Multiple Access Design for Ultra-Dense VLC Networks: Orthogonal vs Non-Orthogonal

Simeng Feng, Rong Zhang, *Senior Member, IEEE*, Wei Xu, *Senior Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*— Small-cell aided ultra-dense networks (UDNs) constitute an efficient solution to the ever-increasing thirst for more data. Thanks to the vast untapped high-frequency spectrum of visible light, visible light communications (VLCs) are a natural candidate for UDN. In this paper, layered asymmetrically clipped optical OFDM (LACO-OFDM) aided ultra-dense VLC (UD-VLC) is investigated in terms of its user association, multiple access (MA), and resource allocation. To handle the severe inter-cell interference (ICI) amongst the densely deployed access points, we propose a novel overlapped clustering technique relying on a hybrid non-orthogonal MA and orthogonal MA scheme for enhancing the performance, with the aid of our dynamic resource allocation conceived. Our simulations show that the proposed LACO-OFDM aided UD-VLC using our hybrid MA scheme is more robust against the ICI, at a price of modestly decreasing the sum throughput.**

***Index Terms*— Ultra-dense network (UDN), visible light communications (VLC), multiple access (MA), non-orthogonal multiple access (NOMA), resource allocation.**

I. Introduction

THE number of connected devices is envisioned to be about 8.6 Billion by 2020, which results in the escalation of tele-traffic [1]. Given this challenge, both academia and industry have invested substantial efforts into exploring the barely tapped higher-frequency spectrum, such as the visible light spectrum [2]. According to the Friis free space equation, higher frequencies however suffer from a higher propagation path loss [3]. Hence, small cell based network design is put forward for 5G and beyond to compensate for the path loss, leading to the ultra-dense network (UDN) concept [4]–[6]. By densely deploying the access points (APs) having a lower power and coverage, the UDN concept is capable of increasing the capacity by enabling light-of-sight (LoS) transmissions [7]. However, due to the short distances between the APs, the interference emanating from neighboring APs makes efficient interference management a crucial issue in UDNs.

As a promising potential next-generation technique, visible light communication (VLC) has attracted substantial interests due to its high-rate wireless connectivity [8]. Due to its predominantly LoS propagation characteristics, VLC provides an efficient connectivity option for UDN [9]. By modulating the intensity of light emitting diodes (LED) above the fusion-frequency of the human eye, VLC is capable of simultaneous communication and illumination [10]. Moreover, visible light cannot propagate through walls, which creates a secure communication environment [11]. Based on this, our paper is focused on an ultra-dense VLC (UD-VLC) multi-user scenario, with an emphasis on access management, interference control and resource allocation.

A. Related Literature

For point-to-point transmission in VLC, besides the classic asymmetrically clipped optical OFDM (ACO-OFDM) and DC-biased optical OFDM (DCO-OFDM), more advanced modulations have been proposed. To elaborate, asymmetrically clipped DC biased OFDM (ADO-OFDM) applies ACO-OFDM to the odd-indexed subcarriers, whilst the even-indexed subcarriers transmit DCO-OFDM [12], which still requires DC bias. In the hybrid ACO-OFDM (HACO-OFDM), ACO-OFDM is transmitted on the odd-indexed subcarriers, further combined with pulse-amplitude-modulated discrete multitone (PAM-DMT) on the even-indexed subcarriers with only the imaginary part being modulated [13], while the real part of the even subcarriers still remains untapped. By contrast, layered ACO-OFDM (LACO-OFDM) improves the ACO-OFDM by actively employing more subcarriers in a layer-based manner, hence attaining a higher rate than ACO-OFDM, despite requiring a lower power than DCO-OFDM [14]. To reduce the peak-to-average power ratio (PAPR) of LACO-OFDM, a tone-injection aided PAPR reduction design was proposed in [15], while an interleaved discrete-Fourier-transform-spread L/e-ACO-OFDM (IDFTS-L/e-ACO) scheme was advocated in [16]. To improve the performance of LACO-OFDM, a two-stage receiver design was conceived in [17], where soft successive interference cancellation (SIC) was invoked at the first stage for suppressing the inter-layer interference (ILI), while the clipping noise was handled at the second stage for suppressing the additive noise.

Manuscript received June 8, 2018; revised October 23, 2018; accepted November 24, 2018. Date of publication November 30, 2018; date of current version March 15, 2019. The work of W. Xu was supported in part by the NSFC under grant 61871109 and the Royal Academy of Engineering under the Distinguished Visiting Fellowship scheme. L. Hanzo would like to acknowledge the financial support of the EPSRC projects EP/Noo4558/1, EP/PO34284/1, of the Royal Society's GRFC Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom. The associate editor coordinating the review of this paper and approving it for publication was H. Elgala. *(Corresponding author: Lajos Hanzo.)*

S. Feng, R. Zhang, and L. Hanzo are with the Next Generation Wireless, School of Electronics and Computer Science, The University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

W. Xu is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 211100, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2884482

Furthermore, a multi-class channel coding scheme was proposed for LACO-OFDM in [18], where a signal-to-noise ratio (SNR) gain of 3.6 dB was attained.

When designing UD-VLC networks, the association between the user equipment (UE) and APs plays a crucial role. In conventional networks, the association is generally implemented based on the highest signal-to-interference-plus-noise ratio (SINR), but this *one-to-one* association was deemed to be inefficient [19]. Given the large number of APs in UDNs, Kamel *et al.* [6] proposed a *one-to-many* association, where the UE is capable of connecting to more than one AP. However, the inter-cell interference (ICI) constitutes a bottleneck of this design. Furthermore, the *many-to-many* user-centric association has been proposed in [20], where a group of neighboring APs jointly form a larger cluster to support multiple UEs. By invoking the transmit pre-coding (TPC) techniques, it constitutes a multi-UE-multi-AP transmission, which has been demonstrated to provide an improved bandwidth efficiency [21]. Moreover, instead of simply focusing on static environments, an anticipatory association has been put forward by considering both the UE's mobility and the dynamically fluctuating wireless traffic [22].

As for the multi-user access, the most straightforward arrangement is time division multiple access (TDMA), where the APs offer access to their associated UEs in orthogonal time slots. However, in UDN, TDMA may not work well, since the number of UEs supported by a given AP may be high. As a parallel development, non-orthogonal multiple access (NOMA) has drawn substantial attention [23]. In contrast to orthogonal multiple access (OMA) techniques, NOMA allows multiple UEs to access the network by multiplexing them in the power domain, where SIC is applied at the receiver side to separate the information of each individual UE [24]. It is clearly beneficial that multiple UEs are capable of simultaneously using the same time- and spectral-slots. Marshoud *et al.* [25] analyzed the bit-error-rate (BER) performance of NOMA-VLC under different channel uncertainties, where the system model relied on a single AP associated with fixed transmit power. Similarly, in single AP deployment, Yin *et al.* [26] demonstrated that both the coverage and ergodic sum rate of the NOMA technique relying on user-pairing is better than that of OMA. To achieve fairness, Yang *et al.* [27] proposed an optimal power allocation solution for maximizing the sum rate. However, these studies ignored the ICI, which is unrealistic in UD-VLC networks [28]. Given the propagation characteristics of VLC, the ICI-level strongly depends on the LoS path, which is detrimental [29].

B. Motivation and Contributions

Against this background, we are inspired to explore the system's performance when LACO-OFDM is invoked in our UD-VLC network. Although TPC aided multi-AP-multi-UE transmission is widely used in VLC network, TPC is incompatible with the SIC-aided demodulation of LACO-OFDM. Therefore, a radical paradigm shift and a new system architecture is timely required in the LACO-OFDM aided UD-VLC network. Motivated by it, a completely new clustering aided MA design is conceived for the LACO-OFDM aided UD-VLC network, where the resource allocation strategy is proposed for dynamically exploiting the flexibility of the LACO-OFDM at the system-level. To the best of our knowledge, this paper is the first one intrinsically amalgamating LACO-OFDM with our holistic VLC network design. The main contributions of this paper are summarized as follows.

1) A Novel Clustering Strategy for LACO-OFDM Is Conceived: Upon considering the commonly existed bidirectional multi-cell connection in UD-VLC, a novel overlapped clustering strategy relying on graph theory is proposed for our LACO-OFDM aided UD-VLC network. Both the channel conditions of the UEs and the transmission coverage of the APs are considered, yielding a *many-to-many* association structure. As demonstrated in our simulations, the proposed *many-to-many* association outperforms the *one-to-many* association both in terms of its throughput and outage probability (OP).

2) An Enhanced MA Technique Is Designed for LACO-OFDM Aided UC-VLC Network: We found that although the overlapping areas of several clusters relying on the proposed *many-to-many* association have the potential of mitigating the ICI imposed by NOMA, the ICI imposed by the densely deployed APs and UEs may still remain excessive. To circumvent this problem, a hybrid NOMA and OMA scheme is proposed, where TDMA-based scheduling and graph-coloring based frequency reuse (FR) may be adaptively invoked, depending on the specific scenarios encountered. Our simulations indicate that the achievable throughput of LACO-OFDM aided UD-VLC relying on the proposed MA design is higher than that of the ACO/DCO-OFDM relying on merged-cell construction. Furthermore, by invoking our hybrid NOMA and OMA scheme, no limitation is imposed on the number of simultaneously served UEs, which cannot be realized in the conventional merged-cell construction due to the TPC employed.

3) A Dynamic Resource Allocation Is Proposed for Exploiting the Flexibility of Our Network: In the LACO-OFDM aided UD-VLC network we constructed, we want to attain the maximum achievable sum-throughput[1] of our new resource allocation strategy. To fully exploit the flexibility of LACO-OFDM, we aim for allocating different layers to different UEs, where for each active layer different modulation-modes may be employed. These decisions are made on the basis of the prevalent communication conditions and practical constraints. Hence, an heuristic dynamic resource allocation strategy is conceived for maximizing the sum-throughput of the UD-VLC network under a range of practical constraints, such as the maximum transmit power, the maximum backhaul-rate and the minimum individual UE rates.

C. Paper Organization

The rest of this paper is organized as follows. The UD-VLC system model specifying the VLC propagation,

[1]Instead of relying on the idealized continuous-input continuous-output memoryless channel (CCMC) capacity metric, we choose the modulation-mode-related throughput as our metric, which reflects the practically achievable throughput of the network.

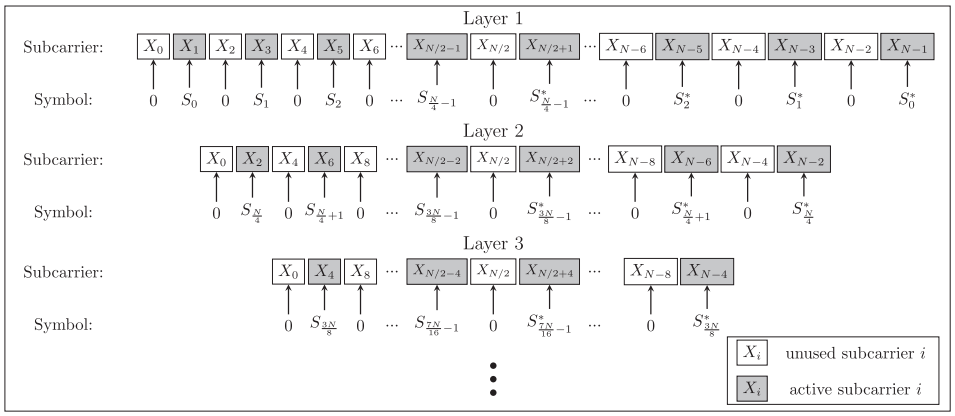


Fig. 1. Schematic of LACO-OFDM relying on N subcarriers.

physical-layer transmission and network association is presented in Section II. Both the conventional and our novel MA schemes are discussed in Section III. Section IV introduces the dynamic resource allocation proposed for the UD-VLC network constructed. Our numerical results are discussed in Section V, whilst our conclusions are drawn in Section VI.

Throughout this paper, we employ calligraphic letters (i.e., $\mathcal{A}$) to represent sets, where $|\mathcal{A}|$ denotes the size of $\mathcal{A}$, while $\binom{a}{b}$ denotes a combination of b items taken from a items. Furthermore, $\mathbb{E}(\bullet)$ and $\mathbb{D}(\bullet)$ represent the expectation and variance, respectively. We let $\sum \boldsymbol{a}$ represent the sum of all elements in vector $\boldsymbol{a}$, and $\mathbf{0}_{(l\times 1)}$ indicates an all-zero vector of length l.

II. System Model

An indoor environment is considered in our UD-VLC network, where the APs are densely deployed on the ceiling facing downward and rely on single-color-LED arrays. We assume that a central controller monitors the entire room and the UD-VLC is supported by a Wi-Fi uplink (UL) and VLC downlink (DL).

A. VLC-Based Propagation Modeling

Our VLC propagation model takes into account both the LoS path and the first-reflection path of visible light. The power of higher-order reflections accompained by the scattering and diffusion is much weaker, and hence it is neglected [30]. According to the Lambertian radiation pattern, the LoS channel gain of UE k fitted with a photo-diode (PD) facing upwards with a pre-designed field-of-view (FoV), ψ_{FoV}, which is provided by AP n, can be formulated as [31]:

$$h_{k,n}^{(0)} = \frac{(\alpha+1)A_s}{2\pi D_{k,n}^2}\cos^{\alpha}(\phi) g_{of}(\psi) g_{oc}(\psi)\cos(\psi), \quad 0 \le \psi \le \psi_{FoV}, \quad (1)$$

where $\alpha = \ln 2/\ln[\cos(\phi_{1/2})]$ is the order of Lambertian emission, and $\phi_{1/2}$ denotes the semi-angle at half power. The physical area of a PD-based receiver at UE-end is A_s, while $D_{k,n}$ denotes the Euclidean distance between UE k and AP n, while ϕ and ψ represent the angles of irradiance and of incidence, respectively. Furthermore, $g_{of}(\psi)$ is the gain of the optical filter and $g_{oc}(\psi) = \varrho^2/\sin^2(\psi_{FoV})$ is the gain of the optical concentrator, where ϱ denotes the refractive index. The channel gain of the first reflected ray is given by: $h_{k,n}^{(1)} = \int_{\text{walls}} dh_{k,n}^{(1)}$, where [32]

$$dh_{k,n}^{(1)} = \frac{\rho(\alpha+1)A_s}{2\pi^2 D_{k,r}^2 D_{r,n}^2} dA_{\text{wall}} \cos^{\alpha}(\phi)\cos(\omega_1)\cos(\omega_2) g_{of}(\psi) g_{oc}(\psi)\cos(\psi), \quad 0 \le \psi \le \psi_{FoV}. \quad (2)$$

Note that $D_{k,r}$ is the Euclidean distance between UE k and the reflection point, while $D_{r,n}$ is the Euclidean distance between the reflection point and AP n. The reflective area is denoted by dA_{wall} and ρ is the reflection coefficient, while ω_1 and ω_2 are the angles of irradiance with respect to the reflection point and to the UE, respectively. Note that the aggregated channel gain between UE k and AP n is expressed as: $h_{k,n} = h_{k,n}^{(0)} + h_{k,n}^{(1)}$.

B. Physical-Layer Transmission

For the sake of balancing the power vs. spectral efficiency, LACO-OFDM is invoked in our UD-VLC as the physical-layer transmission strategy. Fig. 1 demonstrates the operation of LACO-OFDM. A block of N frequency-domain (FD) samples is represented by $\boldsymbol{X} = \left[X_0, X_1, \cdots, X_{N-1}\right]$ and the source symbol is denoted by $\boldsymbol{S} = [S_0, S_1, \cdots, S_{N-1}]$, which is the result of the appropriately mapped transmitted bit stream according to the modulation scheme chosen. Assume that the transmitted signal is composed of a total of L layers. It can be observed that the first layer of LACO-OFDM operates exactly as the conventional ACO-OFDM having N subcarriers, which can only accommodate $N/4$ symbols and leaving $N/2$ unused subcarriers denoted as $[X_0, X_2, X_4, \cdots, X_{N-2}]$. To efficiently exploit the spectrum, those unused subcarriers can then be filled by ACO-OFDM in the second layer, yielding $N/8$ more transmitted symbols. The remaining layers obey the same philosophy, resulting in efficient spectrum utilization. Compared to the traditional ACO-OFDM, the LACO-OFDM associated with L layers is hence capable of transmitting $\sum_{l=2}^{L} N/2^{l+1}$ more source symbols without adding a DC bias. To provide beneficial flexibility, the number of active layers may vary

among the UEs. For each active layer, the modulation mapping scheme may also be different.

C. Network Association

In UD-VLC network, owing to the densification of APs, a given UE may be in the vicinity of multiple APs. Although the channel gain offers a straightforward means of determining the topology, our network association purely relying on it would remain inefficient. This is because the UEs located at the corner or edge of the room are more likely to receive strong reflections from the walls [33]. However, it is not beneficial to associate a UE with an AP providing strong reflections if their Euclidean distance is high. As a further parameter, the transmission distance may also be considered. Inspired by Schaeffer [34], the association between a pair of distinct entities, such as the APs and UEs, allows us to invoke bipartite graph theory. In order to fully explore the coverage overlap of APs, a graph-based clustering strategy is proposed for our association problem in the context of UD-VLC networks.

1) Overlapped Clustering: In order to carry out the association relying on the channel conditions and on the propagation distance, based on graph theory, we build a directed bipartite graph, $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a)$, where the vertex set is given by $\mathcal{V}_a = \mathcal{U} \cup \mathcal{A}$. The UE set $\mathcal{U}$ and AP set $\mathcal{A}$ hold a total of $|\mathcal{U}|$ UEs and $|\mathcal{A}|$ APs, respectively. The edge set $\mathcal{E}_a = \mathcal{E}_a^{\text{ue}\rightarrow\text{ap}} \cap \mathcal{E}_a^{\text{ue}\leftarrow\text{ap}}$ accommodates the intersection of the set, $\mathcal{E}_a^{\text{ue}\rightarrow\text{ap}}$, holding edges spanning from the UE-end to the AP-end and the set, $\mathcal{E}_a^{\text{ue}\leftarrow\text{ap}}$, containing edges emerging from the AP-end to the UE-end. To complete this directed bipartite graph $\mathcal{G}_a$, we conceive three stages detailed as follows.

(a) *Channel-Gain-Guaranteed Matching:* In this stage, the preference of association is based on the channel gain. If the aggregated channel gain $h_{k,n}$ between AP n and UE k is larger than a pre-defined quality threshold H_{thr}, an edge $e_{k\rightarrow n}$ is built with a direction from UE-end to AP-end. Each UE in $\mathcal{U}$ fully explores all potential APs and the resultant edges are held in $\mathcal{E}_a^{\text{ue}\rightarrow\text{ap}}$, which is expressed as:

$$\mathcal{E}_a^{\text{ue}\rightarrow\text{ap}} = \left\{ e_{k\rightarrow n} : h_{k,n} \geq H_{\text{thr}}, \forall k \in \mathcal{U}, \forall n \in \mathcal{A} \right\}. \quad (3)$$

This UE-end matching is exemplified in a 6-UEs-8-APs scenario shown in the Fig. 2(a), where all directed edges in $\mathcal{E}_a^{\text{ue}\rightarrow\text{ap}}$ are denoted by solid arrows.

(b) *Transmission-Distance-Restricted Matching:* In this stage, we conceive a transmission-distance restriction as a complementary matching metric. Explicitly, for a given AP n, an edge $e_{k\leftarrow n}$ will be constructed with a direction emerging from AP n to UE k, if the Euclidean distance between them is below the pre-defined distance threshold D_{thr}. Each AP searches through the set of all the potential UEs and the edges generated in this stage are held in the set $\mathcal{E}_a^{\text{ue}\leftarrow\text{ap}}$ as:

$$\mathcal{E}_a^{\text{ue}\leftarrow\text{ap}} = \left\{ e_{k\leftarrow n} : D_{k,n} \leq D_{\text{thr}}, \forall n \in \mathcal{A}, \forall k \in \mathcal{U} \right\}. \quad (4)$$

This AP-end matching is displayed in Fig. 2(b), where the directed edges in $\mathcal{E}_a^{\text{ue}\leftarrow\text{ap}}$ are given by hollow arrow tips.

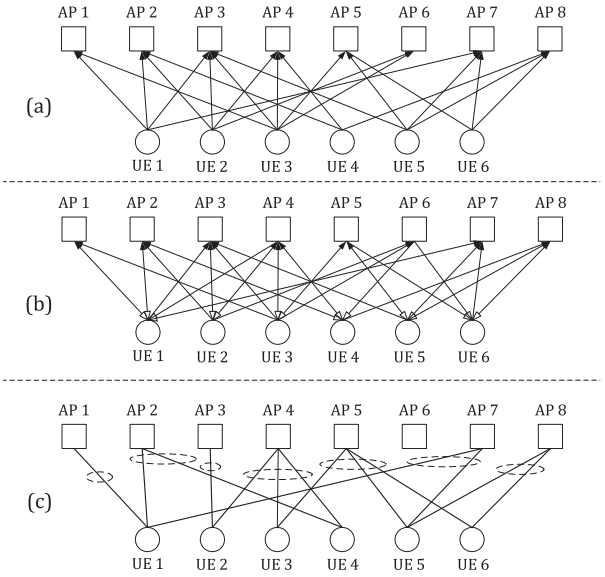


Fig. 2. An example of the proposed overlapped clustering. (a) UE-end channel-strength-guaranteed matching (directed edges are represented by solid arrows); (b) AP-end transmission-distance-restricted matching (directed edges are given by hollow arrows); (c) Confirmation.

(c) *Confirmation:* After the above two stages, the topology ends up with four different types of associations: unidirectional connection from UE-end to AP-end, unidirectional connection from AP-end to UE-end, bidirectional connection and no connection. By satisfying both the channel gain and transmission distance criteria, we confirm the association between the APs and UEs, provided that they are connected bidirectionally, as depicted in Fig. 2(c).

The graph $\mathcal{G}_a$ contributes to our overlapping clusters, where a given AP may have more than one target UE, while a given UE may be associated with multiple APs, ending up with a *many-to-many* matching solution. In the above example, AP 6 has no associated UEs, hence it switches to a "sleep" mode for energy saving.

2) Complexity Analysis: We assume that the number of candidate APs providing adequate links for UE k at Stage (a) is Q_k and the number of candidate UEs located within the maximum tolerance distance from AP n at Stage (b) is K_n. For the first two stages, the complexity order is given by $\mathcal{O}\big(\sum_{k=1}^{|\mathcal{U}|} Q_k\big)$ and $\mathcal{O}\big(\sum_{n=1}^{|\mathcal{A}|} K_n\big)$, respectively. During the last stage, we have to check the possible connections in the second stage for each possible association between the two players paired in the first stage. This implies an extra complexity order of $\sum_{k=1}^{|\mathcal{U}|} \big(Q_k \sum_{n=1}^{|\mathcal{A}|} K_n\big)$. Thus, the total complexity of our proposed overlapping clustering strategy amounts to $\mathcal{O}\left(\sum_{k=1}^{|\mathcal{U}|} Q_k + \sum_{n=1}^{|\mathcal{A}|} K_n + \sum_{k=1}^{|\mathcal{U}|} \left(Q_k \sum_{n=1}^{|\mathcal{A}|} K_n \right) \right)$, which corresponds to a polynomial order. As a benchmark, we compare it to the optimal clustering solution based on the exhaustive search. Since it is a *many-to-many* matching problem, for a given AP q, there are $\sum_{\tau=0}^{|\mathcal{U}|} \binom{|\mathcal{U}|}{\tau}$ possible UE association combinations. For a total of $|\mathcal{A}|$ APs, the computational

complexity order is $\mathcal{O}\left(\left(\sum_{\tau=0}^{|\mathcal{U}|}\binom{|\mathcal{U}|}{\tau}\right)^{|\mathcal{A}|}\right)$. This shows that the optimal association problem scales exponentially with the number of APs, which becomes excessive in UD-VLC when the value of $|\mathcal{A}|$ is high. The proposed overlapped clustering is capable of spectacularly reducing the computation complexity.

III. Multiple Access in UD-VLC

For each constructed cluster, efficiently organizing the multi-user access is an intractable problem in the UD-VLC network. Both the conventional OMA and NOMA are vulnerable to ICI in the context of UDN, leading to an excessive UE OP.[2] To overcome this problem, we propose a hybrid NOMA and OMA scheme for the sake of eliminating the ICI and improving the OP to attain an increased sum-throughput.

A. Conventional NOMA Scheme

Consider a specific cluster $\mathcal{C}_n$, where AP n aims for supporting $|\mathcal{U}_n|$ UEs. We assume that the maximum UE-load of each AP is fixed to Λ. Based on the NOMA principle, $K_n = \min(|\mathcal{U}_n|, \Lambda)$ number of UEs are supported at different levels of power. Note that if we have $|\mathcal{U}_n| > \Lambda$, the AP randomly selects one of the UE combinations based on $\binom{|\mathcal{U}_n|}{\Lambda}$, leaving the rest of the UEs unsupported. Without loss of generality, the selected UEs are sorted[3] based on their channels as $|h_{1,n}|^2 \geq |h_{2,n}|^2 \geq \cdots \geq |h_{K_n,n}|^2$, where the channel gain $h_{k,n}$ between the kth UE and the AP n can be acquired according to Section II-A. Therefore, the NOMA-based information $\boldsymbol{x}_n$ transmitted to all K_n UEs using the superposition coding technique is expressed as:

$$\boldsymbol{x}_n = \sum_{k=1}^{K_n} \sqrt{P_{n,k}} s_{n,k}, \tag{5}$$

where $s_{n,k}$ conveys the desired signal of the kth UE in cluster $\mathcal{C}_n$ and its allocated power is $P_{n,k}$. At the receiver side, the observation at the kth UE is:

$$\begin{aligned} y_{n,k} &= \boldsymbol{x}_n h_{k,n} + w_{n,k} \\ &= \underbrace{\sqrt{P_{n,k}} h_{k,n} s_{n,k}}_{\text{desired signal of the } k\text{th UE}} + \underbrace{\sum_{\substack{q=1 \\ q\neq k}}^{K_n} \sqrt{P_{n,q}} h_{k,n} s_{n,q}}_{\text{intra-cell interference}} + w_{n,k}, \end{aligned} \tag{6}$$

where $w_{n,k}$ represents the sum of both the additive Gaussian white noise (AWGN) and ICI. In order to decode its own signal, the kth UE has to decode and subtract the signal of other UEs having a lower channel gain, based on SIC. By successfully removing part of the intra-cell interference $\sum_{q=k+1}^{K_n} \sqrt{P_{n,q}} h_{k,n} s_{n,q}$, the kth UE then decodes its desired signal. As a result, the residual intra-cell interference, namely $\sum_{q=1}^{k-1} \sqrt{P_{n,q}} h_{k,n} s_{n,q}$, is treated as noise. Therefore, the SINR of the kth UE becomes:

$$\gamma_{n,k} = \frac{|h_{k,n}|^2 P_{n,k}}{\sum_{q=1}^{k-1} |h_{k,n}|^2 P_{n,q} + \sigma^2 + \boldsymbol{I}_{n,k}}, \tag{7}$$

where we have $\sigma^2 = N_0\ B$, N_0 is the noise spectrum, and B is the available bandwidth. The ICI imposed on the kth UE is $\boldsymbol{I}_{n,k} = \sum_{f=1, f\neq n}^{|\mathcal{A}|} I_{n,k}^{[f]}$, which emanates from all other clusters, except for $\mathcal{C}_n$. Explicitly, if there is no association between a specific AP in cluster $\mathcal{C}_f$ and the kth UE in cluster $\mathcal{C}_n$, the ICI arriving from $\mathcal{C}_f$ is given by $I_{n,k}^{[f]} = |h_{k(n),f}|^2\ P_{\max}$, where $h_{k(n),f}$ is the channel gain between the kth UE in $\mathcal{C}_n$ and the AP in $\mathcal{C}_f$ and $P_{\max}$ represents the maximum transmit power of $\mathcal{C}_f$, considering the worst-case scenario. Otherwise, if the kth UE in $\mathcal{C}_n$ is also associated with the AP in $\mathcal{C}_f$, the ICI is given by $I_{n,k}^{[f]} = \frac{\xi_{f,k(n)}-1}{|\mathcal{U}_f|} |h_{k(n),f}|^2\ P_{\max}$, where $1 \leq \xi_{f,k(n)} \leq |\mathcal{U}_f|$ denotes the sorting order of the UE in $\mathcal{C}_f$. In our NOMA regime, only the residual intra-cell interference of the cluster $\mathcal{C}_f$ imposes ICI on the kth UE in $\mathcal{C}_n$, which implies that the overlapping nature of the proposed clustering mitigates the ICI.

B. Hybrid NOMA and OMA Scheme

Apart from the maximum UE-load limitation, the OP of the conventional NOMA scheme may be further degraded, when additional practical constraints are also taken into account. Furthermore, the ICI becomes more severe because of the dense deployment of APs, which further impairs the performance of conventional NOMA. As a remedy, we propose a hybrid NOMA and OMA scheme, which introduces both the FR technique and TDMA-based scheduling, combined with the conventional NOMA scheme. To elaborate, by employing the FR technique, the ICI can be readily controlled by relying on the most appropriate selection of the reuse factor ϱ, while the TDMA-based scheduling is capable of mitigating the UE's OP by re-considering the UEs, which were in outage during the previous time slots.

1) Graph-Coloring Based FR: Again, to cope with the ICI, the classic FR technique is invoked, where the entire spectrum is partitioned into ϱ frequency-resource blocks (FRBs) and allocated to each cluster so that the ICI can be eliminated by ensuring that adjacent APs do not share the same spectrum. Bandwidth partitioning inevitably imposes a sum-throughput reduction. Hence, there is a trade-off between the achievable throughput and the ICI-mitigation. To identify the best frequency reuse pattern, we propose a graph-coloring based FRB allocation strategy, followed by graph building, sorting and coloring stages.

We say that $z_{n,f} = 1$ represents that two clusters $\mathcal{C}_n$ and $\mathcal{C}_f$ are overlapped, if their associated UEs are partially or entirely duplicated, which is indicated by:

$$z_{n,f} = \begin{cases} 1, & \sum_{k\in\mathcal{U}} x_{n,k} x_{f,k} > 0; \\ 0, & \sum_{k\in\mathcal{U}} x_{n,k} x_{f,k} = 0. \end{cases} \tag{8}$$

[2]We define the UE's OP in our context as the probability of having unsupported UEs, which may due to various reasons, such as an inadequate individula rate.

[3]To avoid misunderstanding, UE k has the same meaning as the kth UE after sorting process.

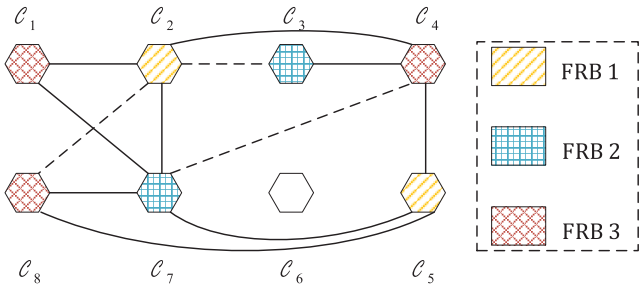


Fig. 3. Graph coloring based FRB allocation example using $\varrho = 3$, according to Fig. 2.

The index of $x_{n,k} = 1$ indicates that UE k is associated with $\mathcal{C}_n$, otherwise it is $x_{n,k} = 0$. Consider the association scenario of Fig. 2 as an example. A graph, $\mathcal{G}_c = (\mathcal{V}_c, \mathcal{E}_c)$, is constructed, where $\mathcal{V}_c$ is the vertex set containing all the clusters and the edge set $\mathcal{E}_c$ holding the relationship between any two vertices. An edge is constructed between any two vertices if $z_{n,f} = 1$ as indicated by the solid lines shown in Fig. 3. Additionally, when the ICI between the non-overlapping clusters is higher than the threshold I_{thr}, the edges are added as indicated by the dashed lines in Fig. 3.

All vertices are then sorted according to their degrees in descending order as: $\deg(v_7) \geq \deg(v_2) \geq \deg(v_4) \geq \deg(v_5) \geq \deg(v_8) \geq \deg(v_3) \geq \deg(v_1) \geq \deg(v_6)$, where the degree of a vertex is defined by the number of edges it connects with. In the ensuing coloring stage, the first ϱ vertices having the highest degree are colored first, and the remaining vertices are assigned a random color, which is different from that of their adjacent vertices. Note that if ϱ is not high enough, it may happen that two adjacent vertices are assigned the same color, yielding ICI.

2) TDMA-Based Scheduling: If the reuse factor is insufficiently high, ICI still exists, leading to excessive OP. The TDMA-based scheduling is therefore invoked within each cluster for reducing the UE's OP with the aid of time domain iterations. To be specific, it consists of two stages: round-robin scheduling and time-resource allocation, which is presented in **Algorithm 1**. To proceed, the sets $\mathcal{U}_n^{\mathbb{S}}$ and $\mathcal{U}_n^{\varnothing}$ are employed to hold the identity of the served UEs and of the unsupported UEs of cluster $\mathcal{C}_n$, respectively, where set $\mathcal{U}_n^{\mathbb{S}}$ is updated in every iteration. During the gth iteration, the UEs supported are saved in the active-UE group $\mathcal{U}_{n,g}^{\mathbb{S}}$.

To start with, the AP in cluster $\mathcal{C}_n$ intends to randomly support $K_{n,1} = \min(|\mathcal{U}_n|, \Lambda)$ number of UEs following the conventional NOMA principle. The successfully served UEs[4] are then held in $\mathcal{U}_{n,1}^{\mathbb{S}}$ as well as in $\mathcal{U}_n^{\mathbb{S}}$. Note that if we have $|\mathcal{U}_n| < \Lambda$ and $\mathcal{U}_{n,1}^{\mathbb{S}} = \mathcal{U}_n$, which means that the AP is able to serve all its associated UEs at the same time, then cluster $\mathcal{C}_n$ does not need the scheduling process. Otherwise, when we arrive at $\mathcal{U}_n^{\mathbb{S}} \neq \mathcal{U}_n$, those unsupported UEs are reconsidered in the following iterations. The set $\mathcal{U}_n^{\mathbb{S}}$ is expanded step-by-step, when more UEs can be supported, until we reach $\mathcal{U}_n^{\mathbb{S}} = \mathcal{U}_n$. If we end up with $\mathcal{U}_{n,g}^{\mathbb{S}} = \emptyset$, the scheduling process is suspended and the remaining unsupported UEs are classified as being in outage, since the AP cannot afford to support them.

[4]Note that the decision of whether a given UE is served or not will be discussed in Section IV, which implies that at each iteration, it may arrive at $|\mathcal{U}_{n,g}^{\mathbb{S}}| \leq K_{n,g}$.

Algorithm 1 TDMA-Based Scheduling

Input: The cluster $\mathcal{C}_n = \{\mathcal{A}_n, \mathcal{U}_n\}$ and its channel matrix H_n
Stage I: Scheduling
Initialisation: $\mathcal{U}_n^{\mathbb{S}} = \emptyset$, $g = 0$, $|\mathcal{U}_{n,g}^{\mathbb{S}}| = +\infty$.
while $\mathcal{U}_n^{\mathbb{S}} \neq \mathcal{U}_n$ & $|\mathcal{U}_{n,g}^{\mathbb{S}}| \neq 0$ **do**
  $g = g + 1$;
  AP n in cluster $\mathcal{C}_n$ attempts to support a number of $\begin{cases} \min(|\mathcal{U}_n|, \Lambda), & g = 1; \\ \min(|\mathcal{U}_n| - \sum_{e=1}^{g-1} |\mathcal{U}_{n,e}^{\mathbb{S}}|, \ \Lambda), & g > 1, \end{cases}$ UEs in NOMA fashion;
  The successfully served UEs at this iteration are held in set $\mathcal{U}_{n,g}^{\mathbb{S}}$;
  $\mathcal{U}_n^{\mathbb{S}} = \mathcal{U}_n^{\mathbb{S}} \cup \mathcal{U}_{n,g}^{\mathbb{S}}$;
end while
$\mathcal{U}_n^{\varnothing} = \mathcal{U}_n \setminus \mathcal{U}_n^{\mathbb{S}}$, flag $= g$;
Stage II: Time-Resource Allocation
if $|\mathcal{U}_{n,\text{flag}}^{\mathbb{S}}| \neq 0$ **then**
  for $g = 1 \rightarrow$ flag **do**
    $t_g = \frac{|\mathcal{U}_{n,g}^{\mathbb{S}}|}{|\mathcal{U}_n|} \times T$;
  end for
else
  for $g = 1 \rightarrow$ flag $- 1$ **do**
    $t_g = \left(\frac{|\mathcal{U}_{n,g}^{\mathbb{S}}|}{|\mathcal{U}_n|} + \frac{|\mathcal{U}_n^{\varnothing}| \times |\mathcal{U}_{n,g}^{\mathbb{S}}|}{|\mathcal{U}_n| \times |\mathcal{U}_n^{\mathbb{S}}|} \right) \times T$;
  end for
end if

For the second stage, we assume that the total time duration available for each cluster is T and that each UE is assigned an equal amount of time resource $T/|\mathcal{U}_n|$. After completing Stage I, if it arrives at $|\mathcal{U}_{n,\text{flag}}^{\mathbb{S}}| \neq 0$, which implies that all UEs in $\mathcal{U}_n$ can be supported, then the time resource t_g allocated to the gth group is proportional to the number of its served UEs, which is equal to $T|\mathcal{U}_{n,g}^{\mathbb{S}}|/|\mathcal{U}_n|$. Otherwise, the time resource of those unsupported UEs $T|\mathcal{U}_n^{\varnothing}|/|\mathcal{U}_n|$ is proportionally reallocated to each active UE group according to $|\mathcal{U}_{n,g}^{\mathbb{S}}|$. To this end, each scheduled UE in cluster $\mathcal{C}_n$ can find its group with respect to the transmission time.

Remark 1: The computational complexity of **Algorithm 1** is jointly determined by the number of iterations and by the complexity of each iteration. Upon considering the worst-case scenario, we have $|\mathcal{U}_{n,g}^{\mathbb{S}}| = 1, \ g = 1, \cdots, |\mathcal{U}_n|$. Therefore, the complexity of the algorithm can be upper-bounded by $\mathcal{O}\left(\sum_{g=1}^{|\mathcal{U}_n|} \zeta_g + |\mathcal{U}_n|\right)$, where ζ_g denotes the complexity of the resource allocation strategy, as we will discuss in Section IV.

IV. Dynamic Resource Allocation Methodology

During the sum-throughput maximization, the decision of whether to support a given UE or not associated with our efficient resource allocation plays a significant role in our UD-VLC network. Explicitly, we focus our attention on the modulation-mode assignment for each UE supported, as well as on the power allocation for each active AP. Since the high-flexibility LACO-OFDM scheme is employed, this allows

the UE to select different numbers of active layers and also allows each active layer to be assigned various modulation-modes. However, this inevitably increases the complexity. Note that these decisions depend not only on the UE's own received signal strength, but also on that of the other UEs in the light of the MA mechanism invoked, hence leading to a complex multi-user-dependent and multi-layer-dependent maximization problem. Furthermore, in practical UD-VLC networks, the APs have a limited transmit power and limited backhaul rate, which may be the bottleneck of the system's performance. Additionally, it is important to guarantee a minimum rate for each UE in order to offer some degrees of fairness. Hence, in this section, this challenging throughput maximization problem is formulated under practical considerations and is then solved by our proposed dynamic resource allocation strategy.

A. Problem Formulation

The available modulation-modes are stored in set $\mathcal{M}$, where mode m is a square QAM with the constellation size of 2^{2m}. To fulfill the target BER, we also assume each active layer has to meet the same target BER[5] and thus the relationship between the BER and the SINR $\gamma_{n,k,l}^{[m]}$ required for the lth layer of UE k in $\mathcal{C}_n$ employing modulation-mode m is indicated as [35]:

$$\text{BER} = \frac{2^m - 1}{m2^m}\text{erfc}\left(\sqrt{\frac{3\gamma_{n,k,l}^{[m]}}{2(2^{2m}-1)}}\right). \tag{9}$$

Recalling (7), the transmit power required by layer l for the modulation-mode m of UE k is:

$$p_{n,k,l}^{[m]} = \frac{\gamma_{n,k,l}^{[m]}\beta_l\Big(\sum\limits_{j=1}^{k-1} P_{n,j}|h_{k,n}|^2 + \sigma^2 + \boldsymbol{I}_{n,k}\Big)}{|h_{k,n}|^2}, \tag{10}$$

where $\beta_l = N_l/N$ is the ratio of the number of active subcarriers in the lth layer to the total number of subcarriers. Based on [15], the power of UE k having L layers is calculated as:

$$\begin{aligned} P_{n,k} &= \mathbb{E}\big[\boldsymbol{S}_k^2\big] = \mathbb{D}\big[\boldsymbol{S}_k\big] + \mathbb{E}\big[\boldsymbol{S}_k\big]^2 \\ &= \sum_{l=1}^{L}\mathbb{D}\big[\boldsymbol{S}_{k,l}\big] + \left(\sum_{l=1}^{L}\mathbb{E}\big[\boldsymbol{S}_{k,l}\big]\right)^2 \\ &= \frac{\pi-1}{\pi}\sum_{l=1}^{L} p_{n,k,l}^{[m_{k,l}]} + \frac{1}{\pi}\left(\sum_{l=1}^{L}\sqrt{p_{n,k,l}^{[m_{k,l}]}}\right)^2. \end{aligned} \tag{11}$$

Equations (10) and (11) imply that the power required by UE k is related not only to the modulation-mode assignment and to the power allocation of each layer itself, but also affected by the residual intra-cell interference in the same cluster and the ICI arriving from other clusters. When further considering the practical constraints, the interdependence between the modulation-mode assignment and the power allocated to each layer of each UE becomes prohibitively complicated. Mathematically, it is formulated for the cluster n as:

$$\max_{\mathcal{M}_n^*,\mathcal{P}_n^*} \sum_{k=1}^{|\mathcal{U}_n|}\sum_{l=1}^{L}\sum_{m=1}^{|\mathcal{M}|} r_{k,l}^{[m]} y_{k,l}^{[m]} \tag{12a}$$

$$\text{s.t.} \sum_{m=1}^{|\mathcal{M}|} y_{k,l}^{[m]} = \{0,1\}, \quad y_{k,l}^{[m]} \in \{0,1\}, \forall k, \forall l, \forall m; \tag{12b}$$

$$\sum_{l=1}^{L}\sum_{m=1}^{|\mathcal{M}|} r_{k,l}^{[m]} y_{k,l}^{[m]} \geq R_{k,\min}, \quad \forall k; \tag{12c}$$

$$\sum_{k=1}^{|\mathcal{U}_n|}\sum_{l=1}^{L}\sum_{m=1}^{|\mathcal{M}|} r_{k,l}^{[m]} y_{k,l}^{[m]} \leq R_{\max}; \tag{12d}$$

$$\sum_{k=1}^{|\mathcal{U}_n|} P_{n,k} \leq P_{\max}. \tag{12e}$$

Notation $r_{k,l}^{[m]}$ represents the lth layer's achievable throughput for UE k using modulation-mode m. The binary indicator $y_{k,l}^{[m]}$ is employed for determining the modulation-mode assignment:

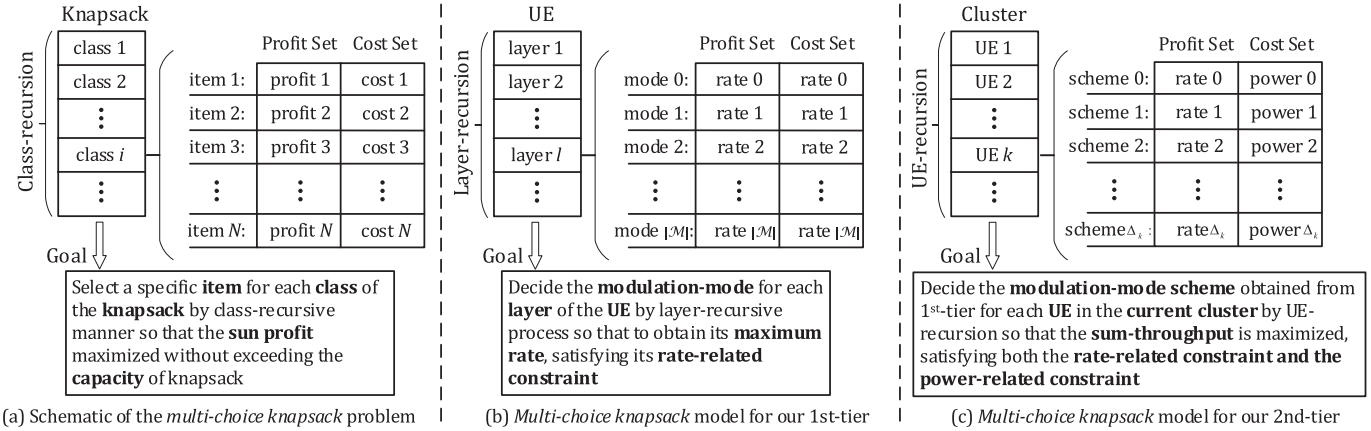
$$y_{k,l}^{[m]} = \begin{cases} 1, & \text{mode } m \text{ is assigned to the } l\text{th layer of UE } k, \\ 0, & \text{mode } m \text{ is not assigned to the } l\text{th layer of UE } k. \end{cases} \tag{13}$$

Note that if we arrive at $y_{k,l}^{[m]} = 0, \forall m$, the lth layer is not activated for UE k. Given $y_{k,l}^{[m]} = 0, \forall m, \forall l$, the UE k is not supported during the current time slot. The feasible modulation-mode assignment and power allocation obtained for every layer of every UE in cluster $\mathcal{C}_n$ is held in the sets $\mathcal{M}_n^*$ and $\mathcal{P}_n^*$, respectively.

To elaborate, constraint (12b) indicates that the UEs are able to adaptively select the layers for their transmission. For the sake of power efficiency, it is not necessary to ensure that the lth layer is utilized only if the $(l-1)$st layer is activated. Constraint (12c) indicates the minimum transmission rate $R_{k,\min}$ required by UE k, while (12d) expresses that the total affordable transmission rate should not exceed the backhaul allowance of $R_{\max}$. The transmit power constraint is presented in (12e), where the total power required should not exceed the maximum transmit power $P_{\max}$.

It can be concluded that the problem in (12) is an interdependent mixed binary integer non-convex problem. It involves both binary modulation-mode-related variables $y_{k,l}^{[m]}$ and continuous power-allocation-related variables $p_{n,k,l}^{[m]}$. In general, the problem is hard to tackle by popular tools. For the sake of tractability, we transfer the original problem to a 2-tier connected maximization problem, where the 1st-tier focuses on the specific modulation-mode assignment for each layer of a given UE, while the 2nd-tier deals with the power allocation for each considered UE. To be more specific, the 1st-tier takes into account the constraints (12c) and (12d), which

[5]In this contribution, the impairment of the ILI in LACO-OFDM is negligible. According to [15], the BER curves of different layer in LACO-OFDM converge to that of the first layer, when BER is below 10^{-4}, which implies that the ILI imposes no more impairment on the performance of LACO-OFDM.



(a) Schematic of the *multi-choice knapsack* problem (b) *Multi-choice knapsack* model for our 1st-tier (c) *Multi-choice knapsack* model for our 2nd-tier

Fig. 4. The schematic diagram of *multi-choice knapsack* problem and its adaptive model for our two tiers maximization problem. Note that, in sub-figure (b), the mode 0 and its associated profit and cost are virtual, which means that the layer will not be assigned a specific modulation-mode if mode 0 is assigned to it, so that no profit contribution and no cost is associated with this layer. The same as the scheme 0 in sub-figure (c).

aims for finding all the potential rate-satisfied modulation-mode assignments for each layer of each UE, while ignoring the power constraint. Feasible solutions of the 1st-tier will then be further considered by the 2nd-tier, which aims for finding the power allocation solutions for multiple UEs in a specific clusters. To this end, the sum-throughput is maximized under the feasible modulation-mode assignment and power allocation.

B. 1st-Tier Single-UE Modulation-Mode Assignment

The main task of the 1st-tier is modeled by a *multi-choice knapsack* problem, as shown in Fig. 4(b). Specifically, by considering the modulation-mode assignment of each layer of a given UE, the rate-requirement can be readily satisfied. Due to the NP-hard nature of the knapsack problem, we invoke dynamic programming for solving it in a layer-recursive manner, which implies that the modulation-mode decision of the lth layer is based on the modulation-mode assigned to the $(l-1)$st layer [36].

1) Sub-Problem Formulation: To explore the modulation-mode assignment, the problem of guaranteeing the transmission-quality of UE k is formulated as:

$$\max_{\mathcal{M}_{n,k}} \sum_{l=1}^{L}\sum_{m=1}^{|\mathcal{M}|} r_{k,l}^{[m]} y_{k,l}^{[m]} \tag{14a}$$

$$\text{s.t.} \sum_{m=1}^{|\mathcal{M}|} y_{k,l}^{[m]} \in \{0,1\}, \quad y_{k,l}^{[m]} \in \{0,1\},\ \forall l,\ \forall m; \tag{14b}$$

$$\sum_{l=1}^{L}\sum_{m=1}^{|\mathcal{M}|} r_{k,l}^{[m]} y_{k,l}^{[m]} \geq R_{k,\min}; \tag{14c}$$

$$\sum_{l=1}^{L}\sum_{m=1}^{|\mathcal{M}|} r_{k,l}^{[m]} y_{k,l}^{[m]} \leq R_{\max}. \tag{14d}$$

It can be found that the 1st-tier is focussed on each individual UE by exploring its potential modulation-mode assignment for each layer.

2) Methodology: To solve this problem, we commence by discretizing the maximum throughput allowance $R_{\max}$ into J levels, while the maximum rate constraint is then assigned to one of the levels, given by $\left\{R_j = jR_{\max}/J:\ j=1,\cdots,J\right\}$. Accordingly, the task of the lth layer-recursion is to obtain the maximized sum-throughput of UE k by searching for the best modulation-mode combinations commencing from layer 1 to l, under all the J rate-constraint levels, as shown in Fig. 5.

Considering a specific rate constraint R_j during the lth layer-recursion, the maximum achievable sum-throughput $\mathrm{T}_{k,l}(R_j)$ of UE k is obtained with the aid of the results of the $(l-1)$st layer-recursion, which is:

$$\mathrm{T}_{k,l}(R_j) = \max_{m} \left\{\mathrm{T}_{k,l-1}\left(\left\lfloor R_j - r_{k,l}^{[m]}\right\rfloor\right) + r_{k,l}^{[m]}\right\} \quad \text{s.t. } R_j - r_{k,l}^{[m]} \geq 0, \quad \forall m \in \mathcal{M}, \tag{15}$$

where we invoke the floor-operator $\lfloor \bullet \rfloor$ to obtain a rounded value corresponding to the nearest lower level of $\left(R_j - r_{k,l}^{[m]}\right)$, since the value of $\left(R_j - r_{k,l}^{[m]}\right)$ is not equal to any of the rate constraint levels. After considering all available modulation-modes for the lth layer, if we still cannot satisfy the condition in (15), we conclude that $T_{k,l}(R_j) = 0$, since no solution was found for UE k to afford a total of l layers under the constraint R_j. Otherwise, the legitimate modulation-mode of the lth layer of UE k under the rate constraint R_j is stored and denoted by $m_{k,l}^{[j]}$. Therefore, the modulation-mode assignment $\boldsymbol{m}_{k,l}^{1\text{st},[j]}$ for UE k having a total of l layers under rate constraint R_j can be collected by combining $m_{k,l}^{[j]}$ and its counterparts in the previously considered $(l-1)$ layers, which are obtained from one of the combinations during the $(l-1)$st layer-recursion. We assume that the modulation-mode combinations of the previous $(l-1)$ layers have already been held in the matrix $\boldsymbol{M}_{k,l-1}^{1\text{st}} \in \mathbb{R}^{[(l-1)\times J]}$ formulated as:

$$\boldsymbol{M}_{k,l-1}^{1\text{st}} = \left[\boldsymbol{m}_{k,l-1}^{1\text{st},[1]};\ \boldsymbol{m}_{k,l-1}^{1\text{st},[2]};\ \cdots;\ \boldsymbol{m}_{k,l-1}^{1\text{st},[J]}\right], \tag{16}$$

By solving (15), the modulation-mode decision of the lth layer $m_{k,l}^{[j]}$ associated with the decisions concerning the previous

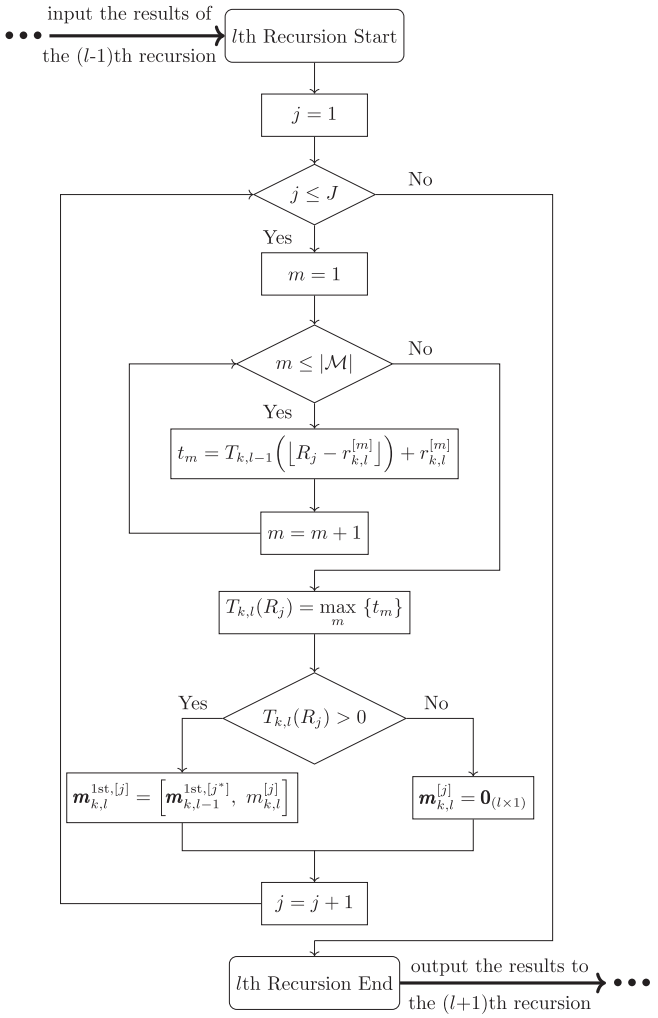


Fig. 5. The flow diagram of the dynamic-programming-based lth layer-recursion.

$(l-1)$ layers $\boldsymbol{m}_{k,l-1}^{1\text{st},[j^*]}$ has been obtained. The two parts together meet the condition indicated in (15). The modulation-mode assignment $\boldsymbol{m}_{k,l}^{1\text{st},[j]}$ of all l layers under the constraint R_j is therefore expressed as:

$$\begin{aligned}\boldsymbol{m}_{k,l}^{1\text{st},[j]} = &\left[\boldsymbol{m}_{k,l-1}^{1\text{st},[j^*]},\ m_{k,l}^{[j]}\right]\\ &\text{s.t. } T_{k,l-1}(R_{j^*}) \ge T_{k,l-1}(R_{j'}),\\ &\quad \forall j',\quad j^* \in \{1,\cdots,j\};\ r_{k,l}^{m_{k,l}^{[j]}}\\ &\quad + T_{k,l-1}(R_{j^*}) \le R_j. \end{aligned} \tag{17}$$

If we arrive at $\mathrm{T}_{k,l}(R_j) = 0$, the modulation-mode vector $\boldsymbol{m}_{k,l}^{1\text{st},[j]}$ is assigned an all-zero vector $\mathbf{0}_{(l\times 1)}$. The matrix $\boldsymbol{M}_{k,l}^{1\text{st}}$ can be fully constructed during the lth layer-recursion upon increasing the rate constraint level for approaching R_J. Note that the layer-recursion based sum-throughput maximization as well as of the modulation-mode assignment for UE k is summarized in Fig. 5.

3) Results: When all L layers of UE k have been considered by obeying the aforementioned operations, the candidate sum-throughput set, $\mathcal{T}_{n,k}$, for UE k as well as the corresponding modulation-mode assignment set $\mathcal{M}_{n,k}$ of the 1st-tier problem is obtained as:

$$\begin{aligned}\mathcal{T}_{n,k} &= \left\{\mathrm{T}_{k,l}(R_j):\ j=1,\cdots,J,\ l=1,\cdots,L,\ k\in\mathcal{U}_n\right\};\\ \mathcal{M}_{n,k} &= \left\{\boldsymbol{M}_{k,l}^{1\text{st}}:\ l=1,\cdots,L,\ k\in\mathcal{U}_n\right\}.\end{aligned} \tag{18}$$

Due to the minimum rate $R_{k,\min}$ required by UE k as shown in (14c), the items in $\mathcal{T}_{n,k}$ which are lower than $R_{k,\min}$ have to be removed to create the corresponding modulation-modes in the set $\mathcal{M}_{n,k}$. The number of remaining throughput items satisfying the rate constraint for UE k is given by Δ_k, where $\Delta_k \le (J \times L)$. Therefore, the set $\mathcal{T}_{n,k}$ is reassigned to $\mathcal{T}_{n,k}^{\mathrm{r}}$ holding the remaining throughput solutions of UE k in the 1st-tier as: $\mathcal{T}_{n,k}^{\mathrm{r}} = \left\{\mathrm{T}_{n,k,1}^{\mathrm{r}}, \mathrm{T}_{n,k,2}^{\mathrm{r}}, \cdots, \mathrm{T}_{n,k,\Delta_k}^{\mathrm{r}}\right\}$, along with the updated modulation-mode scheme set $\mathcal{M}_{n,k}^{\mathrm{r}} = \left\{\mathrm{m}_{n,k,1}^{\mathrm{r}}, \mathrm{m}_{n,k,2}^{\mathrm{r}}, \cdots, \mathrm{m}_{n,k,\Delta_k}^{\mathrm{r}}\right\}$. Note that the variables $\mathrm{T}_{n,k,\delta}^{\mathrm{r}}$ and $\mathrm{m}_{n,k,\delta}^{\mathrm{r}}$ represent the δth candidate throughput and modulation-mode arrangement for UE k, respectively. Hence the modulation-mode assignment for UE k has been successfully explored. The same procedure is applied to all UEs and then the results acquired are delivered to the 2nd-tier.

C. 2nd-Tier Multi-UEs Power Allocation

As indicated in Fig. 4(c), the 2nd-tier can also be modeled as a *multi-choice knapsack* problem, where a total of Δ_k schemes are obtained for UE k from the 1st-tier. Again, the dynamic-programming based allocation strategy is invoked in a recursive UE-specific manner, where the resource allocation of the kth UE is based on the resources allocated to the $(k-1)$st UE, working similar to the layer-recursion in Fig. 5.

1) Sub-Problem Formulation: For simplifying our notation, we assume that in cluster $\mathcal{C}_n$, a total of $|\mathcal{U}_n| < \Lambda$ UEs are associated with AP n. In the proposed hybrid NOMA and OMA scheme, if we have $\mathcal{U}_n^{\textcircled{S}} \neq \mathcal{U}_n$ and $|\mathcal{U}_{n,g}^{\textcircled{S}}| \neq 0$, the corresponding UEs will be re-scheduled in later iterations, as indicated by **Algorithm 1**, whilst using the same maximization methodology. Both the conventional NOMA and the proposed schemes are discussed jointly in the following, where we focus our attention on the first iteration of the proposed scheme. Armed with the results of the 1st-tier, the maximization problem of the 2nd-tier for cluster $\mathcal{C}_n$ is formulated as:

$$\max_{\mathcal{U}_n^{\textcircled{S}},\mathcal{M}_n,\mathcal{P}_n} \sum_{k=1}^{|\mathcal{U}_n|}\sum_{\delta=1}^{\Delta_k} \mathrm{T}_{k,\delta}^{\mathrm{r}} v_{n,k}^{[\delta]} \tag{19a}$$

$$\text{s.t. } \sum_{\delta=1}^{\Delta_k} v_{n,k}^{[\delta]} \in \{0,1\}:\ v_{n,k}^{[\delta]} \in \{0,1\},\quad \forall\delta,\ \forall k; \tag{19b}$$

$$\sum_{k=1}^{|\mathcal{U}_n|}\Big(\sum_{\delta=1}^{\Delta_k} v_{n,k}^{[\delta]}\Big) P_{n,k} \le P_{\max}; \tag{19c}$$

$$\sum_{k=1}^{|\mathcal{U}_n|}\sum_{\delta=1}^{\Delta_k} \mathrm{T}_{k,\delta}^{\mathrm{r}} v_{n,k}^{[\delta]} \le R_{\max}, \tag{19d}$$

where the binary indicator $v_{n,k}^{[\delta]} = 1$ if UE k is supported using the modulation-mode combination $\mathrm{m}_{n,k,\delta}^{\mathrm{r}}$ in the set $\mathcal{M}_{n,k}^{\mathrm{r}}$, otherwise, it is zero. For notational convenience, $P_{n,k}^{[\delta_k]}$

is utilized to represent $\left(\sum_{\delta=1}^{\Delta_k} v_{n,k}^{[\delta]}\right) P_{n,k}$, which quantifies the transmit power required for UE k, when the modulation-mode combination $\mathrm{m}^{\mathrm{r}}_{n,k,\delta}$ is applied. Note that if we have $\sum_{\delta=1}^{\Delta_k} v_{n,k}^{[\delta]} = 0$, none of the modulation-mode combinations is applied to UE k, hence UE k is in outage upon using the conventional NOMA transmission scheme, but it is re-considered in the hybrid NOMA and OMA transmission. As a result, the actually served UEs associated with their most appropriate modulation-mode assignment as well as the allocated power satisfying constraints (19c) and (19d) are held in the sets $\mathcal{U}_n^{(\mathbb{S})}$, $\mathcal{M}_n$ and $\mathcal{P}_n$.

2) Methodology: During the kth UE-recursion, all the Δ_k modulation-modes of UE k are considered for determining whether UE k can or cannot be adequately served, according to the result of the $(k-1)$st UE-recursion. To proceed, we discretize the maximum transmit power constraint $P_{\max}$ in (19c) into I levels, so that the resultant maximum transmit power allowance can be expressed as $\{p_i = iP_{\max}/I : i = 1, 2. \cdots, I\}$. Thus, the objective of the kth UE-recursion is to explore the maximized sum-throughput of a total of k UEs relying on our modulation-mode assignment and power allocation for each of them under all the I power constraint levels. According to (10) and (11), the required transmit power of UE k can be only determined if the power of the preceeding UEs having a higher channel gain than UE k, have already been found. Therefore, to ensure that the required sum-power of a total of k UEs does not exceed a specific power constraint p_i during the kth UE-recursion, the sum-throughput $\mathrm{S}_{n,k}(p_i)$ of all k UEs is attained based on the result of the $(k-1)$st recursion as:

$$\begin{aligned}\mathrm{S}_{n,k}(p_i) = \max_{\delta,j} \ & \left\{\mathrm{S}_{n,k-1}(p_j) + \mathrm{T}^{\mathrm{r}}_{n,k,\delta}\right\} \\ \text{s.t.} \ & \sum \boldsymbol{P}^{[j]}_{n,k-1} + P^{[\delta_k]}_{n,k} \ |^{j}_{k-1} \leq p_i, \quad j \in \{1, \cdots, i\}, \\ & \delta \in \{1, \cdots, \Delta_k\}, \end{aligned} \tag{20}$$

where the vector $\boldsymbol{P}^{[j]}_{n,k-1} \in \mathbb{R}^{[(k-1)\times 1]}$ holds the power required for a total of $(k-1)$ UEs under the maximum transmit power constraint p_j, which is one of the results of the $(k-1)$st UE-recursion. Furthermore, $P^{[\delta_k]}_{n,k} \ |^{j}_{k-1}$ denotes the required transmit power of UE k, whose modulation-mode assignment is decided by the $\mathrm{m}^{\mathrm{r}}_{n,k,\delta}$, while the intra-cell interference imposed is $\boldsymbol{P}^{[j]}_{n,k-1}$.

Note that after attempting every possible combinations for UE k, if there is still no feasible solution for AP n to afford all the k UEs under the power constraint p_i satisfying (20), then we have $\mathrm{S}_{n,k}(p_i) = 0$. Otherwise, the maximized sum-throughput is acquired as $\mathrm{S}_{n,k}(p_i) = \mathrm{S}_{n,k-1}(p_{i^*}) + \mathrm{T}^{\mathrm{r}}_{n,k,\delta_k^{(i)}}$. Correspondingly, to explore the modulation-mode assignment and power allocation, we assume that the modulation-mode assignment result of the $(k-1)$st UE-recursion is already stored in $\boldsymbol{M}^{\text{2nd}}_{n,k-1}$ as:

$$\boldsymbol{M}^{\text{2nd}}_{n,k-1} = \left[\boldsymbol{m}^{\text{2nd},[1]}_{n,k-1};\ \boldsymbol{m}^{\text{2nd},[2]}_{n,k-1};\ \cdots;\ \boldsymbol{m}^{\text{2nd},[I]}_{n,k-1}\right], \tag{21}$$

where we have

$$\boldsymbol{m}^{\text{2nd},[i]}_{n,k-1} = \left[\mathrm{m}^{\mathrm{r}}_{n,1,\delta_1^{(i)}},\ \mathrm{m}^{\mathrm{r}}_{n,2,\delta_2^{(i)}},\ \cdots\ \mathrm{m}^{\mathrm{r}}_{n,k-1,\delta_{k-1}^{(i)}}\right]. \tag{22}$$

The candidate combinations in $\boldsymbol{M}^{\text{2nd}}_{n,k-1}$ containing the modulation-mode assignment for each layer of each of the $(k-1)$ UEs considered satisfy both the individual rate and power constraint. Accordingly, the power allocation result of the $(k-1)$st UE recursion is contained by the matrix $\boldsymbol{P}_{n,k-1}$ formulated as:

$$\boldsymbol{P}_{n,k-1} = \left[\boldsymbol{P}^{[1]}_{n,k-1};\ \boldsymbol{P}^{[2]}_{n,k-1};\ \cdots;\ \boldsymbol{P}^{[I]}_{n,k-1}\right]. \tag{23}$$

For the kth UE-recursion carried out under a specific power constraint p_i, if the legitimate modulation-mode combination $\mathrm{m}^{\mathrm{r}}_{n,k,\delta_k^{(i)}}$ is obtained by solving (20), the modulation-mode assignment for all k UEs under the constraint p_i is formulated as:

$$\begin{aligned}\boldsymbol{m}^{\text{2nd},[i]}_{n,k} = & \left[\boldsymbol{m}^{\text{2nd},[i^*]}_{n,k-1},\ \mathrm{m}^{\mathrm{r}}_{n,k,\delta_k^{(i)}}\right] \\ & \text{s.t.}\ \mathrm{S}_{n,k-1}(p_{i^*}) \geq \mathrm{S}_{n,k-1}(p_{i'}), \\ & \quad \forall i^*, \quad i' \in \{1, 2, \cdots, i\}; \\ & \quad \sum \boldsymbol{P}^{i^*}_{n,k-1} + P^{[\delta_k^{(i)}]}_{n,k} \ |^{i^*}_{k-1} \leq p_i. \end{aligned} \tag{24}$$

After determining $\mathrm{m}^{\mathrm{r}}_{n,k,\delta_k^{(i)}}$ for UE k during the kth recursion under the power limitation p_i, the required transmit power of the k UEs is collected as:

$$\boldsymbol{P}^{[i]}_{n,k} = \left[\boldsymbol{P}^{[i^*]}_{n,k-1},\ P^{[\delta_k^{(i)}]}_{n,k} \ |^{i^*}_{k-1}\right]. \tag{25}$$

In the same manner, the kth recursion is completed, when i approaches I.

3) Solutions: After considering all $|\mathcal{U}_n|$ UEs in the cluster $\mathcal{C}_n$ following the aforementioned procedure, we obtain the candidate sum-throughput set $\mathcal{S}_n$ associated with the modulation-mode assignment combination set $\mathcal{M}_n$ and the power allocation strategy set $\mathcal{P}_n$ for cluster $\mathcal{C}_n$ as:

$$\mathcal{S}_n = \left\{\mathrm{S}_{n,k}(p_i) : \quad i = 1, \cdots, I,\ k \in \mathcal{U}_n\right\}; \tag{26a}$$

$$\mathcal{M}_n = \left\{\boldsymbol{M}^{\text{2nd}}_{n,k} :\ k \in \mathcal{U}_n\right\}; \tag{26b}$$

$$\mathcal{P}_n = \left\{\boldsymbol{P}_{n,k} :\ k \in \mathcal{U}_n\right\}. \tag{26c}$$

As a reminder, the specific values in $\mathcal{S}_n$ exceeding the backhaul-rate constraint $R_{\max}$ in (19d) have to be removed. The resultant sum-throughput values are held in the set $\mathcal{S}_n^{\mathrm{r\&p}} = \left\{\mathrm{S}^{\mathrm{r\&p}}_{n,1}, \cdots, \mathrm{S}^{\mathrm{r\&p}}_{n,\chi_n}\right\}$. Accordingly, the items failing to satisfy the backhaul-rate are also removed from the set $\mathcal{M}_n$ and $\mathcal{P}_n$, respectively, ultimately resulting in the feasible modulation-mode assignment set $\mathcal{M}_n^*$ and the transmit power allocation set $\mathcal{P}_n^*$. Therefore, the solutions we obtained satisfy all the constraints of the original problem. Under the principle of sum-throughput maximization, the solution $\mathrm{S}^{\mathrm{r\&p}}_{n,\epsilon^*}$ is obtained from the set $\mathcal{S}_n^{\mathrm{r\&p}}$ associated with the ϵ^*th modulation-mode assignment obtained from $\mathcal{M}_n^*$ and the ϵ^*th power allocation strategy suggested by $\mathcal{P}_n^*$. It is worth mentioning that the sum-throughout $\mathcal{S}^{\mathrm{r\&p}}_{n,\epsilon^*}$ is contributed to by $|\mathcal{U}_n^{(\mathbb{S})}|$ number of UEs, which may be less than $|\mathcal{U}_n|$. Those currently unsupported UEs will be re-considered according to **Algorithm 1**. Note that after completing all the maximization processes for the current cluster, the sum-throughput values obtained for each

TABLE I
LIST OF PARAMETERS

I-1. Environment-related Parameters	
room size	$15\times15\times3$ m^3
AP height	2.5 m
number of APs	8×8
LED array per AP	60×60
UE height	0.85 m
I-2. VLC Channel-related Parameters	
semi-angle at half-illumination $\phi_{1/2}$	60^o
gain of optical filter $g_{of}(\psi)$	1
gain of optical concentrator $g_{oc}(\psi)$	1
physical area for a PD receiver S	1 cm^2
reflection efficiency ρ	0.75
I-3. Physical-layer-related Parameters	
target BER	10^{-5}
modulation bandwidth B	50 MHz
AWGN power spectral density N_0	10^{-22} A^2/Hz
Maximum number of layers in LACO-OFDM L	4
I-4. Network-construction-related Parameters	
channel quality threshold H_{thr}	10^{-6}
transmission distance threshold D_{thr}	5 m
ICI threshold I_{thr}	10^{-7}
I-5. Resource-allocation-related Parameters	
transmit power constraint $P_{\max}$	10 W/AP
individual minimum rate requirement $R_{k,\min}$	0-150 Mbits/s
backhaul constraint $R_{\max}$	1 Gbits/s
number of rate limit levels J	100
number of power limit levels I	20

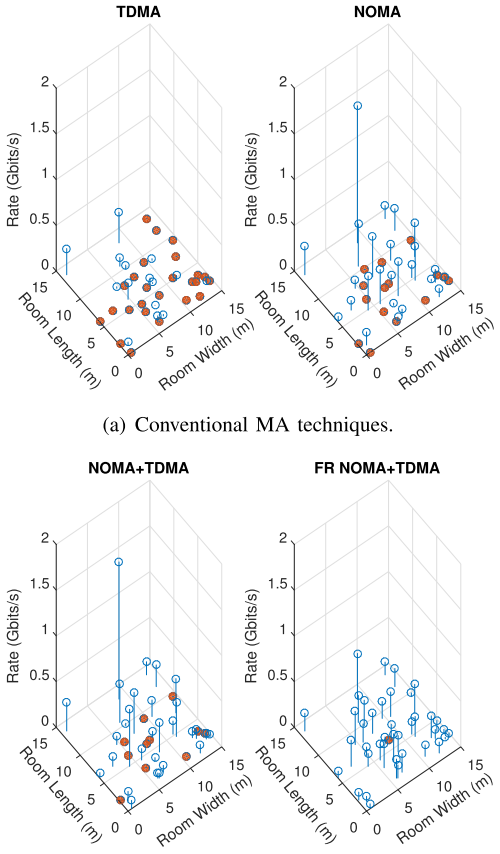
group are further combined with their associated time duration t_g, obtained based on **Algorithm 1**, to realize the achievable throughput within T duration.

D. Complexity Analysis

The optimal resource allocation can be found by exhaustive search, which has an excessive complexity on the order of $\mathcal{O}\left((|\mathcal{M}|+1)^{L|\mathcal{U}|}\right)$. It is hence unrealistic to implement in the UD-VLC due to having a large $|\mathcal{U}|$. By contrast, the complexity order of the modulation-mode assignment during the 1st-tier of our proposed strategy is $\mathcal{O}(JL|\mathcal{M}|)$ for each UE, while the complexity order of the power allocation in the 2nd-tier is $\mathcal{O}\left(I\max_{k\in\mathcal{U}_n}\Delta_k\right)$. Therefore, the computational complexity of the proposed algorithm is significantly reduced to a polynomially increasing order compared to the exponentially increasing order of the optimal exhaustive search.

V. PERFORMANCE EVALUATION

In this section, we characterize our hybrid NOMA and OMA schemes in comparison to the conventional TDMA and NOMA arrangements. The modulation-mode-related sum-throughput is maximized with the aid of the proposed dynamic resource allocation strategy, where each UE can dynamically select upto four layers for LACO-OFDM. The modulation-modes available in the set $\mathcal{M}$ for each layer are BPSK, 4-QAM, 16-QAM, 64-QAM and 256-QAM. Our results are averaged over hundreds independent UE distributions. The rest of the simulation-related parameters are summarized in Table I.



(a) Conventional MA techniques.

(b) Hybrid NOMA and OMA techniques.

Fig. 6. Achievable rate distribution of a total of $|\mathcal{U}| = 40$ UEs randomly located in the indoor environment supported by our UD-VLC network using four different MA schemes under FoV $= 100^\circ$, where the UEs in outage are marked with solid circles.

Let us first consider the achievable rate of multiple UEs in a single distribution. Figure 6 shows the modulation-mode-related rates both in the context of conventional TDMA and of NOMA as well as of the proposed hybrid NOMA and OMA scheme. Note that the left sub-figure of Fig. 6(b) has a reuse factor of $\varrho = 1$, while the right sub-figure has $\varrho = 2$. It can be found from Fig. 6 that the conventional TDMA scheme has the lowest performance in this ultra-dense scenario. The conventional NOMA scheme provides much higher rate for its UEs, as a benefit of its improved spectral and time efficiency, albeit at a cost of an increased UE OP (displayed by solid circles), especially for the corner/neighboring-UEs. The proposed hybrid NOMA and OMA scheme with $\varrho = 1$ is capable of increasing the number of UEs supported, although at a slightly reduced rate. When the FR is adopted, the number of UEs supported is significantly increased in conjunction with a more fair rate distribution. Let us now investigate the performance of the proposed scheme at various reuse factors, FoVs, UE densities and maximum loads.

A. Effects of Reuse Factor

The value of the reuse factor ϱ has a profound impact on both the sum-throughput as well as on ICI elimination.

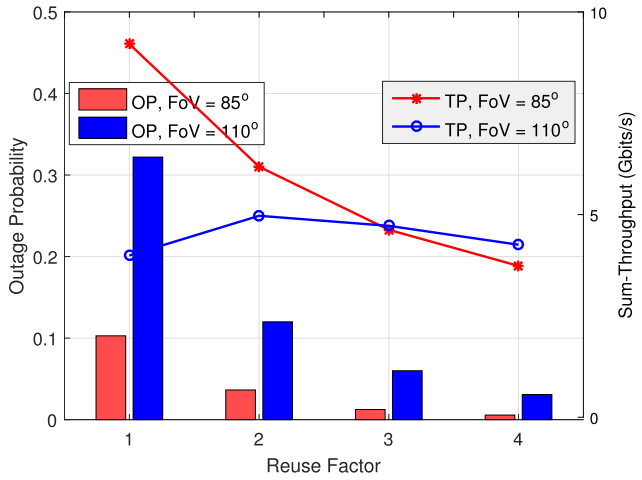
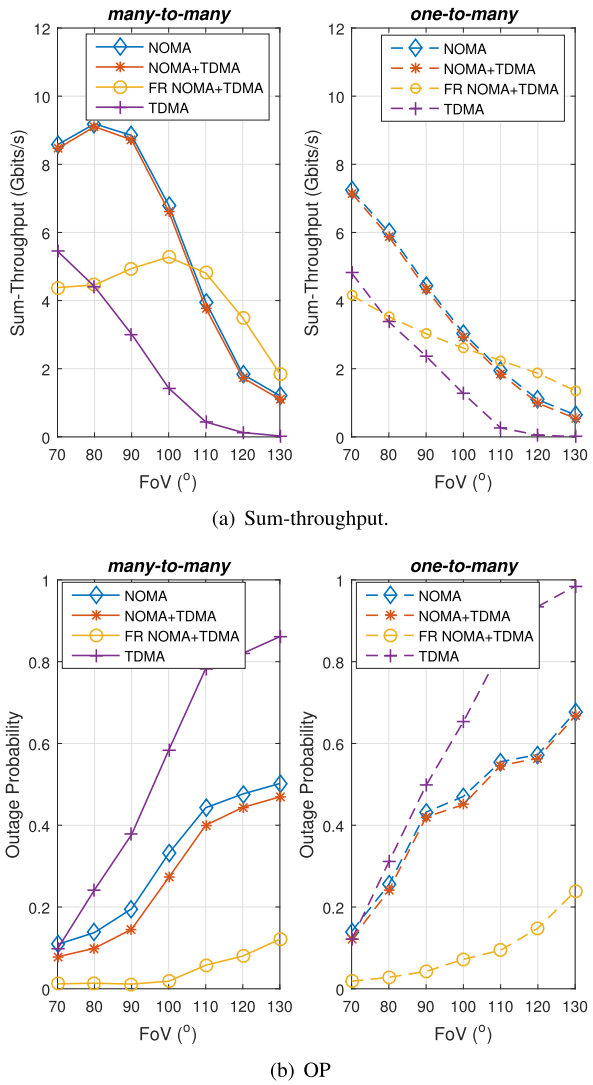


Fig. 7. Effects of various reuse factors on the throughput (TP) and OP based on our hybrid NOMA and OMA system compared under different FoVs, where a total of $|\mathcal{U}| = 25$ UEs have to be supported using the parameters of Table I.

Based on the proposed graph-coloring aided FR technique, we investigate the effect of various reuse factors on the system's performance. To be specific, as shown in Fig. 7, a high reuse factor has a low UE outage as a benefit of the reduced ICI, albeit at a reduced sum-throughput. Furthermore, we observe that for a moderate FoV, further increasing the reuse factor results in a substantial sum-throughput reduction, even though the improvement of the UE outage is not so prominent. Since the ICI is not so strong, it is futile to employ a high reuse factor. By contrast, when the system suffers from serious ICI due to large FoVs, moderately increasing the reuse factor strikes a trade-off between a high sum-throughput and a low OP.

B. Effects of FoV

Let us now investigate the effects of various FoVs on the performance of our UD-VLC network, where the reuse factor of $\varrho = 3$ is adopted. Note that when $\varrho = 1$, no FR is invoked in our proposed scheme. Observe in the left subfigure of Fig. 8(a) that the sum-throughput of TDMA is reduced upon increasing the FoV, hence TDMA almost completely fails in a high FoV scenario. The sum-throughput achieved by the conventional NOMA scheme is initially slightly increased, but then sharply decays upon increasing the FoV. Furthermore, we observe that for FoV $\leq 100^\circ$ the sum-throughput of the hybrid NOMA and OMA scheme associated with $\varrho = 3$ is improved upon increasing the FoV, but remains lower than that of the hybrid NOMA and OMA scheme using $\varrho = 1$ due to the spectrum partitioning. However, this observation is reversed for FoV $\geq 110^\circ$, which indicates that as a benefit of the FR technique, the proposed scheme becomes capable of mitigating the hostile ICI. Furthermore, compared to the throughput achieved by the *one-to-many* association, as shown in the right subfigure of Fig. 8(a), there is a consistent throughput gain for the system relying on the *many-to-many* association, which is an explicit benefit of its overlapped clusters.

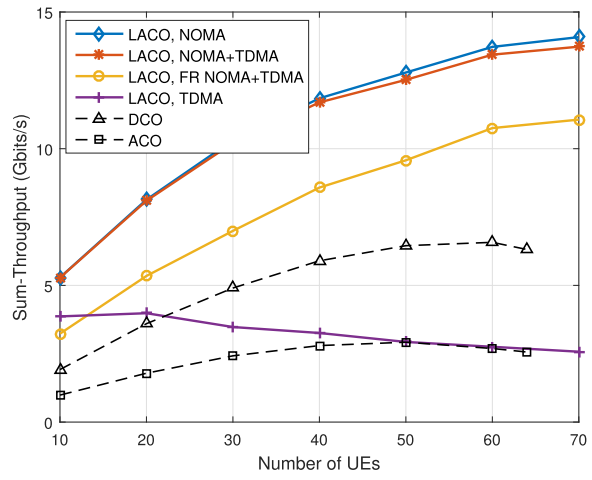


(a) Sum-throughput.

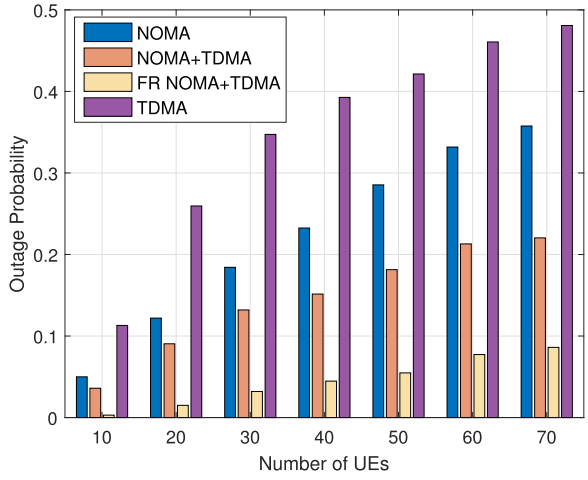
(b) OP

Fig. 8. Achievable sum-throughput and OP of different MA schemes for various FoVs, where $|\mathcal{U}| = 25$ UEs have to be served using the parameters of Table I.

To explore further, the left subfigure of Fig. 8(b) presents the OP for the same range of FoVs, relying on the proposed *many-to-many* association. The OP of the conventional NOMA aided UD-VLC is better than that of the TDMA. As a benefit of TDMA-based scheduling, the proposed hybrid NOMA and OMA scheme using $\varrho = 1$ mitigates the OP, whilst its sum-throughput approaches that of the conventional NOMA scheme. However, the OPs of these schemes remain unsatisfactory. Remarkably, we observe that the OP of the hybrid NOMA and OMA scheme using $\varrho = 3$ remains much lower than that of the other three schemes even in large-FoV-scenarios, which is an explicit benefit of our interference management technique. Since the UE can only be served by a single AP in *one-to-many* association, as expected, the OP of the UD-VLC relying on the *one-to-many* association is higher than that of the system relying on the *many-to-many* association under various FoVs. In the light of the above observations, we suggest that purely relying on the conventional TDMA or NOMA schemes fails to provide a good service in UD-VLC networks. The proposed hybrid



(a) Sum-throughput



(b) OP

Fig. 9. Sum-throughput and OP vs. the number of UEs at FoV $= 85°$.

NOMA and OMA scheme with *many-to-many* association using $\varrho = 1$ is powerful in moderate-ICI environments, while by further adopting the FR technique, the proposed scheme performs well in ICI-contaminated scenarios.

C. Effect of UE Density

It is also interesting to explore the capacity of the UD-VLC networks. In this section, we observe the performance of the LACO-OFDM aided UD-VLC for various UE densities at FoV $= 85°$. As seen in Fig. 9(a), upon increasing the number of UEs $|\mathcal{U}|$, both the conventional NOMA and the proposed scheme using $\varrho = 1$ and 2, respectively, achieve an successively increased sum-throughput, while TDMA achieves a much lower throughput. Furthermore, by invoking NOMA in the LACO-OFDM aided UD-VLC network constructed, its throughput is much higher than that of the ACO/DCO-OFDM aided VLC relying on the multi-AP-multi-UE clustering of [37]. Moreover, due to the employment of TPC in [37], the number of UEs served is limited to the number of APs, as evidenced by Fig. 9(a), where the dashed lines are curtailed at $|\mathcal{U}| = |\mathcal{A}|$. By contrast, this is not the case for our proposed

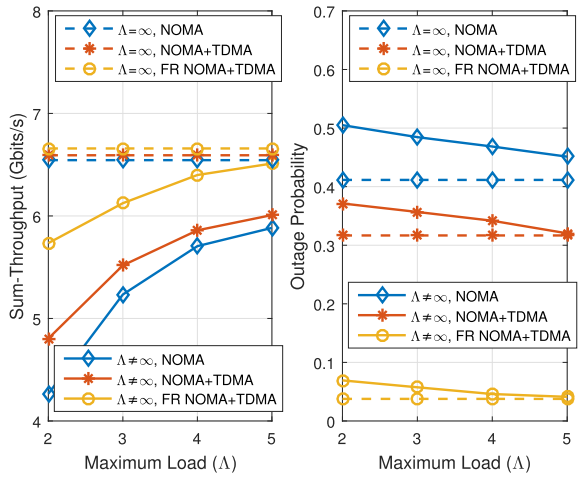


Fig. 10. Achievable sum-throughput and OP as a function of maximum load Λ, investigated in the context of conventional NOMA and hybrid NOMA and OMA schemes with and without FR schemes, compared to the ideal load scenario (denoted by dash lines), where a total of $|\mathcal{U}| = 40$ UEs are aimed to be supported in the system under FoV $= 100°$.

network. Additionally, the OP of the proposed scheme using $\varrho = 1$ or 2 shown in Fig. 9(b) is better, regardless of the UE density. Note that the OP improvement is more remarkable, when the FR technique is invoked, albeit at the cost of a sum-throughput reduction.

D. Effects of UE-Load Restrictions

It is worth mentioning that the above simulation results were obtained under loose UE-load considerations. Since the UE-load restriction has an impact both on the complexity and stability of the transceiver design, we now take into account the effects of various UE-load restrictions imposed on both the conventional NOMA and on the proposed hybrid NOMA and OMA schemes. As seen in Fig. 10, compared to the ideal load scenario (displayed by dash lines), there is a throughput loss for all three schemes under a load constraint, and the OPs are also deteriorated. It implies that the performance erosion of the conventional NOMA is imposed by the load constraint. By contrast, for the proposed scheme operating both with and without FR, the OPs remains nearly unaffected, but there is a throughput loss owing to the resource-partitioning.

VI. Conclusions

A compelling system architecture was designed for LACO-OFDM aided UD-VLC networks. An overlapped clustering strategy based on *many-to-many* association has been proposed for LACO-OFDM, which offered beneficial throughput gains in ICI-contaminated scenarios. With the aid of the proposed dynamic resource allocation algorithm, each AP becomes capable of allocating different transmit power, whilst each UE dynamically exploits its set of legitimate modulation-modes as well as the number of activated subcarriers, fully harnessing the flexibility of LACO-OFDM at the system-level. In pursuit of a further enhanced performance, a hybrid NOMA and OMA scheme has been proposed, which outperforms the conventional TDMA and NOMA schemes in terms of both its sum-throughput and OP. Our simulation results also

demonstrated that the proposed scheme performs even better than the ACO/DCO-OFDM aided VLC network relying on TPC for multi-AP-multi-UE transmission and has no hard limitation on the UE capacity.

ACKNOWLEDGMENT

The data from the paper can be obtained from the University of Southampton ePrints research repository: 10.5258/SOTON/D0730.

REFERENCES

[1] H. Marshoud, S. Muhaidat, P. C. Sofotasios, S. Hussain, M. A. Imran, and B. F. Sharif, "Optical non-orthogonal multiple access for visible light communication," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 82–88, Apr. 2018.

[2] 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," *Proc. IEEE*, vol. 100, pp. 1853–1888, May 2012.

[3] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?" *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1533–1544, Mar. 15, 2016.

[4] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 4th Quart., 2016.

[5] S. Stefanatos and A. Alexiou, "Access point density and bandwidth partitioning in ultra dense wireless networks," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3376–3384, Sep. 2014.

[6] M. Kamel, W. Hamouda, and A. Youssef, "Performance analysis of multiple association in ultra-dense networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3818–3831, Sep. 2017.

[7] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: challenges, methodologies, and directions," *IEEE Wireless Commun. Mag.*, vol. 23, no. 2, pp. 78–85, Apr. 2016.

[8] X. Li, R. Zhang, and L. Hanzo, "Optimization of visible-light optical wireless systems: Network-centric versus user-centric designs," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1878–1904, 3rd Quart., 2018.

[9] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2047–2077, 4th Quart., 2015.

[10] A. Jovicic, J. Li, and T. Richardson, "Visible light communication: Opportunities, challenges and the path to market," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 26–32, Dec. 2013.

[11] B. Li, J. Wang, R. Zhang, H. Shen, C. Zhao, and L. Hanzo, "Multiuser MISO transceiver design for indoor downlink visible light communication under per-LED optical power constraints," *IEEE Photon. J.*, vol. 7, no. 4, Aug. 2015, Art. no. 7201415.

[12] S. D. Dissanayake and J. Armstrong, "Comparison of ACO-OFDM, DCO-OFDM and ADO-OFDM in IM/DD systems," *J. Lightw. Technol.*, vol. 31, no. 7, pp. 1063–1072, Apr. 1, 2013.

[13] B. Ranjha and M. Kavehrad, "Hybrid asymmetrically clipped OFDM-based IM/DD optical wireless system," *J. Opt. Commun. Netw.*, vol. 6, no. 4, pp. 387–396, Apr. 2014.

[14] Q. Wang, C. Qian, X. Guo, Z. Wang, D. G. Cunningham, and I. H. White, "Layered ACO-OFDM for intensity-modulated direct-detection optical wireless transmission," *Opt. Express*, vol. 23, no. 9, pp. 12382–12393, May 2015.

[15] X. Zhang, Q. Wang, R. Zhang, S. Chen, and L. Hanzo, "Performance analysis of layered ACO-OFDM," *IEEE Access*, vol. 5, pp. 18366–18381, 2017.

[16] R. Bai, Z. Wang, R. Jiang, and J. Cheng, "Interleaved DFT-spread layered/enhanced ACO-OFDM for intensity-modulated direct-detection systems," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4713–4722, Oct. 15, 2018.

[17] T. Q. Wang, H. Li, and X. Huang, "Interference cancellation for layered asymmetrically clipped optical OFDM with application to optical receiver design," *J. Lightw. Technol.*, vol. 36, no. 11, pp. 2100–2113, Jun. 1, 2018.

[18] X. Zhang, Z. Babar, R. Zhang, S. Chen, and L. Hanzo, "Multi-class coded layered asymmetrically clipped optical OFDM," *IEEE Trans. Commun.*, to be published.

[19] X. Ge, X. Li, H. Jin, J. Cheng, and V. C. M. Leung, "Joint user association and user scheduling for load balancing in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3211–3225, May 2018.

[20] 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 Commun.*, vol. 22, no. 2, pp. 8–16, Apr. 2015.

[21] X. Li, R. Zhang, and L. Hanzo, "Cooperative load balancing in hybrid visible light communications and WiFi," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1319–1329, Apr. 2015.

[22] R. Zhang, Y. Cui, H. Claussen, H. Haas, and L. Hanzo, "Anticipatory association for indoor visible light communications: Light, follow me!" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2499–2510, Apr. 2018.

[23] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[24] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[25] H. Marshoud, P. C. Sofotasios, S. Muhaidat, G. K. Karagiannidis, and B. Sharif, "On the performance of visible light communication systems with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6350–6364, Oct. 2017.

[26] L. Yin, W. O. Popoola, X. Wu, and H. Haas, "Performance evaluation of non-orthogonal multiple access in visible light communication," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5162–5175, Dec. 2016.

[27] Z. Yang, W. Xu, and Y. Li, "Fair non-orthogonal multiple access for visible light communication downlinks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 66–69, Feb. 2017.

[28] C. Yang, J. Li, Q. Ni, A. Anpalagan, and M. Guizani, "Interference-aware energy efficiency maximization in 5G ultra-dense networks," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 728–739, Feb. 2017.

[29] K. S. Ali, H. Elsawy, A. Chaaban, and M.-S. Alouini, "Non-orthogonal multiple access for large-scale 5G networks: Interference aware design," *IEEE Access*, vol. 5, pp. 21204–21216, 2017.

[30] L. Zeng *et al.*, "High data rate multiple input multiple output (MIMO) optical wireless communications using white led lighting," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1654–1662, Dec. 2009.

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

[32] 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.

[33] S. Feng, X. Li, R. Zhang, M. Jiang, and L. Hanzo, "Hybrid positioning aided amorphous-cell assisted user-centric visible light downlink techniques," *IEEE Access*, vol. 4, pp. 2705–2713, 2016.

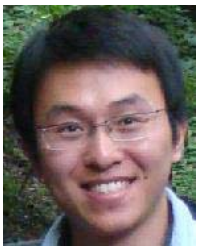
[34] S. E. Schaeffer, "Graph clustering," *Comput. Sci. Rev.*, vol. 1, no. 1, pp. 27–64, Aug. 2007.

[35] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.

[36] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, 4th ed. Belmont, MA, USA: Athena Scientific, 2017.

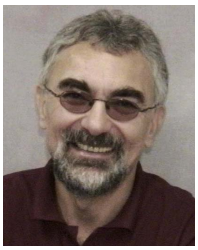
[37] S. Feng, R. Zhang, X. Li, Q. Wang, and L. Hanzo, "Dynamic throughput maximization for the user-centric visible light downlink in the face of practical considerations," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5001–5015, Aug. 2018.

Simeng Feng received the B.Eng. degree in electronic information science and technology from the Ocean University of China and the master's degree in wireless communications from The University of Southampton, U.K., where she is currently pursuing the Ph.D. degree with the Next Generation Wireless Research Group. Her research interests include visible light communications, resource allocation, indoor positioning, and heterogeneous networks.



Rong Zhang (M'09–SM'16) received the Ph.D. degree in wireless communications from The University of Southampton (UoS) in 2009. In 2009, he was a Research Assistant with the Mobile Virtual Centre of Excellence, UoS, one of the U.K.'s largest industrial–academic partnership in ICT. He was a UoS Lead Researcher with the School of Electronics and Computer Science, UoS, where he contributed to a number of international projects. He took his industrial consulting leave from Huawei EU Research and Development as a System Algorithms Expert. He is currently an Assistant Professor with the Southampton Wireless Group, UoS. He has a total of 80+ IEEE/OSA publications, including 55+ journals (20+ of which as the first author). He is a member of the OSA and a RAEng Industrial Fellow. He was a recipient of the prestigious RAEng Industrial Fellowship. He received the prestigious Dean's Publication Award from the Faculty of Physical Sciences and Engineering, UoS. He regularly serves as a reviewer for IEEE/OSA journals and funding bodies and has been several times as a TPC member/invited session chair of major conferences.

Wei Xu (S'07–M'09–SM'15) received the B.Sc. degree in electrical engineering and the M.S. and Ph.D. degrees in communication and information engineering from Southeast University, Nanjing, China, in 2003, 2006, and 2009, respectively. Between 2009 and 2010, he was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Victoria, Canada. He is currently a Professor with the National Mobile Communications Research Laboratory, Southeast University. He has co-authored over 100 refereed journal papers in addition to 35 granted domestic patents and three U.S. patents. His research interests include cooperative communications, information theory, signal processing, and machine learning for wireless communications. He is a Distinguished Visiting Fellow of the Royal Academy of Engineering, U.K., in 2019. He was a co-recipient of the First Prize of the Science and Technology Award in Jiangsu Province, China, in 2014. He received the Best Paper Award from the IEEE MAPE in 2013, the IEEE/CIC ICCC in 2014, the IEEE Globecom in 2014, the IEEE ICUWB in 2016, and the WCSP in 2017. He was an Editor of IEEE COMMUNICATIONS LETTERS from 2012 to 2017. He is currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE ACCESS. He has been involved in technical program committees for international conferences, including the IEEE Globecom, the IEEE ICC, the IEEE WCNC, the IEEE VTC, and the IEEE PIMRC.



Lajos Hanzo (F'04) received the degree in electronics, the Doctorate degree, and the Honorary Doctorate degree from the Technical University of Budapest in 1976, 1983, and 2009, respectively, and the Honorary Doctorate degree from The University of Edinburgh in 2015. In 2016, he was admitted to the Hungarian Academy of Sciences. During his 40-year career in telecommunications, he has held various research and academic posts in Hungary, Germany, and U.K. During 2008–2012, he was a Chaired Professor with Tsinghua University, Beijing. Since 1986, he has been with the School of Electronics and Computer Science, The University of Southampton, U.K., where he holds the Chair in telecommunications. He is currently 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, U.K., the European Research Council's Advanced Fellow Grant, and the Royal Society's Wolfson Research Merit Award. He has successfully supervised 112 Ph.D. students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totaling in excess of 10 000 pages, and published 1760 research contributions at the IEEE Xplore. He is a fellow of the Royal Academy of Engineering, The Institution of Engineering and Technology, and the European Association for Signal Processing. He acted both as the TPC Chair and the General Chair of IEEE conferences, presented keynote lectures, and has been awarded a number of distinctions. 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 ComSoc and VTS. During 2008–2012, he was the Editor-in-Chief of the IEEE Press. For further information on research in progress and associated publications, please refer to http://www-mobile.ecs.soton.ac.uk.