Research Article | Vol. 31, No. 8 / 10 Apr 2023 / Optics Express 13140
Optics EXPRESS

Real-time adaptive optical self-interference cancellation for in-band full-duplex transmission using SARSA(λ) reinforcement learning

XIAO YU,[1] JIA YE,[1,4] LIANSHAN YAN,[1,5] TAO ZHOU,[2] PENG LI,[1,3] XIHUA ZOU,[1] WEI PAN,[1] AND JIANPING YAO[3]

[1]*Center for Information Photonics & Communication, School of Information Science & Technology, Southwest Jiaotong University, Chengdu 611756, China*
[2]*Key Laboratory of Electronic Information Control, Southwest China Research Institute of Electronic Equipment, Chengdu 610036, China*
[3]*Microwave Photonics Research Laboratory, School of Information Technology and Engineering, University of Ottawa, Ottawa ONK1N 6N5, Canada*
[4]*jiaye@home.swjtu.edu.cn*
[5]*lsyan@home.swjtu.edu.cn*

Abstract: Self-interference (SI) due to signal leakage from a local transmitter is an issue in an in-band full-duplex (IBFD) transmission system, which would cause severe distortions to a receiving signal of interest (SOI). By superimposing a local reference signal with the same amplitude and opposite phase, the SI signal can be fully canceled. However, as the manipulation of the reference signal is usually operated manually, it is difficult to ensure a high speed and high accurate cancellation. To overcome this problem, a real-time adaptive optical SI cancellation (RTA-OSIC) scheme using a SARSA(λ) reinforcement learning (RL) algorithm is proposed and experimentally demonstrated. The proposed RTA-OSIC scheme can automatically adjust the amplitude and phase of a reference signal by adjusting a variable optical attenuator (VOA) and a variable optical delay line (VODL) achieved through an adaptive feedback signal, which is generated by evaluating the quality of the received SOI. To verify the feasibility of the proposed scheme, a 5 GHz 16QAM OFDM IBFD transmission experiment is demonstrated. By using the proposed RTA-OSIC scheme, for an SOI at three different bandwidths of 200, 400, and 800 MHz, the signal can be adaptively and correctly recovered within 8 time periods (TPs), which is the required time of a single adaptive control step. The cancellation depth for the SOI with a bandwidth of 800 MHz is 20.18 dB. The short- and long-term stability of the proposed RTA-OSIC scheme is also evaluated. The experimental results indicate that the proposed approach could be a promising solution for real-time adaptive SI cancellation in future IBFD transmission systems.

1. Introduction

In-band full-duplex (IBFD) transmission systems have been widely used for applications such as next-generation communications [1], full-duplex radio relaying [2], and joint radar-communication system [3]. An IBFD transmission system can improve the wireless spectrum efficiency through transmitting and receiving signals simultaneously by utilizing a single channel in the same frequency band. Nonetheless, the main challenge for an IBFD transmission system is the strong in-band self-interference (SI) in the receiving path, which is caused due to signal leakage from the co-site transmitter. Therefore, to avoid in-band SI, an IBFD transmission system needs to be implemented with SI cancellation (SIC), to recover the signal of interest (SOI) free from SI. A basic approach to implement SIC is to generate a replica of the SI signal and to subtract it from the receiving signal. A SIC system can be implemented electronically. However, the

#486889 https://doi.org/10.1364/OE.486889
Journal © 2023 Received 2 Feb 2023; revised 8 Mar 2023; accepted 26 Mar 2023; published 5 Apr 2023

Optics EXPRESS

performance of an electrical SIC scheme is limited due to large nonlinearities, high loss, narrow instantaneous bandwidth, and coarse time-delay tuning [4,5].

Compared with an electrical SIC scheme [5], an optical SIC system with broader instantaneous bandwidth, better time delay accuracy, and wider time delay tunable range, has attracted much attention recently. Numerous optical SIC schemes have been reported. The key to implement a SIC system is to achieve π phase inversion. For example, a wideband co-site SIC scheme was proposed in which the phase reversion was implemented electronically [6,7], thus the system is not all optical, but electrical-optical hybrid. To implement an all-optical SIC scheme, we can opposite-bias two modulators to achieve phase inversion [8–11]. The use of the out-of-phase relationship between two optical sidebands of a phase-modulated signal can also achieve phase inversion [12–14]. The use of balanced photodetection can also achieve phase inversion [15–18]. The interference cancellation of all the schemes mentioned above is accomplished through manual control of a reference signal via waveform or spectrum observation [19]. Thus, the process will take a long time to complete and will limit the practical applications especially for an IBFD system having a fast changing interference signal [20]. To increase the speed, the manual manipulation process must be replaced by a fully automated process that is assisted by a real-time adaptive algorithm [5,19–26]. Nelder-Mead simplex algorithm was employed for automatic optimization in Ref. [20] and [21], while Ref. [22] used Hooke-Jeeves algorithm for adaptive tuning of reference signal parameters. In Ref. [23], particle swarm optimization algorithm was utilized to optimize amplitude, time delay, and phase. Least mean square algorithm was used as a pre-equalizer for parameter optimization in Ref. [24]. Additionally, adaptive algorithms based on regular triangle algorithm with different criteria types were proposed in Ref. [19], [25], and [26] for quick adjustment of amplitude and time delay. Finally, Ref. [5] proposed an optimization algorithm based on neural networks for achieving parameter adjustment.

In recent years, artificial intelligence (AI) has been widely studied, which have been used in the fields such as meteorology [27], biology [28], and optics [29–32]. Adaptive feedback control, pattern recognition, big data analysis, and feature extraction are the most common AI applications. Due to its forecasting and policy-making abilities, reinforcement learning (RL) is an essential branch of AI and it can provide a solution to achieve adaptive feedback control of a complicated system [33] for applications such as self-driving, industrial automation, and robotics control. When the environmental condition changes, the RL strategy algorithm should change as well. Hence, massive strategy algorithms based on the idea of RL are proposed, including Markov decision process, Monte Carlo method [34], temporal difference (TD) [33], Q learning [35], SARSA [36], and others [30–32,37,38]. A system employing these strategy algorithms could explore, reach, and maintain the desired goal in a variety of environmental conditions. In the field of optics, RL is a powerful tool for optimizing convex/nonconvex problems for a given environmental condition [28–31]. The RL can explore the environment and collect its information to tune the parameters of the experimental system for reaching and staying in the desired goal over a long period of time without the need of manual adjustment.

In this paper, a real-time adaptive optical SI cancellation (RTA-OSIC) scheme using a SARSA(λ) RL algorithm is proposed, which is employed to optimize the time delay and attenuation to control a reference signal to achieve adaptive SIC of an IBFD transmission system. The proposed RTA-OSIC scheme can learn from real-time feedback received from an interactive external environment and select the SIC strategy for an IBFD transmission. The proposed approach is evaluated by an experiment. The experimental results show that for a 5 GHz IBFD transmission system, the signal can be adaptively and correctly recovered within 8 time periods (TPs) for an SOI at three different bandwidths of 200, 400, and 800 MHz, where the TP is the required time of a single adaptive control step. The cancellation depth for the SOI with a bandwidth of 800 MHz is 20.18 dB. The short- and long-term stability of the proposed RTA-OSIC scheme is also evaluated.

Optics EXPRESS

2. System model and algorithm

2.1. Adaptive OSIC system

Figure 1 shows the conceptual architecture of the proposed real-time adaptive optical SI cancellation scheme [39]. An electrical signal from the transmitter (Tx) is fed into the electrical/optical (E/O) conversion module to be converted into an optical signal. The optical signal is then split into two branches by an optical coupler (OC), with one serving as a transmitting signal and the other a reference signal. The optical transmitting signal is sent over a length of single-mode fiber (SMF1) to an optical/electrical (O/E) conversion module to generate an electrical signal, which is subsequently radiated to free space via the transmitting antenna. The reference signal is sent to the OSIC module for SI signal cancellation. A hybrid electrical signal from the receiving antenna consisting of an SI signal and a data signal is applied to an electrical/optical (E/O) conversion module to convert the received signal to an optical signal, which is sent to the OSIC module over a second SMF (SMF2). In the OSIC module, the amplitude and phase (time delay) of the reference signal is adjusted by a variable optical attenuator (VOA) and a variable optical delay line (VODL), and then fed into a balanced photo-detector (BPD). The received hybrid optical signal is also injected into the BPD, where the cancellation of the SI signal and the recovery of the SOI are performed. To match the amplitude and time delay of the SI signal, the agent of the algorithm would take the action to control the VOA and VODL according to the quality of the recovered SOI, which is evaluated by bit error rate (BER) and error vector magnitude (EVM). What's more, as the IBFD system is deployed, it is noteworthy that the transmission path of the SI from the transmitter antenna to receiver antenna is a fixed parameter. However, as a result of the fact that the tunable reference signal and the SI signal are from the same source, it becomes evident that regardless of the amplitude, and frequency of the SI signal, the necessary conditions for SIC can be achieved. Thus, it can be concluded that the SIC system does not impose any restrictions on the signal forms of the SI signal.

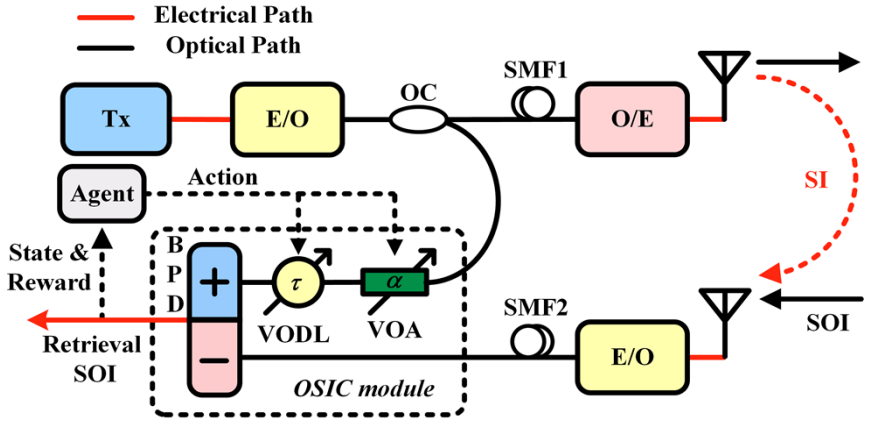

Fig. 1. Conceptual architecture of the proposed real-time adaptive optical SI cancellation scheme. Tx, transmitter; E/O, electrical/optical; O/E, optical/electrical; BPD, balanced photo-detector; OC, optical coupler; SMF, single-mode fiber; SI, self-interference; SOI, signal of interest; VOA, variable optical attenuator; VODL, variable optical delay line.

Figure 2 depicts the schematic diagram of the SARSA(λ) RL agent, which is a real-time intelligence cognitive module. First, the RL agent selects and executes an action according to the previous state. The received signal is then captured by a real-time oscilloscope to calculate the reward at the current state. Finally, the action-value function (Q) is updated via the learning experience by accumulating the rewards at the current state and action, which guide the RL agent

Optics EXPRESS

to choose the optimal action for SI cancellation. The RL agent will gradually converge and stabilize. Thus, a closed adaptive control loop is established to ensure real-time cancellation of the SI signal for the IBFD transmission system. It should be noted that the deployment of an IBFD system can be subject to various changes in its working conditions, notably those concerning the amplitude and time delay of the received SI signal. Hence, it becomes vital to observe that the SIC system may need to be re-calibrated and re-deployed to ensure that it adapts to the new working conditions [23].

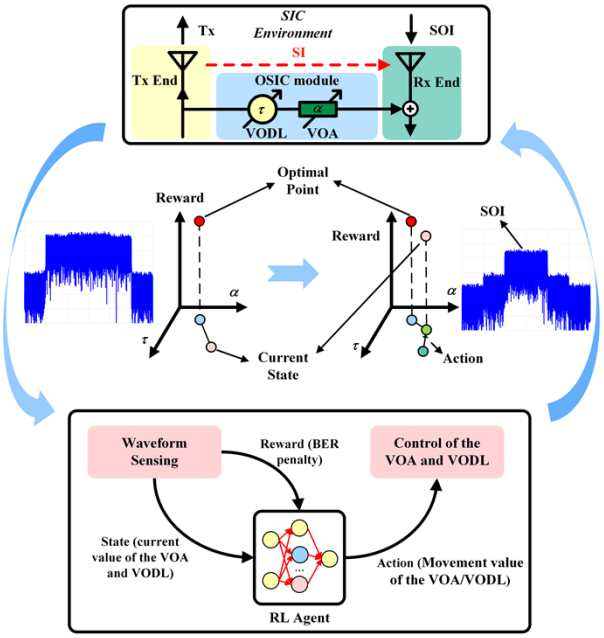

Fig. 2. Schematic diagram of the proposed real-time adaptive optical self-interference cancellation scheme using the SARSA(λ) RL algorithm. Tx, transmitting signal.

2.2. SARSA(λ) reinforcement learning algorithm

A specific SARSA(λ) RL algorithm is utilized in our work to select the optimal strategy for canceling the SI signal, due to the great sensitivity of penalty and the high effectivity of multi-step bootstrapping. The SARSA agent learning represents five core learning factors: the state, the action, the reward, the next state, and the next action, which is an on-policy temporal difference control method to explore an optimal strategy. In the specific SARSA(λ) RL algorithm, the state is defined by two-tuple, including the time delay value of the VODL and the attenuation value of the VOA in the current state. The action is also defined by two-tuple, including the movement values of two dimensions (i.e., the time delay and the attenuation). The movement values of the time delay dimension are fixed at -100 ps, -10 ps, -1 ps, 0 ps, 1 ps, 10 ps, and 100 ps. The movement values of the attenuation dimension are fixed at -0.1 dB, 0 dB, and 0.1 dB. To achieve a large tunable range and high tuning accuracy in the algorithm, careful consideration of movement values in both the time delay and attenuation dimensions is necessary. The adaptive algorithm adjusts only one parameter at a time to achieve the desired tuning accuracy, thereby avoiding the need for a more complex Q table and reducing the amount of trial-and-error required during training. While it is possible to adjust two or more parameters in one control step using the proposed SARSA(λ) RL algorithm, doing so can increase the number of associated actions, making it difficult for the Q table to map the relationship between state and action. Therefore, to

Optics EXPRESS

reduce the complexity of the associated actions and accelerate the convergence of the algorithm, only one action (time delay or attenuation) is selected in one control step. There are seven associated actions for adjusting time delay, including [-100 ps, 0 dB], [-10 ps, 0 dB], [-1 ps, 0 dB], [0 ps, 0 dB], [1 ps, 0 dB], [10 ps, 0 dB], and [100 ps, 0 dB]. In contrast, there are only three associated actions for adjusting attenuation, namely [0 ps, -0.1 dB], [0 ps, 0 dB], and [0 ps, 0.1 dB]. However, the movements of [0 ps, 0 dB] for both time delay and attenuation can be combined as a single action for selection. As a result, there are nine associated actions for policy selection at the current state. Finally, the reward is designed as the $\log_{10}(BER)$ difference between the previous and the current states when the measured thresholds (i.e., the thresholds of the EVM and BER) are unreached. Unless, the reward is set to be 30 to speed up the convergence process of the RL algorithm. Consequently, the reward function can be expressed as

$$\mathrm{r} = \begin{cases} 30 & \text{if goals reached,} \\ \log_{10}(BER_{prev}) - \log_{10}(BER_{curr}) & \text{else} \end{cases} \tag{1}$$

where the subscripts, *prev* and *curr*, represent the previous and the current states, respectively. The goals are set as the current EVM $\geq$ $\mathrm{EVM_{threshold}}$ and the current BER $\geq$ $\mathrm{BER_{threshold}}$. The EVM and the BER thresholds are the maximum acceptable EVM for the recovered signal and hard-decision forward error correction (HD-FEC) limit with 7% overhead, respectively. Therefore, the BER of the current state should be better than the previous BER, indicating that the reward (r) is positive and the agent will have a positive experience, and vice versa.

In the RL algorithm updating rule, the action-value function is updated toward the n step return, which is defined as

$$Q(s_t, a) = r_t + \gamma r_{t+1} + \cdots + \gamma^{n-1} r_{t+n-1} + \gamma^n Q(s_{t+n}, a) \tag{2}$$

where s_t and s_{t+n} represent the current state and previous n state, respectively. a represents all actions in the action space, r is the reward, γ is the decay parameter. The eligibility trace (E) from the backward view provides an online incremental implementation, resulting in the SARSA(λ) algorithm, where $\lambda \in$[0,1]. The eligibility trace is short-term memory and assists the learning process by affecting the weight vector of Q, which will usually last within some steps [33]. Meanwhile, the weight vector of Q is long-term memory and determines the predicted value, which will last for the entire running time. The eligibility trace can help with the issues of long-delayed rewards and non-Markov tasks [40]. In addition, the predicted action of the RL agent a_{n+1} from the state s_{n+1} is made by a policy derived from Q. Therefore, the temporal difference between the current prediction and the next prediction can be written as

$$\delta \leftarrow r + \gamma Q(s_{n+1}, a_{n+1}) - Q(s_n, a_n) \tag{3}$$

where s and a represent the state and action, respectively. The subscripts, n, and $n+1$ represent the current and the next step, respectively. Therefore, the Q table of the SARSA(λ) agent update rule can be expressed as

$$Q(s, a) \leftarrow Q(s, a) + \alpha \cdot \delta \cdot E(s, \mathrm{a}) \tag{4}$$

where α is the learning rate of the SARSA(λ) algorithm.

For training the agent, the hyperparameters used in RL algorithm are shown in Table 1. The initial exploration of ε-greedy is set to be 1 with an exponential decay rate at 0.95 per TP. With the high initial exploration of ε-greedy, the agent will accumulate the positive and negative experiences through exploring the given environmental condition. It's important to note that when the exploration of the ε-greedy strategy is high enough, the agent may prioritize actions that lead to higher rewards. However, as the exploration of ε-greedy decreases due to exponential

Optics EXPRESS

decay rate, the agent will likely opt for safer actions instead of continuing to explore for potentially higher rewards. In addition, the learning rate (α) and the decay parameter (λ) are set to be 0.1 and 0.95, respectively. The eligibility trace decay parameter (γ) is specified as 0.8 to represent the indispensability of every state with the goal reached. The detailed steps for the implemented SARSA(λ) algorithm are described as below:

Algorithm 1. SARSA(λ) RL Algorithm for real-time optical SI cancellation

Initialize $Q(s, a) = 0$ and $E(s, a) = 0$, for all s, a
ε=1, $\lambda = 0.95$, $\gamma = 0.8$,
Observe current state s_0, current BER BER_0
Select current action $a_0(t_t, d_t)$ randomly from action space A
for < TP in number of TPs > **do**
  $t = t + t_t$, $d = d + d_t$
  Observe and calculate next state s_1, EVM and BER_1
  Calculate r using Equation 1.
  if < rand()<ε> **then**
    Select a_1 randomly from action space A
  else
    $a_1 = \arg\max_a Q(s_1, a)$, $a \in$ action space A
  if s_1 is not terminal **then**
    $\delta = r + \lambda \cdot Q(s_1, a_1) - Q(s_0, a_0)$
  else
    $\delta = r - Q(s_0, a_0)$
  $E(s_0, a) = 0$, $a \in$ action space A
  $E(s_0, a_0) = E(s_0, a_0) + 1$
  for all s, a **do**
    $Q(s, a) = Q(s, a) + \alpha \cdot \delta \cdot E(s, a)$
    $E(s, a) = \gamma \cdot \lambda \cdot E(s, a)$
  end for
  $\varepsilon = \varepsilon \cdot 0.95$
  $a_0 = a_1$, $s_0 = s_1$, $BER_0 = BER_1$
end for

Table 1. Hyperparameter used in the RL algorithm

Hyperparameter	Value
exponential decay rate	0.95
the learning rate (α)	0.1
decay parameter (λ)	0.95
eligibility trace decay parameter (γ)	0.8

All the implemented algorithms are written in Python language and run on a desktop with Windows 10 operating system. The model of the CPU is Intel Core i5-10400@2.9 GHz on the desktop. Python programming language is also used to control VOA and VODL as well as to capture oscilloscope data.

3. Experimental setup and result

To validate the feasibility of the proposed RTA-OSIC scheme for SI cancellation in an IBFD transmission system, an experiment is performed based on the setup shown in Fig. 3. In the experiment, an optical carrier (LD1, TNL, TeraXion) with a wavelength of 1550.5 nm and an output power of 12 dBm is fed into MZM1 (Realphoton, IM-40-LN) where the local reference

Optics EXPRESS

signal is modulated on the optical carrier. The other optical carrier is from a tunable narrow linewidth laser diode (LD2, Yenista Optics, OSICS) with a wavelength of 1565.5 nm and an output power of 12 dBm, which is sent to MZM2 (Sumitomo, T.MXH1.5-40PD-ADC) where the SI signal and SOI are modulated on the optical carrier. A 16QAM OFDM signal centered at 5 GHz with a 1 GHz bandwidth is generated by an arbitrary waveform generator (AWG, Keysight, M8195A) and sent to an electrical amplifier (EA, KG-RF-40-Serious). The amplified electrical signal is split into two signals via an electrical splitter, with one serving as the local reference (LR) signal, and the other as the SI signal. After transmitting through two different paths, a time delay and a power attenuation are both introduced between the two electrical signals. The local reference signal is fed into MZM1 via the radio frequency (RF) port and converted to an optical signal. An SOI of 16QAM OFDM centered at 5 GHz is also generated by the AWG. Here, the bandwidth of the desired SOI is intentionally set narrower than that of the SI signal, allowing the SOI to be influenced by the SI signal across the entire frequency spectrum. If the bandwidth of the SI signal is equal to or less than that of the SOI, the SOI may not be fully influenced, resulting in its unsuitability for evaluating SIC systems in IBFD transmission. Clear observation of the recovery process of the SOI signal in the spectral domain during convergence is also enabled by this setup [23,25]. Additionally, scenarios where the signals have different natures, such as the qualities of service, can be simulated by using different bandwidths between the SOI and SI signals [25]. After amplified by an electrical amplifier (EA, SHF 100AP), the SOI is combined with the SI signal at an electrical combiner (EC) to get a hybrid signal (SI signal and SOI), which is applied to MZM2 via the RF port, and converted to an optical signal. The local reference optical signal and the hybrid optical signal are amplified by two erbium-doped fiber amplifiers (EDFAs, Amonics, AEDFA-PA-35-B-FA) to compensate for the coupling and link loss. The amplified local reference optical signal is sent to a multipath optical tunable delay line (MOTDL) module, where the time delay and the attenuation is adjusted [39].

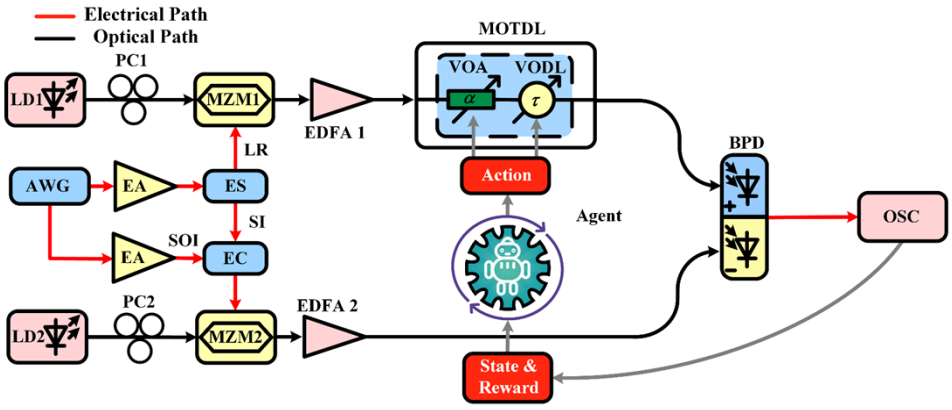

Fig. 3. Experimental setup of the proposed real-time adaptive optical self-interference cancellation scheme using a SARSA(λ) RL algorithm. LD, laser diode; PC, polarization controller; MZM, Mach-Zehnder modulator; EDFA, erbium-doped fiber amplifier; MOTDL, multipath optical tunable delay lines; AWG, arbitrary waveform generator; EA, electrical amplifier; ES, electrical splitter; EC, electrical combiner; LR, local reference signal; BPD, balanced photo-detector; OSC, oscilloscope.

The MOTDL module consists of electrically controlled VOAs and VODLs, where all of them could be controlled by the agent of the SARSA(λ) RL algorithm through the serial ports, in which the current values of the VOA and VODL are regarded as the state. The time delay accuracy of a VODL is 0.1 ps with a tunable range of 12 ns. The attenuation accuracy of a VOA is 0.1 dB

Optics EXPRESS

with a tunable range is 30 dB. It is worth noting that the real-time control and convergence time of the system are ultimately restricted by the relatively slow tuning speed of the VODL integrated within the MOTDL, which operates at the second level of time. The tuning speed can be further enhanced by using the VODL with a faster delay varying speed [23]. After tuning the time delay and attenuation, the local reference optical signal is sent to a BPD via one input port (u2t, BPDV2150RM). The amplified hybrid optical signal is applied to the BPD via the other input port. Thus, the SI signal is cancelled and the SOI free from SI is recovered. Then, the recovered electrical signal is captured by an oscilloscope (OSC, LeCroy, WaveMaster813Zi), which is sensed as the state and reward for the agent of the SARSA(λ) RL algorithm. According to the state and reward, the agent of the SARSA(λ) RL algorithm will learn the positive/negative experience and make a suitable decision to tune the VOA and the VODL. Thus, a closed adaptive control loop is established. For each TP, the optimization time of the adaptive SARSA(λ) RL algorithm in the system depends on the adjusting speeds of the two optimized parameters and the executing time of the adaptive algorithm. On the one hand, the response time of the VOA is usually much faster than the adjusting time of the VODL in the MOTDL. This is due to the fact that the adjusting speed of the VODL usually work at the second level of time, as mentioned earlier. On the other hand, the Q table search time of the SARSA(λ) RL algorithm affects the executing time of the adaptive algorithm. Typically, this time is shorter than the response time of the VODL. Therefore, the adjusting speed of the VODL plays a critical role in the time complexity of each TP, which operates under the second level of time.

To investigate the SI cancellation performance for a wideband radio frequency signal, the SI signal with a bandwidth of 1 GHz and SOI with a bandwidth of 800 MHz are generated by the AWG. The measured RF spectra of the SI signal (blue solid line), SOI (red dash-dotted line), and the SI signal with cancellation (yellow dotted line) are shown in Fig. 4. The initial average powers of the SOI and the SI signal are -59.83 dBm and -59.96 dBm, respectively. After cancellation, the average power of the SI signal is reduced, which achieves a cancellation depth of 20.18 dB within a bandwidth of 800 MHz for the SOI. However, some ripples can be still observed as the transmission response of the devices used in the experimental setup are non-uniform within the working frequency range. The spike at 5 GHz is a spurious carrier from the AWG.

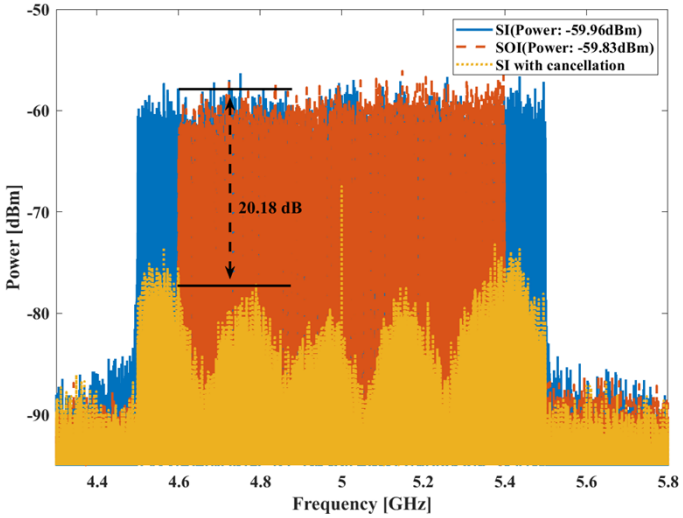

Fig. 4. Measured RF spectra of the SI, the SOI, and the SI with cancellation for the optical SI cancellation system.

Optics EXPRESS

To demonstrate the real-time adaptive SI cancellation ability of the proposed RTA-OSIC scheme, we convert the experimental setup into a real-time adaptive SI canceller by using the implemented SARSA(λ) RL algorithm. The center frequency and the bandwidth of the SOI are, respectively, set as 5 GHz and 400 MHz. Figure 5 shows the training process of the proposed RTA-OSIC scheme. The color bar represents the TPs of the iteration step. It can be seen that the agent of the SARSA(λ) RL algorithm takes actions to explore the higher reward and eventually finds the maximum reward after 108 TPs, which verifies the effectiveness of the proposed OSIC scheme.

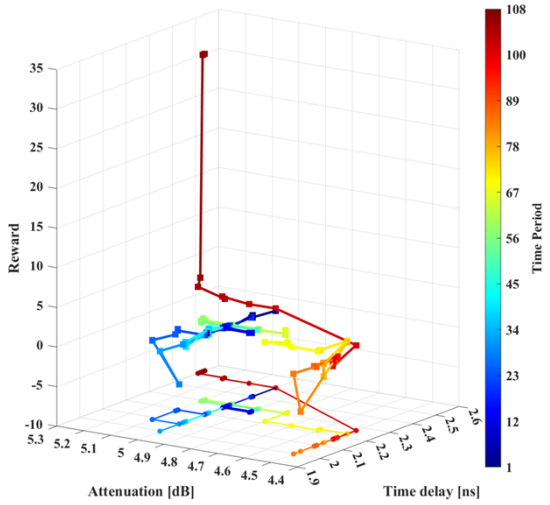

Fig. 5. Training process of the proposed RTA-OSIC scheme.

To test the robustness of the proposed RTA-OSIC scheme, the values of the VOA and the VODL are both deliberately reset to their initial values per 100 TPs, which is called an episode. Figure 6 shows the BER and EVM performance of the recovered SOI and corresponding rewards over 900 TPs. It can be seen that both the BER and EVM of the SOI and the corresponding reward value can be quickly recovered in only several TPs. The reason is that, when the algorithm converges to the optimum cancellation position and arrives at a stable convergence state, the Q and the eligibility trace table of the algorithm will guide the agent of the SARSA(λ) RL algorithm to make the most effective policy action to immediately recover from the system reset. This characteristic ensures the robustness of the proposed RTA-OSIC scheme.

Figure 7 shows that the BERs (blue solid line) and EVMs (red solid line) are measured within the first nine TPs in an episode. The HD-FEC limit (blue dotted line), the required EVM threshold (red dotted line), and the corresponding constellations diagram of the SOI are shown in Fig. 7. The corresponding constellation diagrams become more and more identifiable, while the SI signal is mitigated and the SOI is gradually recovered. It can be seen that the agent of the SARSA(λ) RL algorithm has ability to takes decisive and advantageous actions to quickly cancel the SI signal and to recover the SOI within 8 TPs.

To verify the performance of the proposed RTA-OSIC scheme with different bandwidths, two SOI with one having a bandwidth of 200 M Hz and the other a bandwidth of 800 MHz are generated. Figures 8(a)-(c) and (f)-(h) show the measured BERs, EVMs, and corresponding rewards for the recovered SOI with the bandwidths of 200 MHz and 800 MHz, respectively. After initial adjustment, the BERs and EVMs of the SOI have reached the levels below the designed threshold and the corresponding reward values are stabilized at the maximum value within 8 TPs in each episode, as shown in Figs. 8(a)-(c) and (f)-(h). The recognizable recovered constellation

Optics EXPRESS

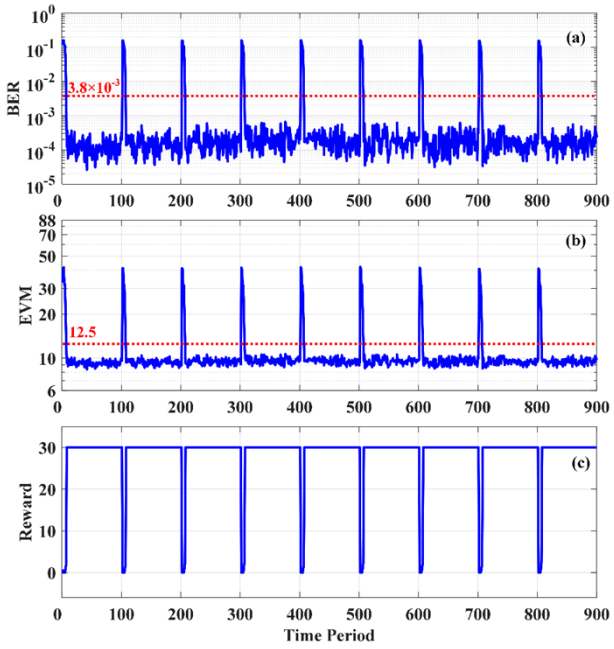

Fig. 6. Measured (a) BERs, (b) EVMs of recovered SOI, and (c) the corresponding rewards over TPs for the proposed RTA-OSIC scheme.

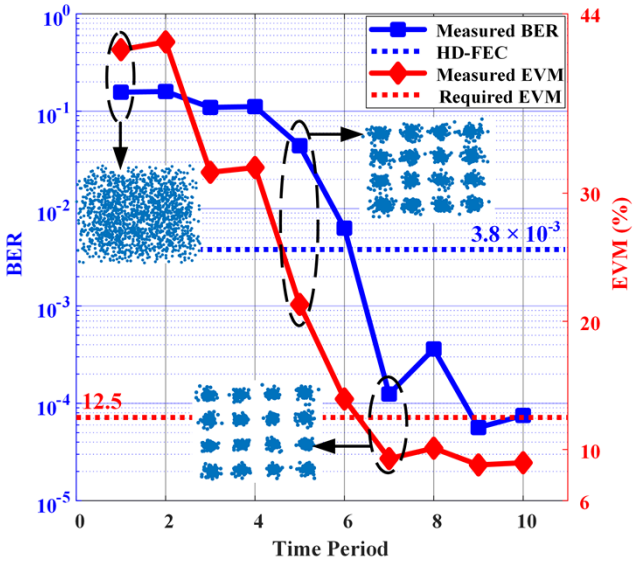

Fig. 7. Measured BERs and EVMs of the recovered SOI.

Optics EXPRESS

diagrams are observed with the EVMs of 8.56% and 10.55%, as shown in Figs. 8(d) and (i). The corresponding measured RF spectra are observed in Figs. 8(e) and (j), with the suppressed SI signal and the visible SOI. It is worth noting that the reward curve may exhibit a dip below the maximum value of 30, even if the BER value meets the established threshold of 3.8×10^{-3}, as depicted in Fig. 8(h). This curious occurrence stems from the fact that the EVM value exceeds its corresponding threshold of 12.5%, as shown in Fig. 8(g).

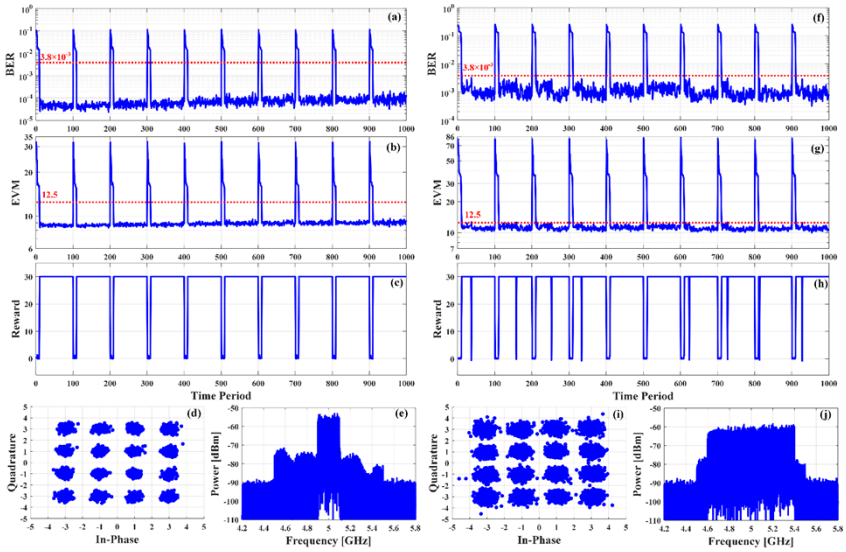

Fig. 8. Measured (a) BERs, (b) EVMs of the recovered SOI with a bandwidth of 200 MHz, and (c) the corresponding rewards over TPs. (d) The recovered SOI constellation with a bandwidth of 200 MHz and (e) the corresponding measured RF spectrum with SI cancellation. The measured (f) BERs, (g) EVMs of the recovered SOI with a bandwidth of 800 MHz, and (h) the corresponding rewards over TPs. (i) The recovered SOI constellation diagram with a bandwidth of 800 MHz and (j) is the corresponding measured RF spectrum with SI cancellation.

To assess the short- and long-term stability of the proposed RTA-OSIC scheme, we investigate the cancellation performance within the different time spans of 30, 50, 100, 500, 1000, and 2000 TPs. Figure 9 shows the measured BERs, EVMs, and the corresponding rewards of the recovered SOI with a bandwidth of 800 MHz in the different time spans for the proposed RTA-OSIC scheme. As shown in Figs. 9(a) and (b), the BER and EVM performance have reached below the designed thresholds again after several initial adjustment and then keep the stable state. In Fig. 9(c), some notched spikes in the reward curve after the maximum reward are observed, which are caused by the dynamic fluctuation of the system. The dynamic fluctuation is induced by the fluctuation of the bias point of the MZMs and output power of the LD and AWG. To eliminate it, the highly stable bias control circuits and adaptive control circuits for the time and power consistencies should be employed in the system [9].

Optics EXPRESS

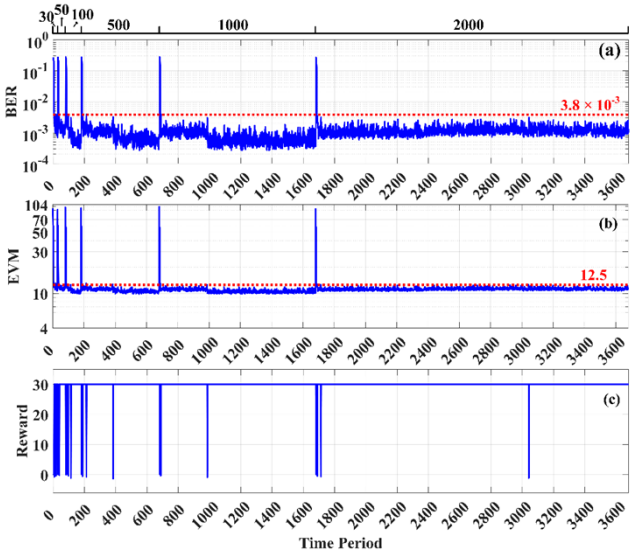

Fig. 9. Measured (a) BERs, (b) EVMs, and (c) the corresponding rewards of the recovered SOI with a bandwidth of 800 MHz in the different time spans for the proposed RTA-OSIC scheme.

4. Conclusion

A real-time adaptive optical SI cancellation scheme using the SARSA(λ) RL algorithm was proposed and experimentally demonstrated. The feasibility of the proposed RTA-OSIC scheme was verified by an experiment. The results showed that SOI could be adaptively and correctly recovered within 8 TPs at three different bandwidths of 200, 400, and 800 MHz. The cancellation depth for the SOI with a bandwidth of 800 MHz was 20.18 dB. Moreover, the short- and long-term stability of the proposed RTA-OSIC scheme was assessed in a real-time environmental condition. Therefore, the proposed RTA-OSIC scheme is a promising solution for wireless IBFD transmission that is free from SI due to its strong ability for real-time adaptive SI cancellation.

Funding. National Key Research and Development Program of China (2019YFB1803500); National Natural Science Foundation of China (61860206006, 62075185); Sichuan International Science and Technology Innovation Cooperation Project (2021YFH0013).

Acknowledgments. The authors would like to thank the anonymous reviewers for their valuable comments that helped improve this paper.

Disclosures. The authors declare no conflicts of interest.

Data availability. 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.

References

1. H. Tabassum, A. H. Sakr, and E. Hossain, "Analysis of massive MIMO-enabled downlink wireless backhauling for full-duplex small cells," IEEE Trans. Commun. **64**(6), 2354–2369 (2016).
2. S. Hong, J. Brand, J. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5 G and beyond," IEEE Commun. Mag. **52**(2), 114–121 (2014).
3. A. Hassanien, B. Himed, and M. G. Amin, "Transmit/receive beamforming design for joint radar and communication systems," in *Radar Conference* (IEEE, 2018), pp. 1481–1486.
4. M. P. Chang, M. Fok, A. Hofmaier, and P. R. Prucnal, "Optical analog self-interference cancellation using electro-absorption modulators," IEEE Microw. Wireless Compon. Lett. **23**(2), 99–101 (2013).
5. K. E. Kolodziej, A. U. Cookson, and B. T. Perry, "RF canceller tuning acceleration using neural network machine learning for in-band full-duplex systems," IEEE Open J. Commun. Soc. **2**, 1158–1170 (2021).

Optics EXPRESS

6. J. J. Sun, M. P. Chang, and P. R. Prucnal, "Demonstration of over-the-air RF self-interference cancellation using an optical system," IEEE Photonics Technol. Lett. **29**(4), 397–400 (2017).
7. Q. Zhou, H. Feng, G. Scott, and M. P. Fok, "Wideband co-site interference cancellation based on hybrid electrical and optical techniques," Opt. Lett. **39**(22), 6537–6540 (2014).
8. X. Han, B. Huo, Y. Shao, and M. Zhao, "Optical RF self-interference cancellation by using an integrated dual-parallel MZM," IEEE Photon. J. **9**(2), 1–8 (2017).
9. Y. Chen and S. Pan, "Simultaneous wideband radio-frequency self-interference cancellation and frequency downconversion for in-band full-duplex radio-over-fiber systems," Opt. Lett. **43**(13), 3124–3127 (2018).
10. Z. Zhu, C. Gao, S. Zhao, T. Zhou, G. Wang, H. Li, and Q. Tan, "Photonics-assisted ultrawideband RF self-interference cancellation with signal of interest recovery and fiber transmission," J. Lightwave Technol. **40**(3), 655–663 (2021).
11. W. Zhou, P. Xiang, Z. Niu, M. Wang, and S. Pan, "Wideband optical multipath interference cancellation based on a dispersive element," IEEE Photonics Technol. Lett. **28**(8), 849–851 (2016).
12. X. Su, X. Han, S. Fu, S. Wang, C. Li, Q. Tan, G. Zhu, C. Wang, Z. Wu, Y. Gu, and M. Zhao, "Optical multipath RF self-interference cancellation based on phase modulation for full-duplex communication," IEEE Photonics J. **12**(4), 1–14 (2020).
13. Y. Zhang, S. Xiao, H. Feng, L. Zhang, Z. Zhou, and W. Hu, "Self-interference cancellation using dual-drive Mach-Zehnder modulator for in-band full-duplex radio-over-fiber system," Opt. Express **23**(26), 33205–33213 (2015).
14. Y. Xiang, G. Li, and S. Pan, "Ultrawideband optical cancellation of RF interference with phase change," Opt. Express **25**(18), 21259–21264 (2017).
15. S. Zhang, S. Xiao, Y. Zhang, H. Feng, L. Zhang, and Z. Zhou, "Directly modulated laser-based optical radio frequency self-interference cancellation system," Opt. Eng. **55**(2), 026116 (2016).
16. Y. Zhang, S. Xiao, Y. Yu, C. Chen, M. Bi, L. Liu, L. Zhang, and W. Hu, "Experimental study of wideband in-band full-duplex communication based on optical self-interference cancellation," Opt. Express **24**(26), 30139–30148 (2016).
17. J. Wang, Y. Wang, Z. Zhang, Z. Zhao, and J. Liu, "Optical self-interference cancellation with frequency downconversion based on cascade modulator," IEEE Photonics J. **12**(6), 1–12 (2020).
18. P. Li, L. Yan, J. Ye, X. Feng, X. Zou, B. Luo, W. Pan, T. Zhou, and Z. Chen, "Photonic-assisted leakage cancellation for wideband frequency modulation continuous-wave radar transceiver," J. Lightwave Technol. **38**(6), 1178–1183 (2020).
19. L. Zheng, Y. Zhang, S. Xiao, L. Huang, J. Fang, and W. Hu, "Adaptive optical self-interference cancellation for in-band full-duplex systems using regular triangle algorithm," Opt. Express **27**(4), 4116–4125 (2019).
20. M. P. Chang, C.-L. Lee, B. Wu, and P. R. Prucnal, "Adaptive optical self-interference cancellation using a semiconductor optical amplifier," IEEE Photonics Technol. Lett. **27**(9), 1018–1021 (2015).
21. M. P. Chang, E. C. Blow, J. J. Sun, M. Z. Lu, and P. R. Prucnal, "Integrated microwave photonic circuit for self-interference cancellation," IEEE Trans. Microwave Theory Tech. **65**(11), 4493–4501 (2017).
22. L. Huang, Y. Zhang, S. Xiao, L. Zheng, and W. Hu, "Real-time adaptive optical self-interference cancellation system for in-band full-duplex transmission," Opt. Commun. **437**, 259–263 (2019).
23. X. P. Hu, D. Zhu, L. Li, and S. L. Pan, "Photonics-based adaptive RF self-interference cancellation and frequency downconversion," J. Lightwave Technol. **40**(7), 1989–1999 (2022).
24. L. Zheng, Z. Liu, S. Xiao, M. P. Fok, Z. Zhang, and W. Hu, "Hybrid wideband multipath self-interference cancellation with an LMS pre-adaptive filter for in-band full-duplex OFDM signal transmission," Opt. Lett. **45**(23), 6382–6385 (2020).
25. L. Zheng, S. Xiao, Z. Liu, M. P. Fok, J. Fang, H. Yang, M. Lu, Z. Zhang, and W. Hu, "Adaptive over-the-air RF self-interference cancellation using a signal-of-interest driven regular triangle algorithm," Opt. Lett. **45**(5), 1264–1267 (2020).
26. Z. Zhang, L. Zheng, S. Xiao, Z. Liu, J. Fang, and W. Hu, "Real-time IBFD transmission system based on adaptive optical self-interference cancellation using the hybrid criteria regular triangle algorithm," Opt. Lett. **46**(5), 1069–1072 (2021).
27. S. Ravuri, K. Lenc, and M. Willson, *et al.*, "Skilful precipitation nowcasting using deep generative models of radar," Nature **597**(7878), 672–677 (2021).
28. K. Tunyasuvunakool, J. Adler, and Z. Wu, *et al.*, "Highly accurate protein structure prediction for the human proteome," Nature **596**(7873), 590–596 (2021).
29. G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," Nature **588**(7836), 39–47 (2020).
30. C. M. Valensise, A. Giuseppi, G. Cerullo, and D. Polli, "Deep reinforcement learning control of white-light continuum generation," Optica **8**(2), 239–242 (2021).
31. Q. Q. Yan, Q. H. Deng, J. Zhang, Y. Zhu, K. Yin, T. Li, D. Wu, and T. Jiang, "Low-latency deep-reinforcement learning algorithm for ultrafast fiber lasers," Photonics. Res. **9**(8), 1493–1501 (2021).
32. Y. Han, S. Xiang, Y. Wang, Y. Ma, B. Wang, A. Wen, and Y. Hao, "Generation of multi-channel chaotic signals with time delay signature concealment and ultrafast photonic decision making based on a globally-coupled semiconductor laser network," Photonics. Res. **8**(11), 1792–1799 (2020).
33. Y. Li, "Deep reinforcement learning: an overview," arXiv, arXiv:1701.07274 (2017).

Optics EXPRESS

34. J. Ma, Z. Piao, S. Huang, X. Duan, G. Qin, L. Zhou, and Y. Xu, "Monte Carlo simulation fused with target distribution modeling via deep reinforcement learning for automatic high-efficiency photon distribution estimation," Photonics. Res. **9**(3), B45–B56 (2021).
35. L. Yan, X. M. Fang, X. B. Wang, and B. Ai, "AI-Enabled Sub-6-GHz and mm-Wave Hybrid Communications: Considerations for Use With Future HSR Wireless Systems," IEEE Veh. Technol. Mag. **15**(3), 59–67 (2020).
36. Q. Zhou, Y. W. Chen, S. Y. Shen, Y. M. Kong, M. Xu, J. W. Zhang, and G. K. Chang, "Proactive real-time interference avoidance in a 5 G millimeter-wave over fiber mobile fronthaul using SARSA reinforcement learning," Opt. Lett. **44**(17), 4347–4350 (2019).
37. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," Nature **518**(7540), 529–533 (2015).
38. T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv, arXiv:1509.02971 (2015).
39. X. Yu, J. Ye, L. S. Yan, T. Zhou, X. H. Zou, and W. Pan, "Photonic-assisted multipath self-interference cancellation for wideband MIMO radio-over-fiber transmission," J. Lightwave Technol. **40**(2), 462–469 (2022).
40. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 2018).