photonics

MDPI

Communication

A Deep Reinforcement Learning Algorithm for Smart Control of Hysteresis Phenomena in a Mode-Locked Fiber Laser

Alexey Kokhanovskiy [1,*], Alexey Shevelev [1], Kirill Serebrennikov [1], Evgeny Kuprikov [1] and Sergey Turitsyn [2]

[1] Laboratory of Nonlinear Photonics, Novosibirsk State University, Pirogova str., 2, 630090 Novosibirsk, Russia
[2] Aston Institute of Photonic Technologies, Aston University, Birmingham B4 7ET, UK
* Correspondence: kay@nsu.ru

Abstract: We experimentally demonstrate the application of a double deep Q-learning network algorithm (DDQN) for design of a self-starting fiber mode-locked laser. In contrast to the static optimization of a system design, the DDQN reinforcement algorithm is capable of learning the strategy of dynamic adjustment of the cavity parameters. Here, we apply the DDQN algorithm for stable soliton generation in a fiber laser cavity exploiting a nonlinear polarization evolution mechanism. The algorithm learns the hysteresis phenomena that manifest themselves as different pumping-power thresholds for mode-locked regimes for diverse trajectories of adjusting optical pumping.

Keywords: fiber mode-locked lasers; reinforcement learning; hysteresis phenomena

Citation: Kokhanovskiy, A.; Shevelev, A.; Serebrennikov, K.; Kuprikov, E.; Turitsyn, S. A Deep Reinforcement Learning Algorithm for Smart Control of Hysteresis Phenomena in a Mode-Locked Fiber Laser. *Photonics* **2022**, *9*, 921. https://doi.org/10.3390/photonics9120921

Received: 14 October 2022
Accepted: 22 November 2022
Published: 30 November 2022

1. Introduction

Machine learning (ML) algorithms have already shown their potential for improving performance of ultrafast photonic systems [1]. In particular, applications of ML in mode-locked fiber lasers target three main areas: self-starting [2], system optimization [3] and characterization [4]. Most of the conventional ML methods optimize static topology of the fitness function and do not take into account the trajectory of adjusting controlling parameters [5–7]. However, fiber mode-locked lasers are complex nonlinear systems that are sensitive to initial conditions and often possess hysteresis phenomena [8,9] that require special approaches to smart control of such devices. Hysteresis and bistability effects may lead to incorrect operation of the optimization algorithms—e.g., genetic or particle swarm optimization algorithms that do not consider dynamic changes in laser cavity parameters. For given control parameters and different tuning strategies, the algorithm will meet different output regimes with corresponding different values of the optimized fitness function. Reinforcement learning (RL) algorithms are attractive candidates for dynamic optimization [10]. In the context of mode-locked lasers, there are already promising applications of deep RL to control the output radiation. Kutz et al. [11] demonstrated using the numerical model of fiber cavity that a deep Q-learning algorithm is capable of learning how to operate with bi-stability to achieve stable mode-locking regimes. In [12], the possibility of stabilizing the mode-locking operation of an NPE laser under temperature and vibratory disturbances by an actor–critic RL algorithm was demonstrated. In our previous work, we have showed that the DDQN algorithm may be efficiently applied to a changeable environment, such as an 8-figure mode-locked laser with a tunable spectral filter [13].

In this Letter we experimentally validate RL's ability to find nontrivial optimization trajectories when adjusting pumping power of the laser cavity for soliton generation with targeted properties.

2. Experimental Setup

We consider here a fiber mode-locked laser based on the nonlinear polarization effect as an experimental platform; see Figure 1. The laser comprises an active Erbium fiber, an

electronically driven polarization controller (EPC), an output coupler, a piece of passive SMF-28 fiber and an optical isolator with blocked fast axis. Active fiber was pumped through the WDM coupler by a single-mode laser diode operating at 978 nm. The fiber laser cavity was placed on a thermal plate inside a thermostat box to eliminate the influence of the ambient temperature fluctuations. The temperature of the plate was controlled by a PID regulator and was set to 32 °C. The goal for the RL algorithm was to learn the behavior of the laser system and to obtain a stable mode-locked regime at a low level of pumping power. We deliberately chose such a relatively simple setup to be able to comprehensively verify the correctness of the algorithm and to demonstrate the feasibility of the method.

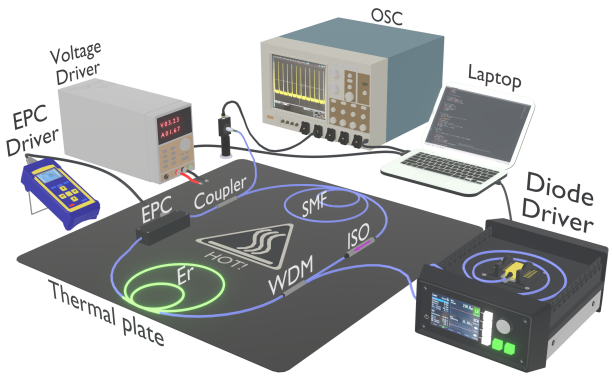

Figure 1. Experimental setup of the fiber mode-locked laser and measuring system.

Anomalous dispersion of the laser cavity provides the conditions for generation of classical optical solitons. A stable mode-locked regime was achieved by proper adjustment of the EPC controlling voltage and reaching the threshold value of the pumping power. The repetition rate of a pulse train was 14.51 MHz with an energy of 1.2 pJ for the pulses. Figure 2a depicts an optical spectrum of the generated solitons with the characteristic Kelly side-bands. A further increase in the pumping power led to multi-soliton generation. For instance, single-soliton generation was obtained at 290 mA (Figure 2b); then, after reaching 310 mA, the output switched to double-soliton regime.

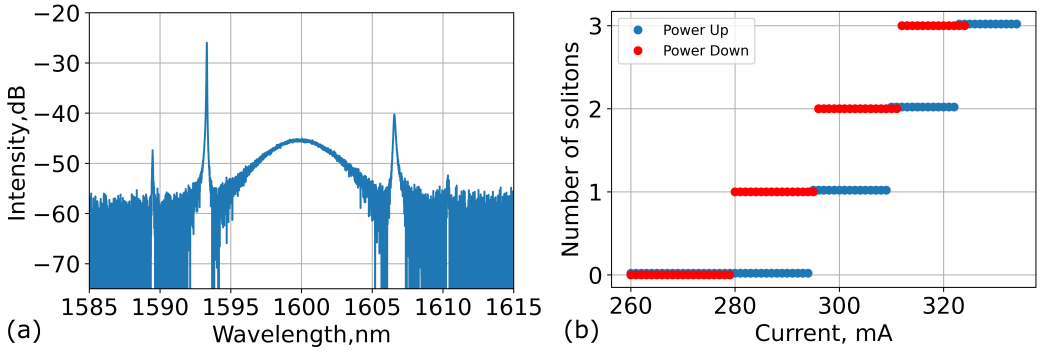

Figure 2. (**a**) Optical spectrum of the generated optical solitons. (**b**) Number of generated solitons as a function of the pumping laser diode current for different direction of the current adjustment.

Multi-soliton generation (also known as soliton quantization) features a hysteresis effect: the number of solitons is different for different directions of adjusting the pumping power [14–16]. Starting from the high level of pumping power, stable soliton generation remained until 280 mW, but it did not exist in case of starting from the low level of the pumping power (Figure 2b). Initiation of a mode-locked lasing regime requires more pumping power than supporting stable soliton generation when it already exists. The goal for the reinforcement learning algorithm was to learn this hysteresis phenomenon

without explicit programming and obtain a stable soliton regime at the lowest level of pumping power.

3. Reinforcement Learning

The RL algorithm operates with the mode-locked laser as a black-box. It does not have information about underlying physics and learns the behavior of the system through the actions and their consequences. In terms of the RL-algorithm, the considered problem is formulated as follows: the laser is an environment where an agent, a neural network, is acting by changing the parameters of the cavity. Fixed parameters of the laser cavity determine the characteristics of the output optical radiation, which are denoted as a state of the environment. The goal for the agent is to maximize the reward gained for the appropriate actions. Reward r is a function that is designed to have maximum at desired mode-locked regime. In our case, we have used the following function:

$$r = \frac{1}{max(ACF)} \tag{1}$$

where $max(ACF)$ is the amplitude of the maximum central peak at the autocorrelation function of the oscilloscope trace. This function provides the highest reward for mode-locked regimes at the lowest pumping power levels. We also have found that the autocorrelation function of an oscilloscope trace provides a convenient way to filter unstable, q-switched and continuous wave regimes (Figure 3). At the first stage, the autocorrelation function of the oscilloscope trace is calculated. Then, the algorithm detects the number of the peaks applying the peak detection algorithm available through Python Scipy package. If the number of the peaks is less than 2, the reward is zero; otherwise, the reward is calculated with Equation (1).

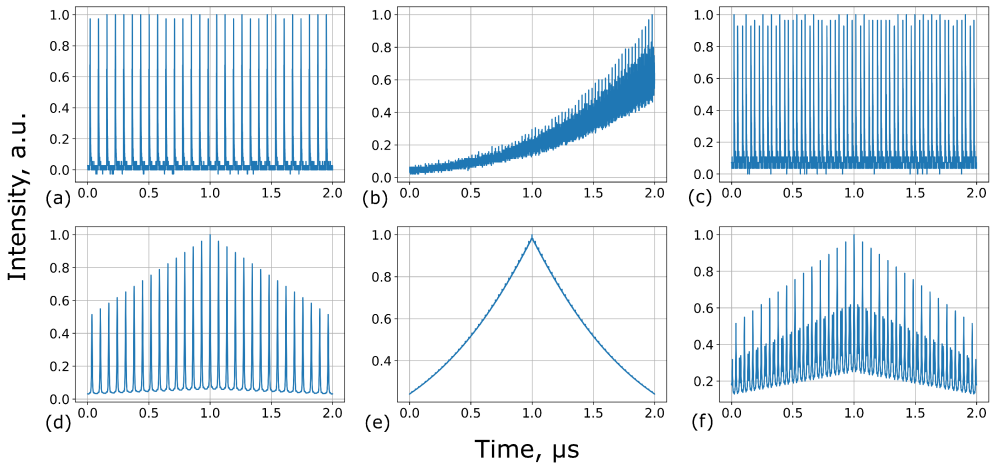

Figure 3. (**a**–**c**) Oscilloscope traces of stable single-pulsed regime; unstable, Q-switched, mode-locked regime; stable double-pulse regime. (**d**–**f**) Corresponding autocorrelation functions of the regimes.

Reinforcement learning algorithms usually require a large amount of training time. In case of an experimental mode-locked laser, there is no straightforward possibility to parallel the training procedure. We have used the double deep Q-learning network (DDQN) as a solution with a balanced trade-off between the stability and training time. For instance, the less complicated deep Q-learning algorithm suffers from an overestimation effect [17], and the actor–critic algorithm requires more training time and is less data efficient.

The scheme of the applied DDQN algorithm is depicted at Figure 4a. Firstly, the feed-forward neural network Q_{net} is initialized with random weights of neurons. Q_{net} predicts the action that has to be done based on the current state. The state in our case is a combination of laser diode current and reward that was measured at this current. The input layer has two neurons; the first corresponds to the the current of the laser diode, and the

second—to the estimated reward. The output layer has two neurons that correspond to the actions: increasing or decreasing the current of the pumping laser diode by 1 mA. Q_{net} also contains three hidden fully connected layers with 32 neurons each with a ReLu activation function. Next, Q_{tnet} is initialized by copying Q_{net}. Due to the low dimensions of the input and output layers, we did not use large neural networks. The final architectures of neural networks were designed by optimization of their performances to the toy numerical model of the hysteresis effect.

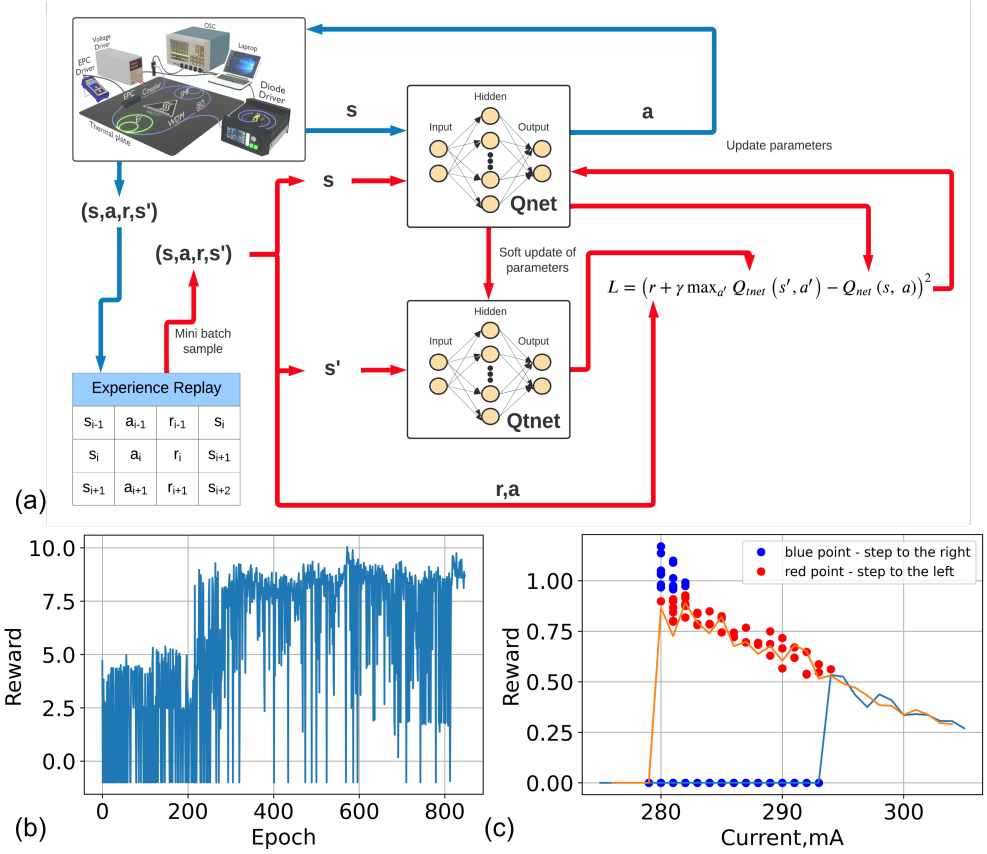

Figure 4. (**a**) Principle scheme of the DDQN algorithm. (**b**) Learning curve of the DDQN algorithm. (**c**) Laser current adjustment by the DDQN algorithm after the training stage. Blue and orange lines correspond to the calculated reward by consequently increasing the current up to 305 mA and decreasing it back down to initial point.

To prevent oscillation and divergence of the algorithm, we implemented the concept of experience replay buffer [18]. It breaks data correlation and makes it possible to use old experience for off-policy algorithms [19]. An experience replay buffer is a table containing the tuple of initial state, action, final state and reward that is obtained by transferring from initial to final state after the action. After Q_{net} and Q_{tnet} initialization, an experience replay buffer with a fixed size of 1000 was filled by records of acting based on Q_{net} predictions. Finally, we introduced the penalty procedure to avoid the situation when the agent receives high cumulative reward by oscillating around some intermediate point at the hysteresis curve. After each action, 10% of the maximum possible reward is deducted from the total reward. The training procedure includes the stage during which the agent makes 100 actions from initial point, trying to maximize the cumulative reward. Every action, the experience buffer was updated by substituting the old records by a new ones. The experience buffer was used as a training dataset for Q_{net}. The weights of Q_{net} were updated every 3 actions, and after 50 actions, the weights of Q_{tnet} were copied from Q_{net}. The new weights were calculated by backpropagation with an adam optimizer [20] at a learning rate of 10^{-4}. The optimizer minimizes the following loss function:

$$L = \left(r + \gamma \max_{a'} Q_{tnet}(s', a') - Q_{net}(s,\ a) \right)^2 \quad (2)$$

where γ—discount factor equal to 0.9; s, s' —current and next states of the environment; a—action. This loss function allows one to find the optimal Q-function which maximizes the cumulative reward during the self-adjusting procedure [21]. The Q-function is defined as the expected discounted reward for an agent making action a from a state s and then following policy π. Mathematically, it can be described in the following way:

$$Q^{\pi}(s, a) = \mathbb{E}_{\pi}[R_t | s_t = s, a_t = a] \quad (3)$$

where R_t is the expected reward that we will receive at the end of the episode. Using two neural networks allows one to solve overestimation bias for predicting the value of the Q-function. To determine the value of the Q-function, the minimum of the predictions of both networks is used. In order to run parallel learning of two networks asynchronously, two identical networks Q_{net} and Q_{tnet} were used. Q_{net} was trained by a back propagation method, and Q_{tnet} copied the weights of Q_{net} every few steps.

To make the agent act different during the learning process, thereby exploring the environment, an ϵ-greedy strategy was used to select an action. During the episode, the agent choose an action predicted by Q_{net} network with probability $(1 - \epsilon)$ or acted randomly with probability ϵ, where ϵ starts from value 0.5. Each epoch consisted of 10 episodes ϵ multiplied by 0.99. The iterations of the algorithm ran while the probability to make random actions was higher than 10^{-4}%. Figure 4b depicts the learning curves of the DDQN algorithms. After each epoch, we tested the trained agent on a laser system and measured the maximum reward it obtained after 100 actions.

There are three key stages of the agent's learning process. First, the agent makes almost random steps and does not explore the mode-locked regime. At this first stage the agent gets almost zero or negative reward. At second stage, the agent finds the mode-locked regime at a high level of pumping power, but then, the actions are close to random. At the last stage, the agent finds the way to decrease the pumping power to increase the reward. One may mention the corresponding three levels with around 0, 2.5 and 8.5 values for the reward. The abrupt drops in the reward at the last stage of the learning curve are caused by excessive step-down of the agent, which leads to breaking of a mode-locked regime. Since the experimental laser exhibits power fluctuations, the threshold value of the pumping power slightly changes. One experimental sampling point consisting of one step of pumping diode current and subsequent measurement of a oscilloscope trace took 0.48 s. Therefore, overall training time during 800 epochs was approximately 110 h.

Figure 4c depicts the strategy of the trained agent. The agent starts tuning a current of the laser diode from the initial point at 280 mA. Blue points correspond to increasing the current level by 1 mA, and red points correspond to decreasing the current level by 1 mA. At the first stage, the agent increases the current step-by-step until 294 mA—when the mode-locked regime emerges. Then, the agent starts to decrease the current while maintaining the mode-locked regime. The crucial point for successive adjustment is to stop decreasing the current on time. Figure 4c demonstrates that the agent has learned the threshold level of the current below which the mode-locked regime breaks up. On reaching the vicinity of the reward maximum, the agent consequently increases and decreases the current maintaining stable mode-locked regime at the lowest current.

4. Conclusions

We have demonstrated the feasibility of using the DDQN algorithm for the self-adjusting operation of a mode-locked fiber laser. The attractive feature of the algorithm is its ability to find a dynamic strategy to tune the pumping power of the laser diode in order to achieve a stable mode-locked regime. We would like to note that this is an initial demonstration and we expect to improve overall performance. For instance, at the

moment, despite the relatively simple task, a large amount of training data was required to teach the agent. In the case of a high-dimensional task, the number of required training samples would increase dramatically. Therefore, we believe that further development of RL algorithms applied to experimental laser systems should be combined with the appropriate numerical models capable of generating synthetic data close to the experimental ones. This would enhance the approach through distributed learning [22] approaches and parallelization of the learning process. Additionally, the transfer learning [23–25] concept may be applied to speed up the learning process; for instance, the already trained network may be applied to other laser system with hysteresis phenomena. We anticipate that our work will stimulate further study of the dynamic design and optimization of complex laser systems, when it is necessary to determine not optimal parameters of the cavity, but also trajectories (in the space of parameters) of reaching them. We believe that results contribute to development of "smart" fiber lasers that can self-adjust their operation to the changing environment.

Author Contributions: Conceptualization, A.K. and S.T.; methodology, A.K. and E.K; software, E.K., A.S. and K.S.; validation, A.K. and A.S.; formal analysis, A.S.; investigation, A.S. and K.S.; resources, K.S.; data curation, A.K.; writing—original draft preparation, A.K.; writing—review and editing, K.S.; visualization, A.S.; supervision, A.K. and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Russian Science Foundation (Grant No. 17-72-30006-P).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Acknowledgments: We are grateful to Nathan Kutz for useful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Genty, G.; Salmela, L.; Dudley, J.M.; Brunner, D.; Kokhanovskiy, A.; Kobtsev, S.; Turitsyn, S.K. Machine learning and applications in ultrafast photonics. *Nat. Photonics* **2021**, *15*, 91–101. [CrossRef]
2. Pu, G.; Yi, L.; Zhang, L.; Luo, C.; Li, Z.; Hu, W. Intelligent control of mode-locked femtosecond pulses by time-stretch-assisted real-time spectral analysis. *Light Sci. Appl.* **2020**, *9*, 1–8. [CrossRef] [PubMed]
3. Wu, X.; Peng, J.; Boscolo, S.; Zhang, Y.; Finot, C.; Zeng, H. Intelligent breathing soliton generation in ultrafast fiber lasers. *Laser Photonics Rev.* **2022**, *16*, 2100191. [CrossRef]
4. Kokhanovskiy, A.; Bednyakova, A.; Kuprikov, E.; Ivanenko, A.; Dyatlov, M.; Lotkov, D.; Kobtsev, S.; Turitsyn, S. Machine learning-based pulse characterization in figure-eight mode-locked lasers. *Opt. Lett.* **2019**, *44*, 3410–3413. [CrossRef] [PubMed]
5. Andral, U.; Buguet, J.; Fodil, R.S.; Amrani, F.; Billard, F.; Hertz, E.; Grelu, P. Toward an autosetting mode-locked fiber laser cavity. *JOSA B* **2016**, *33*, 825–833. [CrossRef]
6. Zibar, D.; Brusin, A.M.R.; de Moura, U.C.; Da Ros, F.; Curri, V.; Carena, A. Inverse system design using machine learning: The Raman amplifier case. *J. Light. Technol.* **2019**, *38*, 736–753. [CrossRef]
7. Iegorov, R.; Teamir, T.; Makey, G.; Ilday, F. Direct control of mode-locking states of a fiber laser. *Optica* **2016**, *3*, 1312–1315. [CrossRef]
8. Wu, H.; Lin, W.; Tan, Y.J.; Cui, H.; Luo, Z.C.; Xu, W.C.; Luo, A.P. Pulses with switchable wavelengths and hysteresis in an all-fiber spatio-temporal mode-locked laser. *Appl. Phys. Express* **2020**, *13*, 022008. [CrossRef]
9. Kuprikov, E.; Kokhanovskiy, A.; Kobtsev, S.; Turitysin, S. Exploiting hysteresis effect for electronic adjusting of fiber mode-locked laser. In Proceedings of the 2020 International Conference Laser Optics (ICLO), Saint Petersburg, Russia, 2–6 November 2020; IEEE: Piscataway, NJ, USA, 2020; p. 1.
10. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
11. Sun, C.; Kaiser, E.; Brunton, S.L.; Kutz, J.N. Deep reinforcement learning for optical systems: A case study of mode-locked lasers. *Mach. Learn. Sci. Technol.* **2020**, *1*, 045013. [CrossRef]
12. Yan, Q.; Deng, Q.; Zhang, J.; Zhu, Y.; Yin, K.; Li, T.; Wu, D.; Jiang, T. Low-latency deep-reinforcement learning algorithm for ultrafast fiber lasers. *Photonics Res.* **2021**, *9*, 1493–1501. [CrossRef]

13. Kuprikov, E.; Kokhanovskiy, A.; Serebrennikov, K.; Turitsyn, S. Deep reinforcement learning for self-tuning laser source of dissipative solitons. *Sci. Rep.* **2022**, *12*, 7185. [CrossRef] [PubMed]
14. Tang, D.; Zhao, L.M.; Zhao, B.; Liu, A. Mechanism of multisoliton formation and soliton energy quantization in passively mode-locked fiber lasers. *Phys. Rev. A* **2005**, *72*, 043816. [CrossRef]
15. Li, R.; Zou, J.; Li, W.; Wang, K.; Du, T.; Wang, H.; Sun, X.; Xiao, Z.; Fu, H.; Luo, Z. Ultrawide-space and controllable soliton molecules in a narrow-linewidth mode-locked fiber laser. *IEEE Photonics Technol. Lett.* **2018**, *30*, 1423–1426. [CrossRef]
16. Komarov, A.; Komarov, K.; Sanchez, F. Quantization of binding energy of structural solitons in passive mode-locked fiber lasers. *Phys. Rev. A* **2009**, *79*, 033807. [CrossRef]
17. Van Hasselt, H.; Guez, A.; Silver, D. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
18. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [CrossRef] [PubMed]
19. Zhang, S.; Sutton, R.S. A deeper look at experience replay. *arXiv* **2017**, arXiv:1712.01275.
20. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
21. Gaskett, C.; Wettergreen, D.; Zelinsky, A. Q-learning in continuous state and action spaces. In Proceedings of the Australasian Joint Conference on Artificial Intelligence, Sydney, Australia, 6–10 December 1999; Springer: Berlin/Heidelberg, Germany, 1999; pp. 417–428.
22. Espeholt, L.; Soyer, H.; Munos, R.; Simonyan, K.; Mnih, V.; Ward, T.; Doron, Y.; Firoiu, V.; Harley, T.; Dunning, I.; et al. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 1407–1416.
23. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
24. Freire, P.J.; Abode, D.; Prilepsky, J.E.; Costa, N.; Spinnler, B.; Napoli, A.; Turitsyn, S.K. Transfer Learning for Neural Networks-Based Equalizers in Coherent Optical Systems. *J. Lightwave Technol.* **2021**, *39*, 6733–6745. [CrossRef]
25. Freire, P.J.; Spinnler, B.; Abode, D.; Prilepsky, J.E.; Ali, A.; Costa, N.; Schairer, W.; Napoli, A.; Ellis, A.D.; Turitsyn, S.K. Domain Adaptation: The Key Enabler of Neural Network Equalizers in Coherent Optical Systems. In Proceedings of the Optical Fiber Communication Conference (OFC) 2022, San Diego, CA, USA, 6–10 March 2022; p. Th2A.35. [CrossRef]