Learning-Aided Network Association for Hybrid Indoor LiFi-WiFi Systems

Jingjing Wang, *Student Member, IEEE*, Chunxiao Jiang, *Senior Member, IEEE*, Haijun Zhang, *Member, IEEE*, Xin Zhang, Victor C. M. Leung, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Given the scarcity of spectral resources in traditional wireless networks, it has become popular to construct visible light communication (VLC) systems. They exhibit high energy efficiency, wide unlicensed communication bandwidth as well as innate security, hence they may become part of future wireless systems. However, considering the limited coverage and dense deployment of light-emitting diode (LED) lamps, traditional network association strategies are not readily applicable to VLC networks. Hence by exploiting the power of online learning algorithms, we focus our attention on sophisticated multi-LED access point selection strategies conceived for hybrid indoor LiFi-WiFi communication systems. We formulate a multi-armed bandit model for supporting the decisions on beneficially selecting LED access points (AP). Moreover, the 'exponential weights for exploration and exploitation' (EXP3) algorithm and the 'exponentially-weighted algorithm with linear programming' (ELP) algorithm are invoked for updating the decision probability distribution, followed by determining the upper bound of the associated accumulation reward function. Significant throughput gains can be achieved by the proposed network association strategies.**

***Index Terms*—Visible light communication (VLC), network association strategies, access control, multi-armed bandit scheme, hybrid LiFi-WiFi indoor networks.**

I. Introduction

Visible light communication (VLC) has drawn substantial research attention relying on the emergence of commercially available light-emitting diode (LED) lamps, which are characterized by low power consumption, long service life as well as excellent energy efficiency [1]. Dedicated efforts have been invested into enhancing the performance of VLC, including their components, devices, protocols, networking techniques, etc. Bearing in mind that the scarcity of spectral resources in traditional radio frequency (RF) wireless networks has been of prime concern, VLC may be invoked as a remedy for supporting high data-rate communications in the license-free spectral domain, especially in indoor scenarios [2]. Hence, VLC may find its way into the downlink of 5G indoor heterogeneous networks for supporting a range of compelling applications [3], such as the construction of intelligent furniture, tele-medicine, etc.

Having access to this new bandwidth is conducive in terms of achieving a high information transmission rate up to 3.5 Gb/s [4]. As a further benefit, VLC mitigates the concern of imposing electromagnetic interference, in particular in places such as hospitals, aircraft, etc., whilst mitigating the risk of eavesdropping in comparison to RF communication. Furthermore, capitalizing on low-cost lighting equipment, we are capable of flexibly constructing wireless networks, especially in shielded areas, where the RF propagation is poor. However, many of the classic RF based wireless communication techniques are not directly applicable to VLC, partly because they rely on the presence or absence of light represented by real-valued positive signals, rather than on complex-valued signals. Therefore, previous studies have focused their attention on the channel modeling, multiplexing and coding techniques of VLC, demonstrating that it is capable of supporting energy-efficient and secure indoor communication [5]. To elaborate, the authors of [6]–[8] discussed the characteristics and channel models of the point-to-point VLC link, whilst others [9]–[11] proposed multiplexing and coding mechanisms for VLC in order to support a high transmission rate as well as a high overall network capacity. However, when the VLC network becomes an inherent part of future wireless communication systems, coordinating with 5G or WiFi networks for example, the multipoint-to-multipoint heterogeneous network topology requires novel network association strategies in order to optimize the resource management [12]. In reality, the bursty nature of the traffic and the power constraint make it impossible for the VLC users to keep track of the whole system's state. Moreover, the dense LED lamp distribution complicates the LED access selection. As a potent member of the reinforcement learning family, bandit theory has been proposed for achieving a beneficial rewards for decision making in high-dynamic environments [13] [14].

These challenges inspired us to conceive this article on the network association strategies of hybrid indoor LiFi-WiFi communication systems. In this paper, we studied the LED source access control relying on multi-armed bandit theory and made the following original contributions.

J. Wang and X. Zhang are with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. E-mail: chinaeephd@gmail.com, zhangxin15@mails.tsinghua.edu.cn.

C. Jiang is with Tsinghua Space Center, Tsinghua University, Beijing, 100084, China. E-mail: jchx@tsinghua.edu.cn.

H. Zhang is with the Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing, 100083, China. E-mail: dr.haijun.zhang@ieee.org.

V. C. M. Leung is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T, Canada. Email: vleung@ece.ubc.ca.

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ U.K. Email: lh@ecs.soton.ac.uk.

This work was funded by Young Elite Scientists Sponsorship Program By CAST (2016QNRC001).

- A new multi-armed bandit game based user access scheme is conceived for improving the link throughput of hybrid indoor LiFi-WiFi systems relying on multiple LED access points (AP), which is beneficial in terms of constructing an environment-aware and learning-aided VLC system.
- The accumulated reward gap function is defined as our metric for characterizing the performance of our AP-selection scheme.
- Furthermore, the 'exponential weights for exploration and exploitation' (EXP3) as well as the 'exponentially-weighted algorithm with linear programming' (ELP) learning techniques are proposed for updating the AP-assignment decision probability distribution of each AP at each time instant. Our ELP based AP selection algorithm is constructed for taking into account both the partially observed conditions of the APs as well as the network topology.
- Finally, based on the concept of trial-and-error learning and on timely adjustment, we derive the theoretical upper bound of the expected value of our proposed accumulated reward gap function of the EXP3- and ELP-based iterative algorithms, followed by our performance analysis.

The remainder of the article is outlined as follows. Section II surveys the VLC aided hybrid network association techniques and challenges. Our system model is detailed in Section III. The EXP3-based decision probability distribution update algorithm is discussed in Section IV, while in Section V we propose the partially observed ELP-based iterative algorithm. In Section VI, simulation results are provided for characterizing both our EXP3- and ELP-based LED AP-selection algorithms, followed by our conclusions in Section VII.

II. State-Of-The-Art

A variety of VLC aided hybrid networking techniques and applications have been proposed in the literature for indoor communication scenarios [15]–[19]. Specifically, in [15], Lee *et al*. presented a hybrid network structure constituted by a VLC downlink (DL) having an extremely low error probability and a Zigbee uplink (UL) as well as a positioning system. In this hybrid visible light and Zigbee radio frequency environment, the position estimation error predominantly caused by the nearby sources of interference was reduced. Relying on an RF UL and free-space optical (FSO) DL, Bouchet *et al*. [16] constructed a Gigabit home network. In [17], Rahaim *et al*. proposed a hybrid indoor communication system integrating a VLC DL and WiFi DL network, which capitalized on VLC DL broadcast channels to supplement classic RF DL/UL communications. A handover mechanism between VLC DL and WiFi DL was also designed to dynamically distribute the resources and to optimize the system's DL throughput. Furthermore, Shao *et al*. [18] characterized the average system delay of two different scenarios in heterogeneous LiFi-WiFi DL/UL networks. In [19], a new metric was defined by Bao *et al*. in order to quantity the DL capacity of the hybrid network. Additionally, a novel VLC DL network relying on a specific frame format was proposed for solving the multiuser access problem.

However, the LED lamps of VLC systems have to be appropriately configured in order to support the users' requests, which we refer to as the access point (AP) selection. As regards to AP selection, a variety of sophisticated strategies have been conceived for both the traditional wireless access networks as well as for VLC based hybrid networks relying for example on price theory [20] [21], on game theory [22] [23] and on Markov decision theory [24] [25] with the objective of optimizing the QoS, such as the throughput, delay, energy consumption, etc. To elaborate, in [26], Zhu *et al*. formulated a hierarchical dynamic gaming framework based on differential game theory and on evolutionary game theory to investigate the issues of dynamic service selection. Taking into account the dynamic service selection provided by their lower-level evolutionary game, they determined the optimal pricing strategy based on Stackelberg's game, which resulted in appealingly rapid convergence. Yang *et al*. [27] found the optimal network access decision rule relying on a multi-dimensional Markov Decision Process (MDP) considering the associated negative network externality, followed by a modified value iteration algorithm, which substantially reduced the system's storage requirement. As for the family of VLC based hybrid networks, Wu *et al*. [28] proposed a two-stage AP selection method benchmarked against a solution relying on fuzzy logic theory. In [29], Soltani *et al*. defined a metric for LED AP connectivity relying both on the received signal to interference plus noise ratio (SINR) and on the tele-traffic in order to balance the AP loading. Furthermore, relying on game theory, Liu *et al*. [30] presented a cooperative AP selection mechanism based on best-response dynamics [31] and best-response strategies [32]. This mechanism utilized the system capacity as the reward function to optimize the AP selection process, which achieved a significant energy consumption improvement. However, the aforementioned studies on AP selection typically relied on exploiting the idealized simplifying assumption of having both accurate reward information and system state knowledge, whilst none of them considered realistic partially observed network conditions or the network topology of multi-LED configurations.

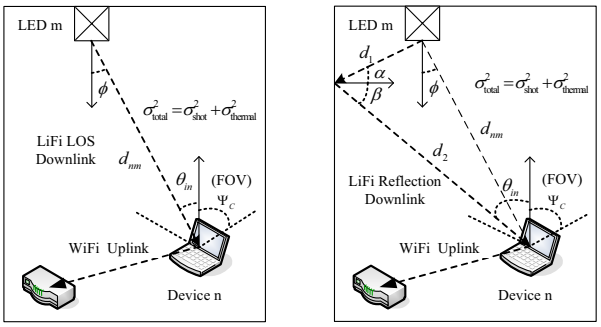
Online learning [33] has been used in situations, where the decision-making dynamically adapts to the uncertain or partially observed environment. Therefore, in this paper, relying on the multi-armed bandit mechanism based online learning algorithm, a multi-LED AP selection strategy is proposed for hybrid indoor LiFi-WiFi communication systems, which considers both realistic partially observed network conditions and the specific network topology, where the possible rate-rewards as well as the AP selection probabilities of all the neighboring LED lamps based on a specific LED connectivity relationship can be tracked in the AP-association decision making process.

III. System Model

In our indoor VLC system, the communication between the devices and the backbone network relies on the VLC DL as well as on the RF WiFi UL, which hence can be viewed as a hybrid LiFi-WiFi network.

A. System Composition

In our system model, we assume that there are M low-energy LED lamps in the indoor space considered. The set of LED lamps is denoted by $\mathcal{M} = \{1, 2, ..., M\}$. The arrival time of each downloading access request from N randomly distributed mobile devices is represented by t, where the set of mobile devices is given by $\mathcal{N} = \{1, 2, ..., N\}$. Hence, the time-interval between the adjacent access request i and $(i + 1)$ can be expressed as $\Delta T_i = t_{i+1} - t_i$. Apart from a WiFi UL module, each mobile device is also equipped with a VLC DL processing module. The distance between the nth mobile device and the mth LED lamp is denoted by d_{nm}. Moreover, we assume that regardless of their positions, the mobile devices are capable of accessing any of the M indoor LED lamps and of downloading packets from the Internet via VLC. When a decision round is due, the access control strategy obeys the decision probability distribution (DPD) of $P = \{p_1, p_2, ..., p_M\}$. We have $\sum_{m=1}^{M} p_m = 1$, where p_m denotes the probability of accessing the mth LED lamp. Furthermore, the service time of each LED lamp obeys the negative exponential distribution with a departure rate ς, while the interval between system access requests, in the same way, obeys the negative exponential distribution with an arrival rate λ.



(a) LOS path (b) Reflected and LOS path

Fig. 1. The model of the indoor VLC link including both the LOS path and the reflected path.

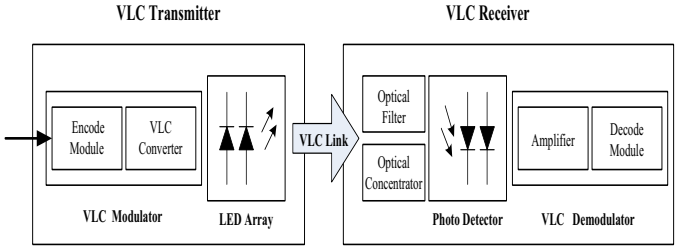


Fig. 2. Structure of the VLC transmitter and receiver.

B. Indoor VLC Link Characteristics

The VLC DL channel is characterized by a diffuse link, where the light beam is radiated within a certain angle. In this subsection, we focus our attention on modeling the indoor VLC channel, including the line of sight (LOS) path (Fig. 1 (a)) as well as a single reflected path (Fig. 1 (b)). The basic channel model can be expressed as:

$$y(t) = \Gamma x(t) \otimes h(t) + n(t), \tag{1}$$

where $x(t)$ and $y(t)$ represent the original optical pulse and the received optical signal, respectively. Moreover, $h(t)$ is the channel impulse response (CIR) and $n(t)$ represents the additive white Gaussian noise (AWGN). Finally, $\otimes$ represents the convolution operation. From the perspective of the received optical power, we calculate the system's direct current (DC) power gain of H, where we have $H = \int_{-\infty}^{+\infty} h(t)dt$. The average optical output power P_t can be estimated as [34]:

$$P_t = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt. \tag{2}$$

Upon considering both the LOS path as well as the reflected path of Fig. 1, we have $H = H^{los} + H^{ref}$, where H^{los} is the LOS path's power gain, whilst H^{ref} indicates the reflected path's power gain. Then, the received optical power can be expressed as:

$$P_r = P_t \times (H^{los} + H^{ref}). \tag{3}$$

Let us now embark on interpreting the above two power gains, where for the sake of simplicity, only a single reflection is considered.

The LED lamp can be deemed to be a Lambertian source, hence the radiation intensity $R(\phi)$ can be expressed as:

$$R(\phi) = \frac{(\gamma + 1)}{2\pi} P_t \cos^{\gamma}(\phi), \tag{4}$$

where ϕ denotes the angle of emergence with regard to the perpendicular direction, while γ is the Lambertian index, which can be derived from the angle of half-power $\phi_{1/2}$, i.e. $\gamma = \ln 2 / \ln(\cos \phi_{1/2})$.

The receiver of a VLC DL system consists of four main parts (Fig. 2), i.e. the optical filter, the optical concentrator, the photo detector[1] (PD) as well as the VLC demodulator. Specifically, the field of view (FOV) at the receiver is denoted as Ψ_c. Moreover, g_f and g_c represent the gain of the optical filter and of the optical concentrator, respectively. Let θ_{in} denote the angle of incidence. Then we have:

$$g_c(\theta_{in}) = \begin{cases} \frac{\gamma^2}{\sin^2(\Psi_c)}, & \text{if } 0 \le \theta_{in} \le \Psi_c, \\ 0, & \text{if } \theta_{in} > \Psi_c. \end{cases} \tag{5}$$

Hence, according to [35] [36], the LOS path's power gain between the nth mobile device and the mth LED lamp can be inferred from:

$$H_{nm}^{los} = \begin{cases} \frac{(\gamma+1)S}{2\pi d_{nm}^2} \cos^{\gamma}(\phi) g_f(\theta_{in}) g_c(\theta_{in}) \cos(\theta_{in}), & \text{if } 0 \le \theta_{in} \le \Psi_c, \\ 0, & \text{if } \theta_{in} > \Psi_c, \end{cases} \tag{6}$$

where S is the physical area of the PD. As for the reflected

[1]The photo detector is applied to convert the optical signal into electronic signal.

path, the single-bounce reflected power gain is formulated as:

$$H_{nm}^{ref} = \begin{cases} \int_A \frac{(\gamma+1)\rho S}{2\pi^2 d_1^2 d_2^2} \Lambda \cos^{\gamma}(\phi) g_f(\theta_{in}) g_c(\theta_{in}) \cos(\theta_{in}) dA, & \text{if } 0 \leq \theta_{in} \leq \Psi_c, \\ 0, & \text{if } \theta_{in} > \Psi_c, \end{cases} \tag{7}$$

where ρ is the reflection coefficient and A represents the area of reflection. Furthermore, d_1 denotes the distance between the mth LED lamp and a reflection point, while d_2 is the distance between the reflection point and the nth mobile device. In Eq. (7), $\Lambda = \cos(\alpha)\cos(\beta)$, where α is the angle of incidence with regard to the reflection point and β denotes the angle of irradiance from the reflection point to the nth mobile device. Geometrically, as shown in Fig. 1 (b), we have $d_{nm}^2 = d_1^2 + d_2^2 - 2d_1 d_2 \cos(\alpha + \beta)$.

In Eq. (1), the AWGN having a total variance of σ_{total}^2 represents the sum of contributions from the shot noise as well as the thermal noise, i.e.,

$$\sigma_{total}^2 = \sigma_{shot}^2 + \sigma_{thermal}^2. \tag{8}$$

According to [35], we assume that q denotes the electronic charge, while B is the equivalent bandwidth. Furthermore, I_{bg} represents the background current and ξ is the responsivity of the PD. Hence, we have:

$$\sigma_{shot}^2 = 2q\xi P_r B + 2qI_2 I_{bg} B \tag{9}$$

as well as

$$\sigma_{thermal}^2 = \frac{8\pi b T_K}{G} \eta S I_2 B^2 + \frac{16\pi^2 b T_K \Gamma}{g_m} \eta^2 S^2 I_3 B^3, \tag{10}$$

where b is Boltzmann constant and T_K is the absolute temperature. Moreover, G denotes the open-loop voltage gain and g_m represents the field effect transistor's (FET) transconductance. The fixed capacitance of the PD per unit area is represented by η, and Γ is the FET's channel noise factor. The noise bandwidth factors in Eq. (9) and Eq. (10) are given by $I_2 = 0.562$ and $I_3 = 0.0868$.

Therefore, the signal to interference plus noise ratio (SINR), namely ζ, of a VLC DL link can be described as [37] [38]:

$$\zeta = \frac{P_r}{\sigma_{total}^2 B + P_I}, \tag{11}$$

where P_I is the received interference power.

IV. Basic Multi-Armed Bandit Model Aided LED AP Selection

First of all, let us consider a simple LED AP selection scenario, where two LED APs in different activity states are considered, which have the initial rate-reward of zero and 1, respectively. Regardless of the specific time-variant state of each lamp at each AP decision making instant, if we randomly select a LED AP, the probability of receiving zero reward will be $1/2$, and the probability of two successive zero rewards received will be $1/4$. By contrast, if we consider a learning aided AP selection scheme, which is capable of adjusting the AP-selection probability relying both on 'learning' the historical reward information, as well as on the AP's state and the environmental information, it will result in a substantial rate-reward. Conceiving a beneficial LED AP selection scheme capable of taking into account a time-variant environment at a low computational complexity is the predominant objective of our treatise.

In this section, the M LED lamps can be viewed as M independent servers, each of which is capable of supporting U downloading services within the maximum available bandwidth B. Different services supported by the same LED are assumed to share the VLC channel in the classic FDMA mode. Without loss of generality, we assume that multiple services access the available bandwidth B without any overhead. When the ith access request arrives at instant t_i, the number of downloading services supported by the M LED lamps are described by the vector $\mathbf{u}(t_i) = [u_1(t_i), u_2(t_i), ..., u_M(t_i)]$. As for the ith AP assignment round, relying on the system's decision probability distribution[2] $P(t_i) = \{p_1(t_i), p_2(t_i), ..., p_M(t_i)\}$ at the instant t_i, only one of the M LED lamps can be selected, such as the mth LED lamp for example, for supporting the ith downloading service. Hence, the received throughput of the ith AP decision round can be formulated as:

$$Q(t_i, m) = \frac{B}{u_m(t_i) + 1} \log_2 \left(1 + \frac{P_r(t_i)}{\frac{\sigma_{total}^2(t_i) B}{u_m(t_i)+1} + P_I(t_i)} \right), \tag{12}$$

where $u_m(t_i) \leq U - 1$. If $u_m(t_i) = U$, for $\forall m = 1, 2, ..., M$, the access request will be declined.

Given that multiple LED lamps may be allocated, how to arrive at an LED AP selection decision with a positive sum-rate contribution is our main concern. In VLC DL systems, the download requests appear at random instants and positions. Hence, we need a sophisticated AP-assignment scheme capable of maximizing the system's sum-rate. As one of the popular online learning algorithms, the 'multi-armed bandit' solution relies on a player at a row of slot machines who has to decide which machines to play on and how many times to play on each. He/she will receive a random reward provided by the selected machine. The objective of the game is to maximize the sum of rewards earned through a sequence of lever pulls [13] [14]. This concept has become one of the popular examples of sequential decision making problems striking an exploration-exploitation trade-off[3], where the players address the fundamental trade-off between the exploration and exploitation in sequential decision making experiments relying on limited information [39].

In the following, we focus our attention on the determination of the system's decision probability distribution at each time instant relying on the multi-armed bandit scheme. First of all, we define the accumulated reward gap function[4] after K access

[2]The decision probability distribution simply represents the specific probability of each of the M APs being selected at the instant t_i for satisfying a randomly positioned user requesting access to the system.

[3]The exploration-exploitation trade-off implies that players must strike a balance between the exploitation of actions that did well in the past and the exploration of actions that might result in a higher reward in the future.

[4]The accumulated reward gap represents the difference between the maximum theoretical reward and the actually acquired reward after sequential decision making experiments.

decisions as $R(K)$, i.e.

$$R(K) = \max_{i_1,i_2,...,i_K} \sum_{k=1}^{K} Q(i_k, t_k) - \sum_{k=1}^{K} Q(a_k, t_k), \quad (13)$$

where $Q(i_k, t_k)$ denotes the user rate associated with the kth decision round in terms of the access decision i_k at the instant t_k, with a_k being the actual access decision. If i_k^* represents the unknown optimal decision associated with the maximum possible individual user rate, the accumulated reward gap function can be rewritten as:

$$R(K) = \sum_{k=1}^{K} Q(i_k^*, t_k) - \sum_{k=1}^{K} Q(a_k, t_k). \quad (14)$$

Again, the decision making hinges on the system's decision probability distribution $P(t_i) = \{p_1(t_i), p_2(t_i), ..., p_M(t_i)\}$. Hence, the expected value of the accumulated reward gap function is given by:

$$\begin{aligned} E[R(K)] &= E\left[\max_{i_1,i_2,...,i_K} \sum_{k=1}^{K} Q(i_k, t_k) - \sum_{k=1}^{K} Q(a_k, t_k)\right] \\ &= \sum_{k=1}^{K} Q(i_k^*, t_k) - E\left[\sum_{k=1}^{K} Q(a_k, t_k)\right]. \end{aligned} \quad (15)$$

To further assist our analysis, we define a simplified-accumulated reward gap function as:

$$R_P(K) = \max_{i_1,i_2,...,i_K} E\left[\sum_{k=1}^{K} Q(i_k, t_k)\right] - E\left[\sum_{k=1}^{K} Q(a_k, t_k)\right], \quad (16)$$

where we have $R_P(K) \leq E[R(K)]$.

Relying on the EXP3 algorithm proposed by Bubeck *et al*. [39], we arrive at the system's decision probability distribution update formula by invoking Algorithm 1. The EXP3 algorithm models a user-based decision probability update process, where the user is unable to directly observe the reward, unless he has made his decision and received the reward. Moreover, the probability update formula of Eq. (17) relies on the reward received before. As for the complexity of our proposed EXP3 algorithm, in each decision round, we have to update the decision probability distribution of each LED lamp according to Eq. (17). Hence, the complexity of each decision round is on the order of $O(M)$. We will demonstrate that this LED AP selection results in a better sum-rate than the random selection, which does not consider any prior knowledge.

The estimate of the normalized user's throughput function is unbiased. For each $m = 1, 2, ..., M$, we have:

$$E[\widehat{Q}_m(X, t_k)] = \sum_{i=1}^{M} p_i(t_k) \frac{\overline{Q}_i(X, t_k)}{p_i(t_k)} I_{t_k}\{X\} = \overline{Q}_m(X, t_k). \quad (18)$$

Without loss of generality, from now on we will use the normalized user's throughput function $\widehat{Q}$ as our metric. Based on the aforementioned definitions and assumptions, according to [39], the upper bound of the simplified-accumulated reward gap function $R_P(K)$ of Eq. (16) can be expressed as:

$$R_P(K) \leq \sqrt{2KM \ln M}, \quad (19)$$

where K is the number of the decision rounds.

Therefore, we arrive at Theorem 1.

Algorithm 1 *EXP3-based Decision Probability Distribution Update*

1: Let $P(t_1)$ be the uniform distribution of the components $\{p_1(t_1), p_2(t_1), ..., p_M(t_1)\}$;
2: Initialize and normalize the user's throughput function $\overline{Q}(t_0) = 0$, and for the kth decision round, set $\overline{Q}(t_k) \in [0, 1]$;
3: Define a non-increasing sequence for the kth decision round, to the value of such as $\delta(k) = \sqrt{\frac{2 \ln M}{kM}}$ for example;
4: At the time instant t_k for the round k, select a single LED lamp, namely X_{t_k}, relying on the $P(t_k)$;
5: For each lamp $m = 1, 2, ..., M$, let the selection indicator function be $I_{t_k}\{X_{t_k}\} = 1$, if $m = X_{t_k}$, otherwise $I_{t_k}\{X_{t_k}\} = 0$;
6: Calculate an estimate of the normalized user's throughput function as $\widehat{Q}_m(X, t_k) = \frac{\overline{Q}_m(X,t_k)}{p_m(t_k)} I_{t_k}\{X_{t_k}\}$;
7: At the time instant t_{k+1}, update the decision probability distribution, yielding $P(t_{k+1})$, where for the lamp m, we have:

$$p_m(t_{k+1}) = \frac{\exp\left(-\delta(t_k)\left(k - \sum_{i=1}^{k} \hat{Q}_m\left(X_{t_i}, t_i\right)\right)\right)}{\sum_{j=1}^{M} \exp\left(-\delta(t_k)\left(k - \sum_{i=1}^{k} \hat{Q}_j\left(X_{t_i}, t_i\right)\right)\right)}. \quad (17)$$

Theorem 1. *In a hybrid LiFi-WiFi communication system supported by M LED lamps, let K be the total number of decision rounds. Hence, relying on the EXP3-based decision probability distribution update algorithm, the simplified-accumulated reward gap function $R_P(K)$ obeys:*

$$R_P(K) \leq \sqrt{2KM \ln M}. \quad (20)$$

V. Neighbor-Observation Aided Multi-Armed Bandit Based LED AP Selection

In a multi-LED indoor LiFi-WiFi hybrid system, the LED lamps may be regularly spaced around the room, which constitutes an undirected-graph based access point topology. The basic multi-armed bandit model ignores both the graph's specific structure as well as the information exchange during the decision making process. In this section, we focus our attention on the more sophisticated neighbor-observation aided multi-armed bandit model, which specifically considers the LED graph structure and relies on neighbor information observation. In the following, we first introduce the ELP algorithm [40], and then derive the upper bound of the accumulated reward gap function's expected value.

Let $\Upsilon(t_k)$ denote the graph structure of the M LED lamps at the instant t_k. Considering a fixed LED configuration, we have

Algorithm 2 *ELP-based Decision Probability Distribution Update*

1: Determine $N_m, m = 1, 2, ..., M$, ε and $s_m, m = 1, 2, ..., M$ according to the graph structure Υ;
2: Initialize and normalize the user's throughput function to $\overline{Q}(t_0) = 0$, and for the kth decision round, we have $\overline{Q}(t_k) \in [0, 1]$;
3: For $\forall m$, initialize a group of weight values according to $w_m(t_1) = 1/M$;
4: At the time instant t_k for the round k, select one LED lamp to access, namely X_{t_k}, relying on the probability:

$$p_m(t_k) = (1-\varepsilon)\frac{w_m(t_k)}{\sum_{m=1}^{M} w_m(t_k)} + \varepsilon s_m; \tag{23}$$

5: Calculate the unbiased estimate of the normalized user's throughput function as $\widehat{Q}_j$ for all $j \in N_{X_{t_k}}$;
6: For $j \in N_{X_{t_k}}$, define $\widetilde{Q}_j(t_k) = \widehat{Q}_j / \sum_{l \in N_{X_{t_k}}} p_l(t_k)$, and for $j \notin N_{X_{t_k}}$, $\widetilde{Q}_j(t_k) = 0$;
7: At the time instant t_{k+1}, update the weight values as $w_m(t_{k+1}) = w_m(t_k) \exp\left[\mu \widetilde{Q}_m(t_k)\right]$.

$\Upsilon(t_1) = \Upsilon(t_2) = ... = \Upsilon(t_K) = \Upsilon$. Each pair of adjacent lamps is connected by optical fiber. We assume that the M LED lamps construct a lattice based network, for example. Moreover, for any lamp $m = 1, 2, ..., M$, we define the specific set of lamps adjacent to lamp m (including the lamp m), namely m's neighbor set by N_m. Let us denote the size of Υ's maximum independent set[5] by Δ. At this stage a pair of auxiliary variables, i.e. s_m and ε are proposed, which obey:

$$s_m = \arg\max_{\forall m, s_m \geq 0, \sum_{m=1}^{M} s_m = 1} \min_{j \in \mathcal{M}} \sum_{l \in N_j} s_l, \tag{21}$$

as well as

$$\varepsilon = \frac{\varphi\mu}{\min_{j \in \mathcal{M}} \sum_{l \in N_j} s_l}, \tag{22}$$

where φ is a fixed parameter related to the neighbor observation, i.e., for $\forall l \in N_j$, $\Pr(\widehat{Q}_l \leq \varphi) = 1$, and $\mu \in (0, 1/2\varphi M)$. To elaborate a little further, s_m represents the weight value characterizing the network topology and the lamps' connectivity relationship, while the weight w_m of Eq. (23) takes into account the rate-rewards received. The optimization algorithm of Eq. (21) aims for balancing the grade of 'exploration' and 'exploitation'. Moreover, ε of Eq. (22) can be viewed as a normalization parameter. Thus, we arrive at the system's decision probability distribution update formula with the aid of Algorithm 2.

Given that the state of the number of services supported by each lamp may change over time, more valuable information should be exploited for assisting our decision making in the LED AP selection process. However, it may become excessively complex to keep track of the time-variant system states' set. The ELP algorithm models the decision probability update process based on 'just sufficient' information exchange among the neighboring LED lamps, which we refer to as 'neighbor observation'. Specifically, according to Eq. (23), the ELP-based AP assignment decision probability update is a function of two parameters, where s_m reflects the LED lamps' connectivity relationship with lamp m, while w_m relies both on the possible rate-rewards and on the AP selection probabilities of all the neighboring LED lamps. In contrast to Eq. (17) of EXP3, more prior information is used for updating the decision probability at the time instant t_{k+1}, which is an explicit benefit of the neighbor observation aided information exchange in the LED network. As for the complexity of our proposed ELP algorithm, the most complex step is to calculate the auxiliary variable s_m as shown in Eq. (21), which can be viewed as a linear programming problem. The worst case complexity of the simplex algorithm is $O(k^M)$. Moreover, steps 4-7 can be completed at a complexity of $O(M)$. Therefore, the complexity of each decision round of the ELP algorithm is $O(k^M)+O(M)$. Fortunately, if the LED network topology is time-invariant, which is usually the case, we only need to solve s_m once. Hence, the complexity is approximately $O(M)$.

In the following, we derive the upper bound of the expected value of the accumulated reward gap function of Eq. (15) based on the ELP algorithm. First of all, let us introduce a pair of Lemmas relying on graph theory [40]. Their proofs can be found in the Appendix.

Lemma 1. *Let Υ be a graph over M nodes and Δ represent its independence number, i.e. the size of Υ's largest independent set. Furthermore, let N_m represent the neighbor set of node m (including the node m) and let $w_1, w_2, ..., w_M$ be the arbitrary positive weights of the M nodes. Hence, we have:*

$$\sum_{m=1}^{M} \frac{w_m}{\sum_{l \in N_m} w_l} \leq \Delta.$$

Lemma 2. *Let Υ be a graph over M nodes and Δ represent its independence number, i.e. the size of Υ's largest independent set and let N_m represent the neighbor set of node m (but including the node m). Thus, there exist M values, namely $v_1, v_2, ..., v_M$ on the M-simplex, which satisfy:*

$$\frac{1}{\min_{m=1,2,...,M} \sum_{l \in N_m} v_l} \leq \Delta.$$

Hence, considering an M-LED lattice based network topology having x rows as well as y columns, yielding $x \times y = M$, each pair of adjacent LED lamps is connected by fiber optic and can exchange decision information. Then, relying on the ELP-based LED AP selection probability update, we arrive at Theorem 2.

Theorem 2. *In a hybrid LiFi-WiFi communication system, M LED lamps are allocated according to a $(x \times y)$-LED lattice network. Let φ be fixed values obeying $\mu \in (0, 1/2\varphi M)$. Furthermore, K is the total number of AP-assignment decision*

[5]In graph theory, the independent set is a set of vertices, in which no two vertices are adjacent. The maximum set of independent members is the set of largest possible size for a given graph. Its size is referred to as independence number Δ.

Algorithm 3 *Bron-Kerbosch Algorithm Based Independent Number.*[6]

1: Initialization: Input the adjacency matrix A of graph Υ, let V be the set of all vertices and define two empty set as R and X;
2: Matrix transformation: Compute the adjacency matrix A' of the complement graph[7] of the graph Υ, and let $N(v)$ represent the neighbor set of node v relying on matrix A';
3: Call the BronKerbosch function:
R=BronKerbosch (R, V, X)
if V and X are both empty:
Report R as the maximal clique[8];
else
Choose a pivot vertex u in $V \bigcup X$;
for each vertex v in $V \setminus N(u)$:
BronKerbosch $(R \bigcup \{v\}, V \bigcap N(v), X \bigcap N(v))$;
$V := V \setminus \{v\}$;
$X := X \bigcup \{v\}$;
4: Calculate the independent number Δ: Select the maximum size of R as Δ;
5: End.

rounds. Hence, relying on the ELP-based decision probability distribution update algorithm, the upper bound of the accumulated reward gap function's expected value is given by:

$$E[R(K)] \leq \begin{cases} \varphi\sqrt{(e-1)Kxy\log(M)}, & x \text{ is even}, \\ \varphi\sqrt{2(e-1)K\left(y\left\lfloor\frac{x}{2}\right\rfloor+\frac{y}{2}\right)}, & x \text{ is odd}, y \text{ is even}, \\ \varphi\sqrt{2(e-1)K\left(y\left\lfloor\frac{x}{2}\right\rfloor+\left\lfloor\frac{y}{2}\right\rfloor+1\right)\log(M)}, & x \text{ is odd}, y \text{ is odd}, \end{cases} \tag{24}$$

where $\lfloor\bullet\rfloor$ denotes the rounding function.

Generally, if we consider an M-LED LiFi-WiFi network obeying the specific AP-connection relationship, then relying on the Bron-Kerbosch algorithm [41] [42], we have Theorem 3. The proof of Theorem 2 and Theorem 3 can be found in the Appendix.

Theorem 3. *In a hybrid LiFi-WiFi communication system, M LED lamps are positioned based on a specific connection relationship. Let K be the total number of AP-assignment decision rounds. Hence, relying on the ELP-based decision probability distribution update algorithm, the accumulated reward gap function's expected value obeys:*

$$E[R(K)] \leq \varphi\sqrt{2(e-1)K\Delta\log(M)}, \tag{25}$$

where Δ can be determined by Algorithm 3.

[6]In computer science, the Bron-Kerbosch algorithm was conceived for finding the maximal number of cliques in an undirected graph. More explicitly, it lists all subsets of vertices having the two properties that each pair of vertices in one of the listed subsets is connected by an edge, while no listed subset can have any additional vertices incorporated while preserving its complete connectivity [41].

[7]In graph theory, the complement graph of the original graph Υ is a graph with the same vertices such that two distinct vertices are adjacent in the complement graph if and only if they are not adjacent in Υ.

[8]The clique is the opposite of independent set, in the sense that every clique corresponds to an independent set in the complement graph.

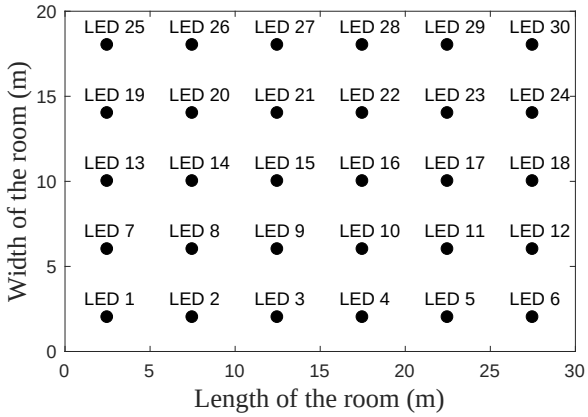


Fig. 3. The location distribution of 30 LED lamps in a room.

Before characterizing the system's performance, let us continue with a simple 'toy-example' to illustrate the benefit of the multi-armed bandit scheme versus the random selection. Considering two LED lamps, namely L_1 and L_2, for simplicity, we assume a pair of arbitrary extreme but constant rate-reward values provided by two lamps. Explicitly, let us assume that if we select L_1, we will obtain reward zero, while the max possible reward 1 for selecting L_2. By contrast, for the random selection, we have the probability of $1/2$ to select each lamp. The expected value of the reward received is 0.5 at each AP-decision time instant. After K decision rounds, the expected value of the accumulated reward can be calculated as $0.5K$, which is only half of the accumulated reward of K based on the optimal selection. However, when we invoke the multi-armed bandit selection scheme, relying on the EXP3 algorithm, once we select L_2 at time instant t, the probability of selecting L_2 at the $t+1$ time instant will increase based on Eq. (17). Moreover, the ELP algorithm allows a 'neighbor observation', where users can infer the potential reward regardless of which lamp they have selected. Hence, during each decision probability update, according to Eq. (23), the probability of selecting L_2 will be increased, while the probability of selecting L_1 will be reduced. After several decision rounds, we almost get the probability of 1 for selecting L_2, which contributes a large reward. Furthermore, according to Theorem 1 and Theorem 3, the square-root-form increase of the accumulated reward gap of the multi-armed bandit scheme is more beneficial than the linear increase of the random selection upon increasing the number of decision rounds K.

VI. Simulation Results

As mentioned before, our proposed learning schemes do not need any global information concerning the system. Only the instant reward of activating the LED selected at the last time slot, or at most its neighbors' information is needed for updating the decision probability distribution. Moreover, our proposed algorithm relies on low-complexity decisions. Hence, our learning based LED AP selection is readily applicable to

TABLE I
THE TABLE OF SIMULATION PARAMETERS

Simulation Parameters	**Value**
The length of the room	30m
The width of the room	20m
The number of LED lamps	30
The height of LED lamps	3m
The height of devices located	1.5m
Dimension of lattice based network	5×6
Boltzmann constant b	1.38×10^{-23}
Electronic charge q	1.6×10^{-19}C
Average optical output power P_t	200W
VLC System bandwidth B	10MHz
Open-loop voltage gain G	10
Capacitance of PD per unit area η	1.12×10^{-6}
FET's noise factor Γ	1.5
FET's trans-conductance g_m	3×10^{-2}
Background current I_{bg}	5.1×10^{-3}A
Responsivity of the PD ξ	0.8
Physical area of the PD S	1cm^2
The arrival rate of access requests λ	0.5

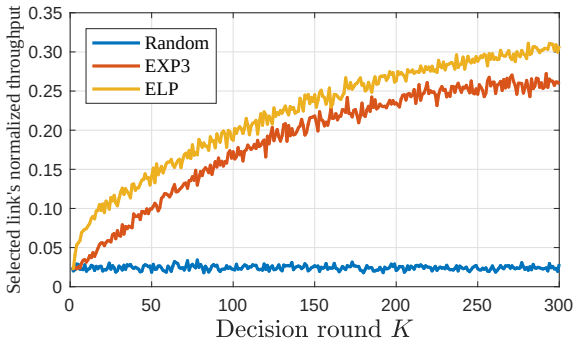


Fig. 4. The normalized throughput of the selected VLC link versus K for different LED AP selection schemes using the parameters of Table I. The results were calculated from Eq. (12).

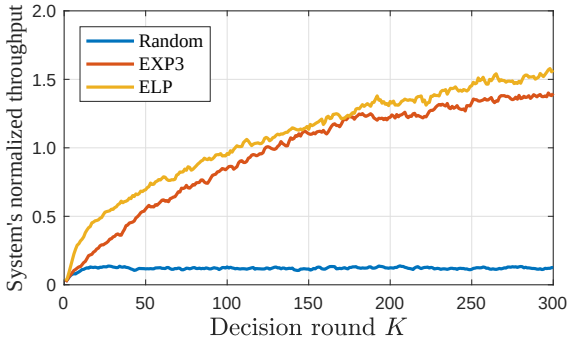


Fig. 5. The system's normalized throughput versus K for different LED AP selection schemes. The results were calculated from Eq. (27).

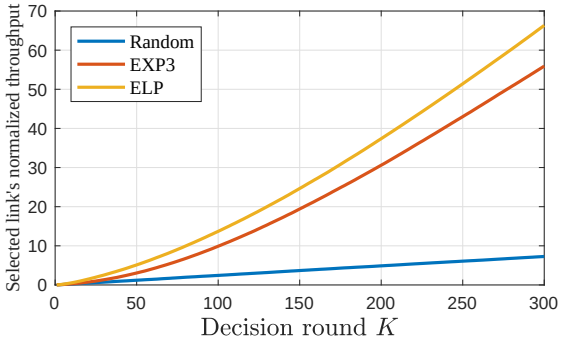


Fig. 6. The accumulated normalized throughput of selected VLC links versus K for different LED AP selection schemes. The results were calculated from Eq. (26).

realistic LiFi-WiFi hybrid networks. However, our assumptions concerning the ideal indoor VLC link characteristics as well as that of having a regular lattice topology of lamps may not hold in practice. In this section, we present our performance results characterizing the proposed EXP3 and ELP algorithms in the context of the hybrid indoor LiFi-WiFi system considered. Moreover, we use Matlab simulations to evaluate our proposed algorithms. In our indoor scenarios, the length of the room is 30m and its width is 20m. We assume that there are 30 LEDs on the 3m-high ceiling constituting a lattice-based network with 5 rows and 6 columns, as shown in Fig. 3. The mobile devices are randomly distributed in the room at a fixed height of 1.5m. Without loss of generality, we assume that each LED transmitter has the same processing capability. Moreover, each downloading service will be satisfied with a departure probability ς at each AP decision-making time instant. The VLC link characteristics detailed in Section III-B are used in our simulations, which are summarized in Table I. Specifically, the Boltzmann constant is $b = 1.38 \times 10^{-23}$, while the electronic charge is $q = 1.6 \times 10^{-19}$C. The average optical output power P_t is 200W and the bandwidth for the VLC system is about $B = 10$MHz. The open-loop voltage gain is $G = 10$. Moreover, the fixed capacitance of the photo detector (PD) per unit area is $\eta = 1.12 \times 10^{-6}$. The FET's noise factor is $\Gamma = 1.5$ and the FET's trans-conductance is $g_m = 3 \times 10^{-2}$. Furthermore, the background current is $I_{bg} = 5.1 \times 10^{-3}$A and $\xi = 0.8$ represents the responsivity of the PD. The physical area of the PD is $S = 1$cm^2 [37].

In our simulations, the normalized throughput of the selected VLC link and of the whole system as well as their accumulated value is proposed for benchmarking the performance of different LED AP selection schemes, respectively. According to the throughput function of the ith decision round defined in Eq. (12), the accumulated normalized throughput of selected VLC links after K decision-making rounds can be formulated

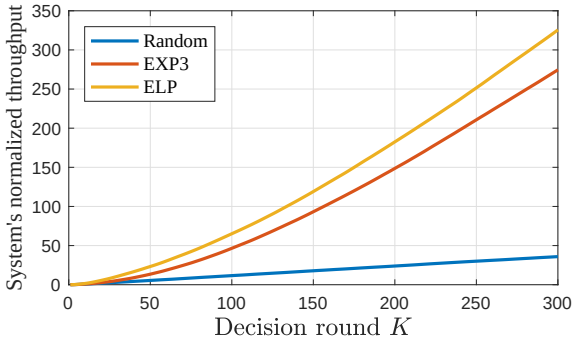


Fig. 7. The system's total accumulated normalized throughput versus K for different LED AP selection schemes. The results were calculated from Eq. (28).

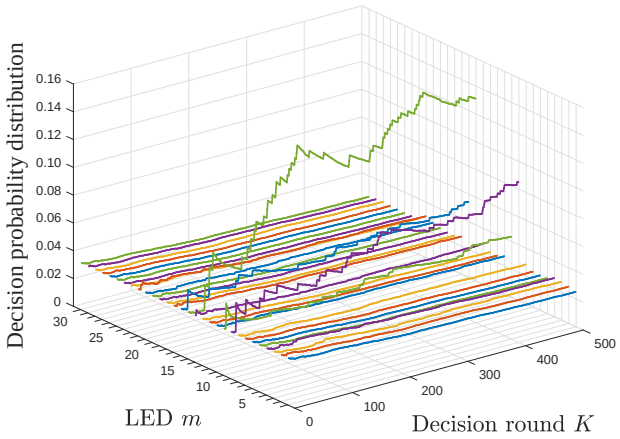


Fig. 8. The decision probability distribution versus K and LED-index based on the EXP3 algorithm. The results were calculated from Eq. (17). The departure probability is $\varsigma = 0.2$, and the location of devices is fixed at (14.5m, 8.4m).

as:

$$\overline{Q}_K = \sum_{i=1}^{K} Q(t_i, m)/Q_{\max}(t_i)$$
$$= \sum_{i=1}^{K} \frac{B}{(u_m(t_i)+1)/Q_{\max}(t_i)} \log_2\left(1 + \frac{P_r(t_i)}{\frac{\sigma^2_{total}(t_i)B}{u_m(t_i)+1} + P_I(t_i)}\right) \quad (26)$$

where $Q_{\max}(t_i)$ is the maximum theoretical throughput function value at the ith AP-decision making time instant, i.e. the maximum value when everything is known, while the mth LED lamp is the actually selected AP. Considering the states of all the LED lamps, the system's normalized throughput function at the ith AP-decision making time instant can be defined as:

$$Q_S(t_i) = \sum_{j=1}^{M} Q_j(t_i)/Q_{j\,\max}(t_i), \quad (27)$$

where $Q_j(t_i)$ is the total throughput of the jth VLC link at time instant t_i. Moreover, the system's accumulated normalized throughput function in terms of K rounds can be formulated as:

$$\overline{Q}_S = \sum_{i=1}^{K}\sum_{j=1}^{M} Q_j(t_i)/Q_{j\,\max}(t_i). \quad (28)$$

We limit the number of AP-decision rounds to $K = 300$ in our simulations, because the accumulated normalized throughput $\overline{Q}_K$ and $\overline{Q}_S$ is a monotonically increasing function of the number of AP-decision rounds K. Moreover, the simulations are repeated 1000 times to generate statistically relevant results.

Fig. 4 and Fig. 5 compare the normalized throughput of the selected VLC links and the whole system relying on the EXP3-based, ELP-based as well as random LED AP selection schemes. Fig. 6 and Fig. 7 show their accumulated values after K decision-making rounds. By contrast, the random selection scheme grants an identical decision probability of accessing any of the M LEDs, namely $1/M$, for each lamp at each decision-making time instant. Here, we assume that the negative exponential departure probability of each downloading service is $\varsigma = 0.2$. Moreover, the initial state of the number of downloading services supported by each lamp,

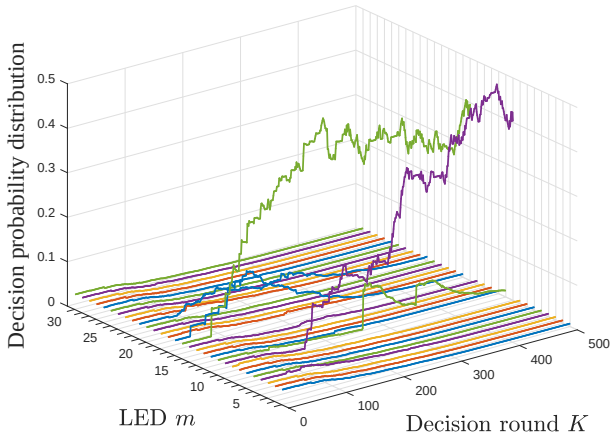


Fig. 9. The decision probability distribution versus K and LED-index based on the ELP algorithm. The results were calculated from Eq. (23). The departure probability is $\varsigma = 0.2$, and the location of devices is fixed at (14.5m, 8.4m).

i.e. $\mathbf{u}(t_1) = [u_1(t_1), u_2(t_1), ..., u_{30}(t_1)]$, is randomly chosen between $[1, 30]$. Upon increasing the number of decision rounds K, the EXP3- and ELP-based selection schemes have a higher accumulated normalized throughput than random selection. Furthermore, relying on more neighbor observation information as well as by exploiting the connection of the LED lamps, the ELP-based AP-selection scheme outperforms that based on EXP3.

Fig. 8 and Fig. 9 portray the decision probability distribution of each AP decision-making time instant. In the same way, the negative exponential departure probability of each downloading service is set to $\varsigma = 0.2$. Moreover, the initial service state of each lamp is randomly chosen from 1 to 30. Without loss of generality, the location of devices is fixed at (14.5m, 8.4m) in order to achieve relatively smooth curves. After 500 decision-making rounds, we can conclude that the decision

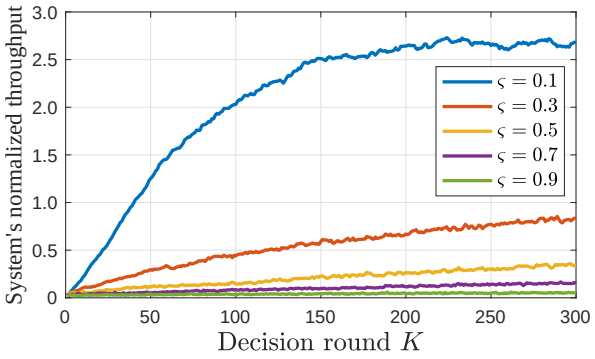


Fig. 10. The EXP3 based system's normalized throughput versus K, parameterized by the departure probability ς. The results were calculated from Eq. (27).

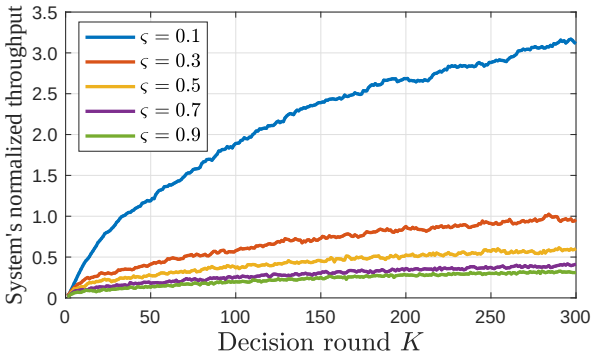


Fig. 11. The ELP based system's normalized throughput versus K, parameterized by the departure probability ς. The results were calculated from Eq. (27).

probability predominantly depends on the service state of each lamp as well as on the normalized throughput of the most recent decision-making time instant.

In Fig. 10 and Fig. 11, we exploit the influence of different departure probabilities on the system's normalized throughput. Fig. 10 illustrates the effect of the negative exponential departure probability on the system's normalized throughput based on the EXP3-aided LED AP selection. Observe from Fig. 10 that the system's normalized throughput is improved upon decreasing the negative exponential departure probability ς. A low departure probability ς results in a high overall throughput for the VLC system serving more devices at that moment. The same conclusion can be obtained from Fig. 11. Comparing each pair of curves associated with the same departure probability in Fig. 10 and Fig. 11, we can conclude that the ELP-based LED AP selection improves the throughput more substantially than the EXP3 algorithm.

Finally, Fig. 12 shows the expected value of the accumulated reward gap function and the theoretical upper bound, which represents the difference between the maximum theoretical accumulated throughput in the condition of everything known and the acquired throughput based on EXP3 and ELP. Here, we assume that the departure probability is $\varsigma = 0.3$ and let $K = 300$. The ELP-based LED AP selection scheme has a narrower reward gap and a lower upper bound than that based

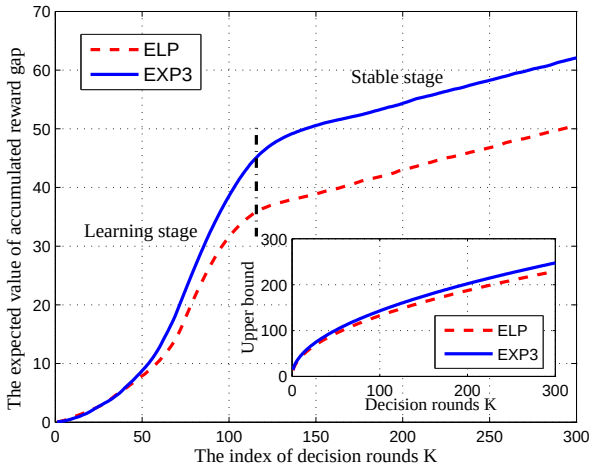


Fig. 12. The expected value of the accumulated reward gap and the upper bound for both EXP3 and ELP. The results were calculated from Eq. (20) as well as Eq. (25).

on EXP3. Moreover, the AP decision making process can be divided into two stages, namely the learning stage as well as the stable stage. With the increasing of K, the growth rate of the accumulated reward gap slows down just because of the 'online learning' capability of our proposed multi-armed bandit schemes.

In summary, the learning based LED AP selection scheme always strives for a better decision with the aid of one by one trial-and-error experiments and adjustments. Its decision generally heads in the optimal direction. Moreover, the learning schemes do not need any global information concerning the system. The information gleaned during the most recent time slot is sufficient for the decision making. By contrast, traditional mathematical tools tend to require perfect global information to reach equilibrium, which is unrealistic for a dynamic system. Furthermore, searching for the equilibrium often results in an NP-hard problem. In the hybrid LiFi-WiFi network considered, the system's state fluctuates as a function of the number of users served by each lamp, and that of the harvested energy, for example. Moreover, it is hard to acquire the exact state information at each decision-making time instant in such a dense LED distribution in the face of a dynamic environment. Hence, our proposed learning based LED AP selection scheme may be beneficial in terms of efficiently improving the overall sum-rate of the hybrid system.

VII. Conclusions

The VLC system has a vast array of compelling applications in indoor communications. In this paper, we focused our attention on the LED AP selection strategies of Lifi-WiFi hybrid networks relying on the multi-armed bandit scheme. Furthermore, in order to construct a capable hybrid Lifi-WiFi system, both the EXP3- as well as the ELP-based LED AP selection algorithms were proposed. Additionally, we conceived the upper bound of the accumulated reward gap function's expected value for both the EXP3 and ELP algorithms considering the LEDs' topology and the neighbor

observation condition. Finally, quantitative performance evaluations were provided. Based on our extensive simulations, our proposed algorithms were shown to achieve significant gains, which confirmed the efficiency of the EXP3- and ELP-based LED AP selection algorithms.

APPENDIX

1. Proof of the **Inequality**: $\forall x \in [0,1]$, $e^x \leq 1 + x + (e-2)x^2$.

Proof: Relying on Taylor's expansion, we have:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + ... \tag{29}$$

Let $x = 1$. Then Eq. (29) can be rewritten as:

$$e - 2 = \frac{1}{2!} + \frac{1}{3!} + ... \tag{30}$$

When $0 \leq x \leq 1$, we obtain:

$$\begin{aligned} &\frac{1}{2!}x^2 + \frac{1}{3!}x^3 + ... = (\frac{1}{2!} + \frac{1}{3!}x + ...)x^2 \\ &\leq (\frac{1}{2!} + \frac{1}{3!} + ...)x^2 = (e-2)x^2. \end{aligned} \tag{31}$$

Combining the Eq. (29) and Eq. (31), $\forall x \in [0,1]$, we have:

$$e^x \leq 1 + x + (e-2)x^2. \tag{32}$$

■

2. Proof of **Lemma** 1: Let Υ be a graph over M nodes and Δ represent its independence number, i.e. the size of Υ's largest independent set. Furthermore, let N_m represent the neighbor set of node m (including the node m) and let $w_1, w_2, ..., w_M$ be the arbitrary positive weights of the M nodes. Hence, we have:

$$\sum_{m=1}^{M} \frac{w_m}{\sum_{l \in N_m} w_l} \leq \Delta.$$

Proof: Assume that there exists M assigned values for $w_1, w_2, ..., w_M$ so that $\sum_{m=1}^{M} w_m / \sum_{l \in N_m} w_l > \Delta$. First, we consider a special case, where the non-zero value is assigned to nodes in the independent set I_S, which indicates that the weight values of the other nodes are zero. Therefore, the left side of the inequality can be rewritten as:

$$\sum_{m=1}^{M} \frac{w_m}{\sum_{l \in N_m} w_l} = \sum_{m \in I_S} \frac{w_m}{w_m} = |I_S|, \tag{33}$$

where $|I_S|$ represents the size of the set I_S. Then, since I_S is one of the independent sets of the graph Υ, we have $|I_S| \leq \Delta$. Therefore, there must be at least two adjacent nodes, namely r and s for example, so that $w_r > 0$ and $w_s > 0$. Let $w_r + w_s = C$, where C is a constant and fix the weight values of the other nodes. Then the left side of the inequality can be divided into the following six parts:

$$\begin{aligned} &\frac{w_r}{C + \sum_{l \in N_r \setminus r,s} w_l} + \frac{C - w_r}{C + \sum_{l \in N_s \setminus r,s} w_l} + \sum_{m:\{r,s\} \bigcap N_m = s} \\ &\frac{w_m}{C - w_r + \sum_{l \in N_m \setminus s} w_l} + \sum_{m:\{r,s\} \bigcap N_m = r} \frac{w_m}{w_r + \sum_{l \in N_m \setminus r} w_l} + \\ &\sum_{m: m \notin \{r,s\}, r,s \in N_m} \frac{w_m}{C + \sum_{l \in N_m \setminus r,s} w_l} + \\ &\sum_{m:\{r,s\} \bigcap N_m = \varnothing} \frac{w_m}{\sum_{l \in N_m} w_l}. \end{aligned} \tag{34}$$

It is plausible that each part is convex in w_r. We may conclude that the maximum of Eq. (34) is achieved when $w_r = 0$ or $w_r = c$. This means that the nodes having non-zero weight values must be non-adjacent, which contradicts to the assumption that there must be at least two adjacent nodes, which have non-zero values. The special case considered at the beginning is the optimal solution. Hence, we have $\sum_{m=1}^{M} w_m / \sum_{l \in N_m} w_l \leq \Delta$ ■

3. Proof of **Lemma** 2: Let Υ be a graph over M nodes and Δ represent its independence number, i.e. the size of Υ's largest independent set and let N_m represent the neighbor set of node m (but including the node m). Thus, there exist M values, namely $v_1, v_2, ..., v_M$ on the M-simplex, which satisfy:

$$\frac{1}{\min\limits_{m=1,2,...,M} \sum_{l \in N_m} v_l} \leq \Delta.$$

Proof: Let χ be a largest independent set of graph Υ. Hence, we have: $|\chi| = \Delta$. Consider the following specific choice for the M values, i.e. for $\forall m \in \chi$, $v_m = 1/\Delta$, otherwise $v_m = 0$. Suppose there exists a node m such that $\sum_{l \in N_m} v_l = 0$, which implies that the node m is not adjacent to any node in χ. According to the characteristic of the independent set, $\chi \cup m$ is also an independent set. It contradicts with the assumption that χ is a largest independent set. Thus, it follows that $\sum_{l \in N_m} v_l \geq 1/\Delta$. Therefore, it is true for Lemma 2. ■

4. Proof of **Theorem** 2 and **Theorem** 3 :

Proof: The weight values in Algorithm 2 can be rewritten as:

$$w_{\mathrm{m}}(t_{k+1}) = \begin{cases} w_m(t_k) e^{\mu \widetilde{Q}_m(t_k)}, & m \in N_{X_{t_k}}, \\ w_m(t_k), & otherwise. \end{cases} \tag{35}$$

Hence, we arrive at $w_m(t_{k+1}) \leq w_m(t_k) e^{\mu \widetilde{Q}_m(t_k)}$. Let $W(t_k) = \sum_{m=1}^{M} w_m(t_k)$. Then we have:

$$\frac{W(t_{k+1})}{W(t_k)} = \sum_{m=1}^{M} \frac{w_m(t_{k+1})}{W(t_k)} \leq \sum_{m=1}^{M} \frac{w_m(t_k)}{W(t_k)} e^{\mu \widetilde{Q}_m(t_k)}. \tag{36}$$

Relying on the inequality $e^x \leq 1 + x + (e-2)x^2$, Eq. (36)

can be reformulated as:

$$\sum_{m=1}^{M}\frac{w_m(t_{k+1})}{W(t_k)} \leq \sum_{m=1}^{M}\frac{w_m(t_k)}{W(t_k)}\left[1+\mu\widetilde{Q}_m(t_k)+(e-2)\mu^2\widetilde{Q}_m^2(t_k)\right]. \tag{37}$$

According to the ELP algorithm, replace $w_m(t_{k+1})/W(t_k)$ by $(p_m(t_k)-\varepsilon s_m)/(1-\varepsilon)$, yielding:

$$\begin{aligned}\sum_{m=1}^{M}\frac{p_m(t_k)-\varepsilon s_m}{1-\varepsilon}\left[1+\mu\widetilde{Q}_m(t_k)+(e-2)\mu^2\widetilde{Q}_m^2(t_k)\right] &\leq 1+\frac{\mu}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m(t_k)\\ &+\frac{(e-2)\mu^2}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k).\end{aligned} \tag{38}$$

Using the inequality $\log(1+x)\leq x$, we arrive at:

$$\begin{aligned}\log\left(\frac{W(t_{k+1})}{W(t_k)}\right) &\leq \frac{\mu}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m(t_k)\\ &+\frac{(e-2)\mu^2}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k).\end{aligned} \tag{39}$$

Summing Eq. (39) over all the K rounds, yielding:

$$\begin{aligned}\sum_{k=1}^{K}\log\left(\frac{W(t_{k+1})}{W(t_k)}\right) &= \log\left(\frac{W(t_{K+1})}{W(t_1)}\right)\\ &\leq \sum_{k=1}^{K}\frac{\mu}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m(t_k)\\ &+\sum_{k=1}^{K}\frac{(e-2)\mu^2}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k).\end{aligned} \tag{40}$$

Moreover, for any $m=1,2,...,M$, we have:

$$\begin{aligned}\log\left(\frac{W(t_{K+1})}{W(t_1)}\right) &\geq \log\left(\frac{w_m(t_{K+1})}{W(t_1)}\right)\\ &= \mu\sum_{k=1}^{K}\widetilde{Q}_m(t_k)-\log(M).\end{aligned} \tag{41}$$

Combining Eq. (40) and Eq. (41), we arrive at the following inequality:

$$\begin{aligned}\mu\sum_{k=1}^{K}\widetilde{Q}_m(t_k)-\log(M) &\leq \sum_{k=1}^{K}\frac{\mu}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m(t_k)\\ &+\sum_{k=1}^{K}\frac{(e-2)\mu^2}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k),\end{aligned} \tag{42}$$

which can be rewritten as:

$$\begin{aligned}\sum_{k=1}^{K}\widetilde{Q}_m(t_k)-\sum_{k=1}^{K}\frac{1}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m(t_k)\\ \leq \frac{\log(M)}{\mu}+\sum_{k=1}^{K}\frac{(e-2)\mu}{1-\varepsilon}\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k).\end{aligned} \tag{43}$$

Furthermore,

$$E\left[\sum_{k=1}^{K}\widetilde{Q}_m(t_k)\right]=\sum_{k=1}^{K}\sum_{i=1}^{M}p_i(t_k)E\left[\widetilde{Q}_m(t_k)\,|\text{select }i\right]. \tag{44}$$

Then, we have:

$$\begin{aligned}E\left[\sum_{k=1}^{K}\widetilde{Q}_m(t_k)\right] &= \sum_{k=1}^{K}\sum_{i\in N_m}p_i(t_k)\frac{\overline{Q}_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}\\ &= \sum_{k=1}^{K}\overline{Q}_m(t_k).\end{aligned} \tag{45}$$

Similar to Eq. (44), we arrive at the following expression:

$$\begin{aligned}&E\left[\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k)\right]\\ &=\sum_{i,m=1}^{M}p_m(t_k)p_i(t_k)E\left[\widetilde{Q}_m^2(t_k)\,|\text{select }i\right],\end{aligned} \tag{46}$$

and because $\widetilde{Q}_m\leq\varphi$, we have:

$$\begin{aligned}&E\left[\sum_{m=1}^{M}p_m(t_k)\widetilde{Q}_m^2(t_k)\right]\\ &\leq\sum_{m=1}^{M}\sum_{i\in N_m}p_m(t_k)p_i(t_k)\frac{\varphi^2}{\left(\sum_{l\in N_m}p_l(t_k)\right)^2}\\ &=\varphi^2\sum_{m=1}^{M}\frac{p_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}.\end{aligned} \tag{47}$$

Relying on Eq. (45) and Eq. (47), we take the expectations of both sides of Eq. (43), i.e.

$$\begin{aligned}\sum_{k=1}^{K}\overline{Q}_m(t_k)-\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{1}{1-\varepsilon}p_m(t_k)\overline{Q}_m(t_k)\\ \leq\frac{\log(M)}{\mu}+\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{(e-2)\mu\varphi^2}{1-\varepsilon}\frac{p_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}.\end{aligned} \tag{48}$$

According to the definition of φ and s_m, ε can be bounded as:

$$\varepsilon=\frac{\varphi\mu}{\min\limits_{j\in\mathcal{M}}\sum\limits_{l\in N_j}s_l}\leq\frac{\varphi\mu}{\min\limits_{j\in\mathcal{M}}\sum\limits_{l\in N_j}\frac{1}{M}}\leq\varphi\mu M\leq\frac{1}{2}. \tag{49}$$

Hence, we have:

$$\varepsilon(1-2\varepsilon)\geq 0. \tag{50}$$

After further manipulations, Eq. (51) can be rewritten as:

$$\frac{1}{1-\varepsilon}\leq 2\varepsilon+1\leq 2. \tag{51}$$

Therefore,

$$\begin{aligned}&\sum_{k=1}^{K}\overline{Q}_m(t_k)-\sum_{k=1}^{K}\sum_{m=1}^{M}(2\varepsilon+1)p_m(t_k)\overline{Q}_m(t_k)\\&\leq\sum_{k=1}^{K}\overline{Q}_m(t_k)-\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{1}{1-\varepsilon}p_m(t_k)\overline{Q}_m(t_k).\end{aligned}\tag{52}$$

Combining Eq. (52) with the Eq. (48), we get:

$$\begin{aligned}&\sum_{k=1}^{K}\overline{Q}_m(t_k)-\sum_{k=1}^{K}\sum_{m=1}^{M}p_m(t_k)\overline{Q}_m(t_k)\\&\leq 2\sum_{k=1}^{K}\sum_{m=1}^{M}\varepsilon p_m(t_k)\overline{Q}_m(t_k)+\frac{\log(M)}{\mu}\\&+\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{(e-2)\mu\varphi^2}{1-\varepsilon}\frac{p_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}\\&\leq 2\varphi\sum_{k=1}^{K}\varepsilon+\frac{\log(M)}{\mu}+\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{(e-2)\mu\varphi^2}{1-\varepsilon}\frac{p_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}.\end{aligned}\tag{53}$$

Substituting the definition of ε into Eq. (53), we have:

$$\begin{aligned}&\sum_{k=1}^{K}\overline{Q}_m(t_k)-\sum_{k=1}^{K}\sum_{m=1}^{M}p_m(t_k)\overline{Q}_m(t_k)\\&\leq 2\varphi\sum_{k=1}^{K}\frac{\varphi\mu}{\min\limits_{j\in M}\sum_{l\in N_j}s_l}+\frac{\log(M)}{\mu}\\&+\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{(e-2)\mu\varphi^2}{1-\varepsilon}\frac{p_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}\\&\leq 2\mu\varphi^2\left(\sum_{k=1}^{K}\frac{1}{\min\limits_{j\in M}\sum_{l\in N_j}s_l}+\sum_{k=1}^{K}\sum_{m=1}^{M}(e-2)\frac{p_m(t_k)}{\sum_{l\in N_m}p_l(t_k)}\right)\\&+\frac{\log(M)}{\mu}.\end{aligned}\tag{54}$$

Invoking Lemma 1 as well as Lemma 2 and exploiting the randomness of m, the upper bound of the accumulated reward gap function's expected value can be written as:

$$\begin{aligned}E[R(K)]&=\sum_{k=1}^{K}\overline{Q}(i_k^*,t_k)-E\left[\sum_{k=1}^{K}\overline{Q}(a_k,t_k)\right]\\&\leq 2\mu\varphi^2(e-1)K\Delta+\frac{\log(M)}{\mu},\end{aligned}\tag{55}$$

where i_k^* is the optimal strategy of the kth decision round, while a_k is the actual selection. Relying on the geometric inequality, let $\mu=\sqrt{\frac{\log(M)}{2\varphi^2(e-1)K\Delta}}$, and we have:

$$E[R(K)]\leq\varphi\sqrt{2(e-1)K\Delta\log(M)}.\tag{56}$$

For the lattice based network, we can obtain the Theorem 2 from Eq. (56). ■

REFERENCES

[1] L. Hanzo, H. Haas, S. Imre, D. O'Brien, M. Rupp, and L. Gyongyosi, "Wireless myths, realities, and futures: from 3G/4G to optical and quantum wireless," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1853–1888, May 2012.

[2] M. Saadi, L. Wattisuttikulkij, Y. Zhao, and P. Sangwongngam, "Visible light communication: opportunities, challenges and channel models," *International Journal of Electronics & Informatics*, vol. 2, no. 1, pp. 1–11, Feb.

[3] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[4] D. Tsonev, S. Videv, and H. Haas, "Light fidelity (Li-Fi): towards all-optical networking," in *SPIE OPTO*. International Society for Optics and Photonics, Dec. 2013, pp. 702–900.

[5] S. H. Lee, S.-Y. Jung, and J. K. Kwon, "Modulation and coding for dimmable visible light communication," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 136–143, Feb. 2015.

[6] D. Ding and X. Ke, "A new indoor VLC channel model based on reflection," *Optoelectronics Letters*, vol. 6, no. 4, pp. 295–298, Jul. 2010.

[7] J. Song, W. Ding, F. Yang, H. Yang, B. Yu, and H. Zhang, "An indoor broadband broadcasting system based on PLC and VLC," *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 299–308, Jun. 2015.

[8] J. Vucic, C. Kottke, S. Nerreter, K.-D. Langer, and J. W. Walewski, "513 Mbit/s visible light communications link based on DMT-modulation of a white LED," *Journal of lightwave technology*, vol. 28, no. 24, pp. 3512–3518, Dec. 2010.

[9] B. G. Guzman, A. L. Serrano, and V. P. G. Jimenez, "Cooperative optical wireless transmission for improving performance in indoor scenarios for visible light communications," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 4, pp. 393–401, Apr. 2015.

[10] I.-C. Lu, C.-H. Yeh, D.-Z. Hsu, and C.-W. Chow, "Utilization of 1-GHz VCSEL for 11.1-Gbps OFDM VLC wireless communication," *IEEE Photonics Journal*, vol. 8, no. 3, pp. 1–6, Jun. 2016.

[11] Y. Wang, L. Tao, Y. Wang, and N. Chi, "High speed WDM VLC system based on multi-band CAP64 with weighted pre-equalization and modified CMMA based post-equalization," *IEEE Communications Letters*, vol. 18, no. 10.

[12] R. Zhang, J. Wang, Z. Wang, Z. Xu, C. Zhao, and L. Hanzo, "Visible light communications in heterogeneous networks: Paving the way for user-centric design," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 8–16, Apr. 2015.

[13] K. H. Schlag, "Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits," *Journal of economic theory*, vol. 78, no. 1, pp. 130–156, Jan. 1998.

[14] S. Maghsudi and E. Hossain, "Multi-armed bandits with application to 5G small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, Mar. 2016.

[15] Y. U. Lee and M. Kavehrad, "Two hybrid positioning system design techniques with lighting LEDs and ad-hoc wireless network," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1176–1184, Nov. 2012.

[16] O. Bouchet, M. El Tabach, M. Wolf, D. C. O'Brien, G. E. Faulkner, J. W. Walewski, S. Randel, M. Franke, S. Nerreter, K.-D. Langer *et al.*, "Hybrid wireless optics (HWO): Building the next-generation home network," in *The 6th International Symposium on Communication Systems, Networks and Digital Signal Processing*, Jul. 2008, pp. 283–287.

[17] M. B. Rahaim, A. M. Vegni, and T. D. Little, "A hybrid radio frequency and broadcast visible light communication system," in *2011 IEEE Global Communications Conference Workshops*, Dec. 2011, pp. 792–796.

[18] S. Shao and A. Khreishah, "Delay analysis of unsaturated heterogeneous omnidirectional-directional small cell wireless networks: The case of rf-vlc coexistence," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8406–8421, 2016.

[19] X. Bao, X. Zhu, T. Song, and Y. Ou, "Protocol design and capacity analysis in hybrid network of visible light communication and OFDMA systems," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 4, pp. 1770–1778, May 2014.

[20] Y. Xing, R. Chandramouli, and C. Cordeiro, "Price dynamics in competitive agile spectrum access markets," *IEEE Journal on selected areas in communications*, vol. 25, no. 3, pp. 613–621, Mar. 2007.

[21] J. Elias, F. Martignon, L. Chen, and E. Altman, "Joint operator pricing and network selection game in cognitive radio networks: Equilibrium, system dynamics and price of anarchy," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4576–4589, Sep. 2013.

[22] I. Malanchini, M. Cesana, and N. Gatti, "Network selection and resource allocation games for wireless access networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2427–2440, Dec. 2013.

[23] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *IEEE International Conference on Computer Communications (INFOCOM)*, Apr. 2013, pp. 998–1006.

[24] J. Lee and S. Bahk, "On the mdp-based cost minimization for video-on-demand services in a heterogeneous wireless network with multihomed terminals," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1737–1749, Sep. 2013.

[25] J. Wang, C. Jiang, Z. Han, Y. Ren, and L. Hanzo, "Network association strategies for an energy harvesting aided super-WiFi network relying on measured solar activity," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3785–3797, Dec. 2016.

[26] K. Zhu, E. Hossain, and D. Niyato, "Pricing, spectrum sharing, and service selection in two-tier small cell networks: A hierarchical dynamic game approach," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1843–1856, Oct. 2014.

[27] Y. Yang, Y. Chen, C. Jiang, C.-Y. Wang, and K. R. Liu, "Wireless access network selection game with negative network externality," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5048–5060, Oct. 2013.

[28] X. Wu, D. Basnayaka, M. Safari, and H. Haas, "Two-stage access point selection for hybrid VLC and RF networks," in *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 1–6.

[29] M. D. Soltani, X. Wu, M. Safari, and H. Haas, "Access point selection in Li-Fi cellular networks with arbitrary receiver orientation," in *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 1–6.

[30] Y. Liu, Z. Huang, W. Li, and Y. Ji, "Game theory-based mode cooperative selection mechanism for device-to-device visible light communication," *Optical Engineering*, vol. 55, no. 3, pp. 030 501.1–030 501.4, May.

[31] A. Matsui, "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, vol. 57, no. 2, pp. 343–362, Aug. 1992.

[32] L. E. Blume, "The statistical mechanics of best-response strategy revision," *Games and Economic Behavior*, vol. 11, no. 2, pp. 111–145, Nov. 1995.

[33] T. Anderson, *The theory and practice of online learning*. Athabasca University Press, 2008.

[34] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

[35] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 100–107, Dec. 2004.

[36] J. Grubor, S. Randel, K.-D. Langer, and J. W. Walewski, "Broadband information broadcasting using LED-based interior lighting," *IEEE Journal of Lightwave Technology*, vol. 26, no. 24, pp. 3883–3892, Dec. 2008.

[37] F. Jin, R. Zhang, and L. Hanzo, "Resource allocation under delay-guarantee constraints for heterogeneous visible-light and RF femtocell," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1020–1034, Feb. 2015.

[38] F. Jin, X. Li, R. Zhang, C. Dong, and L. Hanzo, "Resource allocation under delay-guarantee constraints for visible-light communication," *IEEE Access*, vol. 4, pp. 7301–7312, May 2016.

[39] S. Bubeck, N. Cesa-Bianchi *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[40] S. Mannor and O. Shamir, "From bandits to experts: On the value of side-observations," in *Advances in Neural Information Processing Systems*, 2011, pp. 684–692.

[41] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Communications of the ACM*, vol. 16, no. 9, pp. 575–577, Sept. 1973.

[42] H. Johnston, "Cliques of a graph-variations on the Bron-Kerbosch algorithm," *International Journal of Parallel Programming*, vol. 5, no. 3, pp. 209–238, Sep. 1976.

Jingjing Wang (S'14) received his B.S. degree in Electronic Information Engineering from Dalian University of Technology in 2014 with the highest honors. He currently works for his PhD degree at Complex Engineered Systems Lab (CESL) in Tsinghua University, Beijing. From Jan. 2016 to Mar. 2016, he visited Wireless Networks and Decision Systems (WNDS) Group, Singapore University of Technology and Design as a visiting student. Since Otc. 2017, he has been visiting the Telecommunications Group, University of Southampton as a joint PhD student. His research interests include the resource allocation and network association, learning theory aided modeling, analysis and signal processing, and information diffusion theory. He received Tsinghua GuangHua Scholarship in 2016 and graduate China National Scholarship Award in 2017.

Chunxiao Jiang (S'09-M'13-SM'15) received the B.S. in information engineering from Beihang University in Jun. 2008 and the Ph.D. in electronic engineering from Tsinghua University in Jan. 2013, both with the highest honors. From Feb. 2013 - Jun. 2016, Dr. Jiang was a Postdoc in the Department of Electronic Engineering Tsinghua University, during which he visited University of Maryland College Park and University of Southampton. He is a recipient of the IEEE Globecom Best Paper Award in 2013, the IEEE GlobalSIP Best Student Paper Award in 2015, the IEEE IWCMC Best Paper Award in 2017, and the IEEE Communications Society Young Author Best Paper Award in 2017. Since 2015, Dr. Jiang became a IEEE Senior Member.

Haijun Zhang (M'13, SM'17) is currently a Full Professor in University of Science and Technology Beijing, China. He was a Postdoctoral Research Fellow in Department of Electrical and Computer Engineering, the University of British Columbia (UBC), Vancouver Campus, Canada. From 2011 to 2012, he visited Centre for Telecommunications Research, King's College London, London, UK, as a Visiting Research Associate. Dr. Zhang has published more than 100 papers and authored 2 books. He serves as Editor of IEEE Transactions on Communications, IEEE 5G Tech Focus, and Journal of Network and Computer Applications, and serves/served as a Leading Guest Editor for IEEE Communications Magazine, and IEEE Transactions on Emerging Topics in Computing. He serves/served as General Co-Chair of 5GWN'17 and GameNets'16, Track Chair of ScalCom2015, Symposium Chair of GameNets'14, TPC Co-Chair of INFOCOM 2018 Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Networks, General Co-Chair of ICC 2018 (ICC 2017, Globecom 2017) Workshop on 5G Ultra Dense Networks, and General Co-Chair of Globecom 2017 Workshop on LTE-U. He has served as a TPC member in a numerous international conferences. He received the IEEE ComSoc Young Author Best Paper Award in 2017.

Xin Zhang received his bachelor degree from the Department of Information Engineering, Beijing University of Posts and Telecommunications in 2015. He is pursing his master degree at the Department of Electronic Engineering, Tsinghua University. His research interests include the space information network, MIMO as well as the optimization theory.

Victor C. M. Leung (S'75, M'89, SM'97, F'03) received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (UBC) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at UBC on a Canadian Natural Sciences and Engineering Research Council Postgraduate Scholarship and received the Ph.D. degree in electrical engineering in 1982.

From 1981 to 1987, Dr. Leung was a Senior Member of Technical Staff and satellite system specialist at MPR Teltech Ltd., Canada. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He returned to UBC as a faculty member in 1989, and currently holds the positions of Professor and TELUS Mobility Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering. Dr. Leung has co-authored more than 1000 journal/conference papers, 37 book chapters, and co-edited 13 book titles. Several of his papers had been selected for best paper awards. His research interests are in the broad areas of wireless networks and mobile systems.

Dr. Leung is a registered Professional Engineer in the Province of British Columbia, Canada. He is a Fellow of IEEE, the Royal Society of Canada, the Engineering Institute of Canada, and the Canadian Academy of Engineering. He was a Distinguished Lecturer of the IEEE Communications Society. He is serving on the editorial boards of the IEEE Transactions on Green Communications and Networking, IEEE Transactions on Cloud Computing, IEEE Access, Computer Communications, and several other journals, and has previously served on the editorial boards of the IEEE Journal on Selected Areas in Communications - Wireless Communications Series and Series on Green Communications and Networking, IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, IEEE Transactions on Computers, IEEE Wireless Communications Letters, and Journal of Communications and Networks. He has guest-edited many journal special issues, and provided leadership to the organizing committees and technical program committees of numerous conferences and workshops. He received the IEEE Vancouver Section Centennial Award and 2011 UBC Killam Research Prize. He is the recipient of the 2017 Canadian Award for Telecommunications Research. He co-authored papers that won the 2017 IEEE ComSoc Fred W. Ellersick Prize and the 2017 IEEE Systems Journal Best Paper Award.

Lajos Hanzo (http://www-mobile.ecs.soton.ac.uk) FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 111 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1701 research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk. Lajos has 30 000+ citations and an H-index of 70.