Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/2024MA0069

Deep reinforcement learning-aided minimum current control for the DBSRC based on harmonic analysis method

Ziqiao Yu[1,2], Zhengcheng Li[1,3]

[1]Macau University of Science and Technology, Macau

[2]2009853li011008@student.must.edu.mo
[3]2009853ui011005@student.must.edu.mo

Abstract. Aiming to optimize the modulation efficiency for the dual-bridge series-resonant converter (DBSRC), this paper proposes an deep reinforcement learning (DRL) aided EPS (DEPS) modulation scheme for minimum current operation based on the harmonic analysis method. Using the deep deterministic policy gradient (DDPG) algorithm as an advanced DRL algorithm, the scheme obtains the optimized modulation scheme DEPS through offline training of the agent, which can adopt the extended-phase-shift (EPS) modulation scheme and consider the zero-voltage-switching (ZVS) constraints. Thus, the trained agent of the DDPG which likes an implicit function, can provide optimal phase shift angle for the DBSRC in real-time with the minimum current and soft switching performance in the continuous operation range. Compared with the existing EPS modulation schemes using First Harmonic Approximation (FHA), the DEPS modulation scheme has similar operational efficiency and performance of the converter, while also possessing the ability to obtain modulation angles in real-time based on environmental parameters. Finally, PSIM simulation verifies the effectiveness of the proposed optimization scheme.

Keywords: DRL, DBSRC, DDPG, Harmonic Analysis, Minimum Current.

1. Introduction

Resonant converters operate in near-resonant mode, enabling implementation of soft switching easily for power electronics. Running a resonant converter with minimal resonant tank current is of paramount importance for reducing conduction losses, enhancing overall efficiency, and reducing component stress. The Dual-Bridge Series Resonant DC/DC Converter (DBSRC) is a high-frequency, high-power-density, high-efficiency, soft-switching, and bi-directional DC/DC converter widely employed in renewable energy generation, distribution grids, and static electronic transformers [1].

Like the Dual-Active-Bridge (DAB) converter, DBSRC power flow is regulated using Single-Phase-Shift (SPS) control and Dual-Phase-Shift (DPS) control. SPS is simple, easy to implement, but it only achieves Zero Voltage Switching (ZVS) at a 1:1 voltage ratio and increases circuit current stress and reactive power circulation, reducing efficiency [2].

Traditional circuit control often uses the linear piecewise time-domain (LPTD) model for theoretical analysis, but it requires different expressions for various time intervals and operation modes, complicating the solution of optimal control variables [3]. In AI training, higher-order Fourier series can

DOI: 10.54254/2755-2721/67/2024MA0069

enhance code performance while maintaining high accuracy. To address this, the harmonic analysis method simplifies waveform expressions in the frequency domain [4].

Recent AI advancements have transformed control strategies, enabling real-time optimal control for each required environmental state. Neural networks have been proposed to reduce DAB DC converter current [5], but face issues like lengthy computation and overfitting. In [6], reinforcement learning (RL) has been used for DAB converter efficiency optimization but struggles with state and action discretization, increasing the lookup table size. In subsequent work [7], Tang further introduced a deep reinforcement learning (DRL) based current optimization scheme in the continuous domain to mitigate these issues.

Based on the above analysis, this paper proposes a DRL-based Extended-Phase-Shift (EPS) modulation scheme using harmonic analysis to optimize modulation efficiency in a DBSRC circuit. The Deep Deterministic Policy Gradient (DDPG) algorithm is used, designing the algorithm's state, action space, and reward function for offline training to obtain the DEPS modulation strategy. This strategy considers ZVS constraints, providing the optimal phase shift angle for the DBSRC, achieving minimum current and soft switching. Compared to SPS and EPS control using First Harmonic Approximation (FHA), DEPS offers better operational efficiency and similar performance to EPS. The proposed optimization scheme is validated through PSIM simulation.

2. Resonant Converter Control and Analysis

2.1. DBSRC Fundamental Characteristics

The schematic of the Dual-bridge Series Resonant Converter (DBSRC) is illustrated in **Figure 1**. The circuit comprises two nearly symmetric H-bridge circuits on the left and right side, isolated from each other through LC resonant tank and a high-frequency transformer. Each bridge includes four switches to complete circuit control. The symmetric structure of the circuit enables bidirectional exchange of energy between the voltage sources at both ends, represented by V_X and V_Y. C_X and C_Y are filter capacitors on two sides. The switches on the primary side are denoted by S_1~S_4. As a full-bridge inverter, S_1 and S_4 are switched simultaneously, while S_2 and S_3 are switched at the same time to realize the conversion of direct current source V_X to alternating current power. The same working principle is applied on the secondary side, and the switches are represented by Q_1~Q_4.

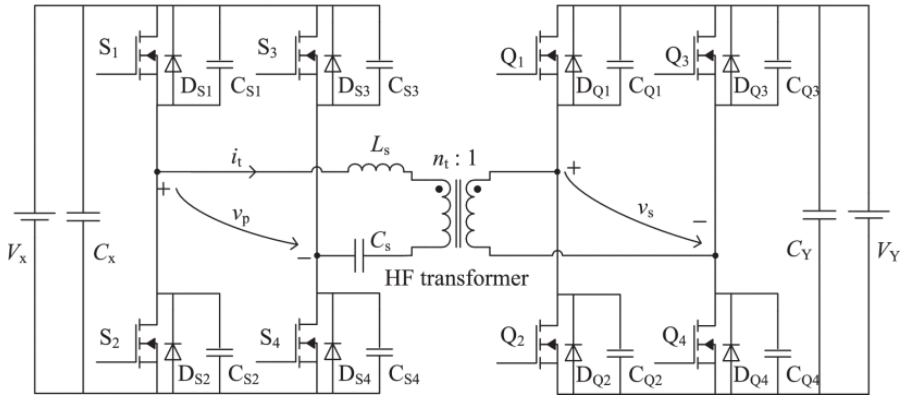

Figure 1. Circuit Schematic of the DBSRC

2.2. Harmonic Analysis of EPS Modulation

The gating scheme of the Extended-Phase Shift (EPS) can be understood by examining the steady-state waveforms depicted in **Figure 2**. δ_1 represents the phase shift angle by which the gating signal S1 leads that of S4. φ is defined as the phase-shift angle by which the turn-ON moment of S_1 and S_4 leads it of Q_1 and Q_4. V_H and V_L represent the amplitude of the primary side voltage and secondary side voltage, respectively.

DOI: 10.54254/2755-2721/67/2024MA0069

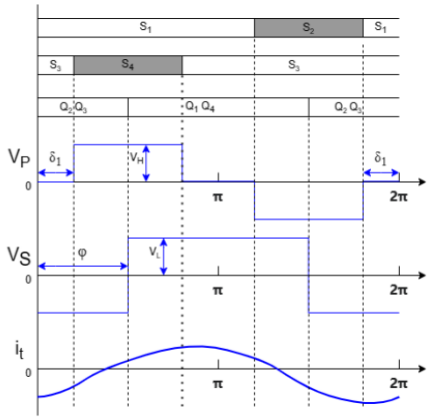

Figure 2. Typical Operating Waveforms of the DBSRC based on EPS

Through the Fourier series expansion, we can derive the expressions for the voltage of the primary side V_P and the secondary side V_S as follows:

$$V_P = \sum_{n=1,3,5\ldots}^{\infty} \frac{4V_H}{n\pi} cos(n\delta_1) sin(n\omega t) \tag{2.1}$$

$$V_S = \sum_{n=1,3,5\ldots}^{\infty} \frac{4V_L}{n\pi} sin[n(\omega t - \varphi)] \tag{2.2}$$

The impedance of the resonant tank is $j\omega_s L_s + \frac{1}{j\omega_s C_s}$, where ω_s is the switching angular frequency. By applying Ohm's law, we can derive the expression for the current in the time domain:

$$i(t) = \sum_{n=1,3,5\ldots}^{\infty} \frac{4}{n\pi(\omega L - \frac{1}{\omega C})} (V_H cos(n\delta_1) cos(n\omega t) - V_L cos[n(\omega t - \varphi)]) \tag{2.3}$$

The expressions for the two fundamental voltage phasors are as follows:

$$\overline{V_{P_1}} = \frac{4V_H}{\pi} cos\delta_1 e^{-j\frac{\pi}{2}} \tag{2.4}$$

$$\overline{V_{S_1}} = \frac{4MV_H}{\pi} e^{-j(\varphi + \frac{\pi}{2})} \tag{2.5}$$

where M is defined as the voltage gain $M = n_t V_Y / V_X$.

The current expression in the phase domain and its root mean square (rms) value can be obtained:

$$\bar{I} = \frac{4V_H}{\pi(\omega L - \frac{1}{\omega C})} (cos\delta_1 - Mcos\varphi + jMsin\varphi) \tag{2.6}$$

$$I_{rms} = \frac{2\sqrt{2}V_H}{\pi(\omega L - \frac{1}{\omega C})} \sqrt{M^2 + cos^2\delta_1 - 2Mcos\delta_1 cos\varphi} \tag{2.7}$$

The average power from either side of the equivalent circuit is given by:

$$P = U_{rms} \cdot I_{rms} \cdot cos(\theta_U - \theta_I) = \frac{8MV_H^2}{\pi^2(\omega L - \frac{1}{\omega C})} sin\varphi \tag{2.8}$$

DOI: 10.54254/2755-2721/67/2024MA0069

Zero Voltage Switching(ZVS), serves as a technique in power electronics aimed at minimizing switching losses. For ZVS to occur, the current must be negative at the turn-ON instant of S_1/S_4 , while positive at the turn-ON moments of Q_1/Q_4 and S_2/S_3. The limitations can be expressed as:

$$\begin{cases} I(\omega t = \delta_1) < 0 \\ I(\omega t = \pi - \delta_1) > 0 \\ I(\omega t = \varphi) > 0 \end{cases} \tag{2.9}$$

3. DRL-Based Optimization

3.1. Algorithm

In Reinforcement Learning (RL), an agent iteratively interacts with its environment to learn the best actions, representing the optimization of current flow as a Markov Decision Process (MDP). Typically, an MDP comprises four components (S, A, P, R), the state space S, action space A, state transition probability P, and reward function R. During each iteration k, the RL agent observes a state $s_k \in S$, selects action $a_k \in S$ based on policy $\pi(a_k|s_k)$, receives reward $r_k \sim R(s_k, a_k)$ and moves to the next state s_{k+1}.The cumulative discounted reward $G_t = \sum_{k=0}^{\infty} \gamma^k * R_{k+t}$, where γ [0~1] is the discount factor. The goal is to maximize G_t. Additionally, the action-value function $Q^{\pi}(s_t, a_t)$ is defined in DDPG algorithm as follows:

$$Q^{\pi}(s_t, a_t) = \mathrm{E}_{\pi}[G(s_t, a_t) + \gamma \mathrm{E}_{a_{t+1} \sim \pi}[Q^{\pi}(s_{t+1}, a_{t+1})]] \tag{3.1}$$

In DBSRC, environmental factors include input voltage V_i, output voltage V_o, and desired active power P_o. The state space S is $[V_i,\ V_o,\ P_o]$, and the action space A is the phase-shift angles$[D_1, D\varphi]$. The reward function R is designed to minimize current and active power:

$$R(D_1, D_{\varphi}) = -[\delta * {I_o}^2 + \xi * (P_o - P_o')^2] \tag{3.2}$$

where P_o' is the target output power, and ξ is the penalty factor. Value of ξ are set to 100 or 150 to balance performance and power deviation. When the Zero Voltage Switching (ZVS) constraint like Equation (2.9) is met, δ and ξ are set to 1 and 100, respectively; otherwise, δ and ξ are set higher to 100 and 150 to discourage suboptimal actions.

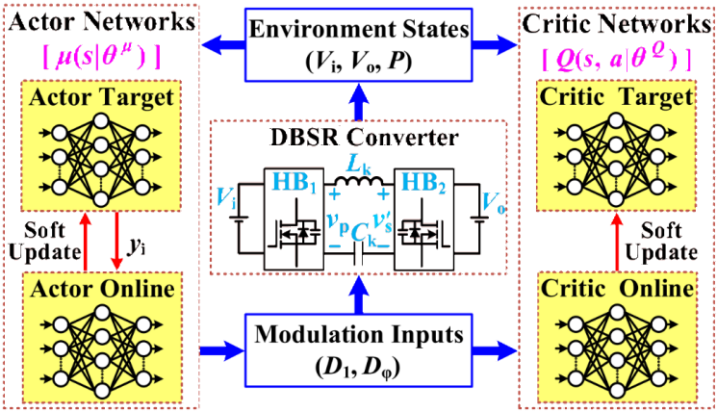

Figure 3. Structure Diagram of the Proposed DDPG Method for DBSRC

Figure 3 utilizes an actor-critic framework for the DDPG algorithm.The critic network's weight θ^Q is updated by minimizing the loss function $\mathcal{L}(\theta^Q) = E_{(s,a)}[(Q(s_t, a_t|\theta^Q) - y_t)^2]$, where $y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{t+1})|\theta^Q)$. This helps the critic network estimate the long-term value of actions.

The actor network's weight θ^u is updated using the policy gradient method, to maximize expected return, as shown as following formular:

$$\nabla_{\theta^u} J^{\theta^{\mu}} \approx \mathbb{E}_{s_t \sim \rho^{\beta}} \left[\nabla_{\theta^u} Q(s, a|\theta^Q)|_{a=\mu(S|\theta^u)} \nabla_{\theta^u} \mu(s|\theta^{\mu}) \right] \tag{3.3}$$

DOI: 10.54254/2755-2721/67/2024MA0069

$$= \mathbb{E}_{s_t\sim\rho^\beta}\left[\nabla_a Q(s,a|\theta^Q)|_{a=\mu^\theta(s)}\nabla_{\theta^u}\mu(s|\theta^\mu)\right]$$

where ρ is the discounted distribution and β denotes the specific policy of the current policy π.

The DDPG algorithm incorporates target networks - actor target network $\mu'\ (s|\theta^{u'})$ and critic target network $Q'\ (s,a|\theta^{Q'})$ - which track the main networks' parameters to provide stable targets for training. The soft update formulas are:

$$\text{Softupdate}\begin{cases}\theta^{Q\prime} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q\prime} \\ \theta^{u\prime} \leftarrow \tau\theta^u + (1-\tau)\theta^{u\prime}\end{cases} \tag{3.4}$$

where $\tau \ll 1$ is the soft update coefficient.

The training process of the DDPG algorithm is outlined in detail in **Table 1**.During training, the algorithm uses environment's input parameters $[V_i, V_o, P_o]$, randomly sampled within specified ranges to generate training data. The DDPG algorithm handles continuous action spaces through deep neural networks. Post-training, the agent can be deployed on a digital signal processor (DSP) to map new environment parameters $[V_i, V_o, P_o]$ to optimal phase shift angles (D_1, D_φ) in real-time, eliminating the need for a lookup table and saving memory and computation time.

The DDPG algorithm's continuous action strategy ensures optimal control under various conditions, making it suitable for online real-time control. Even if environment parameters fall outside the training range, the agent can adapt, ensuring robust performance in unforeseen scenarios.

Table 1. Training Process of the DDPG Algorithm

Algorithm: Offline Learning with the DDPG algorithm

Input: Environments $[V_i, V_o, P_o]$
Output: Phase shift angles (D_1, D_φ)

1: Randomly initialize the weights of actor network θ^μ and critic network θ^Q
2: Initialize target network μ' and Q' by: $\theta^{\mu'} \leftarrow \theta^\mu$, $\theta^{Q'} \leftarrow \theta^Q$

3: **for** episode = 1 to max-episode **do**
4: Observe current state s_1
5: **for** actor = 1,2,...10 **do**
6: Select action $a_t = \mu(s_t|\theta^\mu) + n_t$ according to the deterministic policy μ and exploration rate
7: Execute action a_t, observe the reward r_t and observe new state s_{t+1}
8: Store transition (s_t, a_t, r_t, s_{t+1}) in replay buffer R
9: Sample a random mini-batch of transitions (s_i, a_i, r_i, s_{i+1}) from replay buffer R
10: $y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{t+1})|\theta^Q)$
11: Update critic net by using the loss function $\mathscr{L}(\theta^Q)$:

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

12: Update actor net by using the policy gradient:

$$\nabla_{\theta^u} J^{\theta^\mu} \approx \mathbb{E}_{s_t\sim\rho^\beta}\left[\nabla_{\theta^u} Q(s,a|\theta^Q)|_{a=\mu(s|\theta^u)}\nabla_{\theta^u}\mu(s|\theta^\mu)\right]$$
$$= \mathbb{E}_{s_t\sim\rho^\beta}\left[\nabla_a Q(s,a|\theta^Q)|_{a=\mu^\theta(s)}\nabla_{\theta^u}\mu(s|\theta^\mu)\right]$$

13: Update target net by using the policy gradient:

$$\nabla_{\theta^u} J^{\theta^\mu} \approx \mathbb{E}_{s_t\sim\rho^\beta}\left[\nabla_{\theta^u} Q(s,a|\theta^Q)|_{a=\mu(s|\theta^u)}\nabla_{\theta^u}\mu(s|\theta^\mu)\right]$$
$$= \mathbb{E}_{s_t\sim\rho^\beta}\left[\nabla_a Q(s,a|\theta^Q)|_{a=\mu^\theta(s)}\nabla_{\theta^u}\mu(s|\theta^\mu)\right]$$

14: end for
15: end for

DOI: 10.54254/2755-2721/67/2024MA0069

4. Experimental Analysis

4.1. Training Parameter

To verify the feasibility and correctness of the proposed DRL-based EPS (DEPS) modulation scheme for minimum current operation, a 200W DBSRC simulation model is established in PSIM. The system specifications are summarized in **Table 2**. The specific structure of the DBSRC is illustrated in **Figure 1**.

Table 2. System Specification of the DBSRC in Simulation

Parameter	value
Input voltage V_i (V)	64~96
Output voltage V_o (V)	88~104
Switching frequency f_s (kHz)	100
Inductor L_k (μH)	41.18
Capacitor C_k (nF)	120.57
Turns ratio of the transformer (N: 1)	0.585:1
MOSFET in primary side Rds(on) (Ω)	0.06
MOSFET in secondary side Rds(on) (Ω)	0.036

The training parameters adopted in the DDGPG algorithm are listed in **Table 3**. Two hidden layers with 100 and 100 neurons are designed for both the critic and actor networks, respectively. The whole training process can be accomplished in approximately 30 minutes.

Table 3. Training Parameters Adopted in DDPG Algorithm

Parameter	value
Learning rate of critic network (λ_c)	0.002
Learning rate of actor network (λ_a)	0.001
Soft update coefficient (τ)	0.001
Replay buffer size (M)	100000
Max Episode (N)	100000
Step size of each episode	10
Mini-batch size (m)	32

4.2. Performance Evaluation

To determine ZVS, we check for reverse conduction in the MOSFET switch current, indicated by a small negative spike in the waveform. In **Figure 4**, we take the MOSFET switch current under DEPS for example, The first row at $M = 0.95$, (a) to (d), depicts power decreasing from 200W to 50W. The second row at $M = 0.54$, (e) to (h), shows the same power levels. These figures confirm that the sets of phase angles DEPS achieve ZVS while meeting the rated power. In SPS and EPS, the current waveform diagram is similar to this and will not be listed.

DOI: 10.54254/2755-2721/67/2024MA0069

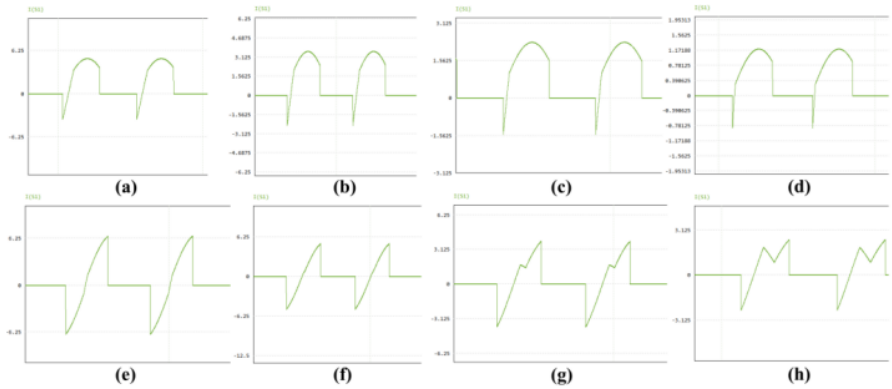

Figure 4. MOSFET Switch Current of DBSRC under DEPS Modulation

Figure 5 presents the relationship curve between the rms value of the DBSRC inductor current theoretically and the power transmission under different voltage conversion ratios. It can be observed that when $M = 0.95$, for the entire power range, the three modulation methods, SPS, EPS, and DEPS, have nearly identical effective values of inductor current. This is because under this voltage ratio, DEPS selects a modulation angle similar to SPS, resulting in almost identical performance.

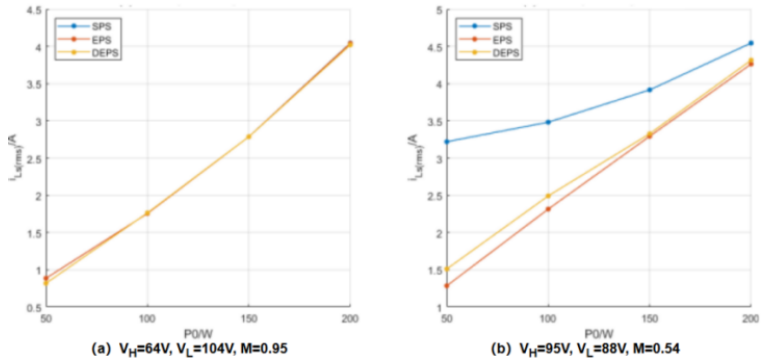

Figure 5. RMS Current of Inductor with Transmission Power

When $M = 0.54$, the situation changes. Since SPS can only modulate one angle, D_{φ}, it can only achieve the desired ZVS and rated power effect by using higher current. Due to the non-pure square waveforms of EPS and DEPS, they can adjust the duty cycle through D_1 to achieve smaller currents while maintaining the rated power. When the rated power is 50W, the current under DEPS is slightly higher than that under EPS. This difference is caused by two factors: the training accuracy of the program did not reach perfection, resulting in DEPS not precisely stopping at the optimal point when searching for the same optimal point as EPS. Another factor is some errors were introduced when manually adjusting the angle.

DOI: 10.54254/2755-2721/67/2024MA0069

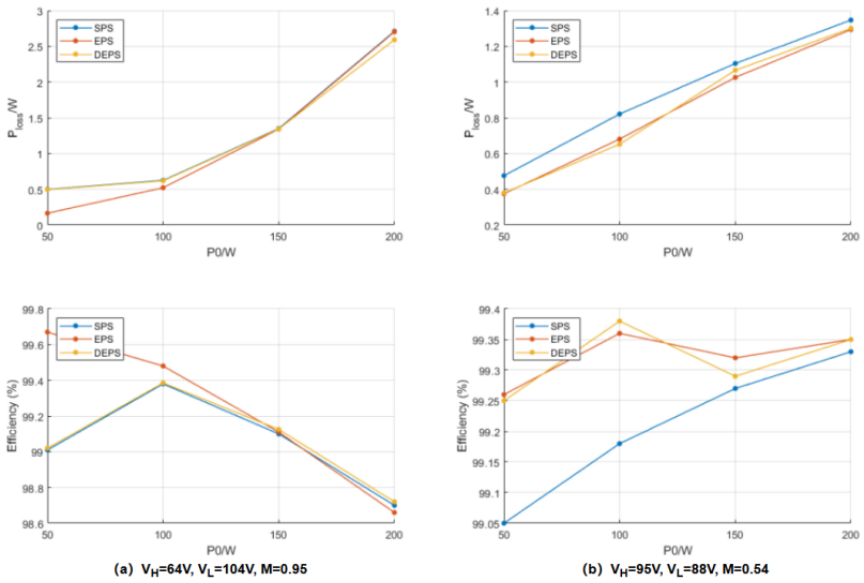

Figure 6. Loss and Efficiency of DBSRC with Transmission Power

The losses and efficiency characteristics of the DBSRC under two different voltage conversion ratios are shown in **Figure 6**. It can be seen that, under the current conditions, compared to the traditional SPS, the DEPS modulation strategy exhibits better efficiency characteristics throughout the entire power range. It is worth noting that in both voltage ratio scenarios, the efficiency is very high, reaching around 99%. This is because in the simulation, for convenience, we only set the conduction resistance of the MOSFETs, used ideal transformers, and did not consider heat dissipation.

Although DEPS and EPS yield similar performance without significant improvement, our DRL model can automatically adjust optimal phase angles in real-time based on environmental changes, achieving minimal current control seamlessly. In contrast, the EPS method requires recalculation for each environmental change, which is cumbersome. For future applications, connecting the model to an actual DBSRC to read real-time environmental data and fine-tune it based on actual conditions could enhance performance compared to the ideal EPS harmonic approximation.

5. Conclusion

This paper proposes a minimum current control scheme for DBSRC based on harmonic analysis and deep reinforcement learning, which can improve the efficiency characteristics of the converter in real-time according to environmental parameter changes. The scheme uses the DDPG algorithm to find the optimal modulation strategy for DBSRC. The paper focuses on describing how the DDPG algorithm is used to select the optimal control variables for DAHB and provides a detailed derivation of the SPS and EPS modulation formulas using Harmonic Analysis. Additionally, PSIM simulations are employed to compare the performance of the SPS, EPS, and DEPS modulation strategies under different operating conditions. The results indicate that under the DEPS modulation strategy, the power losses of DBSRC are smaller than the optimal value obtained under SPS across the entire power range. In most cases, it is similar to the EPS modulation strategy, but as a model, it can achieve real-time modulation, unlike EPS, which requires program calculation for each environmental parameter to obtain the modulation angle. This validates the effectiveness and correctness of the proposed efficiency optimization scheme.

One significant issue with the current program is the problem of accuracy, meaning that the model does not converge to a specific optimal point mathematically after training, but fluctuates around it. In the future, we can address this issue by increasing the number of layers and neurons in the neural network. Additionally, the current training using generated data has the issue of deviating from reality, as real-world circuits cannot be as ideal as in the simulation. Therefore, in the future, we can first run a set of

DOI: 10.54254/2755-2721/67/2024MA0069

theoretically optimal EPS control strategies, then read the port data, and use deep reinforcement learning to train with actual data. This approach can achieve better performance than the theoretical optimum.

References

[1] Li, Xiaodong, and Ashoka KS Bhat. "Analysis and design of high-frequency isolated dual-bridge series resonant DC/DC converter." IEEE Transactions on Power Electronics 25.4 (2009): 850-862.

[2] Lin Xuefeng, Xu Jianping, Zhou Xiang. Soft-switched high step-up DC-DC converter with coupled inductor of resonance[J]. Transactions of China Electrotechnical Society, 2019, 34(4): 747-755.

[3] 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 Trans. Power Electron., vol. 33, no. 6, pp. 5302–5316, Jun. 2018.

[4] B. Zhao, Q. Song, W. Liu, G. Liu, and Y. Zhao, "Universal highfrequency-link characterization and practical fundamental-optimal strategy for dual-active-bridge DC-DC converter under PWM plus phase-shift control," IEEE Trans. Power Electron., vol. 30, no. 12, pp. 6488–6494, Dec. 2015.

[5] S. Zhao, F. Blaabjerg, and H. Wang, "An overview of artificial intelligence applications for power electronics," IEEE Trans. Power Electron., vol. 36, no. 4, pp. 4633–4658, Apr. 2021.

[6] Tang, Yuanhong, et al. "Reinforcement learning based efficiency optimization scheme for the DAB DC–DC converter with triple-phase-shift modulation." IEEE Transactions on Industrial Electronics 68.8 (2020): 7350-7361.

[7] Tang, Yuanhong, et al. "Artificial intelligence-aided minimum reactive power control for the DAB converter based on harmonic analysis method." IEEE Transactions on Power Electronics 36.9 (2021): 9704-9710.