SLIPT in Joint Dimming Multi-LED OWC Systems with Rate Splitting Multiple Access

Sepideh Javadi[†], Sajad Faramarzi[§], Farshad Zeinali[†], Hosein Zarini[§§], Mohammad Robat Mili[†], Panagiotis D. Diamantoulakis[⋆], Eduard Jorswieck[††], George K. Karagiannidis[⋆,⋆⋆]

[†] Pasargad Institute for Advanced Innovative Solutions (PIAIS), Tehran, Iran
[§] Dept. of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran
[§§] Dept. of Computer Engineering, Sharif University of Technology, Tehran, Iran
[⋆] Dept. of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
[⋆⋆] Artificial Intelligence & Cyber Systems Research Center, Lebanese American University (LAU), Lebanon
[††] Institute for Communications Technology, Technische Universitat Braunschweig, Braunschweig, Germany

ICC 2024 - IEEE International Conference on Communications | 978-1-7281-9054-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICC51166.2024.10622982

***Abstract*—Optical wireless communication (OWC) systems with multiple light-emitting diodes (LEDs) have recently been explored to support energy-limited devices via simultaneous lightwave information and power transfer (SLIPT). The energy consumption, however, becomes considerable by increasing the number of incorporated LEDs. This paper proposes a joint dimming (JD) scheme that lowers the consumed power of a SLIPT-enabled OWC system by controlling the number of active LEDs. We further enhance the data rate of this system by utilizing rate splitting multiple access (RSMA). More specifically, we formulate a data rate maximization problem to optimize the beamforming design, LED selection and RSMA rate adaptation that guarantees the power budget of the OWC transmitter, as well as the quality-of-service (QoS) and an energy harvesting level for users. We propose a dynamic resource allocation solution based on proximal policy optimization (PPO) reinforcement learning. In simulations, the optimal dimming level is determined to initiate a trade-off between the data rate and power consumption. It is also verified that RSMA significantly improves the data rate.**

***Index Terms*—Optical wireless communication (OWC), simultaneous lightwave information and power transfer (SLIPT), joint dimming (JD), rate splitting multiple access (RSMA), proximal policy optimization (PPO).**

I. Introduction

Benefiting from illumination and communication at the same time, optical wireless communication (OWC) systems are envisioned to be a promising technology to compensate for the shortcomings of conventional radio-frequency (RF) communication systems. This technology requires a light-emitting diode (LED) for signal transmission at the transmitter and photodiode (PD) for signal decoding at the receiver [1]–[4]. So far, advanced optical wireless techniques have been investigated to unleash the potentials of OWC systems, such as multi-LED transmitters [5], joint dimming (JD) [6], as well as simultaneously lightwave information and power transfer (SLIPT) [7]–[9]. In the former case, multiple LEDs are incorporated in an LED array, commonly known as a spatial multiplexing OWC system. This extension brings about remarkable data rate and extended coverage over the OWC systems compared to the single LED. The second case, as a multi-domain control scheme, relies on both analog dimming (AD) and spatial dimming (SD).

Introduced first in [6], JD control jointly optimizes the direct-current (DC) bias level (in AD), as well as the number of glared LEDs (in SD) to satisfy a required dimming level [10], [11]. The latter case, i.e., the SLIPT technology, is a promising solution for low-battery devices [7], which enables OWC receivers to simultaneously obtain illumination, information, and energy harvesting (EH) via a PD, a solar panel, or both.

The concept of wireless information and power transfer has been throughly investigated in the literature [12]–[16]. However, the coexistence of alternating current (AC) and DC signals in the photo-current of LEDs has encouraged the researchers to study the SLIPT technology [17]. Previously, the performance of an OWC system with multi-LED transmitter and SLIPT technology was considered in [18] and [19] to minimize the energy consumption and maximize the data rate, respectively. These techniques consume significant energy. Furthermore, they were proposed under the assumption of orthogonal resource allocation which limits the data rate. In this paper, we study the performance of [18] and [19] by exploring rate splitting multiple access (RSMA) [20]–[23], which employs non-orthogonality of resources to enable higher data rate and massive access. Additionally, we adopt an efficient JD control scheme, where the number of active LEDs are controlled, thereby reducing the energy consumption of [18] and [19].

To analyze the performance of this system, a data rate maximization problem is formulated by jointly optimizing the transmit beamforming, LED selection, and RSMA rate adaptation. This problem ensures the power budget of the multi-LED transmitter, as well as receivers' quality-of-service (QoS), dimming level and EH level. More specifically, we transform the problem into a Markov decision process (MDP) and propose a real-time dynamic solution methodology, based on proximal policy optimization (PPO) reinforcement learning. Through numerical simulations, we compute the optimal dimming level

978-1-7281-9054-9/24/$31.00 ©2024 IEEE

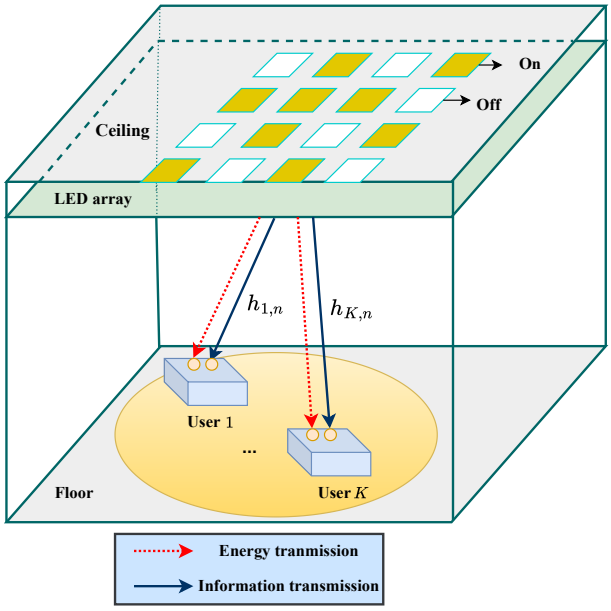

Fig. 1: An OWC system with an LED array, serving K single-PD users for illumination, communication and EH.

for a trade-off between the data rate and power consumption. Additionally, leveraging RSMA enhances the data rate of this system, compared to the non-orthogonal multiple access (NOMA).

The remainder of this paper is organized as follows. The system model of the proposed downlink SLIPT-assited JD multi-LED OWC network with RSMA is introduced in Section II, while Section III evaluates the performance of the proposed DRL-based approach in detail by employing PPO reinforcement learning. In Section IV, the effectiveness of the proposed OWC system is verified by simulation results. Finally, Section V concludes the paper.

Notation: The vectors and matrices are specified by boldface lower-case and upper-case letters, respectively, while diag(.) represents the diagonalization operation. The absolute value of scalar a, the transpose, and the Hermitian transpose of the vector $\mathbf{a}$ are denoted by $|a|$, $\mathbf{a}^T$, and $\mathbf{a}^H$, respectively. Finally, the expectation operator and the set of real numbers are denoted by $\mathbb{E}(.)$ and $\mathbb{R}$, respectively, while $[.]$ represents the round off operation.

II. System Model

As illustrated in Fig. 1, we consider the downlink transmission of an OWC network, in which the transmitter is equipped with an LED array, including a set $\mathcal{N} = \{1, 2, ..., N\}$ of N LEDs and communicates with a set $\mathcal{K} = \{1, 2, ..., K\}$ of K single-PD users. The users are supposed to be distributed randomly, whereas the LEDs are independently modulated via separate drivers, yet all are connected to a central controller that collects channel feedback and performs resource management.

A. Signal Model

We adopt RSMA as the state-of-the-art multiple access scheme. On this basis, the transmit lightweight data stream for each user has a two-fold structure, including a common message in addition to a private message. The former, i.e., the common message of the transmit lightweight data stream has the same content for all users, whereas the latter, i.e., the private message is exclusively encoded for each user. In other words, one common lightweight transmit data stream, as well as K private ones, form a superimposed transmit signal, carrying $K+1$ messages. Let $s^{(\mathrm{c})}$ and $s_k^{(\mathrm{p})}, \forall k \in \mathcal{K}$, denote the common message shared among all users and the private message of the k-th user, respectively, such that $\mathbb{E}\{|s^{(\mathrm{c})}|^2\} = 1$ and $\mathbb{E}\{|s_k^{(\mathrm{p})}|^2\} = 1, \forall k \in \mathcal{K}$. We devise a linear precoding scheme [5] prior to signal transmission, to handle the inter-user interference. Next, a DC bias $\mathbf{i}_{\mathrm{DC}} = [i_{\mathrm{DC}}, ..., i_{\mathrm{DC}}]^{N\times 1}$ is added to the precoded signal before transmission. This bias regulates the brightness of the LEDs and guarantees that the amplitude of the transmitted signal has a real non-negative value. Accordingly, the lightweight transmit data stream of all LEDs, denoted by $\mathbf{x} = [x_1, x_2, ..., x_N]^T$, can be expressed as

$$\mathbf{x} = \mathbf{w}^{(\mathrm{c})} s^{(\mathrm{c})} + \sum_{k=1}^{K} \mathbf{w}_k^{(\mathrm{p})} s_k^{(\mathrm{p})} + \mathbf{i}_{\mathrm{DC}}, \tag{1}$$

where $\mathbf{w}^{(\mathrm{c})} = \left[w_1^{(\mathrm{c})}, w_2^{(\mathrm{c})}, ..., w_N^{(\mathrm{c})}\right]^T \in \mathbb{R}^{N\times 1}$ and $\mathbf{w}_k^{(\mathrm{p})} = \left[w_{k,1}^{(\mathrm{p})}, w_{k,2}^{(\mathrm{p})}, ..., w_{k,N}^{(\mathrm{p})}\right]^T \in \mathbb{R}^{N\times 1}, \ \forall k \in \mathcal{K}$ specify the common transmit beamforming for $s^{(\mathrm{c})}$ and the k-th private transmit beamforming for $s_k^{(\mathrm{p})}$, respectively. Note that $\mathbf{i}_{\mathrm{DC}}$ has the same value for all LEDs, because of uniformity of illumination in indoor environments [5]. The dynamic range of the LEDs is constraint to avoid signal clipping [27]. In other words,

$$\left|w_n^{(\mathrm{c})}\right| + \sum_{k=1}^{K} \left|w_{k,n}^{(\mathrm{p})}\right| \leq \Xi, \ \forall n \in \mathcal{N}, \tag{2}$$

where $\Xi = \min(i_{\mathrm{DC}} - I_l, I_h - i_{\mathrm{DC}})$, and the notations I_l and I_h indicate the minimum and maximum permissible currents of all LEDs, respectively.

B. Channel Model

In this paper, we only consider the line of sight (LoS) links [6]- [11]. The optical channel gain between the n-th LED and the k-th user, denoted by $h_{k,n} \in \mathbb{R}$, can be modelled as

$$h_{k,n} = \begin{cases} \frac{(m+1)A^{\mathrm{OWC}}}{2\pi d_{k,n}^2} G^{\mathrm{OWC}}(\psi_{k,n}) Z^{\mathrm{OWC}}, & 0 \leq \psi_{k,n} \leq \Psi_c, \\ 0, & \psi_{k,n} > \Psi_c, \end{cases} \tag{3}$$

where $Z^{\mathrm{OWC}} = \cos^m(\phi_{k,n})\cos(\psi_{k,n})$, while A^{OWC} and $d_{k,n}$ denote the physical area of the PD for each user and the distance between the n-th LED and the k-th user, respectively. Moreover, $m = -\frac{\ln 2}{\ln(\cos \Phi_{1/2})}$ specifies the Lambertian emission order with $\Phi_{1/2}$ being the semi-angle at half-power of the LED. Besides, the angles of incidence and irradiance are respectively given by

$\psi_{k,n}$ and $\phi_{k,n}$, while the receiver field of vision (FOV) semi-angle is denoted by Ψ_c. Finally, $G^{\text{OWC}}(\psi_{k,n})$ indicates the gain of the optical concentrator which is defined as follows

$$G^{\text{OWC}}(\psi_{k,n}) = \begin{cases} \frac{n_R^2}{\sin^2(\Psi_c)} & 0 \le \psi_{k,n} \le \Psi_c, \\ 0, & \psi_{k,n} > \Psi_c, \end{cases} \quad (4)$$

with $n_R \ge 0$ being the internal refractive index. Given (x_k, y_k, z_k) and (x_n, y_n, z_n) as the coordinates of the k-th user and n-th LED, respectively, the distance between them can be modeled as $d_{k,n} = \sqrt{(x_n - x_k)^2 + (y_n - y_k)^2 + (z_n - z_k)^2}$.

C. JD Control

Regarding the LED array with multiple LEDs and the incorporation of power transfer capability, its energy consumption becomes considerable. In this paper we invoke an efficient JD scheme to control the energy consumption [10]. To this purpose, let $\mathbf{A}$ denote a binary LED selection matrix, where $\mathbf{A} = \text{diag}(\mathbf{a}) \in \{0,1\}^{N\times N}$, and $\mathbf{a} = [a_1, \ldots, a_N]^T \in \{0,1\}^{N\times 1}$, i.e.,

$$a_n = \begin{cases} 1, & \text{if the LED } n \text{ is active,} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

By this definition, we can declare the number of active LEDs as $N_{\text{a}} = \sum_{n=1}^{N} a_n$.

The JD control scheme includes both AD and SD at the same time, such that the number of glared LEDs in SD, as well as the uniform DC-bias level of AD are jointly optimized. To achieve this, a predetermined target dimming level η is considered, based on which we can round-off the number of glared LEDs as follows [24]

$$N_a = \big[\eta N\big]. \quad (6)$$

Accordingly, the uniform DC-bias level i_{DC} is given by [11]

$$i_{\text{DC}} = \frac{\eta N(I_0 - I_l)}{N_a} + I_l, \quad (7)$$

wherein $I_0 = \dfrac{I_l + I_h}{2}$ specifies the original DC-bias that corresponds to AD with all-glared LEDs and a dimming level of $\eta = 100\%$. Although N_{a} determines the number of glared LEDs, it does not specify the index of active LEDs at the LED array. Hence, it is required to formulate a network-wide resource allocation problem, so as to optimize the binary LED selection matrix $\mathbf{A}$.

D. Data Rate

At the receiver side, all users first decode the common received lightweight data stream by considering all private streams as noise. Then, the private received lightweight data stream will be decoded by treating other private streams as noise [23]. On this basis, the common and private lightweight received data rate for the k-th user can be expressed as

$$R_k^{(\text{c})} = \log_2\left(1 + \frac{|\mathbf{h}_k^H \mathbf{A}\mathbf{w}^{(\text{c})}|^2}{\sum_{j=1}^{K} |\mathbf{h}_k^H \mathbf{A}\mathbf{w}_j^{(\text{p})}|^2 + \sigma_k^2}\right), \ \forall k \in \mathcal{K}, \quad (8)$$

and

$$R_k^{(\text{p})} = \log_2\left(1 + \frac{|\mathbf{h}_k^H \mathbf{A}\mathbf{w}_k^{(\text{p})}|^2}{\sum_{j=1, j\neq k}^{K} |\mathbf{h}_k^H \mathbf{A}\mathbf{w}_j^{(\text{p})}|^2 + \sigma_k^2}\right), \ \forall k \in \mathcal{K}, \quad (9)$$

respectively, where $\mathbf{h}_k = [h_{k,1}, h_{k,2}, ..., h_{k,N}]^T$ denotes the channel gain vector of user k. To ensure that the common stream is successfully decoded, the data rates for common data should satisfy a rate adaptation r_k^*, such that

$$\min_k R_k^{(\text{c})} \ge \sum_{k=1}^{K} r_k^*, \ \forall k \in \mathcal{K}. \quad (10)$$

Then, the aggregate system lightweight received data rate can be expressed as

$$R^{\text{Agg}}(\mathbf{w}^{(\text{c})}, \{\mathbf{w}_k^{(\text{p})}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*) = \sum_{j=1}^{K} (r_k^* + R_k^{(\text{p})}), \quad (11)$$

where $\mathbf{r}^* = [r_1^*, r_2^*, ..., r_K^*]^T \in \mathbb{R}^{K\times 1}$.

E. Energy Harvesting

The coexistence of AC and DC signals in the photo-current of LEDs, enable users to harvest energy from the DC component, which is blocked by a capacitor. The total harvested energy at the k-th user from the DC signal of all active LEDs can be expressed as [25]

$$P_k^{\text{Har}} = \sum_{n=1}^{N} \tau a_n V_t h_{k,n} i_{\text{DC}} \ln\left(1 + \frac{\sum_{n=1}^{N} h_{k,n} i_{\text{DC}}}{I_s}\right), \quad (12)$$

where V_t, τ, and I_s are the thermal voltage, the fill factor, and the dark saturation, respectively. Accordingly, the total optical power consumption of the system can be computed as [26]

$$P^{\text{tot}}(\mathbf{w}^{(\text{c})}, \{\mathbf{w}_k^{(\text{p})}\}_{k\in\mathcal{K}}, \mathbf{A}) = \zeta \sum_{n=1}^{N} a_n \left(w_n^{(\text{c})} + \sum_{k=1}^{K} w_{k,n}^{(\text{p})}\right) + P^{\text{DC}} - \sum_{k=1}^{K} P_k^{\text{Har}}, \quad (13)$$

where $P^{\text{DC}} = \varphi N_a i_{\text{DC}}$. Moreover, $\zeta \ge 1$ specifies the power for the amplifier efficiency factor, whereas φ denotes the conversion factor.

F. Problem Formulation

Compared to [18] and [19], this paper explores the potentials of RSMA and JD control scheme to increase the data rate and control the power consumption, respectively. Particularly, we optimize the aggregate system lightweight received data rate, the beamforming design, LED selection and RSMA rate adaptation, such that the power budget of the LED array, as well as the QoS and EH thresholds are preserved for

all users. Mathematically, the abovementioned network-wide optimization problem is formulated as follows

$$\begin{aligned}
\mathbf{P}_1: \quad & \max_{\mathbf{w}^{(\mathrm{c})},\{\mathbf{w}_k^{(\mathrm{p})}\}_{k\in\mathcal{K}},\mathbf{A},\mathbf{r}^*} R^{\mathrm{Agg}}(\mathbf{w}^{(\mathrm{c})},\{\mathbf{w}_k^{(\mathrm{p})}\}_{k\in\mathcal{K}},\mathbf{A},\mathbf{r}^*) \\
\text{s.t.} \quad & \mathrm{C}_1: \quad \min_k R_k^{(\mathrm{c})} \geq \sum_{k=1}^{K} r_k^*, \ \forall k \in \mathcal{K}, \\
& \mathrm{C}_2: \quad r_k^* + R_k^{(\mathrm{p})} \geq \mathrm{QoS}, \ \forall k \in \mathcal{K}, \\
& \mathrm{C}_3: \quad P^{\mathrm{tot}} \leq P_{\max}, \\
& \mathrm{C}_4: \quad P_k^{\mathrm{Har}} \geq P_{\min}^{\mathrm{Har}}, \ \forall k \in \mathcal{K}, \\
& \mathrm{C}_5: \quad a_n \in \{0,1\}, \quad \forall n \in \mathcal{N}, \\
& \mathrm{C}_6: \quad \eta = \frac{N_a(i_{\mathrm{DC}} - I_l)}{N(I_0 - I_l)} \times 100\%, \\
& \mathrm{C}_7: \quad \left|w_n^{(\mathrm{c})}\right| + \sum_{k=1}^{K} \left|w_{k,n}^{(\mathrm{p})}\right| \leq \Xi, \ \forall n \in \mathcal{N}, \qquad (14)
\end{aligned}$$

where $P_{\max}$, $P_{\min}^{\mathrm{Har}}$, and QoS represent the power budget of the LED array, the minimum homogeneous EH requirement and the minimum homogeneous QoS for all users, respectively. More specifically, C_1 assures the successful signal decoding at all users; C_2 satisfies the QoS for all users; C_3 respects the power budget of the LED array; C_4 specifies the minimum EH requirement for all users; C_5 confines each LED to be either active or inactive; C_6 defines a target required dimming level η to be satisfied; Finally, C_7 respects the dynamic range of all LEDs.

Concerning the complex domain of $\mathbf{w}^{(\mathrm{c})}$ and $\{\mathbf{w}_k^{(\mathrm{p})}\}_{k\in\mathcal{K}}$, the binary domain of $\mathbf{A}$, and the continuous domain of $\mathbf{r}^*$, one can clearly claim that this problem is non-convex in the form of mixed integer non-linear programming (MINLP) and belongs to the class of non-deterministic polynomial (NP)-hard problems. The straightforward brute-force method, i.e., the exhaustive search for attaining its globally optimal solution is implausible, considering the coupling and scalability of the problem. Moreover, the classical convex optimization-based solutions mostly rely on time consuming and computationally expensive convex transformations, whereas the wireless environment is quietly dynamic and real-time resource allocation mechanisms are preferred. Instead, we propose a real-time dynamic solution to this problem based on reinforcement learning.

III. Proposed DRL-based Approach

In this section, the non-convex problem $\mathbf{P}_1$ with both discrete and continuous variables is firstly reformulated into a model-free MDP, and then a DRL algorithm based on the PPO framework is designed to solve the problem $\mathbf{P}_1$ [20], [28], [29].

A. MDP formulation

A 4-tuple $(\mathbf{s}_t, \mathbf{a}_t, r(\mathbf{s}_t, \mathbf{a}_t), \mathbf{s}_{t+1})$ is constructed by the MDP formulation, where the current state, the action, the reward function, and the next state are denoted by $\mathbf{s}_t$, $\mathbf{a}_t$, r_t, and $\mathbf{s}_{t+1}$, respectively. The PPO approach enables the agent to interact with the environment (i.e., the OWC-assisted network), to observe the current state $\mathbf{s}_t$ from the state space $\mathcal{S}$ and to select the action $\mathbf{a}_t$ from the action space $\mathcal{A}$ according to the specific policy with the ultimate aim of maximizing the clipping surrogate objective function $\mathcal{L}^{\mathrm{CLIP}}(\cdot)$ that is defined latter. Moreover, based on the formulation of problem $\mathbf{P}_1$, the state, the action, and the reward function are elaborated in the following.

1) State: The current state $\mathbf{s}_t \in \mathcal{S}$ at time step t constitutes of the main environmental information related to problem $\mathbf{P}_1$ in such a way that allows the policy to enhance and to adapt itself to the dynamic environment. More specifically, the state $\mathbf{s}_t$ of the considered system is the set of the common and private rates, the harvested energy, the common and private beamforming vectors as follows

$$\begin{aligned}
&\mathbf{s}_t = \\
&\Big\{\{R_k^{(c)}\}_{k\in\mathcal{K}}, \{R_k^{(p)}\}_{k\in\mathcal{K}}, \{P_k^{\mathrm{Har}}\}_{k\in\mathcal{K}}, \mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}\Big\}. \qquad (15)
\end{aligned}$$

2) Action: In the proposed PPO approach, action $\mathbf{a}_t \in \mathcal{A}$ at time step t refers to the decisions that an agent takes via an interaction with the considered environment. Furthermore, the action at time step t in problem $\mathbf{P}_1$ consists of both discrete and continuous variables.

$$\mathbf{a}_t = \Big\{\mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*\Big\}. \qquad (16)$$

3) Reward Function: In particular, the PPO approach is a reinforcement learning method which trains the agents to take suitable decisions in order to maximize the defined clipping surrogate objective function $\mathcal{L}^{\mathrm{CLIP}}(\cdot)$, which contains the reward function $r(\mathbf{s}_t, \mathbf{a}_t)$. In the optimization problem $\mathbf{P}_1$, the reward function takes both the objective function, i.e., $R^{\mathrm{Agg}}(\mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*)$, as well as the constraints of $\mathbf{P}_1$ into account which can be expressed as

$$r(\mathbf{s}_t, \mathbf{a}_t) = R^{\mathrm{Agg}}(\mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*) + \sum_{j=1}^{7} l_{C_j}, \qquad (17)$$

where $l_{C_j} = \chi_j R^{\mathrm{Agg}}(\mathbf{w}^{(\mathrm{c})}, \{\mathbf{w}_k^{(\mathrm{p})}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*)$, and the index j corresponds to all constraints, i.e., $\forall j \in \{1, 2, ..., 7\}$. Besides, $\chi_j = 1$, if the C_j-th constraint is satisfied and $\chi_j = 0$, otherwise.

B. The PPO-Based Analysis

In this paper, the PPO algorithm is applied to select actions from both discrete and continuous action spaces, thus solving the non-convex problem (14). In particular, the PPO is an actor-critic on-policy gradient method which is used to simplify the complex calculation of earlier policy gradient methods, e.g., trust region policy optimization (TRPO). The detailed process of the proposed PPO-based approach is explained as follows.

The main goal in the reinforcement learning is to maximize the expected cumulative reward by considering a long-term process. Therefore, the cumulative reward at time step t is denoted as $R_t = \sum_{t=0}^{\infty} \lambda^t r(\mathbf{s}_t, \mathbf{a}_t)$, where $\lambda \in [0, 1)$ represents the discount factor. More specifically, both actor and critic

networks are applied to represent the parameterized stochastic policy of action selection denoted as $\pi_{\boldsymbol{\theta}}(\mathbf{a}_t|\mathbf{s}_t)$ and the state-value function $V_{\boldsymbol{\phi}}(\mathbf{s}_t)$, respectively, where $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ represent the parameters of the actor and critic networks, respectively. Then, a surrogate objective function based on PPO approach can be expressed as follows

$$\mathcal{L}\left(\boldsymbol{\theta}, \mathbf{s}_t, \mathbf{a}_t\right) = \mathbb{E}\left[\beta_t(\boldsymbol{\theta})\Omega(\mathbf{s}_t, \mathbf{a}_t)\right], \tag{18}$$

where the probability ratio of the current policy and the old one is represented by $\beta_t(\boldsymbol{\theta}) = \pi_{\boldsymbol{\theta}}(\mathbf{a}_t|\mathbf{s}_t)/\pi_{\boldsymbol{\theta}^{\text{old}}}(\mathbf{a}_t|\mathbf{s}_t)$, while $\boldsymbol{\theta}^{\text{old}}$ denotes the parameter for the old policy in the actor network. Moreover, the advantage function is given by

$$\Omega(\mathbf{s}_t, \mathbf{a}_t) = r(\mathbf{s}_t, \mathbf{a}_t) + \lambda V_{\boldsymbol{\phi}^{\text{old}}}(\mathbf{s}_{t+1}) - V_{\boldsymbol{\phi}^{\text{old}}}(\mathbf{s}_t), \tag{19}$$

where $\boldsymbol{\phi}^{\text{old}}$ represents the critic network parameter for the old state-value estimation function. To ensure that the updated $\pi_{\boldsymbol{\theta}}(\mathbf{a}_t|\mathbf{s}_t)$ satisfies the trust region constraint, a clipping surrogate objective function can be expressed as

$$\mathcal{L}^{\text{CLIP}}\left(\boldsymbol{\theta}, \mathbf{s}_t, \mathbf{a}_t\right) = \tag{20}$$
$$\mathbb{E}\Big[\min\left\{\beta_t(\boldsymbol{\theta})\Omega(\mathbf{s}_t, \mathbf{a}_t), \text{clip}\big(\beta_t(\boldsymbol{\theta}), 1-\varepsilon, 1+\varepsilon\big)\Omega(\mathbf{s}_t, \mathbf{a}_t)\right\}\Big],$$

where the clip function is denoted by $\text{clip}(\cdot,\cdot,\cdot)$, while ε represents a hyper-parameter to restraint $\beta_t(\boldsymbol{\theta})$ to lie in $[1-\varepsilon, 1+\varepsilon]$. More specifically, the clipping surrogate objective function is iteratively maximized in the proposed PPO approach instead of (18). Then, a mini-batch stochastic gradient decent (SGD) method updates the corresponding $\boldsymbol{\theta}$ over Q transitions denoted as $(\mathbf{s}_t^q, \mathbf{a}_t^q, r(\mathbf{s}_t^q, \mathbf{a}_t^q), \mathbf{s}_{t+1}^q)$ sampled from an experience pool, which is given by

$$\boldsymbol{\theta} = \boldsymbol{\theta}^{\text{old}} - \delta_A \frac{1}{Q}\sum_{q=1}^{Q} \nabla_{\boldsymbol{\theta}} \tilde{\mathcal{L}}_q^{\text{CLIP}}\left(\boldsymbol{\theta}, \mathbf{s}_t^q, \mathbf{a}_t^q\right), \tag{21}$$

where δ_A is the learning rate and $\tilde{\mathcal{L}}_q^{\text{CLIP}}\left(\boldsymbol{\theta}, \mathbf{s}_t^q, \mathbf{a}_t^q\right)$ is the realization of $\mathcal{L}^{\text{CLIP}}\left(\boldsymbol{\theta}, \mathbf{s}_t, \mathbf{a}_t\right)$ with the q-th transition, respectively. The mini-batch SGD for updating $\boldsymbol{\phi}$ uses the MSE loss function between $V_{\boldsymbol{\phi}}(\mathbf{s}_t)$ and $\hat{R}(\mathbf{s}_t, \mathbf{a}_t)$ as follows

$$\boldsymbol{\phi} = \boldsymbol{\phi}^{\text{old}} - \delta_C \frac{1}{Q}\sum_{q=1}^{Q} \nabla_{\boldsymbol{\phi}}\big(V_{\boldsymbol{\phi}}(\mathbf{s}_t^q) - \hat{R}(\mathbf{s}_t^q, \mathbf{a}_t^q)\big)^2, \tag{22}$$

where the learning rate is represented by δ_C. Moreover, the target state-value function denoted by $\hat{R}(\mathbf{s}_t, \mathbf{a}_t)$ is given by

$$\hat{R}(\mathbf{s}_t, \mathbf{a}_t) = r(\mathbf{s}_t, \mathbf{a}_t) + \lambda V_{\boldsymbol{\phi}^{\text{old}}}(\mathbf{s}_{t+1}). \tag{23}$$

The details of the proposed PPO-based approach are summarized in Algorithm (1). More specifically, the action $\mathbf{a}_t$ is generated based on the specific policy in the current state $\mathbf{s}_t$ in which the reward $r(\mathbf{s}_t, \mathbf{a}_t)$ is obtained. Furthermore, the transition $(\mathbf{s}_t, \mathbf{a}_t, r(\mathbf{s}_t, \mathbf{a}_t), \mathbf{s}_{t+1})$ is stored in the experience pool such that Q number of transitions from the experience pool are sampled. In the next step, the advantage function $\Omega(\mathbf{s}_t, \mathbf{a}_t)$ in (19) is computed. Finally, the corresponding actor and critic parameters are updated by employing mini-batch SGD. It is worth pointing out that the clipping surrogate objective function in the proposed PPO-based approach ensures that the updated policy satisfies the trust region constraint, thus avoiding the performance collapse.

IV. Simulation Results

In this section, simulation results are presented to assess the performance of the discussed system, within an indoor room of size $8 \times 8 \times 3\ m^3$. The random distribution of users should ensure that the coordinates of each user are within the room's dimensions, specifically $0 \leq x_k \leq 8$, $0 \leq y_k \leq 8$, and $0 \leq z_k \leq 1$. In this system, $N = 6$ number of LEDs are uniformly distributed on a LED array plane, in which the distance between any two adjacent LEDs sharing the same y-coordinate is set to 2 m, while the distance is set to 4 m between any two adjacent LEDs with the same x-coordinate. The simulation parameters are summarized in Table I, unless otherwise stated.

Fig. 2 displays the convergence behaviour of the proposed solution for both RSMA and NOMA schemes. Notably, more average reward is observed for lower number of users, mainly due to lower imposed inter-user interference. This figure also illustrates that RSMA outperforms NOMA for various

Algorithm 1: The Proposed PPO-Based Algorithm

Input: $\left\{\{R_k^{(c)}\}_{k\in\mathcal{K}}, \{R_k^{(p)}\}_{k\in\mathcal{K}}, \{P_k^{\text{Har}}\}_{k\in\mathcal{K}}, \mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}\right\}$,
Output: $\left\{\mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*\right\}$,
Initialization: Initialize the maximum episode E and time step T as well as the parameters of actor and critic networks, i.e., $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, ε, δ_A, and δ_C.
Set $\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}$ and $\boldsymbol{\phi}^{\text{old}} = \boldsymbol{\phi}$,
for *episode=1* **to** E **do**
  Initialize state $\mathbf{s}_t$,
  for *time step* $= 1$ **to** T **do**
    Generate action $\mathbf{a}_t$ according to $\pi_{\boldsymbol{\theta}}(\mathbf{a}_t|\mathbf{s}_t)$ in state $\mathbf{s}_t$, obtain reward $r(\mathbf{s}_t, \mathbf{a}_t)$ and then observe the new state $\mathbf{s}_{t+1}$,
    Store $(\mathbf{s}_t, \mathbf{a}_t, r(\mathbf{s}_t, \mathbf{a}_t), \mathbf{s}_{t+1})$ in the experience pool,
    Calculate the advantage function $\Omega(\mathbf{s}_t, \mathbf{a}_t)$ in (19)
    for $q = 1, 2, 3, ..., Q$ **do**
      Calculate $\nabla_{\boldsymbol{\theta}} \tilde{\mathcal{L}}_q^{\text{CLIP}}\left(\boldsymbol{\theta}, \mathbf{s}_t^q, \mathbf{a}_t^q\right)$ in (21),
      Calculate $\nabla_{\boldsymbol{\phi}}\big(V_{\boldsymbol{\phi}}(\mathbf{s}_t^q) - \hat{R}(\mathbf{s}_t^q, \mathbf{a}_t^q)\big)^2$ in (22),
      Calculate $\hat{R}(\mathbf{s}_t, \mathbf{a}_t)$ in (23),
    end for
    Update $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ in (21) and (22), respectively.
    Update $\boldsymbol{\theta}^{\text{old}} = \boldsymbol{\theta}$ and $\boldsymbol{\phi}^{\text{old}} = \boldsymbol{\phi}$.
  end for
end for

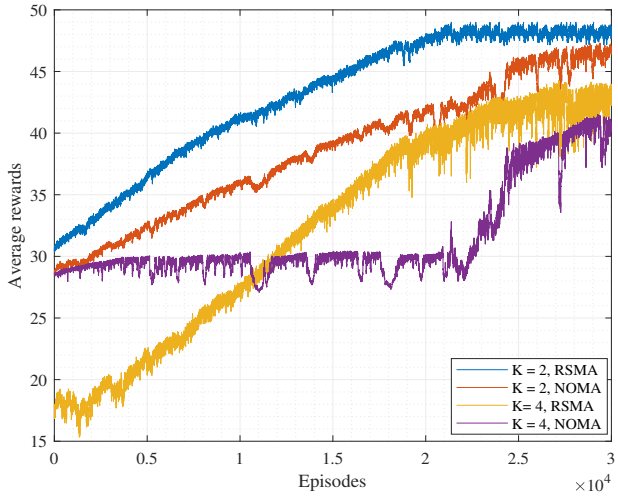

Fig. 2: Convergence: average reward vs the episode number.

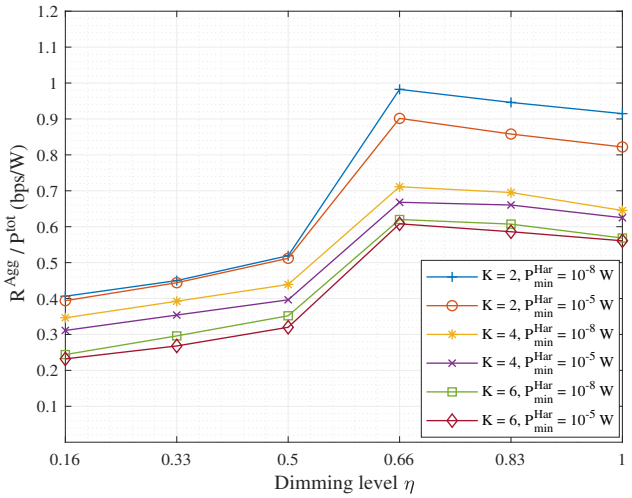

Fig. 3: System EE versus dimming level under $P_{\max} = 20$ Watts and QoS = 3 bits/sec.

number of users at the convergence point, due to more efficient superposition and decoding methodology [20].

Fig. 3 plots the system energy efficiency (EE) versus various dimming levels. The system EE can be defined as $R^{\text{Agg}}(\mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}, \mathbf{A}, \mathbf{r}^*)/P^{\text{tot}}(\mathbf{w}^{(c)}, \{\mathbf{w}_k^{(p)}\}_{k\in\mathcal{K}}, \mathbf{A})$. Through this figure, we evaluate the system performance for different number of users as well as varying the minimum EH requirement for each user. It is evident that the system EE is maximized around the dimming level of 0.66, representing the optimal trade-off between the data rate and energy consumption of the system. Additionally, for the same number of users, a lower system EE is observed when the minimum EH requirement is higher. For instance, for $K = 2$, the baseline related to $P_{\min}^{\text{Har}} = 10^{-5}$ Watts achieves lower system EE, compared to the baseline corresponding to $P_{\min}^{\text{Har}} = 10^{-8}$ Watts.

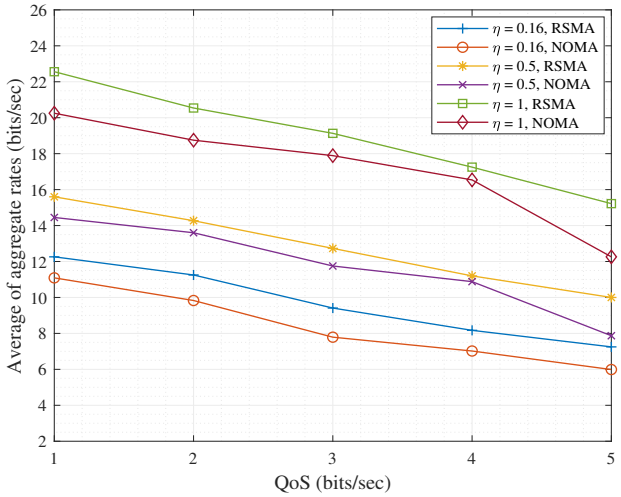

Fig. 4: Average system data rate versus the minimum QoS of users, under $P_{\max} = 20$ Watts, $K = 4$, and $P_{\min}^{\text{Har}} = 10^{-8}$ Watts.

TABLE I: Simulation Parameters

Parameter	Value	Parameter	Value
Ψ_c	60°	n_R	1.5
$\phi_{1/2}$	60°	A^{OWC}	1 cm^2
I_h	10 mA	I_l	0 A
I_0	5 mA	I_s	10^{-9} A
V_t	25 mA	ζ	1.2
φ	1	τ	0.75
N	6	$P_{\max}$	20 W

This is due to the more stringent constraint C_4, leading to a more limited feasible set for problem $\mathbf{P}_1$.

Additionally, Fig. 4 evaluates the average system data rate for various minimum QoS for users. The baselines constitute either RSMA or NOMA, with various dimming levels. We can observe that our proposed scheme with RSMA outperforms the baseline with NOMA. It is evident that increasing the minimum QoS leads to a reduction in the average system data rate for both RSMA and NOMA baseline schemes, under the same justification stated for Fig. 3.

Furthermore, Fig. 5 illustrates the average system data rate versus the minimum EH requirement, where baselines constitute either RSMA and NOMA with different dimming levels. As can be seen, a higher dimming level achieves more average system data rate, regardless of multiple access scheme (i.e., RSMA or NOMA). Thus, RSMA consistently outperforms NOMA in achieving a higher average system data rate for a specific minimum EH requirement.

V. Conclusion

In this paper, the performance of a multi-LED OWC system with SLIPT technology is investigated, where a JD control scheme is proposed to reduce its energy consumption, and RSMA technology is explored to increase its data rate. We formulated a resource allocation problem and proposed a dynamic real-time solution based on reinforcement learning. We numeri-

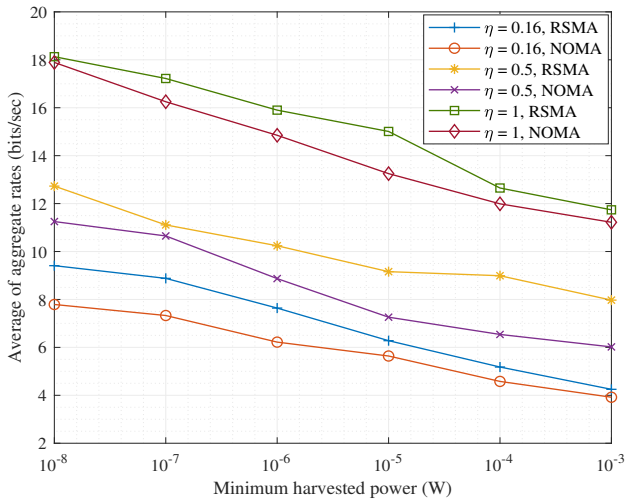

Fig. 5: Average system data rate versus the minimum EH requirement under $P_{\max} = 20$ Watts, $K = 4$, and QoS $= 3$ bits/sec.

cally found the optimal dimming level in this system to achieve a trade-off between the data rate and energy consumption.

VI. Acknowledgment

The work of P. D. Diamantoulakis was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the 3rd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers (Project Number: 7280). The work of G. K. Karagiannidis was implemented in the framework of H.F.R.I call Basic research Financing (Horizontal support of all Sciences) under the National Recovery and Resilience Plan Greece 2.0 funded by the European Union NextGenerationEU (H.F.R.I. Project Number: 15642).

References

[1] M. Obeed, A. M. Salhab, M. -S. Alouini and S. A. Zummo, "On Optimizing VLC Networks for Downlink Multi-User Transmission: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 3, pp. 2947-2976, Mar. 2019.

[2] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100-107, Feb. 2004.

[3] H. Zarini et al., "Multiplexing eMBB and mMTC Services over Aerial Visible Light Communications," in *Proc. IEEE International Conference on Communications (ICC)*, May 2023, pp. 2655-2661.

[4] P. H. Pathak, X. Feng, P. Hu and P. Mohapatra, "Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 4, pp. 2047-2077, Sep. 2015.

[5] H. Qiu, S. Gao and G. Tu, "An Opportunistic NOMA Scheme for Multiuser Spatial Multiplexing VLC Systems," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3017-3021, Sep. 2021.

[6] Y. Yang, Z. Zeng, J. Cheng and C. Guo, "A Novel Hybrid Dimming Control Scheme for Visible Light Communications," *IEEE Photonics J.*, vol. 9, no. 6, pp. 1-12, Dec. 2017.

[7] P. D. Diamantoulakis, G. K. Karagiannidis and Z. Ding, "Simultaneous Lightwave Information and Power Transfer (SLIPT)," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 3, pp. 764-773, Sep. 2018.

[8] S. Ma, F. Zhang, H. Li, F. Zhou, Y. Wang and S. Li, "Simultaneous Lightwave Information and Power Transfer in Visible Light Communication Systems," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 12, pp. 5818-5830, Dec. 2019.

[9] A. M. Abdelhady, O. Amin, B. Shihada and M. -S. Alouini, "Spectral Efficiency and Energy Harvesting in Multi-Cell SLIPT Systems," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 5, pp. 3304-3318, May 2020.

[10] T. Wang, F. Yang, L. Cheng and J. Song, "Spectral-Efficient Generalized Spatial Modulation Based Hybrid Dimming Scheme With LACO-OFDM in VLC," *IEEE Access*, vol. 6, pp. 41153-41162, Jun. 2018.

[11] Y. Yang et al., "Joint LED Selection and Precoding Optimization for Multiple-User Multiple-Cell VLC Systems," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 6003-6017, Apr. 2022.

[12] X. Zhou, R. Zhang and C. K. Ho, "Wireless Information and Power Transfer: Architecture Design and Rate-Energy Tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754-4767, Nov. 2013.

[13] S. Javadi and E. Soleimani-Nasab, "Two-way interference-limited AF relaying with wireless power transfer," in *Proc. Telecommunications Forum (TELFOR)*, Nov. 2016, pp. 1-4.

[14] S. Javadi and E. Soleimani-Nasab, "Outage analysis of cognitive two-way AF relaying systems with wireless power transfer," in *Proc. Iranian Conference on Electrical Engineering (ICEE)*, May 2017, pp. 2066-2071.

[15] S. Javadi and E. Soleimani-Nasab, "Performance analysis of cognitive two-way AF relaying systems with wireless energy harvesting over Nakagami-m fading channels," in *Proc. Iran Workshop on Communication and Information Theory (IWCIT)*, May 2017, pp. 1-6.

[16] E. Soleimani-Nasab and S. Javadi, "Performance analysis of two-way wireless powered Amplify and Forward relaying in the presence of co-channel interference," *IEEE Commun. Surv. Tutor.*, vol. 34, no. 1, pp. 2047-2077, Jan. 2021.

[17] X. Liu, Y. Wang, F. Zhou, S. Ma, R. Q. Hu and D. W. K. Ng, "Beamforming Design for Secure MISO Visible Light Communication Networks With SLIPT," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7795-7809, Dec. 2020.

[18] S. Ma, F. Zhang, H. Li, F. Zhou, Y. Wang and S. Li, "Simultaneous Lightwave Information and Power Transfer in Visible Light Communication Systems," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 12, pp. 5818-5830, Dec. 2019.

[19] Y. Guo, K. Xiong, Y. Lu, B. Gao, P. Fan and K. B. Letaief, "SLIPT-Enabled Multi-LED MU-MISO VLC Networks: Joint Beamforming and DC Bias Optimization," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 3, pp. 1104-1120, Sep. 2023.

[20] C. Meng, K. Xiong, W. Chen, B. Gao, P. Fan and K. B. Letaief, "Sum-Rate Maximization in STAR-RIS-Assisted RSMA Networks: A PPO-Based Algorithm," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 5667-5680, Feb. 2024.

[21] B. Clerckx, H. Joudeh, C. Hao, M. Dai and B. Rassouli, "Rate splitting for MIMO wireless networks: a promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98-105, May 2016.

[22] H. Joudeh and B. Clerckx, "Sum-Rate Maximization for Linearly Precoded Downlink Multiuser MISO Systems With Partial CSIT: A Rate-Splitting Approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847-4861, Nov. 2016.

[23] Y. Mao, B. Clerckx, and V.O. Li, "Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA," J. Wirel. Commun. Netw., vol. 2018, no. 1, pp. 154, Apr. 2018.

[24] Z. Feng, C. Guo and Y. Yang, "A Novel Hybrid Dimming Scheme for MU-MIMO-OFDM VLC System," in *Proc. IEEE International Conference on Communications (ICC)*, May 2019, pp. 1-6.

[25] Y. Guo, K. Xiong, B. Gao, P. Fan and K. B. Letaief, "Energy Efficiency in Secure VLC Networks With SLIPT," *IEEE Wireless Commun. Lett.*, vol. 12, no. 5, pp. 799-803, May 2023.

[26] P. Tennakoon, D. N. K. Jayakody and S. Affes, "Simultaneous Lightwave Information and Power Transfer with Non-orthogonal Multiple Access," in *Proc. International Conference on Information and Automation for Sustainability (ICIAfS)*, Aug. 2021, pp. 214-219.

[27] C. Chen, W. -D. Zhong, H. Yang and P. Du, "On the Performance of MIMO-NOMA-Based Visible Light Communication Systems," *IEEE Photonics Technol. Lett.*, vol. 30, no. 4, pp. 307-310, Feb. 2018.

[28] Schulman, John and Wolski, Filip and Dhariwal, Prafulla and Radford, Alec and Klimov, Oleg, "Proximal policy optimization algorithms," Jul. 2017, *arXiv:1707.06347*. [Online].

[29] R. Saadat Yeganeh, M. Omidi, F. Zeinali, M. Robat Mili, and M. Ghavami, "Sum Throughput Maximization in Multi-BD Symbiotic Radio NOMA Network Assisted by Active-STAR-RIS," Jan. 2024, *arXiv:2401.08301*. [Online].