}(D_1, D_2, D_3) + \varphi \cdot \Delta P(D_1, D_2, D_3) \quad (10)$$

where $P_{A\_Loss}(D_1, D_2, D_3)$ is the power losses function which is described in (3), $\Delta P(D_1, D_2, D_3)$ denotes the power error function which is shown in (9), and φ is the penalty coefficient. Thus, the performance of the DAB DC-DC converter can be evaluated by the objective function F, where the performance will be better if the value of F is smaller. In order to evaluate the quality of the selected action, a reward function $r(s, a)$ is established, which is described as

$$r(s,a) = \begin{cases} 1 & (\Delta F < 0) \\ -|\dfrac{\Delta F}{F_{ref}}| & (F_{ref} > \Delta F \geq 0) \\ -1 & (otherwise) \\ 120 & (F_c \leq F_{\min}) \end{cases} \quad (11)$$

where F_{ref} is the reference value of the objective function F, and $F_{ref} > 0$. F_{min} is the minimum value of the objective function F, which will be described in part 4). Moreover, ΔF is the difference between two adjacent states of the objective function F, which is expressed by

$$\Delta F = F_c - F_p \quad (12)$$

where F_c is the value of the objective function F at the current state, and F_P is the value of the objective function F in the previous state.

$\Delta F > 0$ means the value of the objective function F in the current state is greater than the previous state, thus the action will lead to a negative reward. $\Delta F < 0$ means the value of the objective function F is reduced after action a, and this action will lead to a positive reward. A larger reward value will be given, once the value of the objective function F is less than or equal to the minimum value of the objective function F_{min}, which indicates that the DAB DC-DC converter has reached the optimal state from the initial state.

In this paper, the reward function is used to find the minimal value of the objective function F during the training process of the Q-learning algorithm, thus the optimal phase-shift angles (D_1, D_2, D_3) with minimum power losses and the minimal power error can be obtained after training.

4) Q-value updating and action selection: As an incremental dynamic programming algorithm, the optimal strategy of the Q-learning is determined step by step. For the policy π, the Q-value can be calculated as

$$Q^{\pi}(s,a) = R_s(a) + \gamma \sum_{s'} P_{ss'}[\pi(s)] V^{\pi}(s') \quad (13)$$

where $R_s(a)$ indicates the average value of the reward at the state s, $P_{ss'}[\pi(s)]$ represents the probability of transferring state s under the policy π, $V^{\pi}(s)$ denotes the expected value obtained by following the policy π at state s. After the learning process, the value of $V^{\pi}(s)$ will be converge to $V^{*}(s)$, which can be defined as

$$V^{*}(s_k) \equiv V^{\pi *}(s_k) = \max_a \{R_s(a) + \gamma \sum_{s'} P_{ss'}[a] V^{\pi *}(s')\} \quad (14)$$

where k denotes the number of iterations.

In fact, the state transformation process of agents of the Q-learning algorithm can usually be modeled as a Markov decision process (MDP). Thus, the updating formula of the Q-table can be expressed as [31]

$$Q^{k+1}(s,a) = Q^k(s,a) + \alpha[r^k + \gamma \max_{a' \in A} Q^k(s',a') - Q^k(s,a)] \quad (15)$$

where α is the learning rate, γ is a discounting factor, and $Q^k(s, a)$ is the Q-value under the state s and the action a.

In order to obtain the optimal operation state for the DAB DC-DC converter, the ε-greedy method is used for the behavior selection. During the selection process, as many exploration strategies as possible are used, and the optimal performance state will be saved at each exploration strategy. Moreover, in each phase of the training process, F_{min} is the minimum value of the objective function F in the previous training episode. The minimum value of F_{min} is chosen as the parameter for equation (8), after N times for training process by using the ε-greedy method. Then, the maximum Q-value is used to make action choices during the training process, until the learning strategy of the Q-learning algorithm become converges. The action selecting principle based on the maximum Q-value is denote by

$$a' = \arg\max_{a \in A} Q(s,a). \quad (16)$$

TABLE III
CRITICAL PARAMETERS OF THE Q-LEARNING ALGORITHM

Parameter	Value
Penalty coefficient of F (φ)	1.0
Learning rate (α)	1.0
Discounting factor (γ)	0.8
State quantity (δ)	10^{-3}
Reference value of F (F_{ref})	20
Maximum training times (N_T)	10^6
Step size of each episode (N_i)	2500
Minimum value of F at previous state (F_{min})	120
Exploration times based on ε-greedy (M)	10^3

TABLE IV
TRAINING PROCESS OF THE Q-LEARNING ALGORITHM

Algorithm: Training process of the Q-learning algorithm

Initialize Q-learning parameters F_{min}, P, V_1, V_2.
Create state space, behavior space and Q-value tables.
Set the N_T, α, γ, F_{ref}, , $Mcout$=0
For each episode do
  Initialize D_1, D_2, and D_3.
  Initialize state s and select action a based on ε-greedy
  Set N_i=0
  While (not meet episode end condition) do
    Calculate the F_c using (10)
    Calculate the △F using (12)
    Update the F_p: F_p=F_c
    Calculate the value of last state-action $r(s, a)$ using (11)
    Calculate state s' using (8)
    Update the Q-value using (13)
    If $Mcout$<M do
      Select action a' based on ε-greedy
    Else do
      Select action a' using (14)
    End If
    Update the state and action: s=s', a=a'
    N_i=N_i+1
  End While
  $Mcout$= $Mcout$+1
End For

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

B. Training of the Q-learning Algorithm

The main objective of the Q-learning algorithm is to obtain an optimal control strategy for the whole operation range with the minimal power losses. Thus, it is of primary interest to chosen the critical parameters for the Q-learning algorithm. The critical parameters are summarized in Table III. The training process following the Q-learning algorithm are illustrated in Table IV. In Table IV, "episode" represents each optimization process of the specific operation environment (V_1, V_2 and P_0), that is, the training process from the initial state to the final state.

Moreover, the algorithm mentioned in Table IV consists of two processes. In the first process, the minimum value of $F_{\min}$ is obtained by using the ε-greedy method. The main purpose of the second process is to find the action strategy to obtain the optimal state, which denotes the minimum value of $F_{\min}$. Since the action a is determined by the maximum Q-value ultimately, the action selection based on the maximum Q-value is chosen as the criterion for the second process to ease the training burden and improve the pace of learning.

After the completion of the training process for the Q-

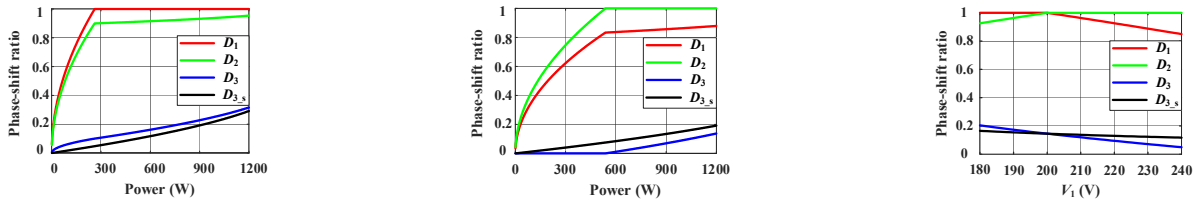


Fig. 3. Curves of the phase-shift angles under different operation conditions after training. (a) V_1=180 V. (b) V_1=240 V. (c) P_o=0.8 kW

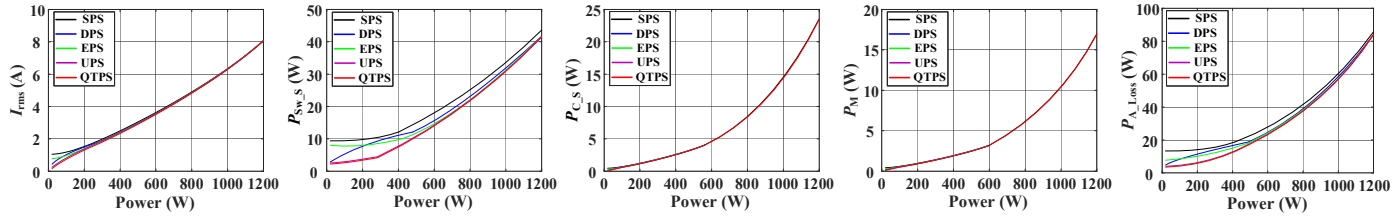


Fig. 4. Power losses and rms current varied with P_o when V_1=180 V and V_2=200 V. (a) I_{rms}. (b) P_{Sw_S}. (c) P_{C_S}. (d) P_M. (e) P_{A_Loss}.

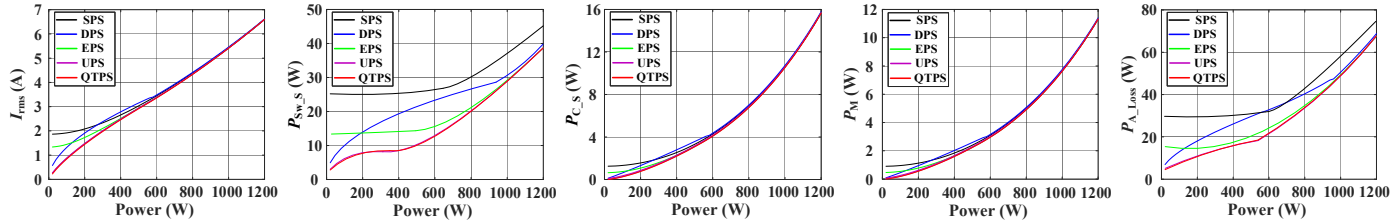


Fig. 5. Power losses and rms current varied with P_o when V_1=240 V and V_2=200 V. (a) I_{rms}. (b) P_{Sw_S}. (c) P_{C_S}. (d) P_M. (e) P_{A_Loss}.

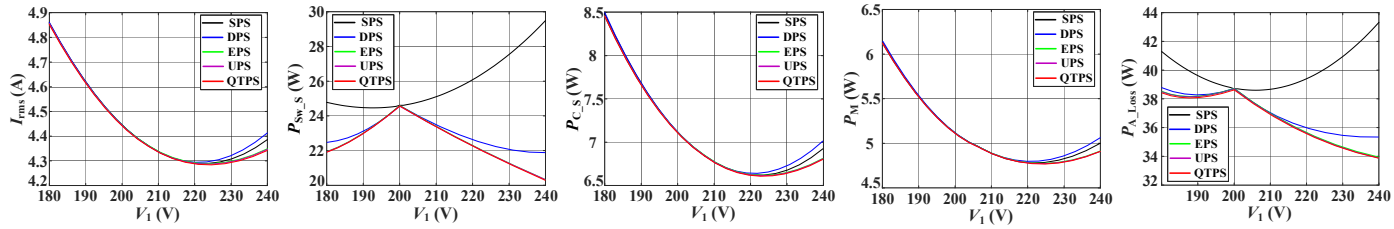


Fig. 6. Power losses and rms current varied with V_1 when P_o=0.8 kW. (a) I_{rms}. (b) P_{Sw_S}. (c) P_{C_S}. (d) P_M. (e) P_{A_Loss}.

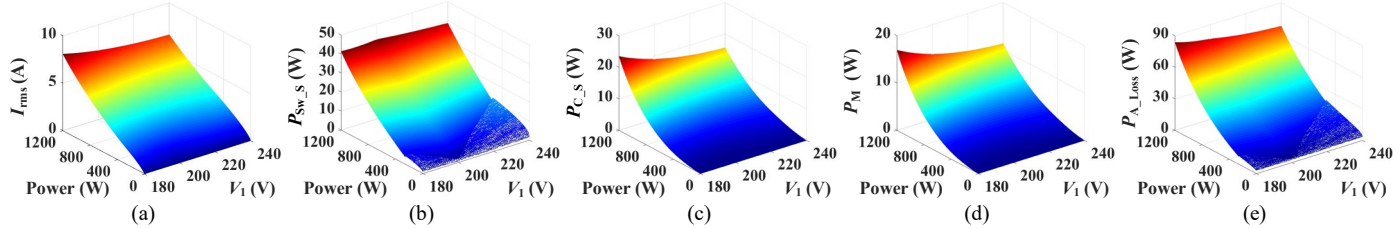


Fig. 7. Power losses and rms current varied with V_1 and P_o. (a) I_{rms}. (b) P_{Sw_S}. (c) P_{C_S}. (d) P_M. (e) P_{A_Loss}.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

learning algorithm, the training results will be stored in a look up table. The inputs into the look up table include the input voltage V_1, the output voltage V_2 and the transmitted power P_0. The corresponding range of the inputs into are shown in Table VI, where V_1 is changing from 180 V to 240 V, V_2 maintains at 200 V and P_0 is changing from 0 W to 1200 W. Moreover, the interval of V_1 and P_0 are set as 0.5 to ensure the control accuracy and reduce the volume of the look-up table. In practice, when the operation environments (V_1, V_2 and P_0) are detected, it will be quantified firstly and then the corresponding action strategy (D_1, D_2 and D_3) will be found directly from this look up table. Notably, if the quantified operation environments (V_1, V_2 and P_0) cannot be found in the look up table, the closest value will be selected and the corresponding action strategy (D_1, D_2 and D_3) can be found directly, which is similar to the method in [25], [26], [31].

To sum up, the Q-learning algorithm is used to solve the efficiency optimization problem for the DAB DC-DC converter in this section. Specifically, the reward function $r(s, a)$ is used to find the minimal value of the objective function F during the training process. By establishing an appropriate algorithm model and selecting the appropriate training parameters, the optimal phase-shift angles (D_1, D_2, D_3) with minimum power losses can be obtained for the whole operation range. Thus, the optimal control strategy can be obtained quickly after the training.

IV. Performance Evaluations And Comparisons Of The Proposed Efficiency Optimization Scheme

The main purpose of the proposed Q-learning optimized triple-phase-shift control (QTPS) is to provide the optimal phase-shift angles (D_1, D_2 and D_3) for the DAB DC-DC converter with the minimum power losses under the whole operation range. In this section, the propose QTPS scheme is evaluated and compared with other modulations through the Matlab simulation. The simulation parameters are chosen in Table VI, and L_k is chosen as 31 μH. The detailed performance evaluations and comparisons will be given as follows.

Fig. 3 shows the variations of the phase-shift angles with the transmitted power P_o and the input voltage V_1 after training, where Fig. 3 (a) illustrates the phase-shift angles varied with the transmitted power P_o when V_1=180 V and V_2=200 V, Fig. 3 (b) illustrates the phase-shift angles varied with the transmitted power P_o when V_1=240 V and V_2=200 V and Fig. 3(c) illustrates the phase-shift angles varied with input voltage V_1 when V_2=200 V and P_o = 0.8 kW. More specifically, D_1, D_2 and D_3 are the phase-shift angles of the propose QTPS, while D_{3_S} is the phase-shift angle of the conventional SPS modulation. As seen from Fig. 3 (a) and Fig. 3 (b), all of the phase-shift angles of the propose QTPS scheme are increased as the transmitted power P_o increase. Moreover, the TPS modulation is adopted under the light load conditions and the EPS modulation is adopted under the heavy load conditions. In can be seen from Fig. 3 (c), D_1 and D_3 are decreased as the increase of V_1, while D_2 is increased as the increase of V_1.

TABLE V
Power Losses Comparison Between Different Approaches

Operating conditions	Initial guess setting	Method	Power losses (W)
V_1=180 V, V_2=200 V, P_o=0.8 kW	D_1=0.5, D_2=0.5, D_3=0.5	Newton	39.2
		GA	38.9
		PSO	38.1
		Q-learning	37.8
V_1=180 V, V_2=200 V, P_o=0.8 kW	D_1=1, D_2=0.5, D_3=0.5	Newton	38.6
		GA	38.4
		PSO	38.0
		Q-learning	37.8
V_1=180 V, V_2=200 V, P_o=0.8 kW	D_1=0.5, D_2=1, D_3=0.5	Newton	39.5
		GA	38.8
		PSO	38.0
		Q-learning	37.8
V_1=240 V, V_2=200 V, P_o=0.8 kW	D_1=0.5, D_2=0.5, D_3=0.5	Newton	35.4
		GA	34.7
		PSO	33.3
		Q-learning	33.1
V_1=240 V, V_2=200 V, P_o=0.8 kW	D_1=1, D_2=0.5, D_3=0.5	Newton	35.8
		GA	34.5
		PSO	33.2
		Q-learning	33.1
V_1=240 V, V_2=200 V, P_o=0.8 kW	D_1=0.5, D_2=1, D_3=0.5	Newton	34.1
		GA	34.4
		PSO	33.2
		Q-learning	33.1

In this paper, one of the advantages of using the Q-learning algorithm to do optimization is its high efficiency in the calculation and does not depend very much on the model knowledge and initial guess setting, as compared to GA and other iterative methods [12], [17], [18], [27], [31]. During the training of the Q-learning algorithm, the initial guess setting for D_1-D_3 are random, which indicated the Q-learning algorithm does not depend on the initial guess setting. Furthermore, the quantitative results of the different approaches under different initial guess settings are presented in Table V. As is seen from Table V, the Q-learning algorithm and the PSO [17] can effectively reduce the power losses for different initial guess setting, while the Newton's method [18] and the GA [12] will suffer from high power losses when the initial guess setting is improper.

According to the comparison results in Table V, both the PSO and the Q-learning algorithm demonstrate similar performance in that both can found the best solutions and are independent of the initial guess settings. However, a new optimization process is required when the environment (V_1, V_2, P) changes for the PSO algorithm. In this paper, the operation range of the DAB DC-DC converter is shown in Table VI, where V_1 is changing from 180 V to 240 V, V_2 maintains at 200 V and P_0 is changing from 0 W to 1200 W. If we assume that the interval of V_1 and P_0 are set as 0.5 V and 0.5 W, respectively. Thus, 288,000 optimization processes are needed by using the PSO algorithm, which will be very time-consuming and almost impossible to solve. However, the Q-learning algorithm can accept and deal with incomplete and uncertain information in dynamic environment, which indicates the optimization strategies can be obtained under different operation environments (V_1, V_2, P) [22], [23], [26].

Fig. 4 shows the power losses under different load conditions for SPS, DPS, EPS, unified-phase-shift (UPS) [16]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

and the proposed QTPS scheme, when V_1=180 V and V_2=200 V. More specifically, Fig. 4 (a) illustrates the rms current I_{rms} follows through the inductor L_k, , Fig. 4 (b) illustrates the switches switching losses P_{Sw_S}, Fig. 4 (c) illustrates the switches conduction losses P_{C_S}, Fig. 4 (d) illustrates the magnetic losses P_M and Fig. 4 (e) illustrates all power losses P_{A_Loss}. As illustrated in Fig. 4 (a), the rms current I_{rms} follows through the inductor L_k of the proposed QTPS and the UPS modulation is slightly smaller than the other three modulations, especially at light load conditions. Moreover, as seeing from Fig. 4 (b), the switches switching losses of the proposed QTPS and the UPS modulation are smaller than other three modulation, especially under the light load conditions. The curves of Fig. 4 (c) and Fig. 4 (d) are similar to Fig. 4 (a), due to the switches conduction losses P_{C_S} and the magnetic losses P_M are proportional to the square of the I_{rms}, Moreover, the switches conduction losses P_{C_S} and the magnetic losses P_M of the proposed QTPS scheme is very close to other four modulations, because of the value of the voltage conversion ratio k is close to 1. Based on these, all power losses of the proposed QTPS is smaller than SPS, DPS, and EPS modulations, especially at the light load conditions.

Fig. 5 shows the power losses under different load conditions for SPS, DPS, EPS, UPS and the proposed QTPS scheme, when V_1=240 V and V_2=200 V. More specifically, Fig. 5 (a) illustrates the rms current I_{rms} follows through the inductor L_k, Fig. 5 (b) illustrates the switches switching losses P_{Sw_S}, Fig. 5 (c) illustrates the switches conduction losses P_{C_S}, Fig. 5 (d) illustrates the magnetic losses P_M and Fig. 5(e) illustrates all power losses P_{A_Loss}. As shown in Fig. 5 (a), the rms current I_{rms} follows through the inductor L_k of the proposed QTPS and the UPS modulation are less than the SPS, DPS and EPS modulations, especially under light load conditions. The curves of Fig. 5 (b) are analogous to Fig. 4 (b), which indicates that the switches switching losses of the proposed QTPS and the UPS modulation are less than the other three modulations. As can be seen from Fig. 5 (c) and Fig. 5 (d), the switches conduction losses and the magnetic losses of the proposed QTPS and the UPS modulation are less than the other three modulations at light load conditions, while are closed to the SPS modulation at heavy load conditions. As is shown in Fig. 5 (d), it is clear that all power losses of the proposed QTPS is smaller than SPS, DPS, and EPS modulations, especially at the light load conditions.

Fig. 6 shows the power losses under different voltage conversion ratios at 0.8 kW for SPS, DPS, EPS, UPS and the proposed QTPS scheme, where the output voltage V_2 is stabled at 200 V and the input voltage V_1 is changing from 180 V to 240V. More specifically, Fig. 6 (a) illustrates the rms current I_{rms} follows through the inductor L_k, Fig. 6 (b) illustrates the switches switching losses P_{Sw_S}, Fig. 6 (c) illustrates the switches conduction losses P_{C_S}, Fig. 6 (d) illustrates the magnetic losses P_M and Fig. 6 (e) illustrates all power losses P_{A_Loss}. As illustrated in Fig. 6 (a), the rms current I_{rms} of the proposed QTPS, UPS and EPS are slightly smaller than the SPS and DPS modulations when the input voltage V_1 is not matched with the output voltage V_2. As seen from Fig. 6 (b), the switches switching losses of the proposed QTPS, UPS and EPS are less than the SPS and DPS modulation, especially under large voltage conversion ratio. Moreover, according to Fig. 6 (c) and Fig. 6 (d), the switches conduction losses and the magnetic losses of the proposed QTPS, UPS and EPS are slightly smaller than the SPS and DPS modulations when the input voltage V_1 is not matched with the output voltage V_2. The curves of Fig. 6 (c) and Fig. 6 (d) are similar to Fig. 6 (a), due to the switches conduction losses and copper losses of the magnetic devices is proportional to I_{rms}^2. From Fig. 6 (e), it is clear that all power losses of the proposed QTPS, UPS and EPS are less than the SPS and DPS modulation, especially under large voltage conversion ratios. Due to the SPS modulation is adopted in the proposed QTPS scheme when input voltage V_1 is matched with the output voltage V_2, the power losses of DPS, EPS, UPS and the proposed QTPS are the same as the SPS.

Fig. 7 shows the rms current and all kinds of power losses of the proposed QTPS scheme under different voltage conversion ratios and different load conditions, where output voltage V_2 is stabled at 200 V, the input voltage V_1 is changing from 180 V to 240V and the transmitted power is varied from 0 W to 1.2 kW. More specifically, Fig. 7 (a) illustrates the rms current I_{rms} follows through the inductor L_k, Fig. 7 (b) illustrates the switches switching losses P_{Sw_S}, Fig. 7 (c) illustrates the switches conduction losses P_{C_S}, Fig. 7 (d) illustrates the magnetic losses P_M and Fig. 7 (e) illustrates all power losses P_{A_Loss}. As is seen from Fig. 7, the rms current and the corresponding power losses will be increased as the increase of the voltage conversion ratio k under the same load conditions. Moreover, the rms current and the power losses will be increased as the increase of the transmitted power P_o under the same voltage conversion ratio k.

To sum up, after training of the Q-learning algorithm, the

TABLE VI
PARAMETERS OF THE DAB DC-DC CONVERTER

Item	Parameter
Rated transmitted power P_{base} (kW)	1.2
Input voltage V_1 (V)	180~240
Output voltage V_2 (V)	200
Transformer turn ratio (n:1)	1:1
Switching frequency f_s (KHz)	100
Power switches S_1~S_4 and Q_1~Q_4	IPP60R099C6 (650 VDC, 37.9 A)

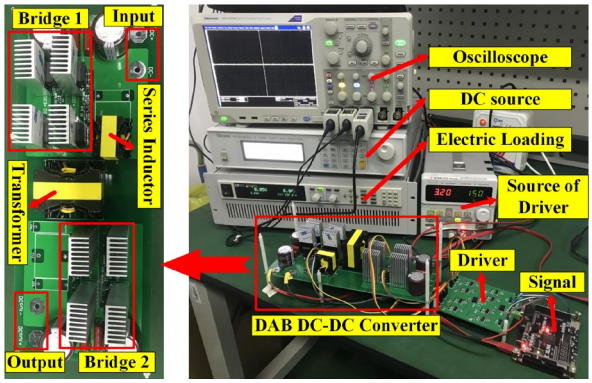


Fig. 8. Experimental hardware platform for 1.2 kW nominal power

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

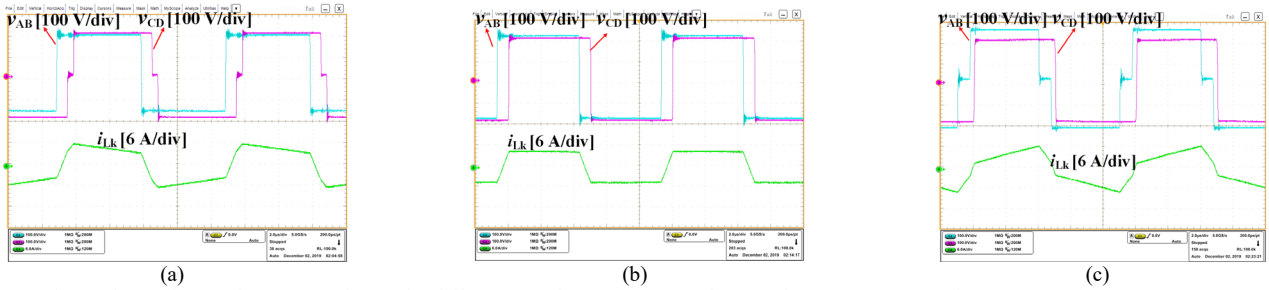


Fig. 9. Experimental waveforms of v_{AB}, v_{CD} and i_{Lk} under different V_1, when V_2=200 V and P_o=0.8 kW. (a) V_1=180 V. (b) V_1=200 V. (c) V_1=240 V.

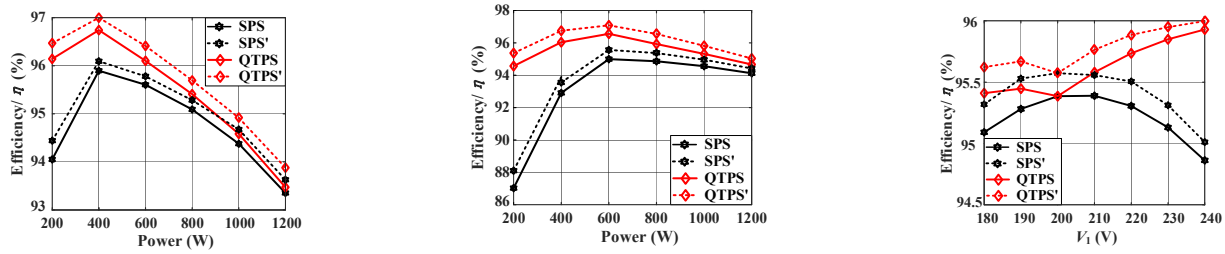


Fig. 10.Theoretical power efficiency and measured power efficiency varied with P_o and V_1. (a) Power efficiency varied with P_o when V_1= 180 V. (b) Power efficiency varied with P_o when V_1= 240 V. (c) Power efficiency varied with V_1 when P_o = 0.8 kW.

optimal phase-shift angles (D_1, D_2 and D_3) can be obtained under whole operation range, thus the power losses and the rms current of i_{Lk} can be reduced, especially under large voltage conversion ratios and light load conditions.

V. Experimental Verification

In this section, a 1.2 kW nominal power hardware topology is built to verify the feasibility of the proposed efficiency optimization scheme with TPS modulation by using RL. The detailed key design parameters of the DAB DC-DC converter are summarized in Table VI. The design procedure of the series inductor L_k and detailed experiment analysis are given in the following.

A. Design Procedure of the Series Inductor L_k

Due to the series inductor L_k should be designed to transmit the maximum transmitted power P_{omax}, thus based on the equation mentioned in (1), L_k can be calculated as:

$$L_k \le \frac{nV_1V_2}{8f_sP_{o\max}} \tag{17}$$

In order to assure a 20% margin of the transmitted power, the maximum transmitted power P_{omax} is chosen as

$$P_{o\max} = 1.2 \cdot P_{base} = 1.44 \text{ kW} \tag{18}$$

Hence, according to the key design parameters listed in Table VI, based on the equations of (17) and (18), L_k can be calculated as:

$$L_k \le 31.25\ \mu H \tag{19}$$

Due to the startup current of the DAB DC-DC converter increases as the decreases of L_k, L_k should be chosen as large enough if equation (19) is met. Thus, L_k is chose as 31 μH.

B. Experiment Analysis

A 1.2 kW nominal power hardware topology is built to prove the correctness of the theory analysis. The experiment hardware platform is shown in Fig. 8. The detail experiment analysis will be given as follows.

Fig. 9 shows the experimental waveforms of v_{AB}, v_{CD} and i_{Lk} under different input voltage V_1, when output voltage V_2=200 V and transmitted power P_o=0.8 kW. More specifically, Fig. 9 (a) illustrates the experiment result when input voltage V_1=180 V, Fig. 9 (b) illustrates the experiment result when input voltage V_1=200 V and Fig. 9 (c) illustrates the experiment result when input voltage V_1=240 V.

As seen from Fig. 9 (b), the conventional SPS modulation is adopted, which indicates that SPS modulation is used when the input voltage V_1 is matched with the output voltage V_2. Furthermore, compared Fig. 9 (b) with Fig. 9 (a) and Fig. 9 (c), the minimum rms value and peak value of i_{Lk} is obtained when V_1=V_2=200 V. Fig. 9 (a), Fig. 9 (b) and Fig. 9 (c) indicate that the rms current and peak current of i_{Lk} increase as the increase of the voltage conversion ratio at the same load condition.

The curves of the measured power efficiency and the theoretical power efficiency varied with transmitted power P_o and input voltage V_1 are depicted in Fig. 10, where Fig. 10 (a) shows the power efficiency varied with the transmitted power P_o when V_1=180 V and V_2=200 V, and Fig. 10 (b) shows the power efficiency varied with the transmitted power P_o when

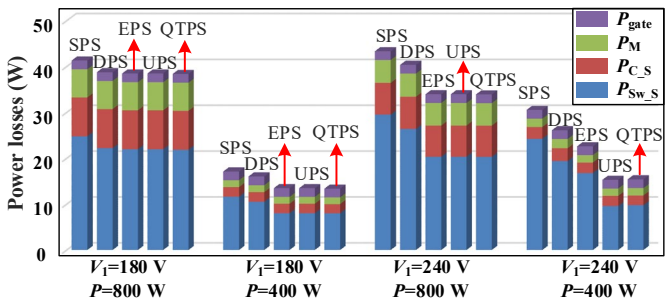


Fig. 11.Detailed power losses comparisons at different V_1 when P_o=0.8 kW and V_2=200 V.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007113, IEEE Transactions on Industrial Electronics

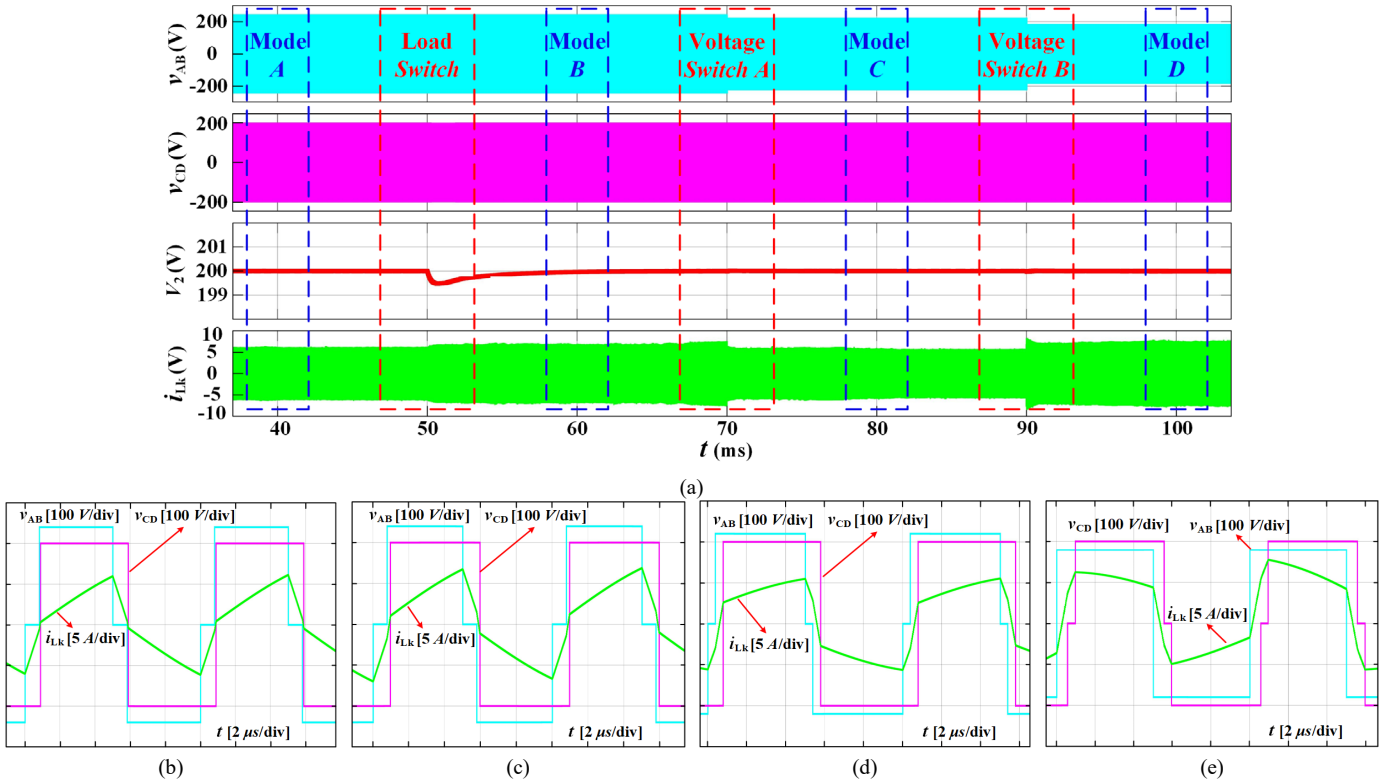


Fig. 12. Dynamic performance for the input voltage variation and the load transition. (a) Simulation waveforms for the input voltage and the load change conditions. (b) Mode *A*: V_1=240 V, *load*=60 Ω. (c) Mode *B*: V_1=240 V, *load*=50 Ω. (d) Mode *C*: V_1=220 V, *load*=50 Ω. (e) Mode *D*: V_1=180 V, *load*=50 Ω.

V_1=240 V and V_2=200 V. More specifically, the curve of SPS indicate the measured power efficiency by using SPS modulation, the curve of SPS' indicate the theoretical power efficiency by using SPS modulation, the curve of QTPS indicate the measured power efficiency for the proposed QTPS scheme, the curve of QTPS' indicate the theoretical power efficiency for the proposed QTPS scheme. According to Fig. 10, the measured power efficiency is slightly lower than the theoretical power efficiency, due to the unknown power losses is not considered in the loss analysis model.

It can be seen from Fig. 10 (a) that the power efficiency of the proposed efficiency optimization scheme is higher than the SPS modulation for the whole range of the transmitted power P_o, especially when P_o is small, while the curves of the power efficiency between the proposed method and the SPS modulation get closer as the increase of P_o after P_o= 400 W. Moreover, the maximum measured power efficiency of the proposed can be reached about 96.7 % at around P_o=400 W. Compared with the SPS modulation, the maximum measured power efficiency has an improvement of 0.8% and the power efficiency is improved by 2.1% at small transmitted power P_o. Furthermore, the curves of the measured power efficiency shown in Fig.10 (b) has the similar trend as Fig. 10 (a), where the maximum measured power efficiency of the proposed can be reached about 96.6 % at around P_o=600 W. Compared with the SPS modulation, the maximum measured power efficiency has an improvement of 1.6% and the power efficiency is improved by 7.5% at small transmitted power P_o.

Based on the comparison results shown in Fig. 10 (a) and Fig. 10 (b), it is obvious that the efficiency of the proposed efficiency optimization scheme with TPS modulation using the RL is higher than the SPS modulation for the whole load conditions when the input voltage V_1 is not matched with output voltage V_2, especially under light load conditions, while the curves of the power efficiency between the proposed method and the SPS get closer as the increase of P_o under heavy load conditions.

Fig. 10 (c) shows the power efficiency varied with input voltage V_1 when V_2=200 V and P_o = 0.8 kW. According to Fig. 10 (c), the efficiency of the proposed efficiency optimization scheme is higher than the SPS modulation, especially under the large voltage conversion ratio conditions. Nevertheless, the curves of the power efficiency is overlapped when V_1=V_2=200 V, due to the SPS modulation is adopted in the proposed DTPS scheme when V_1 is matched with V_2.

The detailed power losses comparisons for the SPS, DPS, EPS, UPS and the proposed QTPS scheme at different V_1 when P_o=0.8 kW and V_2=200 V are illustrated in Fig. 11. According to Fig. 11, for each load condition and each voltage conversion ratio *k*, the power losses under EPS modulation is less than DPS and SPS modulation, while the proposed QTPS scheme and the UPS modulation have the lowest power losses. Moreover, the power losses of the proposed QTPS scheme is very close to the UPS, while the proposed QTPS scheme has slightly lower power losses than the UPS modulation at light load condition. Based on these, the proposed QTPS scheme and the UPS modulation can reduce the power losses and

improve the power efficiency compare to SPS, DPS and EPS modulations, especially under large voltage conversion ratio *k*.

Fig.12 shows the dynamic performance of the proposed QTPS scheme for the input voltage variation and the load transition through the Matlab simulation, where the desired output voltage V_2 is equal to 200 V. Fig. 12 (a) denotes the simulation waveforms for the input voltage and the load change conditions, where the Load *Switch* phase indicates the load resistance is changing from 60 Ω to 50 Ω when the input voltage is maintained at 240 V, the Voltage *Switch A* phase indicates the input voltage is changing from 240 V to 220 V when the load resistance is fixed at 50 Ω, and the Voltage *Switch B* phase indicates the input voltage is changing from 220 V to 180 V when the load resistance is fixed at 50 Ω. According to Fig. 12 (a), the output voltage of the converter can return quickly and keep stable at the desired value 200 V, which prove the fast dynamic performance and good stability of the proposed QTPS scheme. The corresponding steady state waveforms (Mode *A* to Mode *D*) have been zoomed and illustrated in Fig. 12 (b) to Fig. 12 (e), which prove the correctness of the proposed QTPS scheme for different voltage conversion ratio *k* and different load conditions.

To sum up, the proposed efficiency optimization scheme with TPS modulation using RL can improve the power efficiency under whole operation range, especially under the large voltage conversion ratio conditions and light load conditions. However, the SPS modulation is adopted when the input voltage V_1 is matched with output voltage V_2, which indicates the operation performance is the same as the SPS modulation under this condition. Based on these, the experimental results agree well with the theoretical analysis.

VI. Conclusion

This paper proposed an efficiency optimization scheme with triple-phase-shift (TPS) modulation using reinforcement learning (RL) for the dual-active-bridge (DAB) DC-DC converter. More specifically, the Q-learning algorithm, as a typical algorithm of the RL, is applied to train an agent offline to obtain an optimized modulation strategy, and then the trained agent provide control decision online in real-time manner for the DAB DC-DC converter according to the current operation environment of the converter. By using the Q-learning algorithm, the optimal phase-shift angles (D_1, D_2 and D_3) of the triple-phase-shift (TPS) modulation can be obtained, which can achieve the desired maximum power efficiency. Moreover, all possible operation modes of the TPS modulation are considered during the offline training process of the Q-learning algorithm, thus the cumbersome process for selecting the optimal operation mode in the conventional schemes can be circumvented successfully. The correctness of the theoretical analysis and the effectiveness of the proposed optimization scheme are verified by simulation and experimental results. Compared with the conventional SPS modulation, the measured maximum efficiency is improved by 1.6% at around 600 W, and a 7.5% efficiency improvement is observed under light load condition. Based on these merits, the proposed efficiency optimization scheme benefits from the excellent performance under the whole operation range.

References

[1] M. N. Kheraluwala, R. W. Gascoigne, D. M. Divan and E. D. Baumann, "Performance characterization of a high-power dual active bridge DC-to-DC converter," *IEEE Transactions on Industry Applications*, vol. 28, no. 6, pp. 1294-1301, Nov.-Dec. 1992.

[2] Z. Zhang and K. Chau, "Pulse-Width-Modulation-Based Electromagnetic Interference Mitigation of Bidirectional Grid-Connected Converters for Electric Vehicles," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2803-2812, Nov. 2017.

[3] D. Sha, G. Xu and X. Liao, "Control Strategy for Input-Series-Output-Series High-Frequency AC-Link Inverters," *IEEE Transactions on Power Electronics*, vol. 28, no. 11, pp. 5283-5292, Nov. 2013.

[4] G. G. Oggier and M. Ordonez, "High-Efficiency DAB Converter Using Switching Sequences and Burst Mode," *IEEE Transactions on Power Electronics*, vol. 31, no. 3, pp. 2069-2082, Mar. 2016.

[5] L. Wang, Z. Wang and H. Li, "Asymmetrical Duty Cycle Control and Decoupled Power Flow Design of a Three-port Bidirectional DC-DC Converter for Fuel Cell Vehicle Application," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 891-904, Feb. 2012.

[6] D. Costinett, D. Maksimovic and R. Zane, "Design and Control for High Efficiency in High Step-Down Dual Active Bridge Converters Operating at High Switching Frequency," *IEEE Transactions on Power Electronics*, vol. 28, no. 8, pp. 3931-3940, Aug. 2013.

[7] X. Liu *et al*., "Novel Dual-Phase-Shift Control With Bidirectional Inner Phase Shifts for a Dual-Active-Bridge Converter Having Low Surge Current and Stable Power Control," *IEEE Transactions on Power Electronics*, vol. 32, no. 5, pp. 4095-4106, May. 2017.

[8] J. Huang, Y. Wang, Z. Li and W. Lei, "Unified Triple-Phase-Shift Control to Minimize Current Stress and Achieve Full Soft-Switching of Isolated Bidirectional DC–DC Converter," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 7, pp. 4169-4179, July. 2016.

[9] N. Hou, W. Song and M. Wu, "Minimum-Current-Stress Scheme of Dual Active Bridge DC–DC Converter With Unified Phase-Shift Control," *IEEE Transactions on Power Electronics*, vol. 31, no. 12, pp. 8552-8561, Dec. 2016.

[10] G. Xu, D. Sha, J. Zhang and X. Liao, "Unified Boundary Trapezoidal Modulation Control Utilizing Fixed Duty Cycle Compensation and Magnetizing Current Design for Dual Active Bridge DC–DC Converter," *IEEE Transactions on Power Electronics*, vol. 32, no. 3, pp. 2243-2252, Mar. 2017.

[11] S. S. Muthuraj, V. K. Kanakesh, P. Das and S. K. Panda, "Triple Phase Shift Control of an LLL Tank Based Bidirectional Dual Active Bridge Converter," *IEEE Transactions on Power Electronics*, vol. 32, no. 10, pp. 8035-8053, Oct. 2017.

[12] L. Meng, T. Dragicevic, J. C. Vasquez and J. M. Guerrero, "Tertiary and Secondary Control Levels for Efficiency Optimization and System Damping in Droop Controlled DC–DC Converters," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2615-2626, Nov. 2015.

[13] M. Veerachary and A. R. Saxena, "Optimized Power Stage Design of Low Source Current Ripple Fourth-Order Boost DC-DC Converter: A PSO Approach," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1491-1502, Mar. 2015.

[14] A. Tong, L. Hang, G. Li, X. Jiang and S. Gao, "Modeling and Analysis of a Dual-Active-Bridge-Isolated Bidirectional DC/DC Converter to Minimize RMS Current With Whole Operating Range," *IEEE Transactions on Power Electronics*, vol. 33, no. 6, pp. 5302-5316, Jun. 2018.

[15] L. Shih, Y. Liu and H. Chiu, "A Novel Hybrid Mode Control for a Phase-Shift Full-Bridge Converter Featuring High Efficiency Over a Full-Load Range," *IEEE Transactions on Power Electronics*, vol. 34, no. 3, pp. 2794-2804, Mar. 2019.

[16] F. An, W. Song, K. Yang, S. Yang and L. Ma, "A Simple Power Estimation With Triple Phase-Shift Control for the Output Parallel DAB DC–DC Converters in Power Electronic Traction Transformer for

Railway Locomotive Application," *IEEE Transactions on Transportation Electrification*, vol. 5, no. 1, pp. 299-310, March 2019.
[17] H. Shi, H. Wen, Y. Hu and L. Jiang, "Reactive Power Minimization in Bidirectional DC–DC Converters Using a Unified-Phasor-Based Particle Swarm Optimization," *IEEE Transactions on Power Electronics*, vol. 33, no. 12, pp. 10990-11006, Dec. 2018.
[18] Z. Du, L. M. Tolbert, J. N. Chiasson and B. Ozpineci, "Reduced Switching-Frequency Active Harmonic Elimination for Multilevel Converters," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 4, pp. 1761-1770, Apr. 2008.
[19] Y. A. Harrye, K. H. Ahmed and A. A. Aboushady, "Reactive power minimization of dual active bridge DC/DC converter with triple phase shift control using neural network," *2014 International Conference on Renewable Energy Research and Application (ICRERA)*, Milwaukee, WI, 2014, pp. 566-571.
[20] R. Munos, T. Stepleton, A. Harutyunyan, and M. Bellemare, "Safe and efficient off-policy reinforcement learning," *2016 International Conference on Neural Information Processing Systems (NIPS)*, Barcelona, 2016, pp. 1054-1062.
[21] J. Fu, H. He and X. Zhou, "Adaptive Learning and Control for MIMO System Based on Adaptive Dynamic Programming," *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1133-1148, Jul. 2011.
[22] R. Xiong, J. Cao, Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," *Applied energy*, vol. 211, pp. 538-548, Feb. 2018.
[23] L. Xiao, Y. Li, C. Dai, H. Dai and H. V. Poor, "Reinforcement Learning-Based NOMA Power Allocation in the Presence of Smart Jamming," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3377-3389, Apr. 2018.
[24] Q. Wei, F. L. Lewis, Q. Sun, P. Yan and R. Song, "Discrete-Time Deterministic *Q*-Learning: A Novel Convergence Analysis," *IEEE Transactions on Cybernetics*, vol. 47, no. 5, pp. 1224-1237, May. 2017.
[25] Y. Jiang, J. Fan, T. Chai, F. L. Lewis and J. Li, "Tracking Control for Linear Discrete-Time Networked Control Systems With Unknown Dynamics and Dropout," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 4607-4620, Oct. 2018.
[26] Q. Wei, D. Liu and G. Shi, "A novel dual iterative Q-learning method for optimal battery management in smart residential environments," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2509-2518, Apr. 2015.
[27] B. Jang, M. Kim, G. Harerimana and J. W. Kim, "Q-learning Algorithms: A Comprehensive Classification and Applications," *IEEE Access*, vol. 7, pp. 133653-133667, 2019.
[28] Y. A. Harrye, K. H. Ahmed, G. P. Adam and A. A. Aboushady, "Comprehensive steady state analysis of bidirectional dual active bridge DC/DC converter using triple phase shift control," *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, Istanbul, 2014, pp. 437-442.
[29] H. Akagi, T. Yamagishi, N. M. L. Tan, S. Kinouchi, Y. Miyazaki and M. Koyama, "Power-Loss Breakdown of a 750-V 100-kW 20-kHz Bidirectional Isolated DC–DC Converter Using SiC-MOSFET/SBD Dual Modules," *IEEE Transactions on Industry Applications*, vol. 51, no. 1, pp. 420-428, Jan.-Feb. 2015.
[30] D. Graovac, M. Pürschel and A. Kiep, *MOSFET power losses calculation using the data-sheet parameters*, Infineon application, Neubiberg, Germany, 2006.
[31] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, PP.279-292 , Jan.1992.

Yuanhong Tang received the B.S. from Jishou University, Hunan, China, in 2016, the M. S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019. He is currently working toward the Ph.D. degree in control science and engineering at the UESTC. His current research interests include switching-mode power supplies, soft switching techniques, and renewable energy sources.

Weihao Hu (S'06–M'13–SM'15) received the B.Eng. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, both in electrical engineering, and Ph. D. degree from Aalborg University, Denmark, in 2012.

He is currently a Full Professor and the Director of Institute of Smart Power and Energy Systems (ISPES) at the University of Electronics Science and Technology of China (UESTC). He was an Associate Professor at the Department of Energy Technology, Aalborg University, Denmark and the Vice Program Leader of Wind Power System Research Program at the same department. His research interests include artificial intelligence in modern power systems and renewable power generation. He has led/participated in more than 15 national and international research projects and he has more than 170 publications in his technical field.

He is an Associate Editor for IET Renewable Power Generation, a Guest Editor-in-Chief for Journal of Modern Power Systems and Clean Energy Special Issue on Applications of Artificial Intelligence in Modern Power Systems, a Guest Editor-in-Chief for Transactions of China Electrical Technology Special Issue on Planning and operation of multiple renewable energy complementary power generation systems, and a Guest Editor for the IEEE TRANSACTIONS ON POWER SYSTEM Special Section on Enabling very high penetration renewable energy integration into future power systems. He was serving as the Technical Program Chair (TPC) for IEEE Innovative Smart Grid Technologies (ISGT) Asia 2019 and is serving as the Conference Chair for the Asia Energy and Electrical Engineering Symposium (AEEES 2020). He is currently serving as Chair for IEEE Chengdu Section PELS Chapter and he is an IEEE Senior Member.

Jian Xiao received his B.E. degree in Electronic Information Science and Technology from Physics and Electrical Engineering College of Jishou University, Hunan, China, in 2016. He is currently pursuing an M.E. degree in Circuits and Systems at University of Electronic Science and Technology. His research interests include adaptive control, intelligent control, and distributed control.

Zhangyong Chen was born in Sichuan, China, in 1988. He received his B.S. degree in Electrical Engineering and its Automation, and his Ph.D. degree in Electrical Engineering from Southwest Jiaotong University (SWJTU), Chengdu, China, in 2010 and 2015, respectively. From September 2014 to September 2015, he was a Visiting Student in the Future Energy Electronics Center (FEEC), Virginia Tech, Blacksburg, VA, USA. Since January 2016, he was been a Lecturer in the School of Energy Science and Engineering and from Jul.2018, he was been an associate professor in the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. His current research interests include switching-mode power supplies, soft switching techniques, power factor correction converters and renewable energy sources

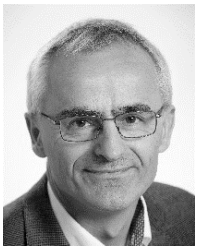
Qi Huang (S'99, M'03, SM'09) was born in Guizhou province in the People's Republic of China. He received his BS degree in Electrical Engineering from Fuzhou University in 1996, MS degree from Tsinghua University in 1999, and Ph.D. degree from Arizona State University in 2003. He is currently a professor at UESTC, the Executive Dean of School of Energy Science and Engineering, UESTC, and the

director of Sichuan State Provincial Lab of Power System Wide-area Measurement and Control. He is a member of IEEE since 1999. His current research and academic interests include power system instrumentation, power system monitoring and control, and power system high performance computing.

Zhe Chen (M'95-SM'98-F'19) received the B.Eng. and M.Sc. degrees from Northeast China Institute of Electric Power Engineering, Jilin, China, and the Ph.D. degree from University of Durham, Durham, U.K.

He is a Full Professor with the Department of Energy Technology, Aalborg University, Denmark. He is the Leader of Wind Power System Research Program in the Department of Energy Technology, Aalborg University and the Danish Principle Investigator for Wind Energy of Sino-Danish Centre for Education and Research. His research areas include power systems, power electronics and electric machines; and his main current research interests are wind energy and modern power systems. He has led many research projects and has more than 400 publications in his technical field.

Dr. Chen is an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS, a Fellow of the Institution of Engineering and Technology, London, U.K., a Chartered Engineer in the U.K. A Fellow of the IEEE.



Frede Blaabjerg (S'86–M'88–SM'97–F'03) was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. From 1988 to 1992, he got the PhD degree in Electrical Engineering at Aalborg University in 1995. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998. From 2017 he became a Villum Investigator. He is honoris causa at University Politehnica Timisoara (UPT), Romania and Tallinn Technical University (TTU) in Estonia.

His current research interests include power electronics and its applications such as in wind turbines, PV systems, reliability, harmonics and adjustable speed drives. He has published more than 600 journal papers in the fields of power electronics and its applications. He is the co-author of four monographs and editor of ten books in power electronics and its applications.

He has received 32 IEEE Prize Paper Awards, the IEEE PELS Distinguished Service Award in 2009, the EPE-PEMC Council Award in 2010, the IEEE William E. Newell Power Electronics Award 2014, the Villum Kann Rasmussen Research Award 2014, the Global Energy Prize in 2019 and the 2020 IEEE Edison Medal. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON POWER ELECTRONICS from 2006 to 2012. He has been Distinguished Lecturer for the IEEE Power Electronics Society from 2005 to 2007 and for the IEEE Industry Applications Society from 2010 to 2011 as well as 2017 to 2018. In 2019-2020 he serves as President of IEEE Power Electronics Society. He is Vice-President of the Danish Academy of Technical Sciences too. He is nominated in 2014-2019 by Thomson Reuters to be between the most 250 cited researchers in Engineering in the world.