Optimizing and Controlling Nonlinear Metrics of Multichannel LED Systems With Ambient Light Interference via Reinforcement Learning

Qi Wang and Jianfei Dong

***Abstract*—Controlling multichannel LED systems is essential in many applications, ranging from stage lighting to multispectral imaging and light therapies. However, the increasing number of channels coupled in a compact package and housing leads to higher nonlinearity in the photo-electro-thermal response. This presents a substantial control challenge for tracking nonlinear lighting indices, which has not yet been solved in the existing literature, especially in the presence of ambient light interference. To address these challenges, this work proposes a spectral optimization and control framework based on reinforcement learning for multichannel LED lighting systems. By comprehensively considering ambient light interference throughout the modeling and training stages, the framework effectively handles the complex nonlinear optimization and control tasks. A seven-channel LED control and measurement system incorporating ambient light interference was established to evaluate its effectiveness in experiments. Through rigorous testing, the proposed framework demonstrated its ability to successfully execute complex lighting control tasks, including achieving desired color rendering index, color difference, and intensity error, in the presence of ambient disturbance.**

***Index Terms*—Antidisturbance, lighting control, multichannel LED, reinforcement learning (RL), spectral optimization.**

I. Introduction

As fourth-generation lighting sources, light-emitting diodes (LEDs) offer superior energy efficiency, facile dimming capabilities, inherent reliability, and a long lifespan, solidifying their position as the predominant global illuminant [1]. Following decades of rapid advancements, the luminous efficiency of white LEDs has far exceeded initial estimates [2]. Concurrently, high-power and high-power density drive equipment has been rapidly developing [3], [4], [5], [6], [7]. Consequently, the foundation for developing higher-quality lighting sources is becoming increasingly solid, and the demand for such sources is growing significantly.

An example of a high quality light source is the one with tunable correlated color temperature (CCT). CCT tunable systems require at least two channels [8], [9], implying the need for more channels to create higher-quality lighting functions. However, similar to typical power electronic systems, LED systems exhibit highly complex nonlinear photo-electro-thermal (PET) response, especially in multiinput multioutput configurations [8], [10]. Previous studies have been focused on addressing these complex nonlinear problems to maintain the stability of luminous flux output [11] or the color point [12], [13], leading to some progress.

However, for complex lighting tasks that require simultaneous optimization of other parameters while adjusting CCT, simply maintaining a stable luminous flux output or color point is insufficient. CCT tunability (or color tunability) is a linear problem [14] that is relatively easy to solve. However, the problem becomes more complex when considering the color rendering index (CRI) [15], [16]. Conventional RGB light sources often exhibit low CRI. To address this, some researchers have explored the CCT tunability of dual-white LED systems [9], [13]. While these systems can achieve a high CRI (if the white LEDs have a sufficiently high CRI) they often suffer from poor CCT accuracy (or high color difference) and may not be able to optimize other parameters or effectively combat strong ambient interference. To handle complex lighting tasks, more channels are necessary, further increasing the complexity of the nonlinear PET response.

Moreover, according to the theory of colorimetry, when the number of channels is less than three, the precise color point can be directly solved using linear algebra [17]. However, when the number of channels exceeds three or when there is a tradeoff between color accuracy and other performance metrics, spectral optimization becomes necessary [16], [18]. This level of optimization often involves more complex nonlinear response, primarily caused by strong nonlinear parameters such as CRI. In other words, the nonlinearity of CRI leads to a nonconvex problem [18], [19].

The combination of these nonlinear response from both factors can make it challenging to achieve complex control tasks in multichannel light source systems. Consequently, the lighting control problem primarily focuses on two aspects, i.e., solving the optimal control parameters and designing a controller to

Received 17 April 2025; revised 9 June 2025; accepted 15 July 2025. Date of publication 22 July 2025; date of current version 27 August 2025. Recommended for publication by Associate Editor S. Y. R. Hui. *(Corresponding author: Jianfei Dong.)*

Qi Wang is with the School of Future Science and Engineering, Soochow University, Suzhou 215222, China, also with the School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Suzhou 215163, China, and also with the Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215222, China (e-mail: wangqi97@mail.ustc.edu.cn).

Jianfei Dong is with the School of Future Science and Engineering, Soochow University, Suzhou 215222, China (e-mail: jfdong@suda.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2025.3591666.

Digital Object Identifier 10.1109/TPEL.2025.3591666

track these optimal parameters. While some researchers have attempted to directly calculate the final control inputs [20], even dual-channel systems can struggle from the strong nonlinear nature of LED PET response [9], [13]. Most previous studies have concentrated on the first step, exploring various methods such as the complex method [15], the optimization vector method [16], quadratic fractional programming [18], differential evolution [21], blackbody curve matching [22], [23], [24], [25], [26], [27], and genetic algorithms (GAs) [28], [29], [30], [31], [32], [33], [34]. Alternatively, some other approaches try to implement color equality constraints to the optimization problem and employing various optimization techniques [35], [36]. However, these methods have their respective pros and cons, as discussed in [37]. Furthermore, the existing methods primarily address the first problem, leaving the second step, i.e., nonlinear PET response controller design, unaddressed. Few researchers have attempted to design nonlinear controllers based on nonlinear PET response for LED systems with more than three channels, especially for complex control tasks involving CRI and ambient interference [13], [38].

To effectively complement each other, the two aspects of optimal parameter determination and controller design must be integrated within a unified framework. A previous research [37] has explored this solution, but without considering the interference of external ambient light. In many real-world scenarios, ambient light interference is present, such as sunlight or other uncontrolled artificial light sources. Therefore, it is essential to develop a framework that can effectively resist ambient light interference while addressing the aforementioned challenges. Current daylight interference control strategies often focus solely on intensity control [39] or maintain a CCT while addressing intensity control [40]. Actually, the challenges posed by ambient light interference in multichannel LED lighting systems deny most classical model-based optimization and control techniques.

To overcome this limitation, a method capable of generating a general policy model for various desired parameters is crucial. The widely used advanced methods, such as convolutional neural networks, recurrent neural networks, and long short-term memory networks, cannot be directly applied to solve this problem because the optimal decision variables for training are not readily available. Additionally, the GA method can only generate fixed optimal solutions, which are subject to changes with varying lighting demands. Reinforcement learning (RL) methods, known for their strength in discovering optimal policies within Markov decision processes (MDPs), present a promising alternative. Unlike traditional optimization methods—which often struggle with nonconvex optimization challenges or focus on single optimal solutions, RL can generate a complete optimal policy model even for nonconvex problems [41]. Given that many spectrum optimization challenges can be formulated as MDPs, using RL to produce an optimal spectral policy is particularly appealing. Therefore, we attempt to introduce a new RL-based spectral optimization and control framework in this work. By comprehensively considering ambient light interference throughout the modeling and training stages, the framework effectively handles complex nonlinear optimization and control tasks. To evaluate the effectiveness of this framework, a seven-channel LED control and measurement system incorporating ambient light interference, represented by a halogen lamp, was developed. The convergence of the control law within this framework was also analyzed. Rigorous testing on a complex control task that required balancing CRI, color difference, and intensity errors demonstrated its effectiveness.

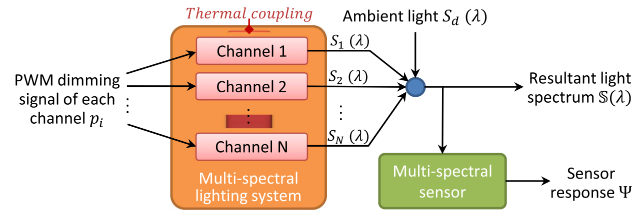


Fig. 1. Response of multichannel LED system with ambient light.

The rest of this article is organized as follows. Section II introduces the preliminaries of the problem and describes the experimental testbed. Section III details the spectral optimization and control framework. Section IV presents and discusses the experiment results. Finally, Section V concludes this article.

II. Preliminaries and Testbed Design

A. Preliminaries

Consider a multispectral lighting system comprising N distinct channels, as illustrated in Fig. 1. Each channel possesses a corresponding measurable absolute spectral output, denoted by $S_i(\lambda)$, where $i \in [1, N]$; and λ represents the wavelength. When the input pulse width modulation (PWM) signal for each channel is $p_i \in [0, 1]$, the corresponding channel output is represented by $S_i(\lambda, p_1, p_2, \ldots, p_N)$. The outputs of these channels are primarily influenced by the corresponding channel inputs and the junction temperature of the LEDs. Due to heat transfer between LED chips, the temperature of each LED is a coupled function of all input powers rather than being solely determined by its own input [10], [12]. This temperature variation contributes to the nonlinear response of the system to input signals.

This nonlinearity leads to two primary errors, i.e., irradiance error and peak offset. While the irradiance error can be compensated by adjusting the input signals, the peak offset remains uncorrectable at the control stage. This is because the peak offset, defined as the spectral peak shift with the junction temperature, is solely a function of temperature [42]. Both factors contribute to variations in visual parameters. However, the use of PWM dimming and effective heat dissipation can mitigate this effect [43]. Therefore, the actual response model can be approximated by a commonly used simplified model [23], [37]

$$S(\lambda) = \sum_{i=1}^{N} S_i(\lambda, p_1, p_2, \ldots, p_N) \tag{1}$$

$$= \sum_{i=1}^{N} \alpha_i(p_1, p_2, \ldots, p_N)\hat{S}_i(\lambda) + S_e(\lambda) \tag{2}$$

where $S(\lambda)$ denotes the output spectrum of the lighting system; $S_e(\lambda)$ represents the modeling error resulting from uncorrectable peak offsets and estimation errors; and $\alpha_i(p_1, p_2, \ldots, p_N)$ and $\hat{S}_i(\lambda)$, respectively, represent the power parameter function of the ith channel to be detailed later and the normalized full power spectrum to be estimated.

The power parameter represents the ratio of output power to normalized maximum output power and is equivalent to the input PWM of each channel under linear conditions, as assumed in most existing work. However, even with PWM control, the output irradiance is a highly nonlinear function of the input, especially for multichannel LED systems. Therefore, the power parameter function is defined as

$$\alpha_i(p_1, p_2, \ldots, p_N) = p_i \cdot (1 + p_{e,i}(p_1, p_2, \ldots, p_N)) \tag{3}$$

where $p_{e,i}$ represents the irradiance deviation associated with the ith channel, which is a nonlinear function of p_l, $l = 1, \ldots, N$. Considering the domain knowledge that the nonlinear parameter function is strongly inversely correlated with the corresponding input signal due to lowered efficiency at higher junction temperatures [8], [44], we can simplify the expression of the nonlinear irradiance deviation as follows, with its derivation postponed to the Appendix A

$$p_{e,i}(p_1, p_2, \ldots, p_N) = \chi - \gamma \cdot p_i - \zeta(p_o) \tag{4}$$

where χ and γ are positive constants, and $\zeta(p_o) \in (0, 1)$ represent the lower efficiency caused by other channel's input.

When ambient light, denoted by $S_d(\lambda)$, is present in the target space to be illuminated, the resultant lighting spectrum, $\mathbb{S}(\lambda)$, becomes the sum of the spectra of both the lighting system and the ambient light disturbance,

$$\mathbb{S}(\lambda) = S(\lambda) + S_d(\lambda). \tag{5}$$

In this work, we assume that the spectrum of the ambient light remains constant, while its intensity may vary. Then, the model can be rewritten as

$$\mathbb{S}(\lambda) = \sum_{i=1}^{N} \alpha_i \hat{S}_i(\lambda) + \alpha_d \hat{S}_d(\lambda) + S_e(\lambda) \tag{6}$$

where α_d and $\hat{S}_d(\lambda)$ represent the power parameter and the estimated normalized full power spectrum of the ambient light. The mixed spectrum $\mathbb{S}(\lambda)$ determines various lighting metrics.

In general lighting, the optimal performance of illumination usually includes the following criteria:

1) achieving a desired CCT;
2) achieving a desired luminous flux, Φ_a; and
3) maximizing the CRI value.

These metrics are required to be achieved simultaneously. Moreover, to render more dynamic lighting atmosphere, the setpoints of CCT can vary, usually in the range of 2500 to 6500 K. This is clearly a multiobjective optimization problem, with the objective function defined as

$$f(\mathbb{S}) = Ra - \varphi_1 \cdot \Delta^2 u'v' - \varphi_2 \cdot |\Phi_a - \Phi|, \tag{7}$$

$$\Delta u'v' = \sqrt{(\Delta u')^2 + (\Delta v')^2}, \tag{8}$$

$$\Phi = 683 \cdot \int_{380}^{780} V(\lambda)\mathbb{S}(\lambda)\, d\lambda \tag{9}$$

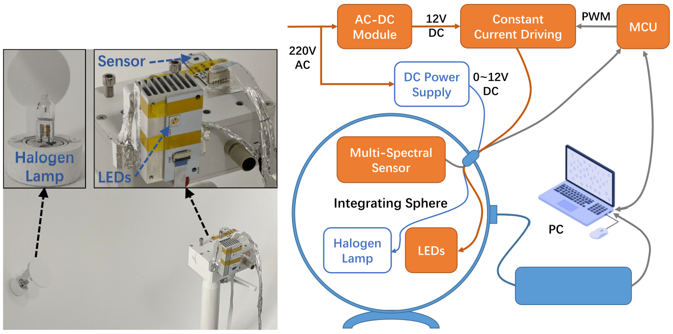
where Ra represents the CRI values calculated according to the standard [45]; $\Delta u'v'$ is the color difference in the CIE 1976 uniform color space; Φ represents the output luminous flux of the light source, with $V(\lambda)$ being human eyes' visual response function; φ_1 and φ_2 are coefficients used to trade off the errors in the CRI, color difference and luminous flux. Generally, an Ra value above 80 is considered acceptable for most general lighting applications; while an Ra exceeding 90 is targeted in scenarios where accurate color rendering is critical, such as art galleries and clothing stores. $\Delta u'$ and $\Delta v'$ represent the differences of respectively the u' and v' coordinates of two color points. $\Delta u'v'$ is often preferred over the McAdam ellipse for color difference calculations due to its computational convenience [46]. A Step size of 0.0011 is defined in the standard, with 5 Step being generally acceptable and 3 Step representing the threshold of human eye.

Consider a multispectral sensor with M distinct channels, each possessing a unique spectral response, $\phi_j(\lambda)$, where $j \in [1, M]$. The sensor output, $\Psi \in R^{1\times M}$, can be expressed as

$$\begin{aligned}\Psi &= [\mathbb{S}(\lambda)\cdot\phi_1(\lambda)\ \mathbb{S}(\lambda)\cdot\phi_2(\lambda)\ \cdots\ \mathbb{S}(\lambda)\cdot\phi_M(\lambda)] \\ &= \left(\sum_{i=1}^{N} \alpha_i \hat{S}_i(\lambda) + \alpha_d \hat{S}_d(\lambda) + S_e(\lambda)\right)\cdot[\phi_1(\lambda)\ \cdots\ \phi_M(\lambda)].\end{aligned} \tag{10}$$

It is evident that the parameters α_i and α_d can be estimated, if $M \geq N + 1$. The accuracy of the model estimation improves with increasing similarity between the disturbance light $S_d(\lambda)$ and the sensor's spectral responses and with a larger number of sensor channels. For disturbance light with significant variations in CCT or spectrum, the error may increase [47]. In such cases, a spectrometer or a multispectral sensor with even more channels is necessary to accurately measure the disturbance light. However, this can also increase the cost of the system, and is not the objective of this work.

Optimization of the control inputs is often hindered by the strong nonlinearity of the LED system, especially when considering the CRI. This is because the core lighting parameters, such as CRI and CCT, are nonlinear functions of the lighting spectra. The nonlinear mapping of these spectral parameters, combined with the nonlinear response of the LED input-output relationship, creates a nonpolynomial-hard optimization control problem that is difficult to solve. Decoupling the optimal parameters from the nonlinear PET response can significantly simplify both the optimization and control problem. Once the optimal power parameters, $\bar{\alpha}_i \in [0, 1]$, are determined, the tracking control can be achieved using traditional controller design approaches. Under these conditions, the core problems transform into estimating and tracking the optimal power parameters. Consequently, the complex lighting control problem can be primarily divided into two parts, i.e., first solving the optimal output parameters, and then computing the control signals to track these optimal parameters.



(a) (b)

Fig. 2. (a) Components installed within the integrating sphere and (b) schematic diagram of the experimental testbed.

B. Testbed Design

A lighting control testbed incorporating ambient light has been established. The lighting source is a high-power seven-channel LED model, offering a substantial number of channels for precise and versatile spectral power distribution (SPD) manipulation. The core component of this model is a high-power LED chip mounted on an aluminum-based printed circuit board (PCB), which is securely attached to a heat sink using thermally conductive silicone to ensure effective heat dissipation. The ambient light is represented by a halogen lamp, controlled by a dc power supply.

A 14-channel color sensor is employed to detect both the ambient and LEDs output irradiance during system operation. The sensor includes 12 spectral channels with different color filters, spanning a wavelength range from 400 to 850 nm, along with a visible channel and a flicker detection channel. The color sensor chip and its associated peripheral driver circuits are designed as a separate PCB, providing flexibility in mounting locations. A high-precision measurement setup, including an integrating sphere, is required for accurate calibration of the multispectral sensor and validation of the control effect. As illustrated in Fig. 2(a), the LED source and sensor are installed at the center of the integrating sphere, while the halogen lamp is mounted on the lamp holder. Additional equipment is placed externally to the integrating sphere to minimize interference with optical measurements.

The integrating sphere is connected to a spectrometer for real-time SPD measurements via optical fiber, with the data transmitted to a PC via a universal serial bus (USB). The multichannel LED chip is driven by a seven-channel constant current driver, maintaining a nearly linear irradiance control process for most channels. The drive current of each individual channel is primarily set at 250 mA, with exceptions for the blue (100 mA), violet (100 mA), and amber (375 mA) channel. These settings were determined based on a previous study [37] to achieve optimal maximal irradiance power ratios for each channel. The constant current driver of each channel is dimmed using PWM signals from a microcontroller unit (MCU), while being supplied by an ac–dc module. The color sensor sends data to the MCU through an inter-integrated circuit bus, and the MCU is connected to the PC via USB. Thus, the PC receives data from both the spectrometer and the sensor simultaneously. A comprehensive list of the components is provided in Table I, while the design scheme of the setup can be found in Fig. 2(b).

TABLE I
COMPONENTS OF THE EXPERIMENTAL SETUP

Type	Brand	Name
LED	ams OSRAM	LZ7-04M100
Multi-Spectral Sensor	ams OSRAM	AS7343
MCU	STMicroelectronics	STM32F103ZET6
AC-DC Module	MEANWELL	LSR75-12
Constant Current Module	PowTech	PT4115
Halogen Lamp	OSRAM	2600 K
DC Power Supply	ZHAOXIN	RXN-3010D
Spectrometer	Ocean Optics	Maya2000pro
Integrating Sphere	LISUN	IS-1.0MA33C

III. Spectral Optimization and Control Scheme

A. Methodology Overview

While previous research has demonstrated excellent control performance of RL-based optimization and control without ambient light [37], these approaches are not able to counteract with ambient light interference. For more practical applications, the framework must be modified to effectively handle ambient light.

As outlined in (6), if the spectra of both the ambient light and LEDs can be reconstructed by combining prior knowledge with sensor data, a trained policy neural network can then generate optimal power parameters, $\bar{\alpha}^t$, for the LED source to meet the control objectives. In this approach, the nonlinear aspects of spectral optimization and input control can be decoupled and solved using different methods. The complex SPD optimization process is handled by the RL method within a simulated virtual environment, while the nonlinearity of the system's PET response is managed by the controller in the real system. The spectrum optimization and control framework is structured as shown in Fig. 3. It consists of two parts: 1) the RL training module; and 2) the controller application module. The RL training module, represented by the dashed box in the figure, includes a simulation environment and an agent that interact to generate the optimal neural network used in the final system. The interface receives user-set parameters, transforms them into the neural network's input format, and transmits them to the network. The policy network (trained in the RL module) uses the input parameters and ambient intensity data (calculated from color sensor signals) to output the control parameters, $\bar{\alpha}^t$. A PID controller then generates the final PWM control signal, p^t, using both, $\hat{\alpha}^t$ (the estimate of the parameter α^t), and the optimal power parameters, $\bar{\alpha}^t$. A calibration matrix, developed from multichannel sensor data, is required for this process and must be calibrated in advance.

B. Determining Training Parameters According to Tasks

For RL training, the simulation environment should be constructed using the LED and ambient light spectra to simulate the real system. The simulation environment can be used to simulate the real system, generating the final SPD and reward when

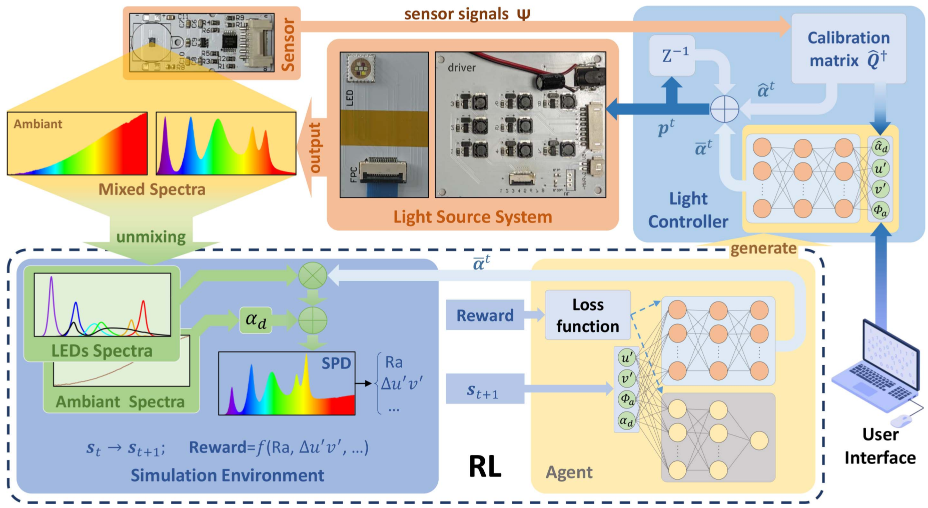


Fig. 3. Scheme of the optimization and control framework.

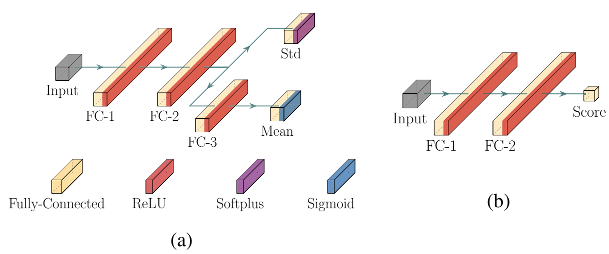


Fig. 4. Schematic of the (a) actor network and (b) critic network.

receiving the optimal power parameters, $\bar{\alpha}^t$, and then sending the reward and the next state s_{t+1} to the agent. The agent contains the required policy network that generates the action (optimal power parameters $\bar{\alpha}^t$), when the current state s_t is input. In this application, the reward function is defined by (7). The state is defined as a vector comprising the CIE 1976 color coordinates corresponding to the reference CCT, (u', v'), the target luminous flux, Φ_a, and the ambient light power parameters, α_d.

The proximal policy optimization algorithm [41] was chosen as the RL method due to its ease of use and successful applications in many cases. Therefore, the agent incorporates both an actor network and a critic network. The structures of these two networks are illustrated in Fig. 4. The actor network's input layer has a size of 4×1, followed by three fully-connected layers: Two layers of size 128×1 (FC1, FC2), and one layer of size 64×1 (FC3), all employing ReLU activation functions. The output layer consists of two layers, i.e., Mean and Std, both with a size of $N \times 1$. The Mean layer employs a Sigmoid activation function; and the Std layer uses a Softplus activation function to ensure that the output values primarily remain within [0,1]. The critic network takes a 4×1-dimensional input, passes it through two fully-connected layers (FC1, FC2) with a size of 128×1 each and the ReLU activation, and outputs a scalar value without any activation function. For our application $N = 7$, the actor and critic neural networks have 26.7k and 17.2k parameters respectively. The influence of network size is detailed in Appendix B.

The main training process was chosen to be grid point training, which had been previously used to train the agent, and had shown its generalization capabilities in a similar problem [37]. The target CCT was set to change from 2500 K to 4900 K at a 100 K interval and from 4900 to 6500 K at a 200 K interval. The ambient interference was set to vary from 0 to 20 lm with a stepsize of 4 lm. The target intensity was set to increase from 40 to 90 lm with a stepsize of 10 lm.

Once training begins, the actor network generates the action, $\bar{\alpha}^t$, based on the input state, s_t, and sends these parameters to the simulation environment. The environment computes the final SPD by utilizing the output of the current neural network, $\bar{\alpha}^t$, and the ambient light power parameter, α_d, as inputs to (6). It then evaluates parameters such as CRI, color difference, and intensity error, determines the reward, and sends the reward and the next state, s_{t+1}, back to the agent. The agent records the rewards, states s_t, actions $\bar{\alpha}^t$, and next states s_{t+1}. For grid-based training, all state points are predefined, while for standard RL, state points are sampled randomly. As sufficient data accumulates in the experience pool, the agent updates its neural networks to maximize the expected reward. Each episode continues until the maximum step limit is reached, and the entire training process terminates upon reaching the predefined limits or maximum episodes. Once training is complete, the actor network can be deployed in the real system.

C. Spectra Modeling

The spectral model of (6) is essential for constructing the RL training environment. The normalized spectra can be unmixed via least squares using test data. For a dataset with H data points, (6) can be rewritten as:

$$\mathbb{S}^k(\lambda) = \sum_{i=1}^{N} \alpha_i^k \hat{S}_i(\lambda) + \alpha_d^k \hat{S}_d(\lambda) + S_e^k(\lambda) \tag{11}$$

where the superscript $k \in [1, H]$ represents the data point index, or in a matrix form

$$\begin{bmatrix} \alpha_1^1 & \alpha_2^1 & \cdots & \alpha_N^1 & \alpha_d^1 \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_N^2 & \alpha_d^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_1^H & \alpha_2^H & \cdots & \alpha_N^H & \alpha_d^H \end{bmatrix} \begin{bmatrix} \hat{S}_1 \\ \hat{S}_2 \\ \vdots \\ \hat{S}_N \\ \hat{S}_d \end{bmatrix} + \begin{bmatrix} S_e^1 \\ S_e^2 \\ \vdots \\ S_e^H \end{bmatrix} = \begin{bmatrix} \mathbb{S}^1 \\ \mathbb{S}^2 \\ \vdots \\ \mathbb{S}^H \end{bmatrix}. \tag{12}$$

While the ambient light may not be freely controllable in some applications, measuring its SPD is feasible by switching OFF the LED light source. Therefore, the coefficient α_d at a sampling instant can be determined simply by dividing the ambient flux Φ_d by the maximum ambient flux $\Phi_d^{\max}$

$$\alpha_d = \frac{683 \cdot \int_{380}^{780} V(\lambda) S_d(\lambda), d\lambda}{683 \cdot \int_{380}^{780} V(\lambda) \hat{S}_d(\lambda), d\lambda} = \frac{\Phi_d}{\Phi_d^{\max}}.$$

However, the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_N$ cannot be directly measured using the spectrometer with the same method due to the internal mutual influence of thermal effects. When other channels are closed, the spectra and corresponding parameters of the measured channel will change simultaneously, and vice versa. Additionally, using this method for all channels would lead to a dramatic increase in the amount of data.

As an approximate, the input PWM signals p_i are utilized to replace α_i, and the unmixed spectra equation becomes

$$\begin{bmatrix} \hat{S}_1 \\ \hat{S}_2 \\ \vdots \\ \hat{S}_N \\ \hat{S}_d \end{bmatrix} = \begin{bmatrix} p_1^1 & p_2^1 & \cdots & p_N^1 & \alpha_d^1 \\ p_1^2 & p_2^2 & \cdots & p_N^2 & \alpha_d^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_1^H & p_2^H & \cdots & p_N^H & \alpha_d^H \end{bmatrix}^{\dagger} \begin{bmatrix} \mathbb{S}^1 \\ \mathbb{S}^2 \\ \vdots \\ \mathbb{S}^H \end{bmatrix} \tag{13}$$

where † represents the pseudoinverse. It should be noted that $H \geq N + 1$ is necessarily for the linear equations to be solvable. The estimated coefficients can then be calculated based on the spectrometer measured data and the equation

$$\begin{bmatrix} \hat{\alpha}_1^k & \hat{\alpha}_2^k & \cdots & \hat{\alpha}_N^k & \hat{\alpha}_d^k \end{bmatrix} = \mathbb{S}^k \begin{bmatrix} \hat{S}_1 \\ \hat{S}_2 \\ \vdots \\ \hat{S}_N \\ \hat{S}_d \end{bmatrix}^{\dagger}. \tag{14}$$

The RL training stage directly determines optimal power parameters, independent of the PET response. However, analyzing the system's nonlinearity is crucial for facilitating the convergence analysis of the subsequent controller design. The range of the nonlinearity deviation for each channel can be approximated using (3) and the estimated coefficients. Equation (3) can be reformulated as follows:

$$p_i^k p_{e,i}^k = \hat{\alpha}_i^k - p_i^k \tag{15}$$

and for $p_i^k \neq 0$

$$p_{e,i}^k = \frac{\hat{\alpha}_i^k - p_i^k}{p_i^k}. \tag{16}$$

Let $p_{e,i}^{\max}$ and $p_{e,i}^{\min}$ represent the maximum and minimum nonlinear deviation that may occur during system operation. As luminous efficiency decreases with increasing junction temperature, $p_{e,i}^{\min}$ occurs when all channel input signals are at full duty cycle (i.e., $p_l = 1$). Conversely, when all channel inputs approach zero (i.e., $p_l \approx 0$), the efficiency degradation becomes negligible ($\zeta(p_o) \approx 0$) and resulting in $p_{e,i}^{\max}$. According to (4) we have

$$p_{e,i}^{\max} = \chi, \tag{17}$$

$$p_{e,i}^{\min} = \chi - \gamma - \zeta(p_o) \tag{18}$$

and γ can be estimated by

$$\gamma = p_{e,i}^{\max} - p_{e,i}^{\min} - \zeta(p_o) \leq p_{e,i}^{\max} - p_{e,i}^{\min}. \tag{19}$$

D. Sensor Matrix Calibration

However, the spectrometer can be expensive for large-scale product applications. A multichannel color sensor can be used as a replacement for the spectrometer in the real control stage. After the unmixed spectra of the ambient and LEDs have been estimated, (10) can be rewritten as

$$\begin{bmatrix} \Psi^1 \\ \Psi^2 \\ \vdots \\ \Psi^H \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_1^1 & \hat{\alpha}_2^1 & \cdots & \hat{\alpha}_N^1 & \alpha_d^1 \\ \hat{\alpha}_1^2 & \hat{\alpha}_2^2 & \cdots & \hat{\alpha}_N^2 & \alpha_d^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\alpha}_1^H & \hat{\alpha}_2^H & \cdots & \hat{\alpha}_N^H & \alpha_d^H \end{bmatrix} \begin{bmatrix} \hat{S}_1 \\ \hat{S}_2 \\ \vdots \\ \hat{S}_N \\ \hat{S}_d \end{bmatrix} \cdot \begin{bmatrix} \phi_1(\lambda) \\ \vdots \\ \phi_M(\lambda) \end{bmatrix}^T + \begin{bmatrix} S_e^1 & S_e^2 & \cdots & S_e^H \end{bmatrix}^T \cdot \begin{bmatrix} \phi_1(\lambda) & \cdots & \phi_M(\lambda) \end{bmatrix}. \tag{20}$$

Let $Q = \begin{bmatrix} \hat{S}_1 \\ \hat{S}_2 \\ \vdots \\ \hat{S}_N \\ \hat{S}_d \end{bmatrix} \cdot \begin{bmatrix} \phi_1(\lambda) \\ \vdots \\ \phi_M(\lambda) \end{bmatrix}^T$. Considering the errors in S_e can be reasonably ignored in a well calibrated system, the parameter $\hat{Q}$ can be effectively estimated as

$$\hat{Q} = \begin{bmatrix} \hat{\alpha}_1^1 & \hat{\alpha}_2^1 & \cdots & \hat{\alpha}_N^1 & \alpha_d^1 \\ \hat{\alpha}_1^2 & \hat{\alpha}_2^2 & \cdots & \hat{\alpha}_N^2 & \alpha_d^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\alpha}_1^H & \hat{\alpha}_2^H & \cdots & \hat{\alpha}_N^H & \alpha_d^H \end{bmatrix}^{\dagger} \begin{bmatrix} \Psi^1 \\ \Psi^2 \\ \vdots \\ \Psi^H \end{bmatrix}. \tag{21}$$

In this manner, the sensor can be used to directly estimate the unmixed spectral power coefficients using the equation

$$\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N, \hat{\alpha}_d] = \Psi \times \hat{Q}^{\dagger}. \tag{22}$$

Note that the matrix, model, and network are tailored to the current configuration; significant changes in LED configuration or ambient light type may require repeating the above process to maintain optimal performance.

E. Control Law and Convergence

Although the target output power parameters $\bar{\alpha}^t$ are generated by the actor network, a controller is necessary to counteract the deviations caused by the nonlinear PET response and maintain the output power parameters $\hat{\alpha}^t$ consistent with the actor network output. Due to the decoupling between the light system's input and output achieved in (6), each channel can be controlled independently, although influenced indirectly by other channels heat. Specifically, an incremental proportional controller, with a proportional gain of 1, is implemented as follows:

$$p_i^0 = \bar{\alpha}_i^0 \tag{23}$$

$$p_i^{t+1} = p_i^t + \bar{\alpha}_i^{t+1} - \hat{\alpha}_i^t \tag{24}$$

where the superscript $t \in [0, \infty]$ denotes the time step. The first-step control error can be expressed as

$$\begin{aligned} e_i^0 &= \bar{\alpha}_i^0 - \hat{\alpha}_i^0 \\ &= p_i^0 - \hat{\alpha}_i^0. \end{aligned} \tag{25}$$

Given that a more accurate data-driven irradiance model of the LEDs is employed, most channels, especially the linear ones, can achieve near-optimal PWM input (corresponding to the optimal power output) in the first step. The initial control error can be analyzed based on (15) and (25). The continuous control error is

$$\begin{aligned} e_i^{t+1} &= \bar{\alpha}_i^{t+1} - \hat{\alpha}_i^{t+1} \\ &= \bar{\alpha}_i^{t+1} - \hat{\alpha}_i^t + \hat{\alpha}_i^t - \hat{\alpha}_i^{t+1} \\ &= p_i^{t+1} - p_i^t + p_i^t \cdot (1 + p_{e,i}^t) - p_i^{t+1} \cdot (1 + p_{e,i}^{t+1}) \\ &= p_i^t \cdot p_{e,i}^t - p_i^{t+1} \cdot p_{e,i}^{t+1}. \end{aligned} \tag{26}$$

When the target remains constant, i.e., $\bar{\alpha}_i^{t+1} = \bar{\alpha}_i^t$, since the first step is highly accurate, $\zeta(p_o)$ can be considered constant for a constant target, therefore, according to (4), we have

$$\frac{\Delta p_{e,i}^t}{\Delta p_i^t} = -\gamma. \tag{27}$$

When the current channel is in a transient state, i.e., $|e_i^t| > 0$ and $|\Delta p_i^t| > 0$, the system is not yet stable, we have

$$e_i^{t+1} = \frac{e_i^{t+1}}{e_i^t} \cdot e_i^t = \frac{\bar{\alpha}_i^{t+1} - \hat{\alpha}_i^{t+1}}{\bar{\alpha}_i^{t+1} - \hat{\alpha}_i^t} \cdot e_i^t \tag{28}$$

using (24) and (26), we have

$$e_i^{t+1} = \frac{p_i^t \cdot p_{e,i}^t - p_i^{t+1} \cdot p_{e,i}^{t+1}}{p_i^{t+1} - p_i^t} \cdot e_i^t. \tag{29}$$

The absolute error ratio will be

$$\begin{aligned} \left|\frac{e_i^{t+1}}{e_i^t}\right| &= \left|\frac{p_i^t \cdot p_{e,i}^t - (p_i^t + \Delta p_i^t) \cdot (p_{e,i}^t + \Delta p_{e,i}^t)}{\Delta p_i^t}\right| \\ &= \left|\frac{p_i^t \cdot \Delta p_{e,i}^t + \Delta p_i^t \cdot (p_{e,i}^t + \Delta p_{e,i}^t)}{\Delta p_i^t}\right| \\ &= \left|p_i^t \cdot \frac{\Delta p_{e,i}^t}{\Delta p_i^t} + p_{e,i}^{t+1}\right| \\ &= \left|-\gamma \cdot p_i^t + p_{e,i}^{t+1}\right|. \end{aligned} \tag{30}$$

If $p_{e,i} \in (-0.5, 0.5)$ and $\gamma \in (0, 0.5)$, we have

$$\left|\frac{e_i^{t+1}}{e_i^t}\right| = \left|-p_i^t \cdot \gamma + p_{e,i}^{t+1}\right| < 1. \tag{31}$$

Therefore, the error will gradually converge until it reaches the sensor's detection resolution. Under these conditions, the lighting efficacy of each channel can exhibit a wide range of variation, from 50% to 150%. This represents a relatively loose constraint for multichannel light sources with PWM control and effective heat dissipation, which typically exhibit nonlinear light decay of no more than 50% [42].

However, it should be noted that if most channels exhibit strong nonlinearity, the controller's stability may need to be proven more rigorously. Such a situation is unlikely to occur unless the LED system has extremely poor heat dissipation or most channels are special band LED chips. Additionally, if the LEDs have aged, the controller should be reanalyzed–or even redesigned–based on the characteristics of the aged devices to ensure steady-state convergence and to guarantee that the optimal output power of each channel remains within the maximum capabilities of the aged devices.

IV. Experiment Results and Analysis

A. Spectra Unmixing Results and Analysis

The accuracy of spectral restoration by the sensor plays a crucial role in both optimization and final control. For the testbed employed in this article, the input PWM signal for each channel in the test datasets was designed with three discrete levels: [0,0.5,1]. Two datasets were collected, each with 0 and 21.4 lm halogen ambient light, resulting in $3^7 \times 2$ data points. Each dataset comprised 2187 entries, corresponding to different combinations of PWM signals applied to each channel. Each combination lasted for 1 s. The sensor data was recorded at 0.4 s and 0.7 s, as the integration time was set to 0.18 s, allowing the color sensor to capture data from six channels simultaneously. Therefore, only 12 spectral channels of the color sensor were used, excluding the NIR and FLICKER channels. The spectrometer's integration time was set to 0.1 seconds. SPD data was recorded at 0.3, 0.5, and 0.7 seconds and then averaged to produce a single data point. This ensured that both SPD and corresponding color sensor data were recorded simultaneously. All color sensor and SPD data were collected in this manner throughout the article. After all combination data was collected, the simplified model was unmixed using (13); and the matrix $\hat{Q}$ was solved using (21). The validation dataset was recorded similarly, but with PWM signals set to intermediate levels: [0.3,0.6,0.9], and the ambient light was set to 16.7 lm.

The unmixed light normalized full power spectra are shown in Fig. 5, where the LEDs SPD curves are plotted as solid lines. The dashed lines depict the SPD curves of the ambient light sources

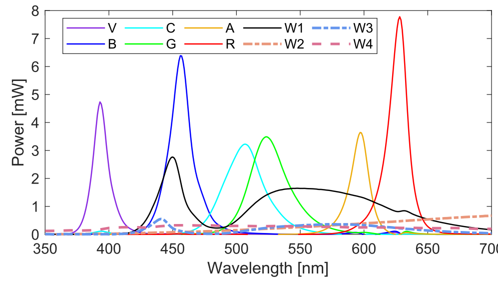


Fig. 5. SPDs of light source system and ambient light. The SPD curves of the LEDs are represented by solid lines of their corresponding colors (excluding white, which is plotted as black). The dashed lines depict the SPD curves of the ambient light sources: W2 for the halogen lamp, W3 for the ambient LED source, and W4 for the CIE standard illuminant D65.

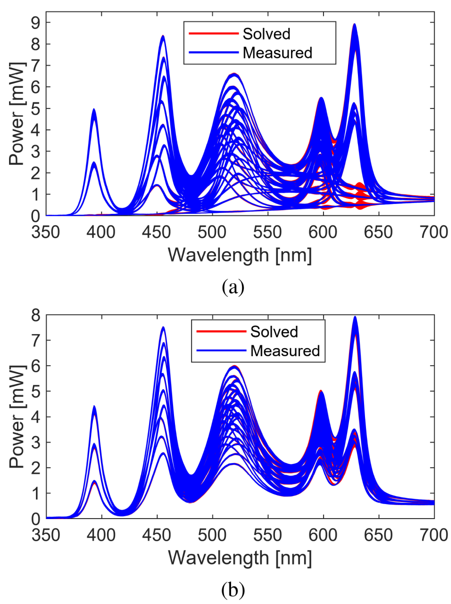


Fig. 6. Solved SPDs (red line) vs. the measured SPDs (blue line) at the (a) fitting dataset and (b) validation dataset.

(W2 for the halogen lamp, W3 for the ambient LED source, and W4 for the CIE standard illuminant D65).

The second fitting dataset, collected under halogen ambient light, and its corresponding solved SPDs (derived from sensor data) were displayed in Fig. 6(a), while the validation dataset and its corresponding solved SPDs were displayed in Fig. 6(b). As shown, the solved SPDs closely resembled the measured SPDs in both figures. Computational analysis revealed that the average color difference and average luminous flux percentage error of the solved spectra were 0.055 Step and 0.144% for the fitting dataset, compared to 0.081 Step and 0.401% for the validation dataset. These results, along with the close resemblance between the solved and measured SPDs, demonstrated the high accuracy of the SPD solution for both the fitting and validation datasets under varying ambient light conditions.

The distributions of nonlinearity deviations for each channel, as calculated by (16), are illustrated in Fig. 7. It was evident that the amber channel exhibited the largest nonlinearity error, reaching up to 24%. In contrast, the blue and green channels showed minimal to no nonlinearity errors, while the remaining channels had minor errors around 10%. This suggested that the model could simulate the system output with high accuracy for most channels, except for the amber channel. But even the amber channel (with $p_{e,i} \in (-0.17, 0.24)$) had a $\gamma \in (0, 0.41)$ that could be estimated by (19). Therefore, the system is suitable for the analysis in Section III-E.

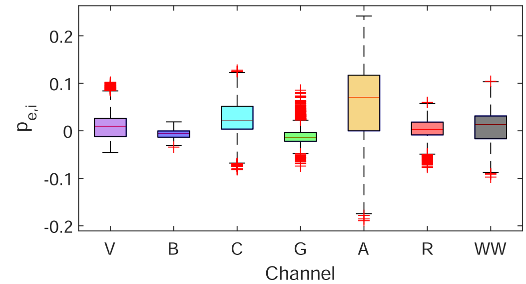


Fig. 7. Nonlinear parameter distributions of LEDs. The error bars for each channel are plotted in their corresponding colors (excluding white, which is plotted as black).

TABLE II
TRAINING SETTINGS FOR THE FOUR GROUPS

Group	Ambient Light	Training Points (State)
Random	2600 K Halogen	Random Points
2600K	2600 K Halogen	Designed Grid Points
4100K	4100 K LED	Designed Grid Points
6500K	Standard Illuminant D65	Designed Grid Points

B. RL Training Results and Analysis

While grid point training methods have been employed in recent research, direct comparisons with randomized training methods have not been thoroughly conducted. To address this, a control group was established, adopting the same training settings and tasks but with randomized inputs. Additionally, two different types of ambient light–4100 K general white LED (W3 in Fig. 5) and CIE standard illuminant D65 (W4 in Fig. 5)–were selected for two other grid-based experiment groups to assess the generalization capabilities of the RL optimization framework with respect to ambient light types. The training settings for the four groups are listed in Table II, and the training progress of these groups is depicted in Fig. 8. During training, the learning rates for the critic and actor neural networks were set to 2e-04 and 1e-04, respectively, except for the 6500 K group, where they were adjusted to 1.6e-04 and 8e-05 to ensure more stable training. ExperienceHorizon[1] was configured to 4752, ClipFactor[2] to 0.2, EntropyLossWeight[3] to 0.01, and MiniBatchSize[4] to 2376.

As evident from the figure, the training trajectories of all groups exhibited a rapid increase during the initial 1000 episodes, followed by a bottleneck phase lasting until around 4000 episodes. The Random group overcame this bottleneck first, while the 4100 K LED group was the last to get it through. This can be attributed to the inherent comprehensive search

[1]Number of steps used to calculate the advantage measures.
[2]Clip factor for limiting the change in each policy update step.
[3]A penalty parameter for enforcing the agent to move out of local optima.
[4]The number of samples used in one update step during training.

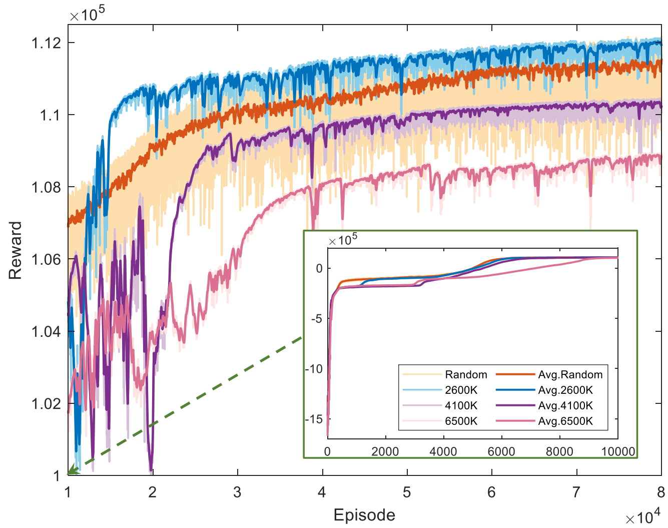


Fig. 8. Training process of the four experimental groups. The translucent lines labeled "Random," "2600 K," "4100 K," and "6500 K" represent the cumulative rewards per episode for each group. The bold lines indicate the 128-episode moving average for the corresponding groups.

space of the random method and the spectral distribution of the LED-type ambient source. After the bottleneck period, all groups resumed their upward trend, albeit at a decreasing slope. This is a natural consequence of the quadratic term present in the reward function. After 40 000 episodes, all groups demonstrated convergence trends, accompanied by a decrease in learning efficiency. While the Random group exhibited the largest reward fluctuations due to its random input, the other grid method groups demonstrated similar swing levels. Notably, the random training method proved to be the most stable when comparing average lines. Both halogen groups achieved higher rewards than other groups, likely benefiting from the lower CCT blackbody spectrum of the halogen light source. Furthermore, the halogen grid group exhibited the highest overall rewards during the training process. Although the agent may have experienced slight improvements with continued training, all groups were halted at 80 000 episodes. This decision was based on the assumption that further improvements in output parameters would not lead to significant practical benefits in real-world controllers. The training time for each group was 85 hours on a 13600 K CPU with a 25% occupancy rate.

To provide a comprehensive comparison between the RL-based method and traditional methods, the GA method was selected as an alternative approach for generating multiple independent target results under the same experimental conditions. The same SPD model and reward function were used, with a 2600 K halogen lamp as the ambient light. The parameters for the GA method were set as follows: PopulationSize = 40, CrossoverFraction = 0.9, and Generations = 800. The target points were defined by varying the CCT from 2500 to 6550 K in 50 K intervals; while the target intensity ranged from 40 lm to 90 lm in 2.5 lm intervals. The ambient light intensity varied from 0 lm to 20 lm in 1 lm intervals, resulting in a total of 36,162 optimization points. The total computational power consumption of the GA method was more than three times that of the RL training, despite the comparable training time. The optimization results were averaged based on both ambient interference intensity and target intensity to facilitate analysis. The color difference, CRI, and intensity error results for the GA method were shown in Fig. 9.

The generated policy networks were then tested under the same target point conditions, with the resulting parameters averaged similarly for both ambient interference intensity and target intensity to allow for easier comparison. The color difference, CRI, and intensity error results for all RL groups were presented in Fig. 10. The final average and worst-case values for color difference, CRI, and intensity error across all test points are summarized in Table III. A comparison of the overall quality metric–specifically, the average reward–clearly showed that the 2600 K group outperformed all others under halogen ambient conditions. This demonstrated that the grid training method (average reward = 94.541) yielded better results than the random training method (average reward = 93.961), and both RL groups outperformed the GA method (average reward = 87.333). Moreover, the average reward decreases as the CCT of the ambient light increases for the grid training results. This trend aligns with the training curves shown in Fig. 8, suggesting the emergence of certain performance limitations under high-CCT ambient conditions. These findings are supported by the detailed comparisons of the results shown in Figs. 9 and 10.

A comprehensive comparison of Figs. 10(a)–(c) and Figs. 10(d)–(f) reveals that the grid training group exhibited smaller color differences and intensity errors while achieving higher CRI values compared to the random training group. This was visually evident in the figures, with the grid training group demonstrating reduced yellow areas ($Step > 2$) for color difference [Fig. 10(d)], deeper red ($Ra > 96$) and fewer blue areas ($Ra < 88$) for CRI [Fig. 10(e)], and more white areas ($Error < 0.2$) for intensity error [Fig. 10(f)]. A comparison between Figs. 9 and 10 further demonstrates that while the GA method performs reasonably well in minimizing color difference, it fails to achieve competitive performance in terms of CRI and intensity error. Specifically, the GA group exhibited almost no deep red regions ($Ra > 96$) for CRI [Fig. 9(b)] and showed extensive deep blue areas for both CRI ($Ra < 85$) and intensity error ($Error > 1$) [Figs. 9(b)–(c)]. These observations indicated that the GA method struggled with the nonlinear optimization challenges associated with CRI, and align well with the quantitative results reported in Table III. Collectively, these findings suggest that the grid training method offers improved generalization and better overall performance than random training, and that the RL-based approach consistently outperforms the GA method.

A comparison among the grid training results in Fig. 10 indicates that the 4100 K LED group showed the minimum deep red areas ($Ra > 96$) for CRI results [Fig. 10(h)], the medium yellow areas ($Step > 2$) for color difference results [Fig. 10(g)], and the medium white areas ($Error < 0.2$) for intensity results [Fig. 10(i)]. These findings confirm that the RL optimization framework is applicable to general LED ambient light, not just full-spectrum blackbody ambient light.

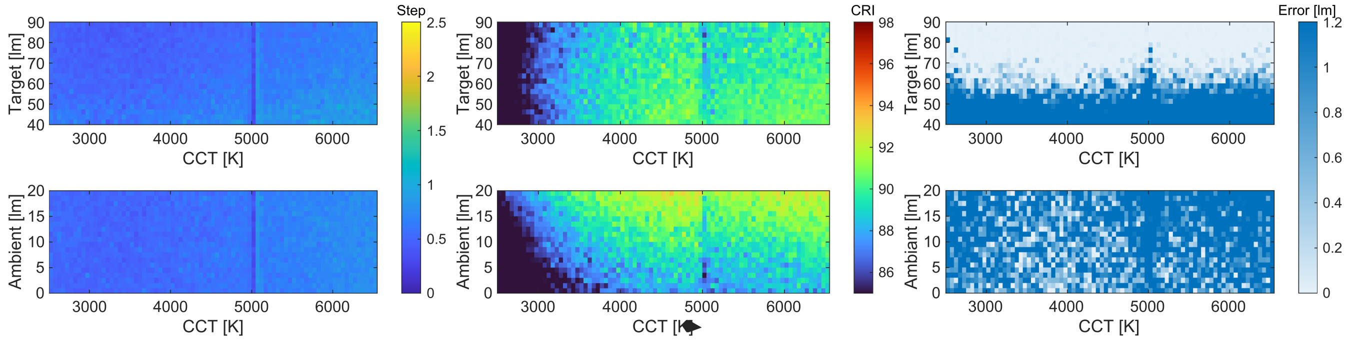


Fig. 9. Average (a) color difference, (b) CRI, and (c) intensity error results for GA2600 K experimental group. The top row of subfigures shows results averaged by ambient interference, while the bottom row shows results averaged by the set target intensity.

TABLE III
AVERAGE AND WORST RESULTS FOR ALL GROUPS

Group	Color Difference (Step)		CRI		Intensity Error (lm)		Average Reward
	Avg.	Max.	Avg.	Min.	Avg.	Max.	
GA2600K	0.599	3.007	88.787	61.517	1.349	72.696	87.333
Random	1.106	4.233	94.687	84.970	0.361	3.899	93.961
2600K	0.742	3.650	94.985	86.540	0.275	3.450	94.541
4100K	0.641	7.176	93.736	86.345	0.298	5.517	93.301
6500K	0.972	7.587	93.254	80.516	0.380	27.369	92.614
GA6500K	0.660	5.042	87.429	60.687	1.714	49.355	85.579

However, the results for the 6500 K group are particularly interesting. This group showed both moderate deep red areas ($Ra > 96$) and a significant proportion of deep blue areas ($Ra < 85$) for CRI results [Fig. 10(k)], the most extensive yellow areas ($Step > 2$) for color difference results [Fig. 10(j)], and the most pronounced blue areas ($Error > 1$) for intensity results [Fig. 10(i)]. The critical values occurred near the 40 lm target intensity, particularly with 20 lm ambient interference for both the 4100 K and 6500 K group. Furthermore, the values approach the tolerance limits (5 Step for color difference and 10% for intensity error) indicating that mitigating ambient light interference when it exceeds 50% may prove challenging. It is evident that some hardware limitations are present, as the light sources struggle to generate high-quality low-CCT SPD results, particularly when higher-CCT ambient light is present. This trend aligns with previous research, which indicated similar challenges when ambient light was not considered [37]. Despite these challenges, the proposed method successfully generated better results than the GA method, which failed to generate better results even under lower-CCT ambient light interference. To further examine this limitation, the GA method was evaluated under D65 ambient light condition to assess the impact of high-CCT ambient light on performance. As shown in Table III, this group yielded the poorest results (average reward = 85.579)—an outcome that is expected given the elevated CCT and the GA method's limited effectiveness in handling such complex optimization scenarios.

These quantitative outcomes, supported by detailed visual results, strongly suggest that the proposed RL-based framework is suitable for a wide range of ambient light conditions, with the grid training approach delivering superior overall performance. Under identical conditions, the RL method consistently outperforms the GA method. Moreover, a key advantage of the policy network approach is its comprehensive nature, encompassing the entire solution space rather than a series of discrete points. The policy network generated 36 162 result points within 5 s, whereas the GA method required 80 h due to its inherent online optimization process. This significant difference in computational time demonstrates that our method is particularly well-suited for applications requiring continuous and rapid results. Furthermore, even on a resource-constrained embedded platform (STM32F103ZET6), the policy network can perform 26 inferences per second.

C. Control Experiment Results and Analysis

Although the optimized policy network was tested extensively in simulations, it is crucial to assess the control precision of the entire framework in practice. Therefore, the generated policy network was deployed on an actual system (as shown in Fig. 2), and two control experiments were conducted. The control trajectories and corresponding lighting performance indices of these experiments are illustrated in Fig. 11. Both experiments followed the same target trajectory, with only the ambient light varying in different profiles. The experiments consisted of four stages: 1) maintaining a constant CCT of 2500 K while increasing intensity from 40 lm to 90 lm in 10 lm steps; 2) maintaining a constant intensity of 90 lm while increasing CCT from 2500 K to 6500 K in 500 K steps; 3) maintaining a constant CCT of 6500 K while decreasing intensity from 90 lm to 40 lm in 10 lm steps; and 4) maintaining a constant intensity of 70 lm while decreasing CCT from 6500 K to 4000 K in 500 K steps, followed by a constant state until the experiment ended. Each step lasted 10 s, with a control interval of 1 s to better display the control effects.

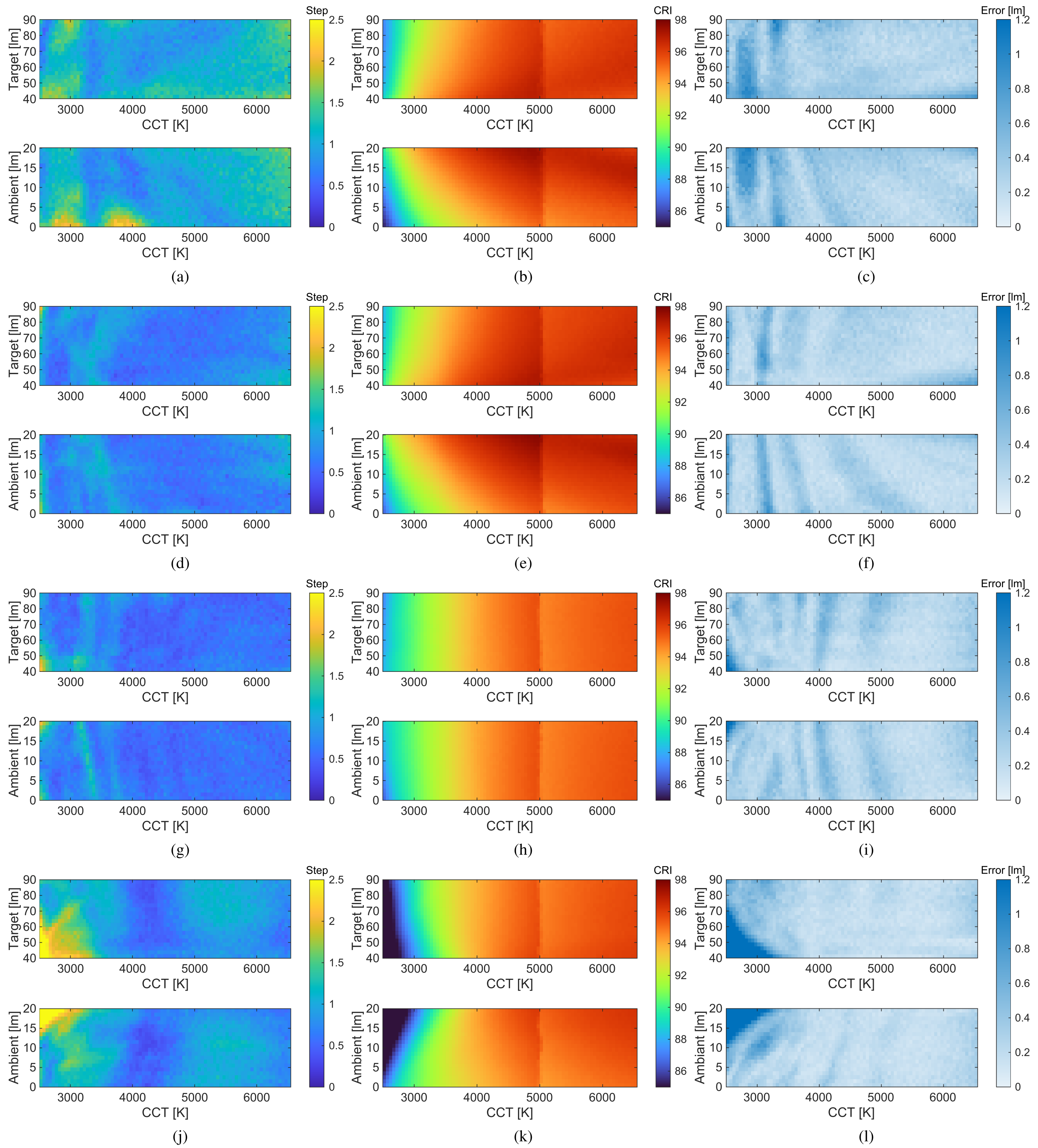
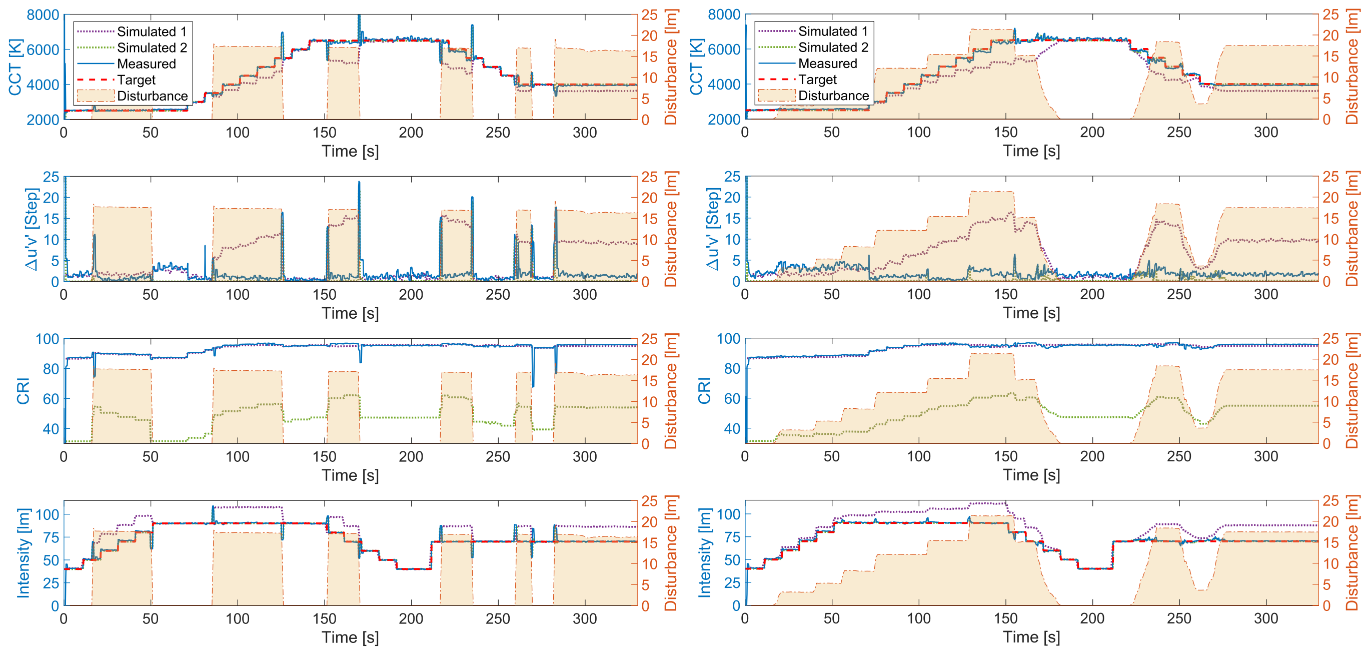


Fig. 10. Average color difference (first column), CRI (second column), and intensity error (third column) results for experimental groups: (a)-(c) halogen random group, (d)-(f) halogen grid group, (g)-(i) 4100 K LED grid group, and (j)-(l) D65 grid group. The top row of subfigures shows results averaged by ambient interference, while the bottom row shows results averaged by the set target intensity.

In the first experiment, the ambient interference remained at a fixed intensity of 17 lm, but it was switched ON and OFF at random time instants, as shown by the yellow shaded area with a dotdash line in Fig. 11(a). From the top subplot of Fig. 11(a), it could be seen that the CCT tracking curve nearly overlapped with the target curve. The corresponding color difference and CRI results are shown in the middle subplot. It was evident that the color difference was below 3 Step during most of the steady stages, with only one stage reaching close to 5 Step (2500 K, 90 lm at 50–70 s). This aligns with the CCT tracking results, where only



(a) (b)

Fig. 11. Control results of the interference experiments. The simulated influence of the ambient light is represented by the purple dotted lines. The simulated control results based on the three-channel RGB system and color theory are plotted as the green dotted lines. The measured results are shown as blue lines, while the target values are indicated by red dashed lines. The ambient intensities are depicted by the shaded yellow areas with dotdash lines. The interference intensity curves correspond to the right vertical axis, whereas all other parameters refer to the left vertical axis.

transitions or sudden interference cause transient deviations, indicating that the tracking is highly accurate. The CRI value remained above 95 for most of the time and above 85 during all steady stages. The intensity tracking was similarly consistent with the set trajectory. Notably, both the color difference and intensity control trajectories exhibit unwanted peaks when the interference occurs, though the controller successfully adjusted the output back to the optimal result within a few seconds. In terms of quantification, the average CRI, color difference, intensity error, and reward during the entire control process were 93.551, 1.698 Step, 1.173 lm, and 90.040.

To demonstrate the influence of ambient light, a simulated control scenario was conducted by directly adding ambient light to the policy network's output, while disabling ambient light consideration in the network input (i.e., by setting the input parameter $\alpha_d = 0$). This configuration emulates a light-source-embedded controller that maintains output according to target parameters while ignoring ambient illumination. The simulation results are depicted by the purple lines in Fig. 11(a). It was evident that the controller performed satisfactorily in the absence of ambient light. However, once ambient light is introduced and the target CCT differs from that of the ambient light (i.e., 2600 K), significant tracking deviations emerged in the CCT and the corresponding color difference due to the mismatch between their CCT values. In addition, noticeable intensity errors occurred regardless of whether the target CCT aligns with the ambient or not. Quantitative analysis revealed that the average CRI, color difference, intensity error, and reward during the simulated control process were 93.443, 4.985 Step, 8.956 lm, and 72.637, respectively. As expected, most of the unwanted peaks in the CCT, color difference, and intensity trajectories observed in the real control experiment (blue lines) coincide with those in the simulation (purple lines), further indicating that ambient light is the primary source of error.

Given the significant impact of sudden interference on control performance, a second experiment was conducted. In this case, the ambient interference was gradually tuned up or down [as shown in Fig. 11(b)]. In this experiment, the unwanted peaks in the CCT tracking, color difference, and intensity were smaller compared to the first experiment. This directly suggested that the peaks in the first experiment were primarily caused by the abrupt changes in the interference. Quantitatively, the average CRI, color difference, intensity error, and reward during the entire control process were 93.640, 1.844 Step, 0.833 lm, and 91.733. The influence of ambient light in the second experiment was also simulated. It was observed that the tracking errors gradually varied in response to the ambient light conditions. The computed results showed that the average CRI, color difference, intensity error, and reward during the simulated control process were 93.334, 6.457 Step, 10.460 lm, and 67.478, respectively. It is evident that the impact of ambient light, when not considered in the control strategy, was even greater than in the first simulation, although the controller's performance under actual conditions was comparatively less affected. Moreover, while ambient light with a high CRI (close to 100) exhibited minimal influence on CRI in both simulations, the CRI performance remained inferior to scenarios where ambient light is explicitly considered in the control strategy. A lower CRI in the ambient light could further degrade the resulting CRI performance. Therefore, it is essential

TABLE IV
AVERAGE AND WORST-CASE RESULTS FOR RGB AND RGBA LED EXPERIMENTS

Group	Color Difference (Step)		CRI		Intensity Error (lm)		Average Reward
	Avg.	Max.	Avg.	Min.	Avg.	Max.	
XYZ26	0.001	0.007	51.596	31.899	0.011	0.020	51.586
XYZ41	0.013	9.753	54.584	31.899	0.008	0.203	54.565
XYZ65	0.201	25.669	49.338	31.899	0.011	0.556	48.823
RL26	1.581	4.472	52.587	31.774	0.172	1.115	51.723
RL41	1.229	9.860	55.286	31.825	0.209	1.267	54.628
RL65	1.479	15.859	49.953	31.905	0.442	15.859	48.830
RGBA65	0.907	7.286	86.815	79.188	0.376	27.550	86.161

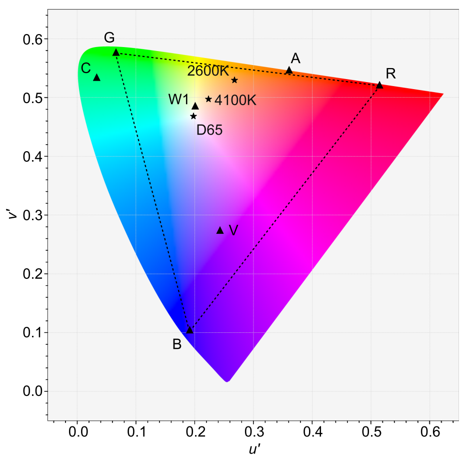


Fig. 12. Gamut of the RGB LED source and the chromaticity coordinates of all spectra used in this article.

to account for the influence of ambient light in the control strategy.

Based on the above analysis, it can be concluded that both experiments demonstrated satisfactory control performance in real systems. Comparing the results of the two experiments, it is clear that abrupt disturbances have a much larger impact on the lighting performance. However, regardless of how the interference is introduced or the system's control target, the controller consistently achieves the desired control objectives.

D. RGB LED System Results and Analysis

To further investigate the observed increase in color difference and intensity error under conditions of lower target CCT and intensity with increasing ambient CCT and intensity, a simplified experiment was conducted using only the RGB channels. In this configuration, the output of the three channels can be directly calculated based on color theory [17], without considering CRI. The chromaticity coordinates corresponding to the SPDs of all seven channels and the ambient lights were plotted in the CIE 1976 $u'v'$ color space, as shown in Fig. 12. It is evident that the RGB channels can cover most of the color gamut similarly to the seven-channel system.

To avoid insufficient optical power, the power of the RGB channels was tripled, and the results were calculated under three ambient conditions using the same procedure. The detailed outcomes are summarized in Table IV. It was clear that both color difference and intensity error increased with rising ambient CCT. No significant limitations were observed when the ambient CCT was 2600 K (XYZ26 group); minor limitations appeared under 4100 K ambient light (XYZ41 group); and the most pronounced limitations occurred under 6500 K ambient light (XYZ65 group). The data further showed that the worst-case scenario consistently arose when the ambient luminous flux was 20 lm and the target flux was 40 lm. As a representative example, detailed results for the XYZ65 group were presented in Fig. 13. These results closely resembled those in Figs. 10(j)–(l), clearly demonstrated that high-CCT ambient lighting introduces unavoidable errors, even when only color and intensity (excluding CRI) were considered.

Under the same experimental conditions, the proposed RL method employing a grid-based training strategy was applied to optimize the RGB lighting system, with CRI included in the objective function. The training parameters remained consistent with those used in previous sections, and training was terminated after 40 000 episodes. The corresponding results are presented in Table IV. The results indicated that the RL method achieved slightly improved results than the color-theory-based approach, which does not take CRI into account. This suggests that while the RL framework is applicable to lighting systems with arbitrary number of channels, the optimization potential becomes limited when fewer than four channels are available. This finding is consistent with the general understanding in the field that at least four channels are typically required to effectively optimize CRI in systems with a wide CCT tuning range. The optimal results for the RGBA channel system (where each channel's SPD was doubled) under D65 ambient light conditions were also included in Table IV as the RGBA65 group. The final performance metrics, particularly the CRI and average reward, were significantly improved (average reward = 86.161), surpassing even the GA6500 K group, which used a seven-channel configuration and achieved an average reward of 85.579. Thus, using more than three channels becomes essential when optimizing many parameters. Alternatively, channel SPDs may be preoptimized to enable satisfactory performance using color-theory-based methods, as demonstrated in [48], [49].

In addition, the simulated control results for the three-channel RGB system using the color theory-based method [17] are depicted as green dotted lines in Fig. 11, providing a direct comparison with the RL-based control strategy. For the first control experiment, the average CRI, color difference, intensity error, and reward were 47.976, 0.441 Step, 0.667 lm, and 46.036, respectively. In the second control experiment, these metrics improved to 48.922, 0.274 Step, 0.383 lm, and 48.440. These results are consistent with the data shown in Table IV, confirming

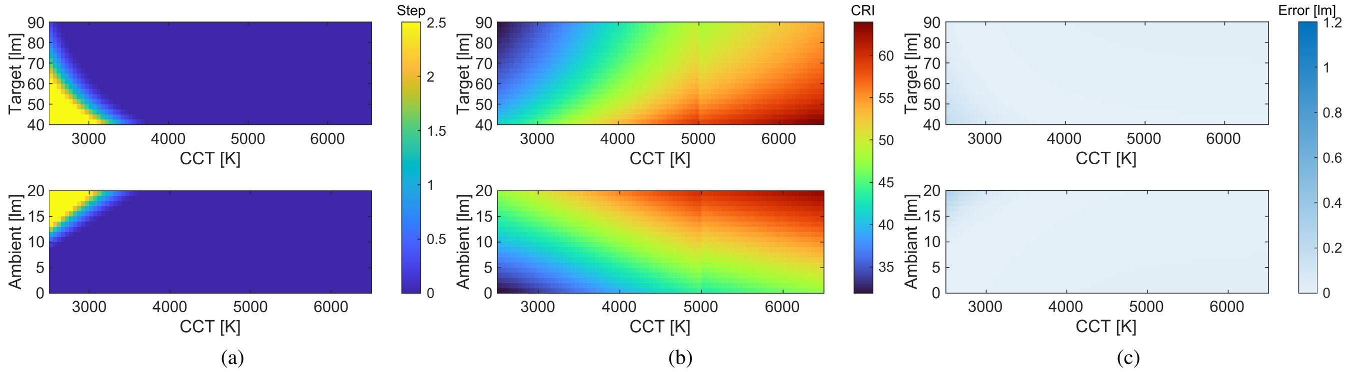


Fig. 13. Average (a) color difference, (b) CRI, and (c) intensity error results for XYZ65 experimental group. The top row of subfigures shows results averaged by ambient interference, while the bottom row shows results averaged by the set target intensity.

the limited CRI performance of the conventional method and limited channels. It should be noted that the nonlinear PET responses of the RGB LEDs was not handled; while the sensor errors were considered using measured data. In real-world conditions, both color differences and intensity errors would likely be worse.

V. Conclusion

In this article, a RL-based multichannel spectral optimization and control framework has been proposed. By comprehensively considering ambient interference during both the modeling and training stages, the proposed framework can successfully perform highly complex nonlinear optimization and control tasks in the presence of ambient light interference. Moreover, the convergence of the control law has been analyzed. The effectiveness of the proposed method has been tested on a seven-channel LED control and measurement system, performing a challenging lighting control task that requires balancing CRI, color difference, and intensity error, under the interference of a halogen lamp.

Due to the limitations of the sensors used, it is difficult to accurately detect rapidly changing ambient spectra, making it challenging to reject their effects. To address this, integrating micro-spectrometers into the framework is a promising direction, enabling real-time monitoring of both ambient spectral variations and LED SPD shifts caused by PET effects or device aging. In addition, an adaptive controller that dynamically adjusts control gains in response to the nonlinear PET behavior can overcome the limitations of the current low-level controller. By combining adaptive control for low-level dynamic compensation with RL for high-level multiobjective optimization, the system may achieve enhanced robustness and stability under dynamic and unpredictable lighting conditions.

Appendix A
Derivation of the Nonlinear Irradiance Deviation

By (4), the nonlinearity can be separated into two components, i.e., the influence of the self-input and that from the other channels. This equation can be readily derived based on the PET model of [8] and [44] as follows:

$$T_{j,i} = T_a + R_{hs}\sum_{l=1}^{N}(P_l^{heat} \cdot p_l) + R_{jc,i}P_i^{heat} \cdot p_i, \tag{32}$$

$$\begin{aligned}\frac{\alpha_i}{p_i} &= 1 + k_{e,i}(T_{j,i} - T_0)\\ &= 1 + [\underbrace{k_{e,i}(T_a - T_0)}_{\chi} + \underbrace{(R_{hs} + R_{jc,i})k_{e,i}P_i^{heat}}_{-\gamma} \cdot p_i\\ &\quad \underbrace{+ R_{hs}k_{e,i}\sum^{l\neq i}(P_l^{heat} \cdot p_l)}_{-\zeta(p_o)}].\end{aligned} \tag{33}$$

Here, $T_{j,i}$ and T_a represent the steady-state junction temperature and the ambient temperature, respectively. R_{hs} and R_{jc} denote the thermal resistance of the corresponding parts; while P_l^{heat} and P_i^{heat} represent the heat power during full power-operation of each channel. T_0 is the average junction temperature; and $k_{e,i}$ is the relative rate of reduction in efficacy with increasing temperature for each channel, i.e., $k_{e,i} < 0$.

For a well-dissipating LED, the values of R_{jc}, R_{hs}, P_l^{heat}, P_i^{heat}, and $k_{e,i}$ are generally considered as constant. Therefore, χ and γ are approximately constant. Comparing (33) with (3), the nonlinear irradiance deviation can be simplified as (4).

Appendix B
Influence of Network Size

The actor and critic network sizes were selected based on pre-experiments under the 2600 K group setting. Each network maintained a consistent internal structure, with hidden layers scaled proportionally. The training trajectories of average reward are shown in Fig. 14. The results show that performance improves with larger network sizes. However, beyond an FC1 size of 96, the final reward plateaus while computational costs continue to rise. Convergence speed and training stability generally trade off—larger networks converge faster but with less stability. An FC1 size of 128 was chosen to balance nonlinear modeling capability under different ambient light conditions and acceptable training time. This choice was made without parallel

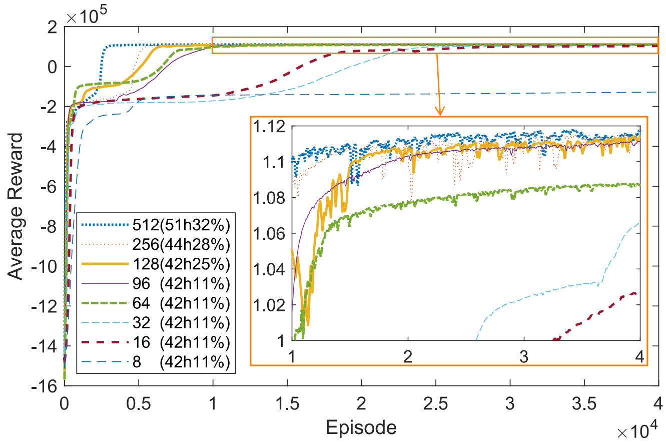


Fig. 14. Preexperiment results. Average reward trajectories for networks of varying sizes, labeled by the size of the FC1 layer. Corresponding training time (e.g., 42 h) and CPU usage (e.g., 11%) are indicated in parentheses.

training. With parallelization, smaller networks may be preferable for faster convergence. Moreover, for platforms with more limited resources than the STM32, smaller networks or pruning techniques could be explored to further reduce computational overhead.

References

[1] E. F. Schubert and J. K. Kim, "Solid-state light sources getting smart," *Science*, vol. 308, no. 5726, pp. 1274–1278, 2005.

[2] T.-Y. Seong, J. Han, H. Amano, and H. Morkoç, *III-Nitride Based Light Emitting Diodes and Applications*, vol. 133. Singapore: Springer, 2017.

[3] S. Martins, L. Roggia, M. Perdigao, and A. Seidel, "Planar magnetic regulator applied to an LLC converter for CCT control," *IEEE Trans. Power Electron.*, vol. 39, no. 10, pp. 13300–13308, Oct. 2024.

[4] Y. Wang, J. Xu, Z. Chen, and F. Qin, "Small signal modeling and transient cross-regulation analysis of peak-current-mode-controlled SIDO buck LED driver," *IEEE Trans. Power Electron.*, vol. 38, no. 10, pp. 12834–12845, Oct. 2023.

[5] Y. Qu and L. Qiu, "I2V2 average current control for modular LED drivers," *IEEE Trans. Power Electron.*, vol. 36, no. 1, pp. 78–82, Jan. 2021.

[6] A. K. Pal and A. K. Singha, "A digital pulse frequency modulation technique for a series resonant converter-based LED driver," *IEEE Trans. Power Electron.*, vol. 40, no. 2, pp. 2797–2808, Feb. 2025.

[7] Y. Huang and D. B. Ma, "A smart silicon carbide LED driver IC with integrated dual-level condition-monitoring mechanism," *IEEE Trans. Power Electron.*, vol. 39, no. 5, pp. 6246–6255, May 2024.

[8] C. P. G. Wong, A. T. L. Lee, K. Li, S.-C. Tan, and S. Y. Hui, "Precise luminous flux and color control of dimmable red-green-blue light-emitting diode systems," *IEEE Trans. Power Electron.*, vol. 37, no. 1, pp. 588–606, Jan. 2022.

[9] H.-T. Chen, S.-C. Tan, and S. Y. Hui, "Nonlinear dimming and correlated color temperture control of bi-color white LED systems," *IEEE Trans. Power Electron.*, vol. 30, no. 12, pp. 6934–6947, Dec. 2015.

[10] H.-T. Chen, D.-Y. Lin, S.-C. Tan, and S. Y. Hui, "Chromatic, photometric and thermal modeling of LED systems with nonidentical LED devices," *IEEE Trans. Power Electron.*, vol. 29, no. 12, pp. 6636–6647, Dec. 2014.

[11] Q. Wang and J. Dong, "Data-driven modeling of the thermal dynamics of a dual-spectral ultraviolet sterilizer and control of the output irradiance," *Appl. Therm. Eng.*, vol. 212, 2022, Art. no. 118615.

[12] C.-W. Tang, B.-J. Huang, and S.-P. Ying, "Illumination and color control in red-green-blue light-emitting diode," *IEEE Trans. Power Electron.*, vol. 29, no. 9, pp. 4921–4937, Sep. 2014.

[13] H.-T. Chen, S.-C. Tan, A. T. L. Lee, D.-Y. Lin, and S. Y. R. Hui, "Precise color control of red-green-blue light-emitting diode systems," *IEEE Trans. Power Electron.*, vol. 32, no. 4, pp. 3063–3074, Apr. 2017.

[14] I.-T. Kim, A.-S. Choi, and J.-W. Jeong, "Precise control of a correlated color temperature tunable luminaire for a suitable luminous environment," *Build. Environ.*, vol. 57, pp. 302–312, 2012.

[15] K. C. Lin, "Approach for optimization of the color rendering index of light mixtures," *J. Opt. Soc. Am A*, vol. 27, no. 7, pp. 1510–1520, 2010.

[16] L. Lohaus, E. Leicht, S. Dietrich, R. Wunderlich, and S. Heinen, "Advanced color control for multicolor LED illumination systems with parametric optimization," in *Proc. IECON*, 2013, pp. 3305–3310.

[17] N. Ohta and A. Robertson, *Colorimetry: Fundamentals and Applications*. Wiley, 2005.

[18] H. Wu, J. Dong, and G. Zhang, "Increasing color saturation by optimizing light spectra constrained on color rendering properties," *J. Opt. Soc. Am. A*, vol. 33, no. 2, pp. 192–204, 2016.

[19] H. Wu, J. Dong, G. Qi, and G. Zhang, "Optimization of LED light spectrum to enhance colorfulness of illuminated objects with white light constraints," *J. Opt. Soc. Am A*, vol. 32, no. 7, pp. 1262–1270, 2015.

[20] M. T. Imam, S. Afshari, and S. Mishra, "An experimental survey of feedback control methodologies for advanced lighting systems," *Energy Build.*, vol. 130, pp. 600–612, 2016.

[21] F. Zhang, H. Xu, and Z. Wang, "Optimizing spectral compositions of multichannel LED light sources by IES color fidelity index and luminous efficacy of radiation," *Appl. Opt.*, vol. 56, no. 7, pp. 1962–1971, 2017.

[22] W. Hertog, A. Llenas, and J. Carreras, "Optimizing indoor illumination quality and energy efficiency using a spectrally tunable lighting system to augment natural daylight," *Opt. Exp.*, vol. 23, no. 24, pp. A1564–A1574, 2015.

[23] X. Zhan, W. Wang, and H. S.-H. Chung, "A novel color control method for multicolor LED systems to achieve high color rendering indexes," *IEEE Trans. Power Electron.*, vol. 33, no. 10, pp. 8246–8258, Oct. 2018.

[24] A. Llenas and J. Carreras, "Arbitrary spectral matching using multi-LED lighting systems," *Opt. Eng.*, vol. 58, no. 3, pp. 35105–35105, 2019.

[25] J. Zhang et al., "Blue light hazard optimization for white light-emitting diode sources with high luminous efficacy of radiation and high color rendering index," *Opt. Laser Technol.*, vol. 94, pp. 193–198, 2017.

[26] J. Nie et al., "Wide color gamut, low non-visual effect, and their stabilities related to luminous properties in four-primary-color display," *Opt. Laser Technol.*, vol. 156, 2022, Art. no. 108565.

[27] J. Nie et al., "Utilization of far-red LED to minimize blue light hazard for dynamic semiconductor lighting," *Leukos.*, vol. 19, no. 1, pp. 53–70, 2023.

[28] G. He, J. Xu, and H. Yan, "Spectral optimization of warm-white light-emitting diode lamp with both color rendering index (CRI) and special CRI of R9 above 90," *AIP Adv.*, vol. 1, no. 3, 2011, Art. no. 032160.

[29] P. Zhong, G. He, and M. Zhang, "Optimal spectra of white light-emitting diodes using quantum dot nanophosphors," *Opt. Exp.*, vol. 20, no. 8, pp. 9122–9134, 2012.

[30] P. Zhong, G. He, and M. Zhang, "Spectral optimization of the color temperature tunable white light-emitting diode (LED) cluster consisting of direct-emission blue and red LEDs and a diphosphor conversion LED," *Opt. Exp.*, vol. 20, no. 105, pp. A684–A693, 2012.

[31] J. Zhang, R. Hu, X. Yu, B. Xie, and X. Luo, "Spectral optimization based simultaneously on color-rendering index and color quality scale for white LED illumination," *Opt. Laser Technol.*, vol. 88, pp. 161–165, 2017.

[32] Z. Guo et al., "Investigation on three-hump phosphor-coated white light-emitting diodes for healthy lighting by genetic algorithm," *IEEE Photon. J.*, vol. 11, no. 1, Feb. 2019, Art. no. 8200110.

[33] Y. J. Saw, V. Kalavally, and C. P. Tan, "The spectral optimization of a commercializable multi-channel LED panel with circadian impact," *IEEE Access*, vol. 8, pp. 136498–136511, 2020.

[34] M. Marín-Doñágueda et al., "Simultaneous optimization of circadian and color performance for smart lighting systems design," *Energy Build.*, vol. 252, 2021, Art. no. 111456.

[35] M. J. Murdoch, "Dynamic color control in multiprimary tunable LED lighting systems," *J. Soc. Inf. Disp.*, vol. 27, no. 9, pp. 570–580, 2019.

[36] M. J. Murdoch, "Characterization and control of a multi-primary LED light lab," *Opt. Exp.*, vol. 25, no. 24, pp. 29605–29616, 2017.

[37] Q. Wang and J. Dong, "Reinforcement learning control of multi-spectral LED systems for rendering optimal lighting performance in indoor environment," *Build. Environ.*, vol. 262, 2024, Art. no. 111811.

[38] A. Goudjil et al., "An interior-points algorithm for color and CCT control of multichannel LED lighting system using a smart 18-channel spectral sensor," *Energy Build.*, vol. 298, 2023, Art. no. 113541.

[39] J. Dong and G. Zhang, "Identification and robust control of the nonlinear photoelectrothermal dynamics of LED systems," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2215–2225, Mar. 2017.

[40] P. Maiti and B. Roy, “Evaluation of a daylight-responsive, iterative, closed-loop light control scheme,” *Light. Res. Technol.*, vol. 52, no. 2, pp. 257–273, 2020.

[41] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *CoRR*, vol. abs/1707.06347, 2017.

[42] H. Morko, *Handbook of Nitride Semiconductors and Devices: Materials Properties, Physics and Growth*. Wiley, Mar. 2008, doi: 10.1002/9783527628438.

[43] M. Dyble, N. Narendran, A. Bierman, and T. Klein, “Impact of dimming white LEDs: Chromaticity shifts due to different dimming methods,” in *Proc. 5th Int. Conf. Solid State Lighting*, SPIE, 2005, vol. 5941, pp. 291–299.

[44] S. Y. Hui and Y. X. Qin, “A general photo-electro-thermal theory for light emitting diode (LED) systems,” *IEEE Trans. Power Electron.*, vol. 24, no. 8, pp. 1967–1976, Aug. 2009.

[45] International Commission on Illumination, *CIE 013.3-1995 Method of Measuring and Specifying Colour Rendering Properties of Light Sources*. Vienna, Austria: CIE, 1995.

[46] International Commission on Illumination, *CIE TN 001:2014 Chromaticity Difference Specification for Light Sources*. Vienna, Austria: CIE, 2014.

[47] P. Myland, S. Babilon, T. Hegemann, S. Klir, and T. Q. Khanh, “Reconstruction of spectral irradiance in a real application with a multi-channel spectral sensor using convolutional neural networks,” *Opt. Exp.*, vol. 31, no. 16, pp. 25724–25746, 2023.

[48] L.-L. Zheng et al., “Spectral optimization of three-primary LEDs by considering the circadian action factor,” *IEEE Photon. J.*, vol. 8, no. 6, 2016, Art. no. 8200209.

[49] F. Zhang, “Peak wavelength selection of chips for three-chip LED light sources with high color fidelity,” *Optik*, vol. 224, 2020, Art. no. 165725.

Qi Wang received the Ph.D. degree in biomedical engineering from the University of Science and Technology of China (USTC), Suzhou, China, in 2025.

From 2023 to 2025, he was a Visiting Research Student with the School of Future Science and Engineering, Soochow University, Suzhou, China. His current research interests include data-driven modeling, optimization, and control of optoelectromechanical systems.

Jianfei Dong received the Ph.D. degree in systems and control from the Delft University of Technology (TUD), Delft, The Netherlands, in 2009.

From 2009 to 2011, he was a Postdoctoral Researcher with TUD. Then, he worked as a Research Scientist with Philips Research, Eindhoven, the Netherlands. From 2012 to 2023, he was with State Key Lab of Solid State Lighting and Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, respectively. Since 2023, he has been a Professor with the School of Future Science and Engineering, Soochow University, Suzhou, China. The current applications focus on LED light sources, optical therapies and diagnosis. His current research interests include data-driven modeling, control and fault diagnosis of complex dynamic systems, e.g., biological and optoelectronic systems, and AI learning techniques for image analysis.