

# stray light suppression of opto-mechanical system based on deep reinforcement learning

ZIYANG ZHANG,<sup>1</sup> YANGGUANG XING,<sup>1\*</sup> YIFAN HUANG,<sup>1\*</sup> JUN CHANG,<sup>1</sup>  
ZEYU WU,<sup>1</sup> ZIWEN DUAN,<sup>1</sup> AND JIACHI SONG<sup>2</sup>

<sup>1</sup> School of Optics and Photonics, Beijing Institute of Technology, Beijing, Beijing, China;

<sup>2</sup> Tianjin Port Pacific International Container Terminal co., LTD, Tianjin, Tianjin, China

\*yuewanping1999@163.com

**Abstract:** Stray light suppression constitutes a vital aspect in the development of opto-mechanical systems, but its complexity and the uncertainty surrounding scattered light necessitate intricate mathematical calculations and ample simulation iterations, along with extensive expertise and time. Consequently, researching stray light suppression in opto-mechanical systems becomes a time-consuming and challenging endeavor. To validate the feasibility of using reinforcement learning for stray light suppression, this study adopts a model-based deep reinforcement learning approach within a Monte Carlo ray-tracing environment to devise suppression strategies. The experimental results demonstrate that the model-based deep reinforcement learning method can propose effective stray light suppression measures tailored to various optical system configurations, resulting in significant improvements in suppression efficiency.

## 1. Introduction

Over the past two decades, the importance of Machine Learning (ML) and Data Science in engineering applications and scientific research has grown dramatically due to continuous advancements. It has been widely used in a variety of fields from biology<sup>[1]</sup>, material science<sup>[2]</sup>, and astronomy<sup>[3,4]</sup> to sociology<sup>[5]</sup>. In recent years, optical research has adopted machine learning techniques for various applications, including enhancing the performance of optical microscopes using deep neural networks<sup>[6]</sup>, implementation of a backpropagation algorithm on photonic neural networks using deep learning<sup>[7]</sup>, using deep learning for the design of photonic structures<sup>[8]</sup> from passive optimization to reverse creation of nano-optical designs using deep learning<sup>[9]</sup>, implementing lensless computational imaging<sup>[10,11]</sup> or computational spectral imaging<sup>[12]</sup> with deep learning and establishing design frameworks for freeform imaging systems using reinforcement learning<sup>[13–15]</sup>. This shows that machine learning is constantly being combined with optical systems and thus promotes the development of optical research.

Stray light suppression research is a crucial part of the development process of an opto-mechanical system, which determines whether an opto-mechanical system can function according to its intended operating results. The presence of stray light can degrade image contrast and signal-to-noise ratio, and in severe cases, signals could be obliterated by stray light. In this paper, we try to propose a new method to solve the problem of stray light suppression in the opto-mechanical system through the combination of machine learning and optical machine system.

The formulation of traditional non-automated stray light suppression schemes is done by the designer based on mathematical formulae, experience and simulation results to find a better stray light suppression effect through manual iteration. Lionel Clermont et al. proposed a stray light control and analysis methods in an off-axis three-mirror anastigmat (TMA) telescope.

The first-order scattered stray light from non-optical surfaces was controlled, and the direct stray light was blocked through usage of elements such as apertures and baffles, both internal and external to the TMA telescope<sup>[16]</sup>. Huang et al. proposed a stray light analysis method and some suppression principles for panoramic annular lens (PAL) by finding stray light paths. In that optical system, stray light caused by light splitting on the two refractive index surfaces of the PAL block and then cutting them off, and the stray light suppression method of reducing scattering, diffraction, and other stray light will be reduced from the optical design stage<sup>[17]</sup>. Song et al. proposed an optimization method for the baffle design of an axial two-mirror telescope. The method of baffle design overcomes the shortcomings of the graphing method, and it can be finished simultaneously with the optical design to obtain the optimization configuration of the telescope<sup>[18]</sup>. Hu et al. proposed a stray light suppression aim to Cassegrain optical structure, which consists of a honeycomb-structured ultrashort outer baffle. The baffle is designed by a constraint formula based on the characteristics and the geometrical design of the primary and secondary baffles considering the bumps, which ensures the same stray light suppression as the conventional baffle while greatly reducing the size.<sup>[19]</sup> Sun et al.'s Ritchey-Chretien optical system uses a built-in baffle, a stray light suppression measure that greatly reduces the length of the outer baffle while maintaining the original stray light suppression<sup>[20]</sup>. Similar stray light suppression is also mostly used in star tracker<sup>[21–24]</sup>. These stray light suppression schemes are all determined through the derivation of mathematical formulas or the combination of experience and simulation results, such as the position of baffles, size of the baffles, the baffles' aperture and length of the light shields.

With the constant progression in the sensitivity and threshold of photoelectric detectors, coupled with ongoing advancements in optical research, the demand for enhanced precision and threshold for stray light suppression and assessment in space optical-mechanical systems is on the rise. A new, more effective strategy is needed for stray light suppression that considers surface attributes and the entire opto-mechanical system.

Despite Machine Learning (ML) being successfully integrated into various fields of optical research, the application of optical system design analysis and ML still faces challenges due to the intricate physical equations in optics and the random, non-linear nature of scattered light in ray tracing. If a suitable ML model can be used to complete the design analysis of the opto-mechanical system, it can simplify the difficulty of the design of the opto-mechanical system and find a better solution by ML.

Reinforcement learning, a subset of machine learning, is a widely employed control method in fields such as autonomous driving. It aims to solve the decision-making process in interactions between intelligent agents and their environments. The approach maximizes the cumulative rewards obtained from the environment by the agent until an optimal strategy to accomplish the set goals is found. Under the influence of reinforcement learning, the agent continuously interacts with the environment, learning its characteristics and leveraging neural network structures to optimize control strategies. When applied to stray light suppression, the agent iteratively interacts with the opto-mechanical structure, learns the system's stray light properties, and consequently, suggests appropriate suppression schemes. This process automates the design of stray light suppression strategies and eliminates the need for multiple manual iterations, thereby identifying the optimal suppression approach. Accordingly, we propose a model-based reinforcement learning control method capable of generating effective stray light suppression strategies for various opto-mechanical structures. We have designed a physical environment for ray tracing simulation based on the Monte Carlo method to facilitate interaction between the agent and the environment and generate rewards that signify the level of stray light suppression.

By translating the opto-mechanical structure into a mathematical model and assigning respective surface properties based on the bidirectional scattering distribution function (BSDF),

we can perform ray tracing according to the decisions. The point source transmittance (PST) acquired from the environment after the agent's decision accurately characterizes the effectiveness of stray light suppression. The experimental results show that the agent can propose an effective suppression strategy when the optical mechanical structure and stray light suppression requirements are known, and the environment can accurately represent the effect of stray light suppression.

This method introduces a new research perspective on stray light suppression, enhances the efficiency of design, lowers the threshold for designing suppression strategies, and provides a robust initial structure for devising high-precision suppression schemes.

## 2. Page layout and length

The research principle of stray light suppression based on deep reinforcement learning is shown in Fig.1. The model is divided into two parts. One part is the environment that interacts with the agent. The environment undertakes the function of converting the decision proposed by the agent into an indicator of the stray light suppression effect as a reward feedback to the agent. The other part is the agent according to the state of the environment at this time  $S_t$ , optimize the stray light suppression scheme. Under  $S_t$  and the guidance of strategy  $\pi(s)$ , the agent selects different stray light suppression schemes as action  $a_t$ , and interacts the scheme with the environment to generate new internal state and feedback of stray light suppression effect. Then, the agent updates the state and the reward  $r_t$  representing the performance index with  $S_{t+1}$ .

In this chapter, we first introduce the PST of the reward source of the model and the formulation of the traditional stray light suppression scheme. Then we elaborate the principle of deep reinforcement learning (RL) composed of environment, RL agent and model, and finally complete the construction of the environment.

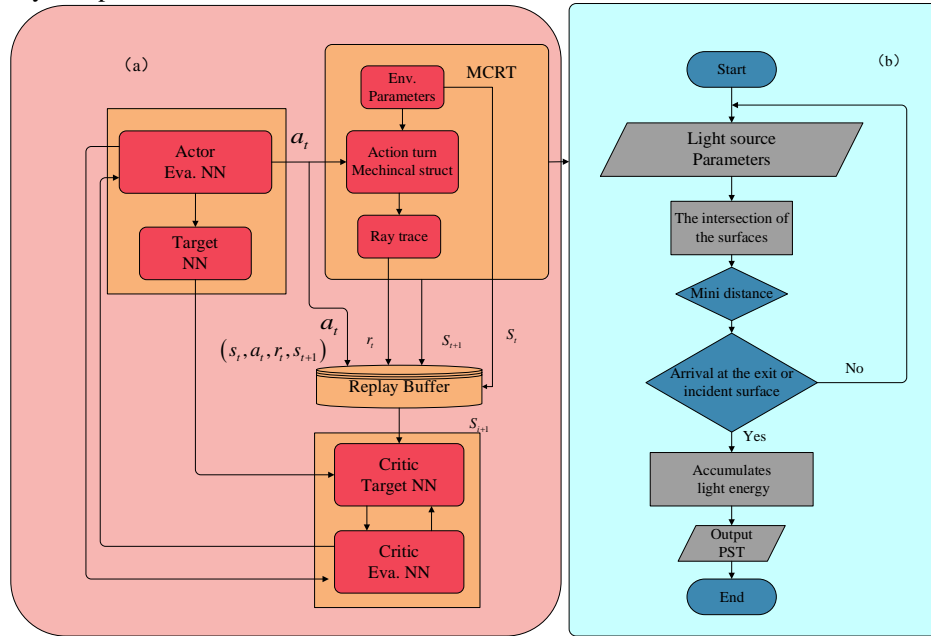


Fig. 1. (a) RL agent, (b) Environment based on Monte Carlo ray tracing

## 2.1 Stray light analysis process

PST often used to evaluate the effect of stray light suppression. PST refers to the ratio of the irradiance received by the detector to the irradiance at the entrance of the optical system at a specific off-axis angle.

$$PST(\theta) = \frac{E_d(\theta)}{E_i(\theta)} \quad (1)$$

$E_d(\theta)$  is the irradiance on the image surface when the incident angle is  $\theta$ , and  $E_i(\theta)$  is the irradiance on the object surface when the incident angle is  $\theta$ . It reflects the suppression effect of the optomechanical system on the stray light at a specified angle  $\theta$ . Although the PST cannot characterize the stray light suppression effect of all incident angles except the field of view, it can also characterize the stray light suppression effect of the optomechanical system at different angles through multiple measurements from multiple angles. Therefore, in this paper, the function curve composed of PST of multiple angles is used to characterize the stray light suppression effect of an optomechanical system.

Taking the PST as the stray light analysis index, the analysis process of stray light is shown in the following figure Fig.2. After determining the distribution of system noise suppression index and the characteristics of stray light source, the basic scheme of stray light suppression is formulated, and the corresponding optomechanical structure design is carried out. Then, according to different scattering surfaces such as mirrors, coatings, etc. The surface scattering properties are represented by the BSDF, and the scattering database of the system is established. According to the scattering database, the corresponding stray light analysis model is established based on the surface properties of each surface and the opto-mechanical structure. Finally, the simulation of stray light suppression effect is carried out. In the design stage, the stray light suppression level of the optomechanical system is simulated to guide the optimization iteration of the next suppression scheme, and the suppression measures are improved to avoid the defects that cannot be saved and have a huge impact on the imaging results in the physical stage.

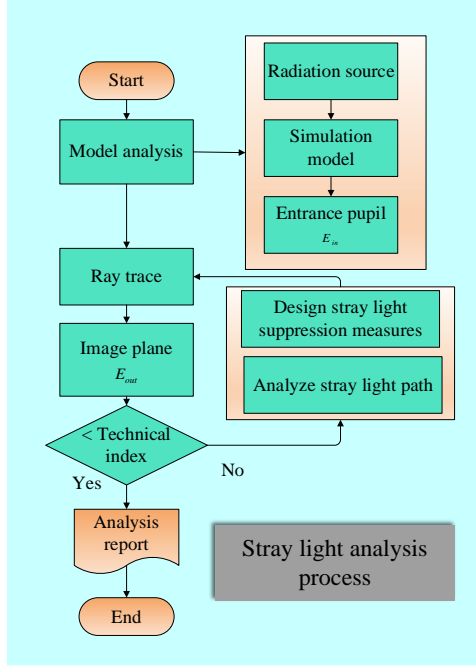


Fig. 2. The process of traditional stray light analysis and suppression

## 2.2 The network and model basis of RL

As shown in Figure 1(a), the RL model used in this paper controls the generation of stray light suppression schemes through the 'actor-critic' architecture.<sup>[25]</sup> The deep reinforcement learning method uses a pair of neural networks, actor networks and critic networks with different goals.

The actor network optimization strategy  $\pi(s)$  determines the probability distribution of state-to-action mapping. The critic's network optimization target state-action value function  $Q(S, a)$  represents the cumulative discount reward of the state-behavior to the value function  $Q(S, a)$  following the current strategy  $\pi(s)$ .

$$Q^\pi(s_t, a_t) = E_{r_t \geq t, s_t \sim \pi} [R_t | s_t, a_t] \quad (2)$$

When the target strategy is deterministic, it is described as a function  $Q^\pi$ . The mapping from action to state avoids falling into internal expectations. The optimal action value function is expressed according to the Bellman equation as follows :

$$Q^\pi(s_t, a_t) = E_{r_t, s_{t+1} \sim E} [r(s_t, a_t) + \gamma E_{a_{t+1} \sim \pi} [Q^\pi(s_{t+1}, a_{t+1})]] \quad (3)$$

Since the reward only depends on the environment, this means that the Q value can be learned through different random behavior strategies. The model uses the common Q-learning off-policy algorithm, which uses a greedy strategy  $\mu(s) = \arg \max_a Q(s, a)$ . The algorithm uses the function approximator of  $\theta^Q$  to optimize the algorithm through minimum loss.

$$L(\theta^Q) = E_{S_t \sim \rho^\beta, a_t \sim \beta, r_t \sim E} \left[ \left( Q(s_t, a_t | \theta^Q) - y_t \right)^2 \right] \quad (4)$$

where  $y_t$  is

$$y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{t+1}) | \theta^Q) \quad (5)$$

The critic network iterates by using the Bellman equation. The actor network consists of three fully connected layers. The neurons in the input and output layers are associated with the dimensions of state parameters and action parameters. The neurons in the hidden layer are a hyperparameter. There is a rectified linear unit ( ReLU ) and a hyperbolic tangent function as the activation function behind the input layer and the hidden layer, respectively.

The critic neural network, which consists of three fully connected layers. The number of neurons in the input layer is the sum of the dimensions of  $S_t$  and  $a_t$ , and the number of neurons in the output layer is 1. As with actor neural networks, the number of neurons in the hidden layer is a hyperparameter. All layers except the last one are followed by a ReLU function as an activation function.

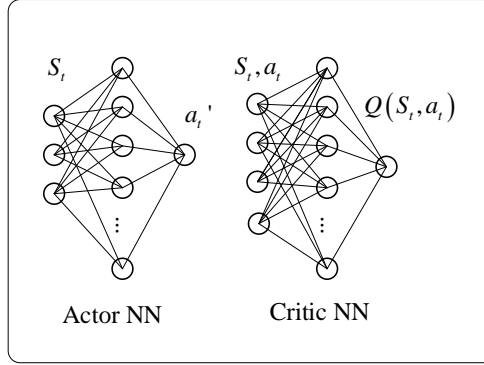


Fig. 3 neural network structure of RL agent.

In the model-based RL training process proposed in this paper, each training round starts from a random state and ends after obtaining the corresponding reward. In the algorithm, the same replay buffer as the standard RL is introduced. At the beginning of training, the agent takes different actions according to the random strategy, and transmits the learned experience back to the replay buffer. We summarize the training process of the model-based deep reinforcement learning algorithm as algorithm 1.

**Table 1 RL algorithm**

Algorithm 1 Research model of stray light suppression based on RL
preliminary environment $E_m$
Initialize the replay buffer $R_b$
Randomly initialize network parameters: $\theta^\mu, \theta^Q, \theta^{\mu'}, \theta^{Q'}$
for episode $\leftarrow 1$ to N
Receive the initial state $S_t$ from the environment $E_m$
Initialize various parameters of stray light suppression measures $\phi(1)$
for t $\leftarrow 1$ to M

---

```

select the action  $a_t = \mu(s_t | \theta^\mu)$ 

Update the action with action noise  $N_a : a_t \leftarrow a_t + N_a$ 

Update stray light suppression measures  $\phi(t+1) \leftarrow a_t$ 

Calculate rewards  $r_t$ 

Get the next state from the environment  $S_{t+1}$ 

Store the experience  $(s_t, a_t, r_t, s_{t+1})$  in  $R_b$ 

If  $t \geq t_s$ 

If  $r \geq r_o$ 

Update action noise  $N_a$ 

End

Randomly select minibatch of experience  $(s_t, a_t, r_t, s_{t+1})$  from the replay buffer.

Compute

$$y_t = r_t + \gamma_r \left[ Q'(s_{t+1}, \mu'(S_{t+1} | \theta^\mu) | \theta^Q) \right]$$


Update the critic evaluation NN by minimize

$$L(\theta^Q) = E \left[ \left( Q(s_t, a_t | \theta^Q) - y_t \right)^2 \right]$$


Update the actor evaluation NN by minimize

$$-\nabla_{\theta^\mu} J = -E \left[ \nabla_{\theta^\mu} Q(s, a | \theta^Q) | s = s_t, a = \mu(s_t | \theta^\mu) \right]$$


Update the target NNs by

$$\theta^{\mu'} \leftarrow k\theta^\mu + (1-k)\theta^{\mu'}$$


$$\theta^{Q'} \leftarrow k\theta^Q + (1-k)\theta^{Q'}$$


end

Updates the corrected state  $s_t \leftarrow s_{t+1}$ 

End

End

```

---

### 2.3 Ray tracing based on Monte Carlo method

The radiative transfer equation is a multivariate integral differential equation involving multiple dimensions. For stray light analysis, in addition to obtaining the correct light path in the opto-mechanical system, it is also necessary to obtain the numerical solution of radiation. The Monte Carlo Ray Tracing (MCRT) method is a statistical method. The process of MCRT method calculating ray is to track and record the emission position, direction, and surface medium information of light beams with sampling significance, and then calculate the radiative transfer factor to obtain new information. MCRT can handle multi-dimensional complex geometry, anisotropic scattering, and other issues at the same time. Based on the MCRT method, the radiative transfer equation can be easily solved.<sup>[26,27]</sup>

This paper adopts the path length method (PL) in the MCRT method, which tracks the forward route of light by judging the direction of ray, finding the minimum distance for ray to reach all surfaces, determining the surface that the ray reaches, and outputting the position, direction, surface medium, energy, and other information carried by the ray when it reaches the surface. Through this method, this paper realizes the non-sequential tracking of ray, in order to achieve

the purpose of converting the stray light suppression scheme proposed by the intelligent agent into the stray light suppression index.

When ray is irradiated on a surface with a certain roughness, due to the uneven distribution of the microscopic morphology of the surface, a part of the incident light is absorbed, and most of it undergoes scattering and reflection phenomena. In conventional MCRT tracing, the scattered light caused by factors such as surface roughness, scratches, pockmarks, and coatings is often ignored. This part of the light is the main component of the source of stray light. This paper applies MCRT to stray light analysis. As the main component of the source of stray light, scattered light must be emphasized. In the process of MCRT tracing, this paper introduces the BSDF that expresses the light scattering characteristics of the object surface. In the face of the BSDF, the measured BSDF data is usually used to accurately characterize the information of the surface.<sup>[28]</sup>

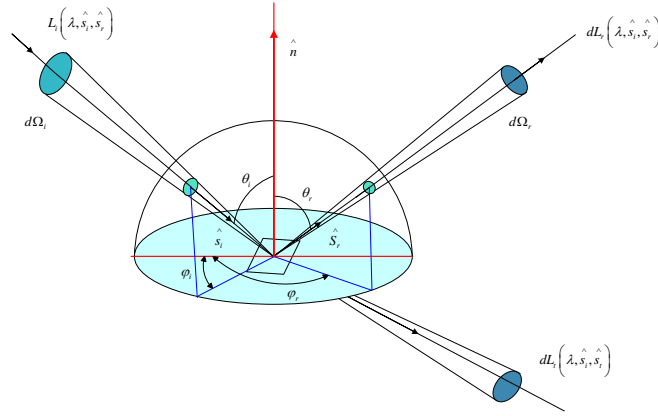


Fig. 4. BSDF model

Considering the statistical error and time cost of MCRT due to the need for a large number of rays to offset the uncertainty in the process of light propagation during MCRT simulation, this paper uses the ABg mathematical model to fit the surface measured BSDF data. The expression is shown as follows.

$$BRDF(\vec{\beta} - \vec{\beta}_0) = \frac{A}{B + |\vec{\beta} - \vec{\beta}_0|^g} \quad (6)$$

## 2.4 RL definition

Based on the proposed model, as shown in Figure 1(a), with state  $S_t$ , action  $a_t$ , and reward  $r_t$  defined as experience, the stray light suppression effect is transformed into an indicator that reinforcement learning can observe.

The action given by the model,  $a_t$ , is a stray light suppression measure, usually achieved by controlling the length of the two levels of the light shield, the number of light-blocking rings, the central coordinate position, and the size of the light aperture. Due to the nonlinear relationship between the PST at different angles and the stray light suppression measures, it is necessary to also use the PST as information to describe the state, so that the agent can more clearly understand the relationship between the PST and the stray light suppression measures. Therefore, the final state  $S_t$  description is a combination of the stray light suppression measures proposed by the intelligent agent, the PST and the reward. The parameters included in its actions and states are shown in the table below.

Table 2 actor's parameters and observation



actor's parameters		observation	
number	parameters	number	parameters
1	Length of external hood	1	Length of external hood
2	Length of inner hood	2	Length of inner hood
3	Caliber of outer baffle	3	Caliber of outer baffle
4	Diameter of inner baffle	4	Diameter of inner baffle
5	ring z coordinate	5	ring z coordinate
6	ring y coordinate	6	ring y coordinate
		7	ring light-passing aperture
		8	Number of blocking rings
		9	PST
		10	reward

In order to allow the intelligent agent to choose better suppression measures and discard worse ones during the training process, and to consider the volume and weight issues in the design of the opto-mechanical structure when the PST tends to be stable, this paper defines the reward function as a combination of the PST at multiple angles and the evaluation of the opto-mechanical structure parameters. The reward function is as follows:

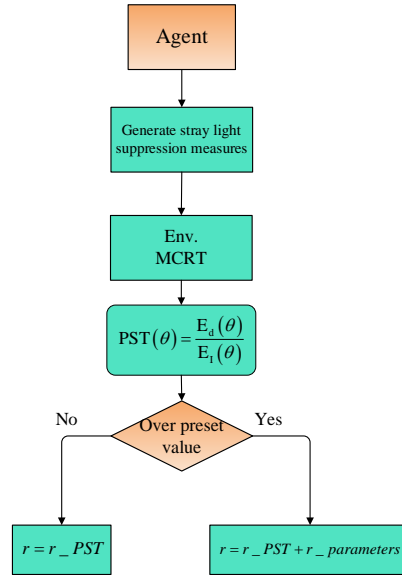


Fig. 5. The definition of reward function

When the stray light suppression effect is less than the preset indicator, it is considered that the current stray light suppression scheme is not good, making the reward smaller. As the number of episodes increases, the stray light suppression effect is greater than the preset indicator, at which point the stray light suppression effect gradually improves. When the stray light suppression effect does not change in order of magnitude, the parameters of the opto-mechanical structure are optimized to achieve better rewards.

### 3. Comparison of model training results based on different optical machine systems

Before the commencement of the training process, this chapter switches to the appropriate physical environment model based on the different optical structures. Through the input of opto-mechanical structure parameters, the reflective optical system can function within this model.

To verify the effectiveness of stray light strategies proposed through reinforcement learning, we utilize stray light analysis software to mimic the actual environment. By adhering to traditional stray light suppression analysis methods, we analyze the reflective optical system, obtaining the suppression effects under the traditional stray light scheme.

We compare these results with the suppression effects of the stray light suppression strategy proposed by the agent and applied in the stray light analysis software to validate the practicality of reinforcement learning for reflective opto-machine systems in optimizing stray light suppression schemes. Section 3.1 demonstrates the suppression effects of the traditional stray light scheme in the reflective opto-mechanical system, 3.2 highlights the agent's training results, and 3.3 contrasts the performance of both.

### 3.1 Traditional Stray Light Suppression Measures

The example of a reflective optical system used in this paper is the Gravitational Wave Optical Detection Telescope. The telescope consists of two secondary surfaces, one free-form surface, and one planar reflector to form an off-axis four-reflector optical structure with a field of view of  $\pm 0.0013^\circ$ . The optical structure system parameters of this gravitational wave optical detection telescope are shown in the table below.

**Table 3 Reflective Optical System**

surface	curvature	thickness	material	conic	Zernike 4	Zernike 5
quadratic aspherical surface	1298.5591	625.7814	Mirror	-1		
quadratic aspherical surface	-49.105	547.872	Mirror	-1.1871		
Zernike surface	Inf	-92.4756	Mirror		-0.0024	-0.0497
plane	Inf	270	Mirror			

**Table 4 reflective optical system**

num	surface	curvature	thickness	GLASS	num	surface	curvature	thickness	GLASS
1	sphere	66.266	2.004	SILICA	9	sphere	32.155	9.998	N-LAK33
2	sphere	66.176	2.915		10	sphere	-67.867	1.776	
3	sphere	45.299	6.017	N-LAK33	11	sphere	21.738	9.998	N-LAK33
4	sphere	381.555	8.993		12	sphere	16.053	2.869	
5	sphere	-50.574	2.001	ZF4	13	sphere	23.575	7.545	ZF4
6	sphere	-202.404	4.905		14	sphere	-43.099	1.127	
7	sphere	163.695	9.997	ZF4	15	sphere	-28.920	3.815	ZF4
8	sphere	29.254	1.039		16	sphere	40.259	5	

The optical system parameters in the table are inputted into the model constructed in this paper, and the optical path diagram after tracing through the model is as follows, which is consistent with the optical path diagram shown by the optical design software.

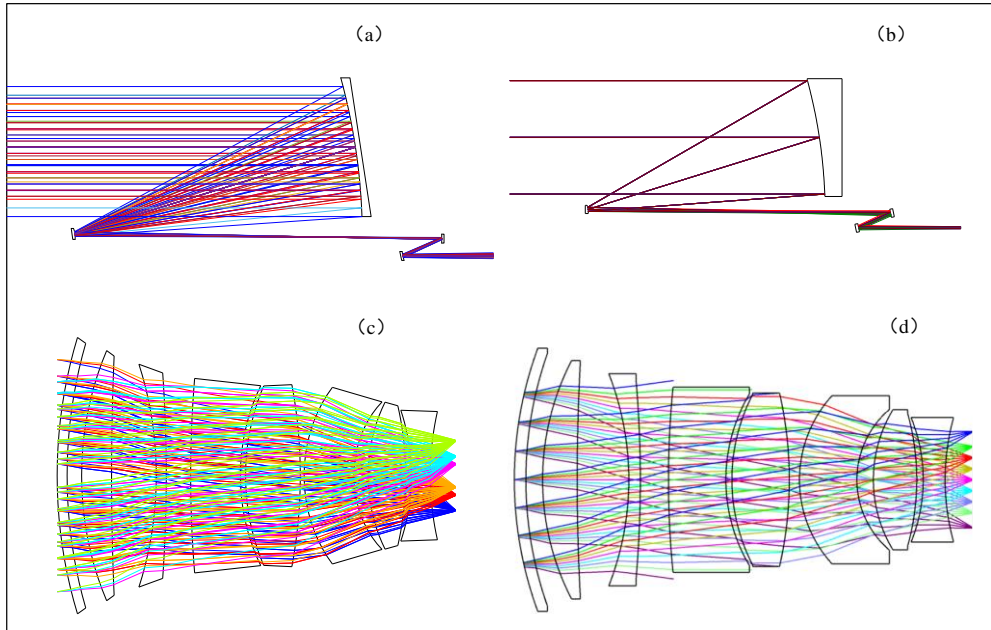


Fig. 6 . Gravitational Wave Telescope Optical Detection System(a) env. (b) Optical Design Software(c) Star sensor optical camera env. (d) Star sensor optical camera Optical Design Software

According to the automatic design of the light stop ring, the position between each light stop ring can be obtained. The Star sensor optical camera system is taken as an example to illustrate. The stray light suppression requirement of the star sensor optical camera is to ensure that the aperture and length are as small as possible when the light above the sun suppression angle is suppressed. It can be obtained that the total length of the star sensor optical camera is not more than 140 mm, and the aperture is not more than 120 mm. According to the stray light suppression index, in the case of knowing the maximum length and maximum diameter and the target solar suppression angle, according to the design method of the traditional baffle, we can get the optical machine structure of the optical camera of the star sensor. The GWOT optical system is also designed along the same lines, and the optical system with added stray light suppression measures can be obtained.

Subsequently, the optical surfaces and mechanical structures are simulated in the stray light analysis software using the measured mirror and black paint properties, respectively. The different optical machine systems' stray light suppression effects from various angles are illustrated in the following figure.

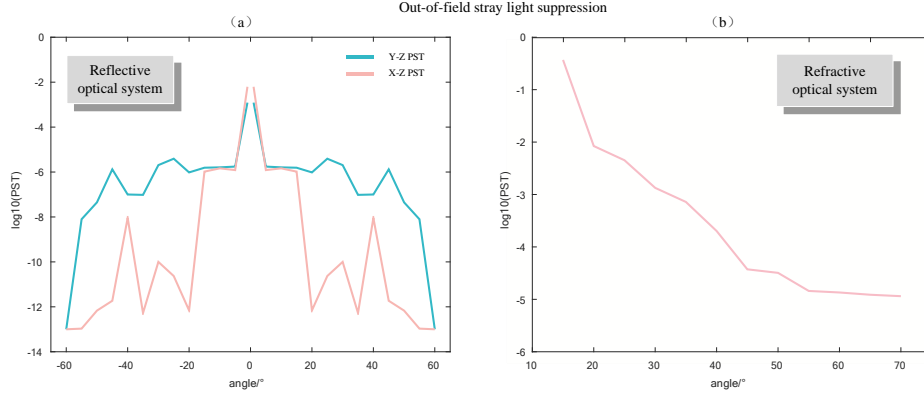


Fig. 7. (a)PST of Gravitational Wave Telescope Optical Detection System, GWTO is an opto-machine system with asymmetry in the y-z plane and symmetry in the x-z plane. (b) PST of Star sensor optical camera, Star sensor optical camera is a rotationally symmetrical optical system.

### 3.2 Model-based run results

In order to solve the problem of long calculation time caused by the number of rays, the size of the experience pool is appropriately reduced, and the learning ratio of the actor and the observer is adjusted. When the learning ratio is  $1e-2$ , although the action randomness is larger, it can better explore the space, but it may converge to the wrong result. When the learning ratio is  $1e-4$ , the speed of action exploration is slow. In the face of continuous space of mechanical structure and more operands, the convergence speed is too slow. Therefore, the learning ratio is  $1e-3$  in this experiment. The hyperparameters of this model are shown in the following table.

**Table 5 hyper parameter**

number	hyper parameter	value
1	Actor learning rate	0.001
2	Critics learning rate	0.001
3	DiscountFactor	0.98
4	ExperienceBufferLength	5000
5	MiniBatchSize	32
6	SampleTime	0.5
7	Actor GradientThreshold	2
8	Critics GradientThreshold	2

The outcomes of the reinforcement learning operation are displayed in the ensuing figure. In controlling stray light suppression measures with RL, a reward surpassing anticipated values indicates satisfactory stray light suppression from the current suppression measures. As the reward progressively stabilizes throughout training, it suggests convergence in the effectiveness of stray light mitigation plans proposed by the reinforcement learning. The relevant network structure parameters are then extracted to replicate the actor-proposed stray light suppression scheme. This scheme, when modeled as machine structural parameters and analyzed using stray light analysis software, forms the reinforcement learning's stray light suppression scheme for this optical system.

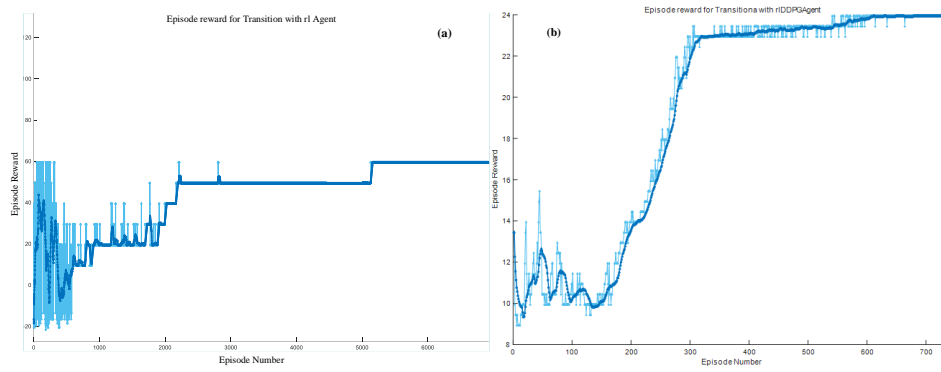


Fig. 8. Results of RL training on a different optomechanical system,(a) GWOT RL results,(b) Star sensor optical camera RL results

At this time, the stray light suppression scheme proposed after the reflective optical structure training is shown in the figure.

Assign the corresponding surface properties to the established optical-mechanical model and import it into the stray light analysis software for simulation. The stray light suppression effect of each angle is shown in the table below.

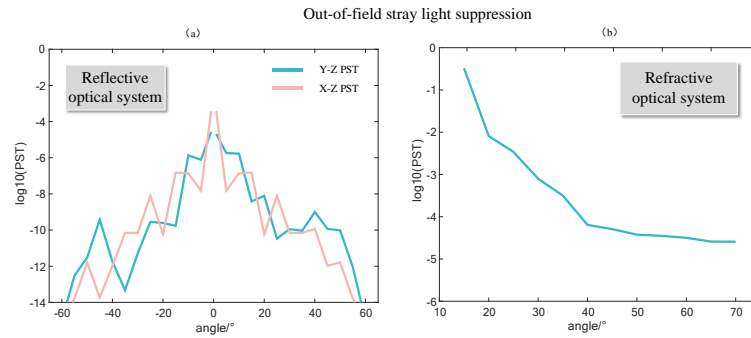


Fig. 9. (a)PST of Gravitational Wave Telescope Optical Detection System, GWTO is an opto-machine system with asymmetry in the y-z plane and symmetry in the x-z plane. (b) PST of Star sensor optical camera, Star sensor optical camera is a rotationally symmetrical optical system.

### 3.3 Comparison of stray light suppression measures between the two design methods

In stray light suppression analysis, this paper simulates the stray light suppression scheme obtained by the model-based RL and the traditional method. According to the results of the simulation analysis, the different stray light suppression situations exhibited by the two methods under different angles are obtained. It can be seen that considering the stray light impact caused by scattering, the stray light suppression effect of the RL method is better than the traditional stray light suppression scheme. After using the RL method, the stray light suppression effect at each angle has been significantly improved, and it is not limited by different optical structures, and it shows good suppression effect in reflective optical systems. The suppression effects presented by the two optical systems under different suppression schemes are shown in the following Fig.12. The use of reinforcement learning (RL) methods has significantly improved the suppression effect of stray light at all angles. This method demonstrates effective suppression in reflective optical systems, off-axis optical systems, and non-spherical optical systems.

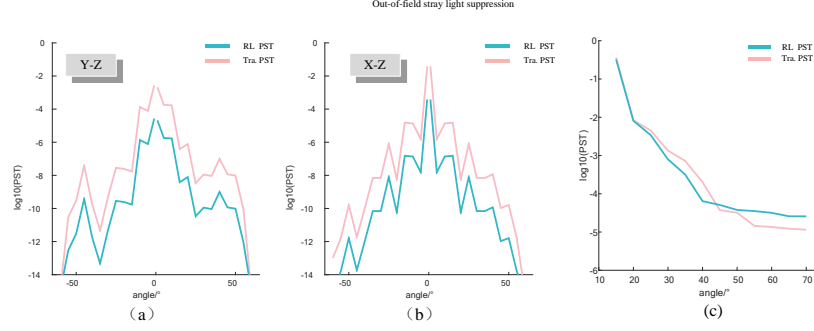


Fig. 7 The comparison of the traditional stray light suppression scheme and the RL stray light suppression scheme in different planes, (a) comparison PST of Gravitational Wave Telescope Optical Detection System in y-z plane, (b) comparison PST of Gravitational Wave Telescope Optical Detection System in x-z plane, (c) comparison PST of Star sensor optical camera System.

Therefore, based on the results presented, this paper posits that model-based deep reinforcement learning (RL) can be effectively applied to the stray light analysis and suppression process in different optical systems. The RL approach, compared to traditional methods that primarily consider reflection factors, shows a notable superiority in stray light suppression. From a reinforcement learning perspective, the agent successfully identifies the stray light features of the different optical system through its interaction with the environment, thus adopting a suitable approach that enhances stray light suppression. Utilizing model-based deep reinforcement learning for stray light analysis in optical systems is indeed feasible.

#### 4. Conclusion

In summary, this study has adopted model-based reinforcement learning to investigate the optimization of stray light suppression in universally applicable optical systems. To ensure the successful integration of stray light suppression scheme optimization with reinforcement learning, we utilized Monte Carlo ray tracing and ABg's BRDF model to establish an environment similar to the results of stray light analysis software. This environmental construction accelerated the episode speed of reinforcement learning, provided more direct feedback on stray light suppression effects, and improved network optimization efficiency.

The agent, throughout the model training process, has learned the characteristics of stray light in the optical-mechanical system and proposed corresponding stray light suppression measures that meet the expected stray light suppression effects while minimizing the weight and volume of the optical-mechanical structure. In the third and fourth sections of this paper, deep reinforcement learning was employed to analyze off-axis four-mirror structures and transmissive optical structures, respectively, and corresponding stray light suppression schemes were proposed. The stray light suppression schemes for the two different optical structures both demonstrated good suppression effects, validating the applicability of the model-based reinforcement learning stray light suppression study designed in this paper for refractive and reflective optical systems; co-axial and off-axis optical systems. The research results of this paper indicate that reinforcement learning, as a branch of machine learning, has the ability to propose effective stray light suppression schemes for different optical-mechanical structures. This method provides a new perspective for studying stray light suppression, achieves automated design of stray light suppression schemes, enhances the efficiency of stray light suppression design, lowers the threshold for formulating stray light suppression schemes, and offers a good initial structure for formulating higher-precision stray light suppression schemes.

## 5. References

1. M. Lapeyrolerie, M. S. Chapman, K. E. A. Norman, and C. Boettiger, "Deep reinforcement learning for conservation decisions," *Methods Ecol. Evol.* **13**(11), 2649–2662 (2022).
2. X. Zheng, X. Zhang, T. Chen, and I. Watanabe, "Deep Learning in Mechanical Metamaterials: From Prediction and Generation to Inverse Design," *Adv. Mater.* 2302530 (2023).
3. A. Szenicer, D. F. Fouhey, A. Munoz-Jaramillo, P. J. Wright, R. Thomas, R. Galvez, M. Jin, and M. C. M. Cheung, "A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance," *Sci. Adv.* **5**(10), eaaw6548 (2019).
4. D. B. Dhuri, S. Bhattacharjee, S. M. Hanasoge, and S. Kiran Mahapatra, "Deep-learning Reconstruction of Sunspot Vector Magnetic Fields for Forecasting Solar Storms," *Astrophys. J.* **939**(2), 64 (2022).
5. A. Heuillet, F. Couthouis, and N. Díaz-Rodríguez, "Explainability in deep reinforcement learning," *Knowl.-Based Syst.* **214**, 106685 (2021).
6. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437 (2017).
7. S. Pai, Z. Sun, T. W. Hughes, T. Park, B. Bartlett, I. A. D. Williamson, M. Minkov, M. Milanizadeh, N. Abebe, F. Morichetti, A. Melloni, S. Fan, O. Solgaard, and D. A. B. Miller, "Experimentally realized in situ backpropagation for deep learning in photonic neural networks," *Science* **380**(6643), 398–404 (2023).
8. W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, "Deep learning for the design of photonic structures," *Nat. Photonics* **15**(2), 77–90 (2021).
9. N. Wang, W. Yan, Y. Qu, S. Ma, S. Z. Li, and M. Qiu, "Intelligent designs in nanophotonics: from optimization towards inverse creation," *PhotonX* **2**(1), 22 (2021).
10. Z. Tian, Z. Ming, A. Qi, F. Li, X. Yu, and Y. Song, "Lensless computational imaging with a hybrid framework of holographic propagation and deep learning," *Opt. Lett.* **47**(17), 4283 (2022).
11. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* **6**(8), 921 (2019).
12. L. Huang, R. Luo, X. Liu, and X. Hao, "Spectral imaging with deep learning," *Light Sci. Appl.* **11**(1), 61 (2022).
13. T. Yang, D. Cheng, and Y. Wang, "Direct generation of starting points for freeform off-axis three-mirror imaging system design using neural network based deep-learning," *Opt. Express* **27**(12), 17228 (2019).
14. T. Yang, D. Cheng, and Y. Wang, "Designing freeform imaging systems based on reinforcement learning," *Opt. Express* **28**(20), 30309 (2020).
15. W. Chen, T. Yang, D. Cheng, and Y. Wang, "Generating starting points for designing freeform imaging optical systems based on deep learning," *Opt. Express* **29**(17), 27845 (2021).
16. L. Clermont and L. Aballea, "Stray light control and analysis for an off-axis three-mirror anastigmat telescope," *Opt. Eng.* **60**(05), (2021).
17. Z. Huang, J. Bai, T. X. Lu, and X. Y. Hou, "Stray light analysis and suppression of panoramic annular lens," *Opt. Express* **21**(9), 10810 (2013).
18. N. Song, "Baffles design for an axial two-mirror telescope," *Opt. Eng.* **41**(9), 2353 (2002).
19. X. H. Xiaodong Hu, W. W. Weike Wang, Q. H. Qiang Hu, X. L. Xing Lei, Q. W. Qing Wei, Y. L. Yuanzheng Liu, and J. W. Jiliang Wang, "Design of CASSEGRAIN telescope baffles with honeycomb entrance," *Chin. Opt. Lett.* **12**(7), 072901–072904 (2014).
20. L. Sun, Q. Cui, N. Xie, and J. Wang, "Design of a built-in baffle for a Ritchey–Chretien optical system," *Appl. Opt.* **57**(35), 10264 (2018).
21. T. Sun, F. Xing, J. Bao, S. Ji, and J. Li, "Suppression of stray light based on energy information mining," *Appl. Opt.* **57**(31), 9239 (2018).
22. S. Lee, R. Saleem, and S.-S. Lee, "Micro star tracker with a curved vane for a short baffle length and sharp stray light attenuation," *Appl. Opt.* **59**(13), 4131 (2020).
23. Y. Meng, X. Zhong, Y. Liu, K. Zhang, and C. Ma, "Optical system of a micro-nano high-precision star sensor based on combined stray light suppression technology," *Appl. Opt.* **60**(3), 697 (2021).
24. G. Wang, F. Xing, M. Wei, and Z. You, "Rapid optimization method of the strong stray light elimination for extremely weak light signal detection," *Opt. Express* **25**(21), 26175 (2017).
25. 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," (2019).
26. H. Liu, H. Zhou, D. Wang, and Y. Han, "Performance comparison of two monte carlo ray-tracing methods for calculating radiative heat transfer," *J. Quant. Spectrosc. Radiat. Transf.* **256**, 107305 (2020).
27. M. Yarahmadi, J. Robert Mahan, and K. J. Priestley, "Uncertainty Analysis and Experimental Design in the Monte Carlo Ray-Trace Environment," *J. Heat Transf.* **141**(3), 032701 (2019).
28. Z. Ma, H. Wang, Q. Chen, Y. Xue, Y. Pan, Y. Shen, and H. Yan, "Implementation of empirical modified generalized Harvey–Shack scatter model on smooth surface," *J. Opt. Soc. Am. B* **39**(7), 1730 (2022).