Automatic Optimization of Electro-Optic Frequency Comb Based on Deep Reinforcement Learning

Zixuan Li
The State Key Laboratory of Information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
email: zixuanli@bupt.edu.cn

Shifan Chen
The State Key Laboratory of Information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
email: chensf2001@163.com

Yunping Bai*
The State Key Laboratory of Information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
*email: baiyunping@bupt.edu.cn

Yue Zhou
The State Key Laboratory of Information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
email: yuezhou@bupt.edu.cn

Xingyuan Xu
The State Key Laboratory of Information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
email: xingyuanxu@bupt.edu.cn

Kun Xu
The State Key Laboratory of Information Photonics and Optical Communications
Beijing University of Posts and Telecommunications
Beijing, China
email: xukun@bupt.edu.cn

Abstract—We propose a novel automatically optimized electro-optic frequency comb (AO-EOFC) based on deep reinforcement learning, which utilizes a deep Q-learning network algorithm to replace blind and inefficient manual optimization in traditional ways.

Keywords—Electro-Optic frequency comb, Deep Q-Learning Network, Electro-Optic modulator, Microwave photonics, Optical communication equipment.

I. Introduction

Due to the advantages of adjustable repetition rate, high-power sideband and reconfigurable spectrum, electro-optic frequency combs (EOFCs) play a fundamental role in optical communications, frequency measurement, atomic clock, distance measurement, arbitrary waveform generation, photonic neuromorphic computing and so on [1-3]. The most significant measures to evaluate the performances of an EOFC involve the number of comb lines, spectral bandwidth and flatness [4]. One of the crucial factors determining these performances of an EOFC is the phase relationships of the input RF signals among multiple electro-optic modulators (EOM) consisting of the system. In theory, the optimal performances of an EOFC can be achieved as phases of all input RF signals keep identical [5], which can be considered phase synchronization.

However, due to the inherent physical manufacturing differences and non-ideal performances within microwave transmission components, transmission channels of multiple input RF signals inevitably have a specific delay difference. In an EOFC, it is hardly possible to directly realize phase synchronization in a practical hardware system [6-8]. The common existing solution is to add phase shifters in front of each input RF signal and carry out phase synchronization through manual blind tuning. This way is inefficient and time-consuming in the scenario of the EOFC based on multiple cascaded modulators. Besides, phase synchronization needs to be repeatedly performed when the frequency of input RF signals is adjusted because the phase differences among multiple RF signals will vary from one working frequency to another working frequency at the same transmission channel delay differences. Compared to other classes of optic frequency combs, the prominent advantage of the EOFC is its adjustable repetition rate. The problem induced by the delay differences significantly weakens this advantage of the EOFC system. As such, in the EOFC system, it is of great significance to construct a phase synchronization method with fast tuning, high efficiency, and multivariable extension at arbitrary working frequency.

Deep reinforcement learning (DRL), capable of making decisions by interacting with the dynamic environment to achieve a specific goal, has been successfully applied to many scenarios where multiple complex and uncertain variable factors are included, such as the adaptive optimization method for optical communications [9] and power excursions compensation of erbium-doped fibre amplifiers (EDFAs) in wavelength-division multiplexing systems [10]. The features of DRL are naturally suitable for the adaptive optimization of the EOFC system. Therefore, compared to the approach to manually tuning the phase shifter, utilizing a reinforcement learning algorithm to manipulate phase relationships of the input RF signals is undoubtedly an advantageous solution to phase synchronization of the EOFC system, which enables phase matching at any working frequency in more accurately and elegantly manner.

In this paper, we innovatively employ DRL in tuning and optimizing the phase relationships of the input RF signals of the EOFC assisted by multiple cascaded EOMs. Specifically, we construct an automatically optimized electro-optic frequency comb (AO-EOFC) using a deep Q-learning network (DQN) algorithm to replace the manual tuning approach. The proposed DQN-based method includes a training process and a testing process. During the training process, the statistical feature of the spectral amplitude of the generated EOFC is regarded as the input state of the DQN, and the output action of the network is the specific tuning value of the phase shifter inside the EOFC system. The comb lines and spectral bandwidth of the EOFC will accordingly vary in the adjusting and optimizing process, and the corresponding reward will be recorded. The parameters of the DQN network with reduced loss function are constantly updated employing

2023 Asia Communications and Photonics Conference/2023 International Photonics and Optoelectronics Meetings (ACP/POEM) | 979-8-3503-1261-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACP/POEM59049.2023.10369497

empirical playback [11]. Stop until the number of internal frequency comb lines within a specific power variation reaches the theoretical optimal value. In the test process, the trained offline network can be directly used in the EOFC hardware system, where the original input phase states of the input RF signals are determined by a certain range of random numbers. Through the network, the phase values of the input RF signals will be continuously and automatically adjusted so that the performances of the finally generated EOFC are close to the ideal and optimal case as quickly as possible. Besides, the trained offline network can work at different working frequencies, resulting from the original phase states of the input signals being initially random.

II. Principle

The schematic diagram of the adaptive optimization scheme for the cascaded electro-optical modulation optical frequency comb is shown in Fig. 1. The laser generates the input continuous optical carrier, and then the optical carrier passes through a Mach-Zehnder modulator (MZM) and three phase modulators. The four EOMs are driven by the same RF signal through a power divider that divides the input signal into four ports, and three phase shifters are connected between the power divider and the phase modulator to adjust the phase of the RF signal in order to achieve a cascaded electro-optical modulation system.

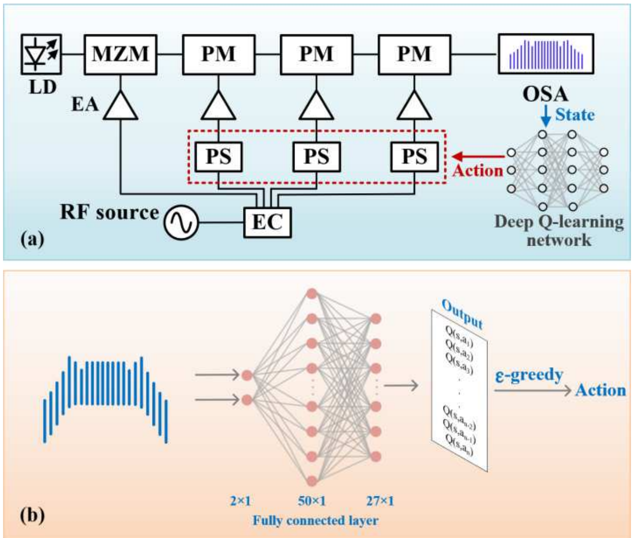

Fig. 1. (a). Optical frequency comb generated by cascaded electro-optic modulation based on DQN algorithm. (b). The optical frequency comb selects the action that the phase shifter should perform through DQN to achieve adaptive optimization.

The used MZM has a push-pull structure, and its DC bias voltage is set at the quadrature bias point. In this scenario, the output optical field is as follows:

$$E_{out}(t) \propto E_{in}(t) \cdot [e^{i \cdot (V_{bias} + \beta_{mzm} \cdot \cos(\omega_{rf} t))} + e^{-i \cdot \beta_{mzm} \cdot \cos(\omega_{rf} t)}] \cdot e^{-i \cdot \beta_{pm}(\cos(\omega_{rf} t + \varphi_1))} \cdot e^{-i \cdot \beta_{pm}(\cos(\omega_{rf} t + \varphi_2))} \cdot e^{-i \cdot \beta_{pm}(\cos(\omega_{rf} t + \varphi_3))} \tag{1}$$

Where β_{mzm} and β_{pm} is the modulation index of MZM and PM, respectively. V_{bias} denotes the DC bias voltage of MZM, ω_{rf} represents the RF voltage, and φ_{i} describes the phase of the input RF signal loaded on the i-th PM.

The key step to implementing automatic adaptive EOFC is to choose the input state of DQN, resulting from an efficient input state capable of significantly affecting the performance of the DQN. In the proposed system, feature extraction is regarded as the method to fulfill the design of the input state of the network. First, we transform the output optical signal into the frequency domain by Fourier transform. Next, we extract the absolute value of the spectrum and calculate the mean and variance of the comb line amplitudes according to Eq. 2 and Eq. 3.

$$\left|E(\omega)\right|_{\_avg} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|E(\omega)\right|d\omega \tag{2}$$

$$\left|E(\omega)\right|_{\_\text{var}} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|E(\omega)-\left|E(\omega)\right|_{\_avg}\right|^{2}d\omega \tag{3}$$

In the proposed framework, we choose the feature values as the input state of the entire network. Compared to other manners of describing the input state, the feature extraction can effectively extract more concise and representative feature representations from the original frequency domain information. These feature representations can better describe the characteristics of the optical frequency comb under current phase relationships, which can significantly improve the training efficiency and performance of the DQN. The outputs of the DQN are the Q-values for the required action space. Besides, this proposed network merges all phase-shifting actions of phase shifters into the same action space, which contribute to avoiding the computational complexity and cost brought by multiple agents.

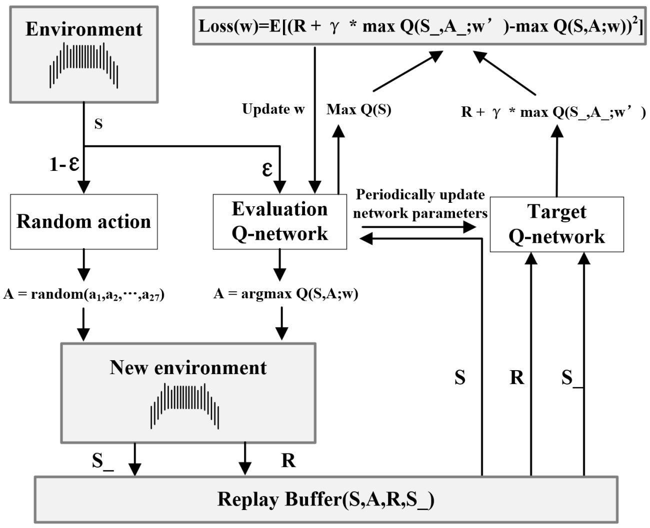

Fig. 2. The framework of Deep Q-learning Network. S is the current state; A represents the chosen action; R donates the reward; S_ is the next state after choosing the action of A; γ is the discount factor.

As is shown in Fig. 2, to overcome bootstrap, the training process involves updating two networks. The parameters of the evaluation network are updated with each iteration. After a certain period, these parameters from the whole evaluation network will be copied to the target network. The above two steps will be periodically performed. Once the state values are determined, a ε - greedy strategy is utilized to determine the next action to execute, leading to a new state. If the number of comb lines in the new state increases compared to the previous state, a certain reward R is assigned. [S, A, S_, R] is then stored as a set of experiences in the Replay Buffer.

When the Replay Buffer reaches its set capacity limit, a batch of data will be sampled from it for training updates of the model parameters. If not, the collection of training data [S, A, S_, R] will be continuously performed until the capacity of the Replay Buffer is reached. After extracting a batch of data from the Replay Buffer, the target network will work out the maximum Q-value for the next state S_. At the same time, the maximum Q-value and R of S can be obtained by the evaluation network. Next, on the basis of the previous two steps, the loss function $L(\omega)$, with the network parameter ω as the independent variable, can be calculated. Afterwards, the loss function will be minimized by updating the parameters ω' of the target network. After a certain number of steps, the parameter ω' of the target network is synchronized to the parameter w of the evaluation network, resulting in updating the network parameters.

III. Simulation Results

In order to demonstrate the capacity of the proposed approach for improving comb lines and spectral bandwidth, we perform a simulated test according to the principle of AO-EOFC. In the simulation, the output power of the laser is set to 10 mW, and the MZM and PMs are all driven by the same RF signal with a frequency of 20 GHz. The DC voltage of MZM is biased at the quadrature bias point. The modulation depths of PMs are all set to 10. The output optical spectrum of the EOFC is sent into the DQN, and the simulation is divided into a training process and a testing process.

Both the evaluation network and the target network adopt a three-layer structure, including 2, 50, and 27 neurons in each layer, respectively. The ReLU function is employed as the activation function. The capacity of the Replay Buffer is set to 1000. The target network synchronizes its parameters with the evaluation network after performing every 100 iterations. The value ε is set to 0.9. The learning rate of the network is set as 0.01, and the entire agent is trained for 500 episodes.

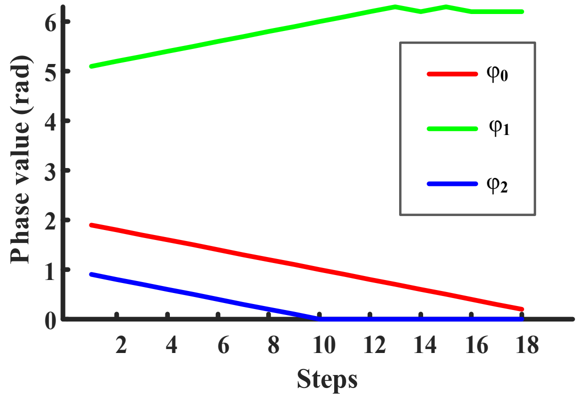

Fig. 3. In the process of optimizing phase relationships, the phase responses of the three input RF signals at different steps.

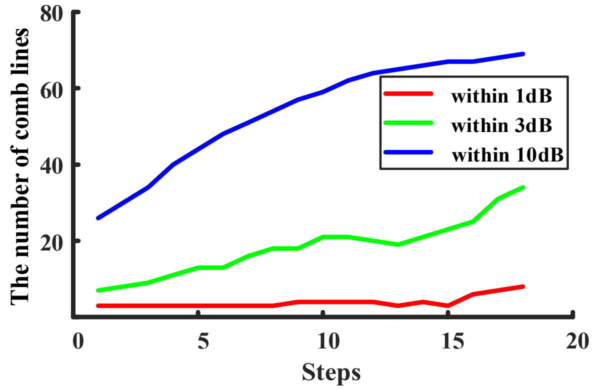

Fig. 4. The numbers of comb lines within 1-dB, 3-dB and 10-dB flatness at different steps.

After performing training, we obtain an offline network, which can be directly applied to the phase adaptive tuning of the EOFC system. The measured simulated results shown in Fig. 3, Fig. 4 and Fig. 5 show the automatic optimizing process from the random initial phase values of input RF signals to the optimal performances by utilizing this trained offline network. Specifically, Fig. 3 displays the change in the phase values of input RF signals with the increase in the number of steps. It

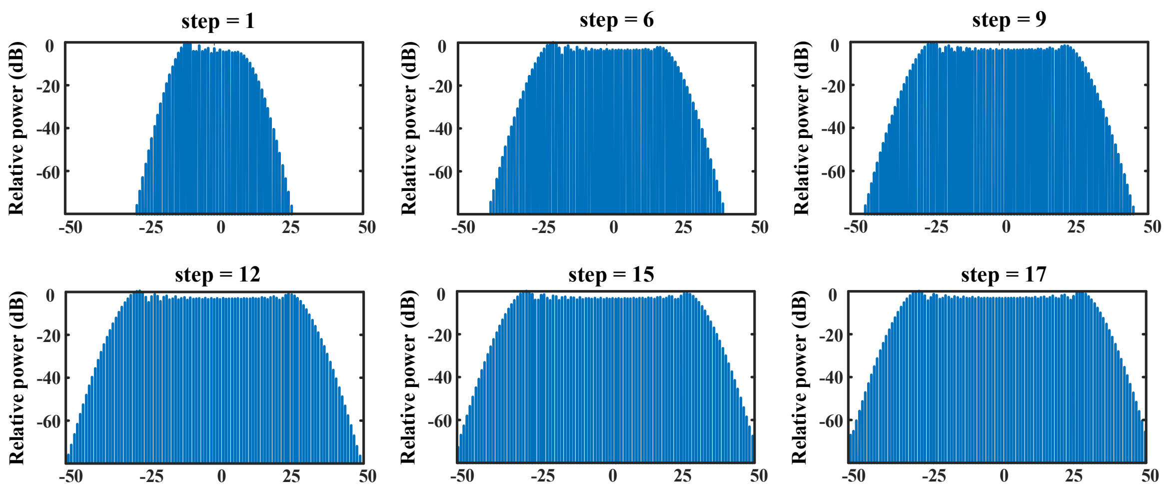

Fig. 5. The simulated comb lines and spectral bandwidth of EOFC in the process of optimizing phase relationship when the DQN is performed.

can be observed that they all converge to either 0 or 2π eventually. Fig. 4 shows the numbers of optical frequency comb lines within 1dB, 3dB, and 10dB, respectively, as the number of steps increases. The measured results indicate that the comb lines of the EOFC are gradually enhanced in the process of running DQN.

As far as the spectral bandwidth in Fig. 5, under the condition of given random original phases, there are 3, 7, and 26 comb lines within the three flatness ranges, respectively. By contrast, at the desired and terminating states, the number of comb lines within 1dB, 3dB, and 10dB flatness increases to 8, 34, and 69, respectively. The simulated results in Fig. 5 can effectively verify the feasibility of the proposed AO-EOFC.

Lastly, we also measure the consumed optimization times at different working frequencies. The measured results shown in Fig. 6 indicate that the trained offline network can always realize the adaptive optimization of phase relationships within a certain time as the frequency of the input RF signal vary from 5 GHz to 40 GHz, thus achieving the optimal performances of the EOFC. As such, a phase synchronization method featuring of fast tuning, high efficiency, and multivariable extension at arbitrary working frequency is successfully constructed for the adaptive optimization of the EOFC.

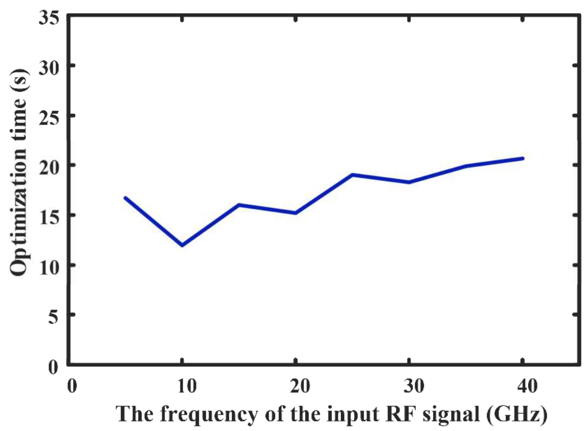

Fig. 6. In the testing process, the optimization times are required at different working frequencies.

IV. CONCLUSION

In this work, a novel AO-EOFC is proposed to replace the traditional manual way. In the proposed system, a deep reinforcement learning algorithm based on the DQN model is employed in optimizing the phase synchronization in the multiple-EOM cascaded system. The algorithm extracts the features from the generated optical frequency comb as states and automatically selects the effective action of the phase shifter to optimize the phase value, thus realizing the phase synchronization of the input RF signals, and finally generating flat and wideband optical frequency combs that are almost close to the theoretical optimal values. Simulation results show that the used DQN can successfully fulfil adaptive optimization of the system. Multiple RF signals with random initial phases will be automatically tuned to implement synchronization after performing the proposed approach, and wider and flatter optical frequency combs are finally generated. The proposed scheme has the potential to be applied in high-performance EOFC systems.

REFERENCES

[1] R. Zhuang, K. Ni, G. Wu, T. Hao, L. Lu, Y. Li, et al. “Electro-Optic Frequency Combs: Theory, Characteristics, and Applications,” Laser&Photonics, 2200353, Rev. 2023.

[2] Y. Sun, J. Wu, M. Tan, X. Xu, Y. Li, R. Morandotti, et al. “Applications of optical microcombs,” Advances in Optics and Photonics, vol. 15, no. 1, March 2023.

[3] Y. Bai, X. Xu, M. Tan, Y. Sun, Y. Li, J. Wu, et al. “Photonic multiplexing techniques for neuromorphic computing,” Nanophotonics, vol. 12, pp. 795-817, March 2023.

[4] V. Torres-Company, J Lancis, P. Andres, “Lossless equalization of frequency combs”, Opt. letters, vol. 33, pp. 1822-1824, August 2008.

[5] M. Song, S. Han, J. Park, H. Choi, S. Kim, T. Tran, et, al. “Flat-top supercontinuum generation via Gaussian pulse shaping”, Opt. Express, vol. 29, pp. 12001-12009, April 2021.

[6] R. Wu, V. Torres-Company, D. E. Leaird, A. M. Weiner, “Supercontinuum-based 10-GHz flat-topped optical frequency comb generation,” Opt. Express, vol. 21, pp. 6045-6052, March 2013.

[7] H. Kim, A. J. Metcalf, O. E. Sandoval, D. E. Leaird, and A. M. Weiner, “Broadband and ultra-flat optical comb generation using an EO comb source and a programmable pulse shaper,” in CLEO: 2014, San Jose, CA, paper JW2A.72, Jun. 2014.

[8] M. Song, S.-P. Han, S. Kim, and M. Song, “Flat-top Supercontinuum Generation Based on Electro-optic Optical Frequency Comb,” in Conference on Lasers and Electro-Optics, paper JTu2A. 11, May 2019.

[9] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: A brief survey,” IEEE Signal Process. Magazine, vol. 34, no. 6, pp. 26-38, Nov. 2017.

[10] X. Liu, J. Wu and S. Chen, “Efficient Hyperparameters optimization Through Model-based Reinforcement Learning and Meta-Learning,” IEEE 22nd International Conference on High Performance Computing and Communications; Yanuca Island, Cuvu, Fiji, 2020, pp. 1036-1041, 2020.

[11] M. Volodymyr, K. Koray, S. David,G Alex, A. Ioannis, W. Daan, et, al. “Playing Atari with Deep Reinforcement Learning”, Computer Science, arXiv.1312.5602, Dec. 2013.