Joint User Association and Beamforming in Integrated Satellite-HAPS-Ground Networks

Shasha Liu, Student Member, IEEE, Hayssam Dahrouj, Senior Member, IEEE, and Mohamed-Slim Alouini, Fellow, IEEE

Abstract—This article proposes, and evaluates the benefit of, one particular hybrid satellite-high-altitude-platform-station (HAPS)-ground network, where one HAPS connected to one geo-satellite assists the ground base stations (BSs) at serving ground-level users. The paper assumes that the geo-satellite is connected to the HAPS using free-space-optical backhaul links. The HAPS, equipped with multiple antennas, aims at transmitting the geo-satellite data to the users via radio-frequency (RF) links using spatial-multiplexing. Each ground BS, on the other hand, is equipped with multiple antennas, but directly serves the users through the RF links. The paper then focuses on maximizing the network-wide throughput, subject to HAPS payload connectivity constraint, HAPS and BSs power constraints, and backhaul constraints, so as to jointly determine the user-association strategy of each user (i.e., user to geo-satellite via HAPS, or user to BS), and their associated beamforming vectors. We tackle such a mixed discrete-continuous optimization problem using an iterative approach, where the user-association is determined using a combination of integer linear programming and generalized assignment problems, and where the beamforming strategy is found using a weighted-minimum-mean-squared-error approach. The simulations illustrate the appreciable gain of our proposed algorithm, and highlight the prospects of augmenting the ground networks with beamforming-empowered HAPS for connecting the unconnected, and super-connecting the connected.

Index Terms—High altitude platform station (HAPS), satellite-HAPS-ground network, backhaul, user association, beamforming, throughput.

I. INTRODUCTION

A. Overview

LARGE-SCALE ground-level connectivity has a major impact on current telecommunications infrastructures, which today support billions of people, and tens of billions of devices [2]. As this demand is expected to grow at an even faster pace over the next few years, a major practice of telecom operators is to densify the terrestrial network infrastructures [3]. Such densification may, however, not always be able to satisfy the data ultra-hungry devices and their ambitious quality-of-services requirements in high-interference regimes, and is also not feasible to be realized in rural and remote areas. Augmenting ground-level communications with spatial networks, e.g., satellites at the GEO-layer (i.e., geo-satellites), and the stratospheric layer (i.e., High-Altitude-Platform-Station (HAPS)), is expected to revolutionize the physical layer paradigm of the sixth generation of wireless systems and beyond (6G and beyond). Satellite-ground networks can provide service in remote areas. However, due to the high launch costs and limited orbit, satellite communications are too expensive to promote [4]. Space-air-ground integrated network (SAGIN) emerges as a powerful architecture to provide seamless, reliable, and affordable connectivity [5]. Generally speaking, SAGIN consists of space, air, and ground segments. In particular, the space segment can achieve global connectivity and guarantee coverage, the air segment with characteristics of easy deployment, low cost, and extensive coverage can provide high-quality broadband services [6], and the ground segment can support high-data-rate applications. Such an integrated network, in fact, boosts several timely applications, including Intelligent Transportation System (ITS) [7], wireless communications for public protection and disaster relief (PPDR) [8], and Internet of Things (IoT). In this article, we focus on a specific air segment, i.e., HAPS. By adding HAPS to the traditional space-ground network, some of the shortcomings and challenges of the existing networks can be well-resolved, especially those related to 6G networks challenges and goals towards connecting the unconnected, and ultraconnecting the connected [2], [3]. HAPS enhances terrestrial communications due to its better path-loss profile as compared to higher layers platforms, e.g., cubesats [9]. Compared with Unmanned Aerial Vehicles (UAVs), HAPS can cover a more expansive area due to its elevated altitude, and wider beam coverage [10]. Specifically, [11] points out that the coverage footprint of a HAPS could be 20-50 km for metropolitan areas. Further, a HAPS can provide a wide coverage radius of 50-500 km with LoS communication [11]. In addition, HAPS is located at the stratospheric layer, which provides several appealing deployment characteristics, e.g., the ability to maintain a quasi-stationary state and achieve global connectivity [11], [12], [13]. The true assessment of such deployment remains, however, a strong function of the joint resource allocation across the HAPS and ground base stations (BSs), and so this article proposes one particular framework for optimizing integrated

0018-9545 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
satellite-HAPS-ground networks under specific physical connectivity constraints.

In the context of our paper, the HAPS functions as an intermediate station to improve the capacity and the signal-to-interference-plus-noise ratio of the served users, so as to improve the performance of current satellite-ground networks performance [14]. This article considers a vertical heterogeneous network (VHetNet) comprising one geo-satellite, one HAPS, and several terrestrial BSs, where both the HAPS and ground BSs are equipped with multiple antennas to simultaneously serve multiple users. The paper assumes that the satellite and HAPS are connected through free-space optical (FSO) links, which have a wide bandwidth, and are secure, license-free, and suitable for deploying point-to-point (P2P) communication in the space [2], [15]. Moreover, since HAPS is located in the stratosphere layer, the FSO links connecting the geo-satellite to HAPS do not go through the troposphere layer, which makes the geo-satellite-to-HAPS FSO links less vulnerable to weather conditions, e.g., turbulence, rain, fog, etc. [2]. FSO links, however, require strong alignment between the sender and receiver [16], which is not suitable for the mobile nature of ground users communications. To this end, the paper assumes that both the HAPS and ground BSs communicate with their respective ground users using radio-frequency (RF) links, where each user can be served by either the ground BSs or by the HAPS. Unlike ground BSs, however, HAPS payload consists of three subsystems: flight control system, energy management subsystem, and communication payload system [11], which poses an additional constraint on the HAPS connectivity capability. More specifically, HAPS requires a flight control system to handle mobility and maintain a quasi-stationary state, an energy management subsystem for energy storage and distribution, and a communication payload system to handle the communication between the HAPS and other entities. Therefore, our paper accounts for such HAPS particularity, by adding the HAPS payload connectivity constraint to the optimization problem at hand. Given their high operational altitude and large platform, the capacity and coverage of HAPS are much better than ground BSs and suitable for spatial multiplexing technology. Therefore, different from ground BSs which only serve a small area of users, HAPS helps connecting the unconnected and ultraconnecting the connected. Further, the transmissions across the FSO satellite-HAPS links are assumed to occur over different optical bands, and do not interfere with each other. The considered network performance becomes, therefore, a strong function of the intra-HAPS interference, intra-BS interference, inter-BS interference, and HAPS-BS interference (hereafter denoted by inter-layer interference). The paper then attempts at managing such multi-mode multi-layered interference by means of associating users with BSs or HAPS, and determining their corresponding beamforming vectors so as to maximize the considered integrated satellite-HAPS-ground network throughput.

B. Related Work

The problem considered in this article is related to the optimization of vertical heterogeneous networks, and particularly to the resource allocation problem in integrated satellite-HAPS-ground networks. The tackled problem is also related to user scheduling and beamforming problems, which are studied extensively in the past literature, both individually and jointly, especially in the context of classical interference networks optimization.

Optimizing system throughput in interference networks is often a non-convex optimization problem, and so managing the wireless networks radio resources remains a challenging problem in general [17]. Many recent techniques, therefore, aim at devising numerically reasonable optimization algorithms that promise to offer major performance improvements as compared to conventional systems strategies. For example, the user association scheduling problem is considered in several classical networks, e.g., [18], [19], [20], [21], all of which focus on terrestrial networks optimization only. Similarly, the joint user association and power assignment problem are addressed in [22]; please also see the references therein. The user-association subproblem considered in the current paper, however, involves a more intricate coupling of cross-mode cross-layered interference and HAPS connectivity constraints, and so the paper leverages techniques such as linear integer programming [23] and generalized assignment problems [24] to develop reasonable heuristics for dealing with the problem discrete intricacies.

The problem of beamforming optimization is also extensively studied in the literature of wireless networks, either using Lagrangian-duality [25], semidefinite programming (SDP) [26], weighted-minimum mean squared error (WMMSE) [27], [28], or fractional programming (FP) in [29]. The joint user association and beamforming problem is also addressed in [30], [31] under specific terrestrial systems scenarios. Further, reference [32] proposes a cooperative multigroup multicast transmission in integrated terrestrial-satellite networks that employs relation based algorithm and iterative beamforming design algorithm to jointly allocate resource and maximize the multigroup capacity. In [33], the joint cache placement, content delivery, and multicast beamformings in the integrated satellite-terrestrial network (ISTN) are considered. From a methodology perspective, WMMSE and FP are noticeably popular solutions to maximize the sum-rate in conventional terrestrial networks [27], [28], [29]. Given the structure of our problem formulation, a part of the current paper proposes a modified version of WMMSE to best account for the physical constraints stemming from the FSO backhaul link constraint and the multi-mode multi-layered interference in the context of maximizing the system sum-rate under fixed user-association strategy.

The problem considered in this article is strongly coupled with the latest advances and studies of HAPS networks, which come at the forefront of sky connectivity latest trends. For example, reference [11] presents a comprehensive overview about the vision and framework of HAPS networks. Reference [11] further highlights the prospects of HAPS systems in radio resource management, which our current paper studies under one particular system architecture. In fact, the HAPS used in the current paper also acts as a super-macro BS, which is well motivated through reference [3] that illustrates the role of HAPS in serving both remote and metropolitan dense areas. The current paper builds
The paper compares the proposed algorithm to classi-
As a one step forward toward achieving the 6G digital
and
stud-
by accounting for the joint user association and beam-
considers power assignment and transmission proto-
This article proposes an iterative algorithm to solve the
problem in an integrated satellite-HAPS-ground network com-
partly resource allocation problem in VHetNets, especially
particular VHetNet consisting of
in an integrated HAPS-mobile telecommunications (IMT)
system. References [14], [38] are particularly related to our
system model; however, our paper goes beyond both [14],
[38] by accounting for the joint user association and beam-
forming problem in a cross-mode cross-layered interference
setup. More specifically, on the one hand, reference [14] studies
a hybrid RF/FSO VHetNet consisting of satellites, HAPS, and
ground BSs, and focuses on the systematic performance
analysis of the networks. On the other hand, reference [38]
proposes an integrated satellite-airborne-ground network and
optimizes the user access, power assignment, and HAPSs’ lo-
cation under an orthogonal frequency division multiple access
scheme.
Considering the limitations of terrestrial networks, some ex-
treme regions, including oceans, glaciers, and mountains, cannot
be covered, which can be best addressed through the deploy-
ment of satellite-ground networks. However, such networks still face a non-negligible delay, and are oftentimes subject to
large path-loss profile. To achieve ubiquitous global connectiv-
ity, this article considers one particular VHetNet consisting of
one geo-satellite, one HAPS, and several BSs, as a means to
connect the unconnected and ultra-connect the connected. In addition, differently from both [38] and [14], our paper adopts
a multiple-antenna scheme at the HAPS and at the ground
BSs, and optimizes the user association and spatial multiplexing
strategies so as to efficiently serve the ground users subject to
practical system-level constraints.

C. Contributions
Unlike the aforementioned papers, this article focuses on one
particular resource allocation problem in VHetNets, especially
introduced to improve digital inclusion among users. More
specifically, we study the joint user association and beamforming
problem in an integrated satellite-HAPS-ground network com-
prising one satellite, one HAPS, and several ground BSs, where
the geo-satellite is connected to the HAPS using FSO links. The
ground users, each equipped with a single antenna, can then be
served via RF links either by the HAPS or by one of the ground
BSs, all whilst accounting for the FSO link quality between the
HAPS and the satellite, and for the HAPS payload capabili-
ties. The paper then addresses the mixed-integer optimization
problem of maximizing the network sum-rate subject to power,
backhaul, and payload constraints using an iterative modular
approach. That is, we iterate between solving the user association
strategy for fixed beamforming, and solving the beamforming
problem for fixed user association. The paper contributions can
then be summarized as follows:

- As a one step forward toward achieving the 6G digital
inclusion objective, i.e., connecting the unconnected and
ultraconnecting the connected, this article proposes one
particular satellite-HAPS-ground multi-antenna network
architecture, specifically designed to augment ground com-
 munications through user scheduling and spatial multi-
plexing. The system performance becomes, therefore, a
strong function of the emerging intra-HAPS interference,
 intra-BS interference, inter-BS interference, and HAPS-
BS interference. A complex mixed discrete-continuous
non-convex optimization problem is then formulated to
jointly pair users with BSs and HAPSs, and to design the
beamforming vectors of the associated users so as to max-
imize the network sum-rate while accounting for HAPS
payload connectivity constraint, FSO backhaul constraints,
and HAPS and BS maximum power constraints.

- This article proposes an iterative algorithm to solve the
non-convex optimization problem. That is, we iteratively
optimize each of the optimization parameters by fixing
other variables. Given fixed beamforming vectors, the user
association strategy is first determined by linearizing the
original problem by adding additional constraints so as to
enable the utilization of integer linear programming (ILP),
followed by a generalized assignment problem (GAP)-
type solution. Then, given the fixed user association vari-
bles, the beamforming optimization problem is solved by
accounting for the problem-specific constraints. That
is, given the formulated problem constraints, the beam-
forming vectors at the HAPS and at the BSs are derived
based on a series of problem reformulations that eventually
enable the use of a weighted minimum mean square error
(WMMSE)-type solution.

- The paper compares the proposed algorithm to classi-
cial techniques using Monte-Carlo simulations. The pa-
per results illustrate the appreciable sum-rate gain of the
proposed joint user association and beamforming algo-
 rithm as compared to classical techniques for various
network parameters. The simulations particularly high-
light the numerical potential of the proposed integrated
satellite-HAPS-ground networks optimization framework
for connecting the unconnected, and super-connecting the
connected, especially at the high interference regime, and
and under beefed-up HAPS capabilities (i.e., power, number of
antennas, quality of FSO backhauling, etc.).

- The paper draws a handful of design guidelines and rec-
ommendations for deploying HAPS in both remote and
metropolitan dense areas.

The rest of the paper is organized as follows. Section II
presents the system model and problem constraints. The problem
IV. ALGORITHM FOR GROUND

\[ I = \text{is used to denote Rician small-scale fading (with} \{ R \}\text{, the value} \{ R \}\text{be the number of antennas at BSs} = B \text{users. The paper assumes that the set of transmitters is denoted by} \mathcal{U} = \{ 1, 2, ..., N \}, \text{where the} \mathcal{U} \text{set of users by} \mathcal{U} = \{ 1, 2, ..., N \}. \text{Let} N_{\text{A}} \text{be the number of antennas at BSs} \text{and HAPS} (i.e.,} i = 0 \text{for HAPS,} i = 1, 2, ..., N_{B} \text{for ground BSs). The paper assumes that the satellite communicates with HAPS via FSO links, while HAPS and BSs connect to the users via RF links. Each user can be served either by the HAPS or by one of the ground BSs. Being served by the HAPS means that the required data is sent from the satellite to HAPS via the FSO link, and then the HAPS sends it to the ground via the RF link. The paper directly, communicates with the user directly. The paper adopts a space division multiplexing scheme, where all RF links use the same central frequency, and where the HAPS and BSs adopt multi-user beamforming to serve multiple users simultaneously. Fig. 1 shows an example of the considered network, which consists of one geo-satellite, one HAPS, 3 ground BSs, and 9 users. Fig. 1 also illustrates the information flow from the terrestrial gateway to the geo-satellite. The paper, in fact, assumes that such data feeding happens over different time-scales than the considered downlink transmission, i.e., it does not interfere with the considered satellite-HAPS-ground network. Furthermore, the paper assumes full signal synchronization at the user terminals, which is best realized through system-level coordination among the different layers of the considered VHetNet, similar to [38]. Moreover, we consider a scenario where the central processor (CP) can acquire channel information. Further, in this scenario, we assume that the CP has the ability to schedule the data placement for both the BSs and the HAPS [33]. Thus, by leveraging the distributed data placement and channel information, the CP can jointly process the beamforming decisions both at the HAPS and at the BSs. We note, however, that the statistical update of data placement and acquisition of channel information are beyond the scope of the current paper, and are left as a future research direction. Further, unlike the coordinated multipoint (CoMP) approach, the CP in our paper simply serves as an algorithmic implementation module, only capable of providing the user association and beamforming vectors decision to both ground BSs and the HAPS based on the acquired channel information primarily fed through these entities. As such communication overhead, i.e., the optimization coordination between the HAPS and BSs is limited to control signaling, the transmission delay and capacity limitations between the HAPS and CP for such signaling purposes is considered negligible, as also treated in most of the recent literature on the topic, e.g., please see [38] and references therein. We next present the channel model and rate expressions of the hybrid space-air-ground system under study.

II. SYSTEM MODEL AND PROBLEM CONSTRAINTS

A. System Model

Consider an integrated satellite-HAPS-ground network consisting of one geo-satellite, one HAPS, \( N_{B} \) ground BSs, and \( N_{U} \) users. The paper assumes that the set of transmitters is denoted by \( \mathcal{I} = \{ 0, 1, 2, ..., N_{B} \} \), where the \( 0 \)th transmitter points to the HAPS. We also denote the set of users by \( \mathcal{U} = \{ 1, 2, ..., N_{U} \} \). Let \( N_{A} \) be the number of antennas at BSs and HAPS (i.e., \( i = 0 \) for HAPS, \( i = 1, 2, ..., N_{B} \) for ground BSs). The paper assumes that the satellite communicates with HAPS via FSO links, while HAPS and BSs connect to the users via RF links. Each user can be served either by the HAPS or by one of the ground BSs. Being served by the HAPS means that the required data is sent from the satellite to HAPS via the FSO link, and then the HAPS sends it to the ground user via the RF link. The ground BS, however, communicates with the user directly. The paper adopts a space division multiplexing scheme, where all RF links use the same central frequency, and where the HAPS and BSs adopt multi-user beamforming to serve multiple users simultaneously. Fig. 1 shows an example of the considered network, which consists of one geo-satellite, one HAPS, 3 ground BSs, and 9 users. Fig. 1 also illustrates the information flow from the terrestrial gateway to the geo-satellite. The paper, in fact, assumes that such data feeding happens over different time-scales than the considered downlink transmission, i.e., it does not interfere with the considered satellite-HAPS-ground network. Furthermore, the paper assumes full signal synchronization at the user terminals, which is best realized through system-level coordination among the different layers of the considered VHetNet, similar to [38]. Moreover, we consider a scenario where the central processor (CP) can acquire channel information. Further, in this scenario, we assume that the CP has the ability to schedule the data placement for both the BSs and the HAPS [33]. Thus, by leveraging the distributed data placement and channel information, the CP can jointly process the beamforming decisions both at the HAPS and at the BSs. We note, however, that the statistical update of data placement and acquisition of channel information are beyond the scope of the current paper, and are left as a future research direction. Further, unlike the coordinated multipoint (CoMP) approach, the CP in our paper simply serves as an algorithmic implementation module, only capable of providing the user association and beamforming vectors decision to both ground BSs and the HAPS based on the acquired channel information primarily fed through these entities. As such communication overhead, i.e., the optimization coordination between the HAPS and BSs is limited to control signaling, the transmission delay and capacity limitations between the HAPS and CP for such signaling purposes is considered negligible, as also treated in most of the recent literature on the topic, e.g., please see [38] and references therein. We next present the channel model and rate expressions of the hybrid space-air-ground system under study.

B. FSO Backhaul Capacity

In this article, the geo-satellite and the HAPS are connected by the FSO link. Further, we assume that the transmissions across the FSO satellite-HAPS link for different users occur over different optical bands. Therefore, the data rate via the FSO link for every user served by the HAPS is denoted by \( R_{FSO} \), the value of which strongly depends on the atmospheric attenuation (e.g., absorption, scattering, etc.). The data rate between the HAPS and the satellite \( R_{FSO} \) can then be written as [39]:

\[
R_{FSO} = \frac{P_{t}\eta_{R}10^{-\frac{L_{p}}{10}}10^{-\frac{L_{atm}}{10}}A_{R}}{A_{B}E_{p}\eta_{b}}, \tag{1}
\]

where \( P_{t} \) denotes the transmit power of the satellite, \( \eta_{R} \) and \( \eta_{b} \) stand for the optical efficiencies of the transmitter and receiver, respectively, \( L_{p} \) is the pointing loss, \( L_{atm} \) is the atmospheric attenuation, \( A_{R} \) and \( A_{B} \) are the area of the FSO receiver and beam, respectively i.e., \( A_{R} = \frac{\lambda^{2}}{4\pi^{2}d_{i,n}^{2}} \) is the geometrical loss, which characterizes the path loss, \( E_{p} \) denotes the photon energy, and \( \eta_{b} \) represents the receiver sensitivity.

C. RF Channel Model

According to [38], the channel coefficient between the \( n \)th antenna of the \( i \)th transmitter (\( t = 0 \) for HAPS, \( i = 1, 2, ..., N_{B} \) for BSs) and the \( j \)th user, denoted by \( h_{i,j,n} \), is given by

\[
h_{i,j,n} = \left( \frac{c}{4\pi d_{i,j,n}f_{c}} \right) A_{i,j,n}F_{i,j,n}, \tag{2}
\]

where \( c \) is the speed of light, \( f_{c} \) is the carrier frequency, \( d_{i,j,n} \) is the distance between the \( n \)th antenna of the \( i \)th transmitter and the \( j \)th user. Considering the influence of potential signal blockage in terrestrial networks, for the links from BSs to ground users, we use \( A_{i,j,n} \) to denote the log-normal shadowing, and \( F_{i,j,n} \) to denote the Rayleigh small-scale fading. However, different from ground BSs, in the case of HAPS, \( A_{i,j,n} \) is omitted, and \( F_{i,j,n} \) is used to denote Rician small-scale fading (with factor \( \kappa_{HAPS} \)), mainly due to the strong line-of-sight between the HAPS and the ground users. This, however, is adopted without loss of generality as the optimization framework rather...
depends on the values of channel vectors. More specifically, in the rest of the paper, we simply denote the general form of RF channel vector between transmitter \(i\) and user \(j\) \((i = 0\) for HAPS, \(i = 1, 2, \ldots, N_B\) for BSs) as \(h_{ij} \in \mathbb{C}^{N_A}\), where \(h_{ij} = [h_{ij,1}, h_{ij,2}, \ldots, h_{ij,n}, \ldots, h_{ij,N_A}]^T\). In the simulations part of the paper, the adopted parameters are explicitly provided so as to distinguish the cases of HAPS to users and ground BSs to users.

### D. User Association Scheme

The paper considers the practical consideration that user \(j\) request may (or may not) be available at transmitter \(i\). To this end, we introduce the binary variable \(\gamma_{ij}\) which is defined as 1 if data required by user \(j\) is available at transmitter \(i\), and zero otherwise, \(\forall i \in \mathcal{I}\) and \(\forall j \in \mathcal{U}\). We note that the variable \(\gamma_{ij}\) is fixed in the context of our paper, and are known to the optimizer. The paper then assumes that each ground user can be served by one transmitter at most (i.e., either by the HAPS or by one of the ground BSs.). Furthermore, to account for the HAPS payload connectivity constraint, we denote by \(K_0\) the maximum number of users that the HAPS can serve. Introduce a binary variable \(\alpha_{ij}\), which is equal to 1 if the user \(j\) is served by the transmitter \(i\) and 0 otherwise, which yields the following connectivity constraints:

\[
\sum_{j=1}^{N_j} \gamma_{0j} \alpha_{0j} \leq K_0, \tag{3}
\]

\[
\sum_{i=0}^{N_B} \gamma_{ij} \alpha_{ij} \leq 1, \forall j \in \mathcal{U}. \tag{4}
\]

### E. Rates Expressions

This article considers that multiuser downlink transmit beamforming is employed at both the HAPS and the ground BSs. Let \(s_{ij}\) represents the information signal for user \(j\) when served by transmitter \(i\), \(\forall i \in \mathcal{I}\) and \(\forall j \in \mathcal{U}\), and let \(w_{ij} \in \mathbb{C}^{N_A}\) be the beamforming vector associated with \(s_{ij}\). Therefore, the received signal at the \(j\)th user, denoted as \(y_j\), is given by

\[
y_j = \sum_{b=0}^{N_B} \sum_{u=1}^{N_i} \gamma_{bu} \alpha_{bu} s_{bu} h_{ij}^H w_{bu} + z_j, \tag{5}
\]

where \(h_{b} = [h_{b,1}, h_{b,2}, \ldots, h_{b,n}, \ldots, h_{b,N_A}]^T\) is the vector channel from transmitter \(b\) to user \(j\), and where \(z_j\) is the additive white circularly symmetric Gaussian complex noise with variance \(\sigma^2\) on each of its real and imaginary components. The above expression (5) implicates four types of interference, namely, intra-HAPS interference, intra-base-station interference, interlayer interference, and inter-base-station interference.

We next present the rate expressions of each user \(j\) according to the two types of user-association possibilities. If user \(j\) is served by transmitter \(i\), the associated signal-to-interference-plus-noise ratio (SINR), denoted by \((\text{SINR}_{ij})\), can be expressed as:

\[
\text{SINR}_{ij} = \frac{|h_{ij}^H w_{ij}|^2}{\sum_{u=1, u \neq j}^{N_i} \sum_{b=0}^{N_B} \gamma_{bu} \alpha_{bu} |h_{bj}^H w_{bu}|^2 + \sigma^2}. \tag{6}
\]

The user achievable RF rate can then be written as:

\[
R_{ij}^{RF} = \beta \log_2 \left(1 + \frac{|h_{ij}^H w_{ij}|^2}{\sum_{u=1, u \neq j}^{N_i} \sum_{b=0}^{N_B} \gamma_{bu} \alpha_{bu} |h_{bj}^H w_{bu}|^2 + \sigma^2} \right), \tag{7}
\]

where \(\beta\) is the transmission bandwidth. In the case where the \(j\)th user is served by a ground BS \(i\) (i.e., \(i \in \mathcal{I} = \{0\}\)), the data rate of user \(j\) is as in (7), which can be written as:

\[
R_{ij}^{\text{Ground$_{-}$BS}} = R_{ij}^{RF}, \forall i = 1, 2, \ldots, N_B. \tag{8}
\]

In the other case where the \(j\)th user is served by the HAPS (i.e., \(i = 0\)), the data rate via RF link (i.e., the rate expression in (7) for \(i = 0\)) can be written as:

\[
R_{ij}^{\text{HAPS$_{-}$RF}} = R_{ij}^{RF}. \tag{9}
\]

Further, the data rate of user \(j\) served by the HAPS, denoted by \(R_{ij}^{\text{HAPS}}\), can be written as follows:

\[
R_{ij}^{\text{HAPS}} = \min\{R_{ij}^{\text{HAPS$_{-}$RF}}, R_{ij}^{\text{FSO}}\}. \tag{10}
\]

For ease of presentation, we provide a list of the expressions used in this article in Table I. The paper next presents the considered optimization problem, together with the algorithm devised to address the problem intricacies.

### III. Problem Formulation and Proposed Solution

The paper focuses on maximizing the network sum-rate by optimizing the user association strategy, and the beamforming vectors at both the HAPS and the ground BSs, subject to HAPS payload connectivity constraint, maximum transmit power constraints and backhaul constraints. Let \(P_{\text{max}}\) be the maximal allowable power at the HAPS and BSs, \(\forall i \in \mathcal{I}\). Our optimization problem can then be mathematically written as follows:

\[
{\max_{\alpha_{ij}, \gamma_{ij}}} \sum_{i=1}^{N_I} \sum_{j=1}^{N_J} \gamma_{ij} \alpha_{ij} R_{ij}^{\text{Ground$_{-}$BS}} + \gamma_{0j} \alpha_{0j} R_{0j}^{\text{HAPS}}, \tag{11a}
\]

s.t.

\[
(8) - (9), (10), \tag{11b}
\]

\[
\sum_{j=1}^{N_J} \gamma_{0j} \alpha_{0j} \leq K_0, \tag{11c}
\]

\[
\sum_{i=0}^{N_B} \gamma_{ij} \alpha_{ij} \leq 1, \forall j \in \mathcal{U}, \tag{11d}
\]

\[
\alpha_{ij} \in \{0, 1\}, \forall i \in \mathcal{I}, \forall j \in \mathcal{U}, \tag{11e}
\]

\[
\sum_{j=1}^{N_J} \gamma_{ij} \alpha_{ij} w_{ij}^H w_{ij} \leq P_{i}^{\text{max}}, \forall i \in \mathcal{I}, \tag{11f}
\]

where the optimization in problem (11) is jointly over the binary association variables \(\alpha_{ij}\), and the continuous beamforming variables \(w_{ij}, \forall i \in \mathcal{I} \text{ and } j \in \mathcal{U}\). The objective function (11a) is the sum-rate of the integrated network, which consists of the rates of the users served by ground BSs and the rates of the users served by the HAPS. Constraint (8) is the achievable data rate of a user served by the ground BS, constraint (9) is the achievable data.
rate of a user served by the HAPS via RF link, and (10) is the achievable data rate of a user served by the HAPS subject to FSO backhaul constraint. Constraint (11c) guarantees that the HAPS can serve at most $K_0$ users, and constraint (11d) guarantees that a user can be served by one transmitter at most. Finally, constraint (11f) imposes maximal power constraints on both the HAPS and the ground BSs.

Problem (11) is a mixed-integer non-convex optimization problem. Finding the joint global optimal solution to the problem would require an exhaustive search of exponential complexity, i.e., finding the beamforming vectors for each potential user association, and then choosing the association with the highest throughput, which is unfeasible for any reasonably sized network. The paper, therefore, next solves the problem through an iterative approach. That is, first, for fixed beamforming vectors, the user association strategy is determined by linearizing the original problem so as to enable the utilization of integer linear programming, followed by a generalized assignment problem-type solution. Then, for fixed user association, the beamforming vectors are found through a series of problem reformulations that enable the use of WMMSE-type solutions.

### A. User Association Strategy

This part focuses on solving problem (11) over the user association strategy $\alpha_{ij}$ by fixing the beamforming vectors at the HAPS and the ground BSs. By first substituting the min term of (10) in the objective function, we rewrite problem (11) as the following binary optimization problem:

$$
\max_{\alpha_{ij}} \sum_{i=1}^{N_B} \sum_{j=1}^{N_i} \gamma_{ij} \alpha_{ij} R_{ij}^{\text{Ground\_BS}} + \sum_{j=1}^{N_i} \gamma_{0j} \alpha_{0j} \min \{R_{0j}^{\text{HAPS\_RF}}, R_{0j}^{\text{FSO}}\},
$$

where the optimization is over the binary variable $\alpha_{ij}$. Problem (12) remains, however, a complex non-linear discrete optimization problem, the global optimal solution of which would require an exhaustive search of exponential complexity. We next address such intricacies by first linearizing (12), and then adopting a GAP-based heuristic which proves to be an adequate numerical solution in the context of our problem formulation.

1) Integer Linear Problem Formulation: To linearize problem (12), we first replace the $\min\{\ldots\}$ term in (12) with an auxiliary variable $t_{0j}$ given by:

$$
t_{0j} = \min \{R_{0j}^{\text{HAPS\_RF}}, R_{0j}^{\text{FSO}}\}. \tag{13}
$$

In order to decouple the variables $\alpha_{ij}$ from (6) and linearize problem (12), we add one auxiliary additional constraint as follows:

$$
0 \leq w_{ij} H_{ij} w_{ij} \leq \gamma_{ij} \alpha_{ij} M_{ij}, \quad \forall i \in \mathcal{I}, \quad \forall j \in \mathcal{U}, \tag{14}
$$

where $M_{ij}$ is a sufficiently large constant, added as an artifact for linearizing (12). We note that $\alpha_{ij}$ is a binary variable. Constraint (14) can guarantee that if user $j$ is not served by transmitter $i$, the beamforming vector $w_{ij}$ must be zero vector. Therefore, subject to such bounding constraint, we rewrite (7) as:

$$
\hat{R}_{ij}^R = \beta \log_2 \left(1 + \frac{H_{ij}^H W_{ij} W_{ij}^H}{\sum_{u=1, u \neq j}^{N_B} H_{uj}^H W_{uj} W_{uj}^H + \sigma^2}\right). \tag{15}
$$

Problem (12) can, therefore, be reformulated as follows:

$$
\max_{\alpha_{ij}} \sum_{j=1}^{N_U} \gamma_{0j} \sum_{i=1}^{N_B} \left(\gamma_{ij} \alpha_{ij} \hat{R}_{ij}^R + \gamma_{0j} \alpha_{0j} t_{0j}\right), \tag{16a}
$$

subject to:

$$
\gamma_{0j} \alpha_{0j} \leq K_0, \quad \forall j \in \mathcal{U}, \tag{16b}
$$

$$
\sum_{j=1}^{N_U} \gamma_{ij} \alpha_{ij} \leq 1, \quad \forall j \in \mathcal{U}, \tag{16c}
$$

where $\gamma_{ij}$, $\alpha_{ij}$, and $\hat{R}_{ij}^R$ are defined in Table I.
The optimization problem (16) becomes an integer linear problem (ILP), which can be solved using off-the-shelf available algorithms, e.g., [23], [40]. ILP solvers, however, often provide suboptimal solutions to (16), and so we next improve upon the ILP solution by proposing an additional heuristic that exhibits appealing numerical prospects, as illustrated in the simulations section of the paper.

2) Integer Linear Problem and Generalized Assignment Problem (ILP-GAP): To further improve upon the ILP-based solution proposed above, the paper goes one step beyond by proposing an additional heuristic that relies on maximizing an auxiliary interference-free function of the original objective function of the optimization problem (7). Such heuristic allows us to use of the ILP-based solution as an initial point to solve a generalized assignment problem of reasonable computational complexity; see [22], [24], [41] and references therein. The simulation results of our paper later illustrate the numerical prospect of our proposed heuristic ILP-GAP scheme, as it outperforms the classical user association techniques.

More specifically, decouple the user association dependency by approximating the rate expression (7) with an interference-free term as:

$$R_{ij}^{RF} = \beta \log_2 \left(1 + \frac{|h_{ij}w_{ij}|^2}{\sigma^2} \right).$$

We now reformulate problem (12) as a GAP. More specifically, given the set of users $\mathcal{U}$ and the set of transmitters $\mathcal{I}$ (i.e., knapsacks), if the $j$th user associates with the $0$th knapsack (i.e., user $j$ is connected to the HAPS), the profit is $t_{0j}$. Otherwise, if the $j$th user associates with the $j$th knapsack ($i \neq 0$), the profit becomes $R_{ij}^{RF}$. Hence, problem (16) can be reformulated as follows:

$$\max_{\alpha_{ij}} \sum_{j=1}^{N_U} \sum_{i=1}^{N_B} \gamma_{ij} \alpha_{ij} R_{ij}^{RF} + \gamma_{0j} \alpha_{0j} t_{0j},$$

s.t. $\sum_{j=1}^{N_U} \gamma_{0j} \alpha_{0j} \leq K_0,$

$\sum_{i=0}^{N_B} \gamma_{ij} \alpha_{ij} \leq 1,$ $\forall j \in \mathcal{U},$

$\sum_{j=1}^{N_U} \gamma_{ij} \alpha_{ij} w_{ij}^H w_{ij} \leq P_i^{\max}, \quad \forall i \in \mathcal{I},$

$\alpha_{ij} \in \{0, 1\}, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{U},$  

where constraint (18b) is the HAPS payload connectivity constraint, constraint (18c) guarantees that a user can be assigned to one transmitter only, and constraint (18c) denotes the power constraints at every transmitter $i \in \mathcal{I}$.

The above problem (18) can be readily cast as a GAP [24], which can be solved using a handful of efficient algorithms. In this article, we utilize the branch and bound techniques for its provable performance guarantees [23]. The solution of GAP is numerically manageable, yet strongly dependable on the initialization strategy [22], [24], [41]. Our paper, therefore, first calculates the $t_{0j}$ and then adopts the solution reached by solving the ILP (16) as the initial point, owing to its good numerical prospects. We note that the variable $t_{0j}$ eventually gets updated after solving the above GAP. The steps of such iterative process, i.e., GAP and updating $t_{0j}$ (in this order), prove to be an efficient solution to solve the complicated user association problem (12) as shown in the simulations section and are summarized in Algorithm 1 description below.

**Algorithm 1: Determine the User Association.**

1) Fix the beamforming vectors $w_{ij} = \sqrt{\frac{P_i^{\max}}{N_B}} b_{ij}$.

2) Solve the integer linear problem (16) and update $\alpha_{ij}$ to get the initial point.

3) Set $m = 0$.

4) Use (13) to compute $t_{0j}^m$ and update $\alpha_{ij}$.

5) Define $R_{\text{optimization}} = R_{\text{initial}}$.

6) Set $m = m + 1$.

7) Solve the general assignment problem (18) and update $\alpha_{ij}^m$.

8) Calculate the sum rate $R_{\text{sum}}^m$ and $t_{0j}^m$. If $R_{\text{sum}}^m > R_{\text{optimization}}$, then $R_{\text{optimization}} = R_{\text{sum}}^m$.

9) Go to step 7 and stop at convergence (i.e., when $|R_{\text{sum}}^m - R_{\text{sum}}^{m-1}| \leq \epsilon$).

B. Beamforming Vectors Optimization

We now focus on finding the beamforming vectors by fixing the user association variables $\alpha_{ij}$, which are determined in the previous subsection. Problem (11) can now be rewritten as:

$$\max_{w_{ij}} \sum_{i=1}^{N_B} \sum_{j=1}^{N_U} \gamma_{ij} \alpha_{ij} R_{ij}^{\text{Ground\_BS}}$$

$$+ \sum_{j=1}^{N_U} \gamma_{0j} \alpha_{0j} \min \{R_{ij}^{\text{HAPS\_RF}}, R_{ij}^{\text{FSO}}\},$$

s.t. $\sum_{j=1}^{N_U} \gamma_{ij} \alpha_{ij} w_{ij}^H w_{ij} \leq P_i^{\max}, \quad \forall i \in \mathcal{I},$

where the optimization is over the beamforming vectors $w_{ij}$ (i.e., $i = 0$ for HAPS, $i = 1, 2, \ldots, N_B$ for ground BSs). The above problem (19) is a non-convex optimization problem due to the cross-mode cross-layered interference coupling in the SINR’s expressions, as well as the min term stemming from the FSO backhaul constraints. The paper next tackles the difficulties of problem (19) by proposing a modified version of
WMMSE [27] that best accounts for the current problem physical constraints.

1) WMMSE Reformulation: We first note that the minimum term in the optimization objective in (19) makes our problem different from the classical WMMSE formulation [27]. We, therefore, next provide a series of problem reformulations with proper outer loops updates, so as to develop a WMMSE-like solution for solving problem (19).

First, based on the values of αij determined in the previous subsection, one can readily determine the set of users served both the HAPS (i = 0), and the set of users served both the ground BSs (i = 1, . . . , N_B). To this end, we define \( U_i = \{ j \in U \mid \gamma_{ij} \alpha_{ij} = 1 \} \) as the set of users served by transmitter i (i = 0 for HAPS, i = 1, 2, . . . , N_B for ground BSs). Problem (19) can now be reformulated as:

\[
\max_{\mathbf{w}_i} \sum_{i \in \mathcal{I}} \sum_{j \in U_i} R^\text{Ground BS}_i \sum_{j \in \mathcal{I}} \min\{ R^\text{HAPS-RF}_i, R^\text{FSO}_i \},
\]

s.t. \( \sum_{j \in \mathcal{I}} w_{ij}^H w_{ij} \leq P^\text{max}_i, \forall i \in \mathcal{I}. \) (20a)

(20b)

Then, introduce an auxiliary variable \( \tau_{ij} \) defined as:

\[
\tau_{ij} = \min\{ R^\text{HAPS-RF}_i, R^\text{FSO}_i \}, \forall j \in U_0.
\]

\( \tau_{ij} \) can, therefore, be written as:

\[
\tau_{ij} = \begin{cases} R^\text{FSO}_i, & R^\text{FSO}_i \leq R^\text{HAPS-RF}_i, \\ R^\text{HAPS-RF}_i, & R^\text{FSO}_i > R^\text{HAPS-RF}_i. \end{cases}
\]

(22)

We now introduce another auxiliary variable \( \lambda_{ij} \), which can be regarded as the weight of the rate-terms of user \( j \) served by the transmitter \( i \), i.e., \( \tau_{ij} \), within the objective function of problem (19). For \( i \neq 0 \) (i.e., in the case of ground BSs), \( \lambda_{ij} = 1 \). For \( i = 0 \) (i.e., in the case of HAPS), \( \lambda_{ij} \) can be defined as:

\[
\lambda_{ij} = \begin{cases} 1, & \tau_{0j} = R^\text{HAPS-RF}_j, \\ 0, & \tau_{0j} = R^\text{FSO}_j. \end{cases}
\]

(23)

We note that the above equation (23) is mainly due to the fact that if \( \tau_{0j} \) is equal to the constant FSO link rate, the optimization problem would no longer depend on the value of \( w_{0j} \), and so we can omit such constant from the objective function. Problem (20) can now be re-written as follows:

\[
\max_{\mathbf{w}_{ij}} \sum_{i \in \mathcal{I}} \sum_{j \in U_i} \lambda_{ij} R_{ij} + \sum_{j \in U_0} \lambda_{0j} \tau_{0j},
\]

s.t. \( \sum_{j \in \mathcal{I}} w_{ij}^H w_{ij} \leq P^\text{max}_i, \forall i \in \mathcal{I}. \) (24a)

(24b)

At this stage, we note that our problem reformulation (24) now emulates, to some extent, a sum-rate maximization problem subject to transmit power constraints, i.e., similar to the classical WMMSE formulation [27]. In the context of our paper, problem (20) has the equivalent optimal solution with the following WMMSE minimization problem:

\[
\min_{\mathbf{r}_{ij}, \mathbf{w}_{ij}, \mathbf{w}_{ij}} \sum_{i \in \mathcal{I}} \sum_{j \in U_i} \lambda_{ij} (\text{Tr}(\mathbf{r}_{ij} \mathbf{e}_{ij}) - \log \rho_{ij}),
\]

s.t. \( \sum_{j \in U_i} \mathbf{w}_{ij}^H \mathbf{w}_{ij} \leq P^\text{max}_i, \forall i \in \mathcal{I}, \) (25a)

\[
\sum_{j \in \mathcal{I}} w_{ij}^H w_{ij} \leq P^\text{max}_i, \forall i \in \mathcal{I}. \] (25b)

where \( P^\text{max}_i \) is the maximum power of BS \( i \), \( \rho_{ij} \) denotes the mean squared error (MSE) weight for user \( j \) served by transmitter \( i \) (i.e., \( \forall j \in U_i \)), and \( \mathbf{u}_{ij} \) is the receive beamforming vector at the user \( j \) when served by transmitter \( i \). Finally, \( \mathbf{e}_{ij} \) is the MSE at the user \( j \) when served by transmitter \( i \) defined as:

\[
\mathbf{e}_{ij} = (\mathbf{I} - \mathbf{u}_{ij}^H \mathbf{h}_{ij}^H)(\mathbf{I} - \mathbf{u}_{ij}^H \mathbf{h}_{ij}^H)^H + \sum_{(b,l) \neq (i,j)} \mathbf{u}_{bl}^H \mathbf{h}_{bl} \mathbf{h}_{lj}^H + \sigma^2 \mathbf{u}_{ij}^H \mathbf{u}_{ij}, \forall i \in \mathcal{I},
\]

\[
\forall j \in U_i. \] (26)

2) Beamforming Algorithm (Algorithm 2): The reformulated problem (25) is convex in each of the optimization variables \( \rho_{ij}, \mathbf{u}_{ij}, \mathbf{w}_{ij} \). Therefore, one can solve (25) via finding one variable by fixing two other variables. More specifically, \( \forall j \in \mathcal{I}, \forall i \in \mathcal{U}_i \), the optimal receiver \( \mathbf{u}_{ij} \) under fixed \( \mathbf{w}_{ij} \) and \( \rho_{ij} \) is an MMSE receiver defined by:

\[
\mathbf{u}_{ij} = \mathbf{u}_{ij}^{\text{mmse}} = \frac{\mathbf{h}_{ij}^H \mathbf{w}_{ij}}{\sum_{b \in \mathcal{I}} \sum_{l \in U_b} \mathbf{h}_{bl}^H \mathbf{w}_{bl} \mathbf{h}_{lj}^H + \sigma^2}.
\]

(27)

Similarly, the optimal MSE weight \( \rho_{ij} \) under fixed \( \mathbf{u}_{ij} \) and \( \mathbf{w}_{ij} \) can be written as:

\[
\rho_{ij} = \mathbf{e}_{ij}^{-1}, \forall i \in \mathcal{I}, \forall j \in \mathcal{U}_i.
\]

(28)

Lastly, finding the optimal transmit beamformer \( \mathbf{w}_{ij} \) under fixed \( \rho_{ij}, \mathbf{u}_{ij} \) can be cast as a convex quadratic optimization problem, which can be solved efficiently [42]. The above updates of \( \rho_{ij}, \mathbf{u}_{ij}, \mathbf{w}_{ij}, \) i.e., (26)–(28), are eventually executed in an iterative way together with the proper updates of \( \tau_{ij} \) and \( \lambda_{ij} \) according to (22) and (23), respectively, which enables finding the beamforming vectors \( \mathbf{w}_{ij} \) efficiently, as presented in Algorithm 2. Such algorithm is in fact guaranteed to converge to a stationary point of (20), as further illustrated in the next lemma.

Lemma 1: The solution obtained by Algorithm 2 converges to a stationary point of (20).

Proof: The steps of the proof of Lemma 1 are included in Appendix A of the paper.

\[ \blacksquare \]

C. Overall Algorithm and Convergence

Now that both the discrete and continuous variables of problem (11) are determined, as per Algorithms 1 and 2, respectively, the paper adopts an iterative algorithm to optimize both variables alternatively. Specifically, the solution involves three loops: two inner loops and one outer loop. The first inner loop solves the user association strategy, and the second inner loop updates the beamforming vectors at HAPS and ground BSs. The outer loop, finally, combines two inner loops to optimize user association and beamforming. Since each of the two loops provides a non-decreasing function in the network sum-rate (which is bounded
Algorithm 2: Determine Beamforming Vectors.

1) Fix the user association variables.
2) set \( m = 0 \).
3) Fix the initial beamforming vectors \( \mathbf{w}_{ij} = \frac{\sqrt{P^{\text{max}}}}{N_A} \mathbf{h}_{ij}, \)
4) Calculate the \( \tau_{ij}^{m} \) and determine the \( \lambda_{ij}^{m} \), according to (22), (23).
5) Fix \( \mathbf{w}_{ij} \), and update \( \mathbf{u}_{ij} \), according to (27).
6) Fix \( \mathbf{u}_{ij} \) and \( \mathbf{w}_{ij} \), and update \( \rho_{ij} = \mathbf{e}_{ij}^{-1} \).
7) Calculate and update the optimal transmit beamformer \( \mathbf{w}_{ij} \) under fixing \( \rho_{ij}, \mathbf{u}_{ij} \).
8) Compute the sum-rate \( R^{m}_{\text{sum}} \).
9) set \( m = m + 1 \).
10) Go to step 4 and stop at convergence (i.e., when \( |R^{m}_{\text{sum}} - R^{m-1}_{\text{sum}}| \leq \epsilon \)).

Algorithm 3: Overall Algorithm.

1) Generate initial beamforming vectors \( (\mathbf{w}_{ij}) \).
2) Repeat
3) Fix the beamforming vector of BSs and HAPS.
4) Implement Algorithm 1 to update the \( \alpha_{ij} \).
5) Implement Algorithm 2 to update the beamforming vectors of all users.
6) Compute the sum-rates \( R_{\text{sum}} \) of the network.
7) Stop at convergence.

by the network capacity), the overall algorithm is guaranteed to converge, as also validated later through the paper simulations. The steps of the overall algorithm are shown in Algorithm 3 description below.

D. Computational Complexity

To best characterize the computational complexity of the proposed algorithm, we note that Algorithm 3 solves two problems sequentially. The user association problem can be solved by an integer linear program and a GAP-based algorithm. Linear programming has a polynomial-time solvable computational complexity \( O(n^k) \), where \( n \) is the number of variables \( \alpha_{ij} \) and \( k \) is the degree of complexity. Our paper adopts a GAP-solver based on branch-and-bound (BnB) solutions [23], and so the GAP step complexity is in the order of \( O(\xi^n) \), where \( 1 < \xi < 2 \). The beamforming solution, on the other hand, relies on WMMSE [27], the per-iteration computational complexity of which is upper-bounded by \( O(N_U^2 N_A + N_U^2 N_A^2 + N_U^2 N_A^3 + N_U) \), where \( N_A \) is the maximum number of antennas across both HAPS and ground BSs.

To summarize, our proposed algorithm includes three loops. When the numbers of iterations of the two inner loops are \( T_1 \) and \( T_2 \), respectively, and that of the outer loop is \( T_3 \), the computational complexity of the overall algorithm becomes \( T_3 [O(n^k) + T_1 O(\xi^n) + T_2 O(N_U^2 N_A + N_U^2 N_A^2 + N_U^2 N_A^3 + N_U)] \), which is reasonably dependable on the particular GAP solution complexity.

IV. SIMULATION RESULTS

This section evaluates the performance of the proposed algorithm for various network scenarios, so as to illustrate the numerical gains of the developed joint user association and beamforming optimization framework in the context of integrated satellite-HAPS-ground networks. We next present the baseline and simulation results with different parameters and network scenarios.

A. Baseline Approaches

As mentioned above, the major complexity of the proposed solution originates from the user association strategy, especially the GAP algorithm. To this end, the paper now presents two alternative low-complexity methods that depend on the distance and channel values, respectively. Similar approaches can be found in [43].

1) Baseline 1 (Distance Dependent Approach): This method assigns user \( j \) to transmitter \( i \) (\( \forall i \in \mathcal{I} \) and \( j \in \mathcal{U} \)) based on their mutual distance, denoted by \( d_{ij} \). Let \( \mathbf{D} \) be \( (N_B + 1) \times N_U \) matrix whose entries, \( d_{ij} \), denote the distance between the transmitter \( i \) and the user \( j \), i.e., the \( (i,j) \)th entry of the matrix \( \mathbf{D} \) is \( d_{ij} = d_{ij} \). At each step, find the smallest entry of matrix \( \mathbf{D} \), call it \( \mathbf{D}_{\text{min}^{-}-\text{min}} \). User \( j^{\text{min}} \) then maps to transmitter \( \mathbf{D}_{\text{min}^{-}-\text{min}}\), as long as each ground BS does not serve more than its number of antennas and the HAPS payload connectivity constraint is satisfied. Next, delete the \( (\mathbf{D}_{\text{min}^{-}-\text{min}}) \)th column of the matrix, so that user \( j^{\text{min}} \) cannot be associated with other transmitters in subsequent steps. Repeat the above procedure until all users are connected to transmitters or all transmitters resource constraints (11c), (11d) are violated.

2) Baseline 2 (Channel Dependent Approach): Unlike the distance dependent approach, this method assigns users to the transmitters based on the channel gain between the transmitters and the users, denoted by \( c_{ij} = ||\mathbf{h}_{ij}||^2 \), \( \forall i \in \mathcal{I} \) and \( j \in \mathcal{U} \). Let \( \mathbf{C} \) be \( (N_B + 1) \times N_U \) matrix whose entries are the channel gains between transmitter \( i \) and user \( j \) denoted by \( c_{ij} \), i.e., the \( (i,j) \)th entry of the matrix \( \mathbf{C} \) is \( c_{ij} = c_{ij} \). At each step, find the largest entry of matrix \( \mathbf{C} \), call it \( C_{\text{max}^{-}-\text{max}} \). User \( j^{\text{max}} \) then maps to transmitter \( j^{\text{max}} \), as long as the resource constraints of transmitter \( j^{\text{max}} \) are satisfied. Next, delete the \( (\mathbf{C}_{\text{max}^{-}-\text{max}}) \)th column of the matrix, so that user \( j^{\text{max}} \) cannot be associated with other transmitters in subsequent steps. The procedure then gets repeated as in the distance dependent approach above.

In particular, the paper compares the proposed joint optimization solution adopting ILP and GAP (IG) as user association strategy and WMMSE as beamforming design method in high backhaul capacity (HBC-IG-WMMSE) to 8 different benchmarks: 1- joint optimization with channel-dependent (CD) and WMMSE in high backhaul capacity (HBC-CD-WMMSE), 2- joint optimization with distance-dependent (DD) and WMMSE in high backhaul capacity (HBC-DD-WMMSE), 3- joint optimization solution with ILP-GAP and WMMSE in low backhaul capacity (LBC-IG-WMMSE), 4- joint optimization with channel-dependent and WMMSE in low backhaul capacity (LBC-CD-WMMSE), 5- joint optimization with distance-dependent and WMMSE in low backhaul capacity (LBC-DD-WMMSE), 6- joint optimization with ILP-GAP and WMMSE in low backhaul capacity (LBC-CD-WMMSE), 7- joint optimization with distance-dependent and WMMSE in low backhaul capacity (LBC-DD-WMMSE), 8- joint optimization with ILP-GAP and WMMSE in low backhaul capacity (LBC-CD-WMMSE).
TABLE II
ALGORITHMS ABBREVIATION

| Algorithm                          | Definition                                                                 |
|------------------------------------|---------------------------------------------------------------------------|
| HBC-IG-WMMSE                       | Joint optimization with ILP-GAP (IG) and WMMSE in high backhaul capacity   |
| HBC CD-WMMSE                       | Joint optimization with channel-dependent (CD) and WMMSE in high backhaul capacity |
| HBC-DD-WMMSE                       | Joint optimization with distance-dependent (DD) and WMMSE in high backhaul capacity |
| LBC-IG-WMMSE                       | Joint optimization with ILP-GAP (IG) and WMMSE in low backhaul capacity     |
| LBC-CD-WMMSE                       | Joint optimization with channel-dependent (CD) and WMMSE in low backhaul capacity |
| LBC-DD-WMMSE                       | Joint optimization with distance-dependent (DD) and WMMSE in low backhaul capacity |
| HBC-IG                             | ILP-GAP (IG) approach in high backhaul capacity                           |
| HBC CD                             | Channel-dependent (CD) approach in high backhaul capacity                  |
| HBC DD                             | Distance-dependent (DD) approach in high backhaul capacity                 |

TABLE III
SIMULATION PARAMETERS

| Parameter Name                  | Parameter Value |
|--------------------------------|-----------------|
| The height of HAPS, $z_HAPS$    | 18 km           |
| The height of geo-satellite, $z_{satellite}$ | 36000 km |
| Bandwidth of BSs, $\beta$       | 10 MHz          |
| Central carrier frequency, $f_c$ | 3 GHz           |
| Rician factor, $k_{HAPS}$       | 5               |
| Noise power, $N_0$              | $-174$ dBm/Hz   |
| Ground BS antenna, $N^BS$       | 1               |
| HAPS antennas in medium network, $N^HAPS_{mid}$ | 20 |
| HAPS antennas in large network, $N^HAPS_{large}$ | 40 |
| Maximum power of urban BS, $P_{max,urban}^{BS}$ | 0 dBw |
| Maximum power of suburban BS, $P_{max,suburban}^{BS}$ | 3 dBw |
| Maximum power of rural BS, $P_{max,rural}^{BS}$ | 7 dBw |
| Maximum power of HAPS in mid network, $P_{max,HAPS}^{mid}$ | 20 dBw |
| Maximum power of HAPS in big network, $P_{max,HAPS}^{large}$ | 23 dBw |
| Standard deviation of ground-level shadowing, $\sigma_a$ | 5dB |

(LBC-DD-WMMSE), 6- ILP-GAP approach in high backhaul capacity (HBC-IG), 7- channel-dependent approach in high backhaul capacity (HBC-CD), and 8- distance-dependent approach in high backhaul capacity (HBC-DD). For completeness, we also summarize the above algorithms in Table II. Note that, in the context of the current paper, we define the low backhaul capacity regime to be in the order of $0 - 10^3$ bps, the medium backhaul capacity regime in the order of $10^3 - 10^7$ bps, and the high backhaul capacity regime in the order of $10^7$ bps, respectively.

B. Medium Networks Simulation

We first simulate a network of medium size with a ground footprint of $5 \times 5$ km, similar to Fig. 1. We herein assume that $N_U$ users are distributed in two different subareas. Subarea 1 contains 12 BSs with coordinates: $x$: (0 km to 1 km) and $y$: (0 km to 1 km) and contains 60% of the total number of users. The remaining area is the subarea 2 and contains 40% of the total number of users, with no deployed BS. In this case, subarea 1 can be considered as an urban area, while subarea 2 can be considered as a suburban area. The satellite is fixed at the coordinates [2.5, 2.5, 36000] km. The HAPS is also fixed at the coordinates [2.5, 2.5, 18] km. Table III presents the values of the parameters used in the simulations (unless mentioned otherwise), and highlights the difference between HAPS-to-user channels and ground BSs-to-user channels. The standard deviation of ground-level shadowing $\sigma_a$ is 5 dB. However, we omit the shadowing in HAPS-to-user channels, and consider Rician fading with factor $k_{HAPS} = 5$. The FSO-related parameters are adopted from [39]. For illustration purposes, the data-availability variables $\gamma_{ij}$ are set to 1 throughout the simulations section. Further, for illustration purposes, we set the maximum number of users that the HAPS can serve to the number of its corresponding antennas (i.e., $K_0 = N^A_0$).

We first illustrate the impact of the number of users on the network performance in Fig. 2, which plots the sum-rate versus the total number of users. Fig. 2 shows how the proposed solution provides a substantial gain compared to other algorithms, especially when the number of users increases. This is particularly the case since the interference level becomes larger in denser networks, and so the role of the proposed resource allocation scheme in mitigating interference becomes
shows the sum-rate versus the total number of users; our proposed joint optimization attains a considerable improvement at high backhaul capacity. This is due to the fact that, unlike low backhaul capacity regimes, high backhaul capacity regimes match a fully enabled HAPS, which unleashes the full power of the HAPS towards serving more users, thereby increasing the network total throughput. 

We can notice that the performance of HBC-DD-WMMSE is better than HBC-CD-WMMSE, when the number of users exceeds 40. This is due to the fact that the payload constraint of HAPS also only allows HAPS to associate with 20 users (i.e., the number of HAPS antennas). Therefore, when the number of users exceeds 40, only a subset of users can be served. Further, due to the location of HAPS, the distance-dependent user association solution gives priority to map users in the rural area with HAPS. However, given that the channel-dependent user association strategy is based on channel gains, the users in the rural area are no longer preferred by the HAPS. Hence, compared with HBC-CD-WMMSE, the HBC-DD-WMMSE solution maps more users in the rural area and fewer users in the urban area to HAPS, in an effort to mitigate the interference level. However, at low backhaul capacity, because of the limitation of the FSO link, the HAPS are idle. The sum-rate becomes, therefore, decided by the users served by ground BSs. In this article, we consider shadowing and rayleigh small scale fading between users and ground BSs, which makes LBC-DD-WMMSE inferior to HBC-CD-WMMSE.

In addition, Fig. 3 shows the user behavior in terms of their association with the HAPS and ground BSs, when the number of users is 50. Notably, based on Fig. 3, we observe that within the urban areas, 3 users are catered to by ground BSs, while 13 users in the suburban regions are served by HAPS. The remaining 34 users in this case are not served. Consequently, the average rate allocated per active user amounts to 1200 Mbps divided by 16, i.e., 75 Mbps. Since the adopted bandwidth is 10 MHz, the achievable data rate expression per user, i.e., \( 75 = 10 \log_2(1 + \text{SINR}) \), yields an average achievable SINR of 22.57 dB.

To further highlight the performance of our proposed algorithms, we now add the zero-forcing (ZF) beamforming as an additional beamforming design baseline. In this context, ZF beamforming is implemented on a per-transmitter basis, i.e., it helps to eliminate the intra-mode interference [44], namely, intra-HAPS interference, and intra-BS interference in the current paper setup. Fig. 4 shows the sum-rate versus the total number of users, when the number of HAPS antennas is 50. Since ZF beamforming cannot handle inter-mode interference, when the number of users increases, we can notice that the sum-rate relying on ZF beamforming increases firstly and then decreases. The rationale behind this algorithmic behavior is that when the number of users initially increases, the BSs tend to serve users in the urban area, while HAPS tends to serve users in the rural area. In this case, the intra-mode interference level becomes dominant, and so ZF beamforming depicts an increasing rate trend. However, as the number of users continues to grow, HAPS starts serving more users in the urban area, which makes the inter-mode interference more dominant, thereby causing the ZF beamforming strategy to become less efficient. Having said that, we note that Fig. 4 shows that our proposed WMMSE approach always outperforms the ZF beamforming design strategies, because of the proposed approach ability to mitigate both inter-mode interference and intra-mode interference.

In order to illustrate the computational complexity of our proposed algorithms, Table IV presents the respective runtimes of two algorithms, namely HBC-IG-WMMSE and HBC-IG-ZFBF for different numbers of users and different numbers of HAPS antennas. The runtime results in Table IV are generated using MATLAB R2022b based on an Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50 GHz processor. As indicated in the table, the general trend of the algorithms is that as the number of users and the number of antennas increase, the algorithms runtimes also increase, as their respective computational complexities also increase. Further, Table IV shows that our proposed joint optimization algorithm, i.e., HBC-IG-WMMSE, which involves an iterative implementation of integer programming, generalized assignment problem solution, and a WMMSE-based solution displays a longer runtime due to its relatively higher computational complexity as compared to the algorithm that uses ZFBF for setting the beamforming vectors. This is because the complexity of the WMMSE step is higher than ZFBF. HBC-IG-WMMSE runtime, however, comes with a noticeable superior gain in sum-rate performance, as illustrated earlier in Fig. 4.

Judging from Fig. 2, when the number of users is 50, one can notice that the network resources are fully utilized. Therefore, we

more pronounced. Fig. 2 also shows how the joint optimization attains a considerable improvement at high backhaul capacity. This is due to the fact that, unlike low backhaul capacity regimes, high backhaul capacity regimes match a fully enabled HAPS, which unleashes the full power of the HAPS towards serving more users, thereby increasing the network total throughput. We can notice that the performance of HBC-DD-WMMSE is better than HBC-CD-WMMSE, when the number of users exceeds 40. This is due to the fact that the payload constraint of HAPS also only allows HAPS to associate with 20 users (i.e., the number of HAPS antennas). Therefore, when the number of users exceeds 40, only a subset of users can be served. Further, due to the location of HAPS, the distance-dependent user association solution gives priority to map users in the rural area with HAPS. However, given that the channel-dependent user association strategy is based on channel gains, the users in the rural area are no longer preferred by the HAPS. Hence, compared with HBC-CD-WMMSE, the HBC-DD-WMMSE solution maps more users in the rural area and fewer users in the urban area to HAPS, in an effort to mitigate the interference level. However, at low backhaul capacity, because of the limitation of the FSO link, the HAPS is idle. The sum-rate becomes, therefore, decided by the users served by ground BSs. In this article, we consider shadowing and rayleigh small scale fading between users and ground BSs, which makes LBC-DD-WMMSE inferior to LBC-CD-WMMSE.

In addition, Fig. 3 shows the user behavior in terms of their association with the HAPS and ground BSs, when the number of users is 50. Notably, based on Fig. 3, we observe that within the urban areas, 3 users are catered to by ground BSs, while 13 users in the suburban regions are served by HAPS. The remaining 34 users in this case are not served. Consequently, the average rate allocated per active user amounts to 1200 Mbps divided by 16, i.e., 75 Mbps. Since the adopted bandwidth is 10 MHz, the achievable data rate expression per user, i.e., \( 75 = 10 \log_2(1 + \text{SINR}) \), yields an average achievable SINR of 22.57 dB.

To further highlight the performance of our proposed algorithms, we now add the zero-forcing (ZF) beamforming as an additional beamforming design baseline. In this context, ZF beamforming is implemented on a per-transmitter basis, i.e., it helps to eliminate the intra-mode interference [44], namely, intra-HAPS interference, and intra-BS interference in the current paper setup. Fig. 4 shows the sum-rate versus the total number of users, when the number of HAPS antennas is 50. Since ZF beamforming cannot handle inter-mode interference, when the number of users increases, we can notice that the sum-rate relying on ZF beamforming increases firstly and then decreases. The rationale behind this algorithmic behavior is that when the number of users initially increases, the BSs tend to serve users in the urban area, while HAPS tends to serve users in the rural area. In this case, the intra-mode interference level becomes dominant, and so ZF beamforming depicts an increasing rate trend. However, as the number of users continues to grow, HAPS starts serving more users in the urban area, which makes the inter-mode interference more dominant, thereby causing the ZF beamforming strategy to become less efficient. Having said that, we note that Fig. 4 shows that our proposed WMMSE approach always outperforms the ZF beamforming design strategies, because of the proposed approach ability to mitigate both inter-mode interference and intra-mode interference.

In order to illustrate the computational complexity of our proposed algorithms, Table IV presents the respective runtimes of two algorithms, namely HBC-IG-WMMSE and HBC-IG-ZFBF for different numbers of users and different numbers of HAPS antennas. The runtime results in Table IV are generated using MATLAB R2022b based on an Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50 GHz processor. As indicated in the table, the general trend of the algorithms is that as the number of users and the number of antennas increase, the algorithms runtimes also increase, as their respective computational complexities also increase. Further, Table IV shows that our proposed joint optimization algorithm, i.e., HBC-IG-WMMSE, which involves an iterative implementation of integer programming, generalized assignment problem solution, and a WMMSE-based solution displays a longer runtime due to its relatively higher computational complexity as compared to the algorithm that uses ZFBF for setting the beamforming vectors. This is because the complexity of the WMMSE step is higher than ZFBF. HBC-IG-WMMSE runtime, however, comes with a noticeable superior gain in sum-rate performance, as illustrated earlier in Fig. 4.

Judging from Fig. 2, when the number of users is 50, one can notice that the network resources are fully utilized. Therefore, we
we now use $N_U = 50$ so as to better characterize the impact of backhaul capacity on the network sum-rate performance, by plotting the sum-rate versus the backhaul capacity in Fig. 5. The figure shows that our proposed solution always attains the highest sum-rate as compared to all other classical strategies. The figure also illustrates the significant gap between joint optimization and user association strategy, which highlights the importance of the beamforming step in mitigating the interference, i.e., beyond the initial user association step. When the backhaul capacity exceeds the data rate of the RF link, the $\min\{\ldots\}$ term in (10) becomes equal to the data rate of the RF link. This is the reason why the sum-rate becomes invariant at the high backhaul regime. We also can notice that the proposed user association strategy does well in the lowest backhaul capacity. Due to the FSO backhaul constraint (10), when the data rate of the FSO is as low as 0, the user can reasonably not choose the HAPS to connect to, which is depicted through the behavior of the ILP-GAP algorithm. However, the users association strategies of the two baseline algorithms, i.e., DD and CD, are only based on distance and channel gains, so users may still choose to associate with HAPS at low backhaul capacity, which introduces high interference to the RF network and, at the same time, exacerbates the network sum-rate.

To illustrate the impact of the HAPS antennas on the network performance, Fig. 6 plots the sum-rate versus the total number of HAPS antennas with $N_U = 50$. The figure shows that, as the number of antennas increases, the sum-rate resulting from the solutions that rely on user association only slightly decreases, while the sum-rate resulting from the solutions which implement the additional beamforming optimization step increases. This is because if the HAPS has more antennas, the HAPS can serve more users, which introduces more interference. Since the user association strategy can not alone reduce the high level of interference stemming from the HAPS newly deployed antennas on its own, the sum-rate slightly decreases. However, the additional beamforming optimization step can significantly mitigate the interference due to the empowered spatial multiplexing capabilities. Fig. 6 further shows that the proposed joint approach always outperforms all other baseline solutions for all the simulated scenarios. The figure particularly shows the gain harvested through augmenting the ground networks with HAPS capabilities, which is shown through the substantial gain at high backhaul capacity (i.e., when the rate of the HAPS to ground users RF link is inferior to the FSO link capacity) as compared to the low backhaul capacity (i.e., when the HAPS potential is rather limited by the FSO link). In fact, at the low backhaul capacity, increasing the number of antennas does not change the sum-rate as the HAPS remains idle in this case. Specifically, $R_H^0_{APS}$ is limited by FSO link rate $R_{FSO}$, which is relatively small. Therefore, if HAPS serves users at low backhaul capacity, the contribution to the network sum-rate becomes negligible, and so the optimization rather favors connecting the users with ground BSs instead, i.e., regardless of the number of antennas at the HAPS. On the opposite, at the high backhaul capacity regime, the active operation of the HAPS becomes a major driver in pushing the network throughput upward.

Fig. 7 shows the sum-rate of the network versus the maximum power of the HAPS, when the number of HAPS antennas is set to $N_A^0 = 40$ and the number of users is 50. We can notice that the sum-rate of the baseline algorithms that do not apply the WMMSE step decreases when the maximum power of HAPS increases. Since the beamforming vectors are not optimized under such a realm, the interference including intra-HAPS interference and HAPS-BS interference increases while increasing the

---

**TABLE IV**

| Scheme 1 ($N_U^0 = 20$) | $N_U = 20$ | $N_U = 40$ | $N_U = 60$ |
|------------------------|-----------|-----------|-----------|
| HBC-IG-WMMSE           | 34.3431s  | 44.8239s  | 137.5462s |
| HBC-IG-ZFBF            | 24.7392s  | 36.3879s  | 135.6578s |

| Scheme 2 ($N_U^0 = 40$) | $N_U = 20$ | $N_U = 40$ | $N_U = 60$ |
|------------------------|-----------|-----------|-----------|
| HBC-IG-WMMSE           | 35.7519s  | 74.1602s  | 174.0206s |
| HBC-IG-ZFBF            | 25.8201s  | 64.3281s  | 164.6160s |
power of the HAPS, which decreases the overall sum-rate. Such baseline algorithms, in fact, cannot handle the high interference regimes well, which further emphasizes the significance of the joint optimization approaches proposed in this article. Not only does Fig. 7 reaffirm the superiority of our proposed joint optimization, but the figure also illustrates how the additional power capability at the HAPS helps increase the network sum-rate. This is because when the maximum power of HAPS increases, the number of users served by the HAPS increases. Given the strong capability of the proposed scheme at mitigating the cross-mode cross-layered interference, the sum-rate does indeed increase with the joint optimization scheme. Such a result is further highlighted by depicting the fraction of users served by the HAPS out of the total number of users (denoted by $\delta$) in the high backhaul capacity as shown in Fig. 8. The figure shows that as the maximum power of HAPS increases, the number of users served by the HAPS increases, which reflects the ability of HAPS to serve more users in the high-power regime. However, as the power of the HAPS increases from 30 dBw to 40 dBw in Fig. 7, we can clearly see a slowdown in the growth rate of the network sum-rate. This is because the HAPS payload constraint limits the number of users served by the HAPS. In addition, the induced multi-mode multi-layered interference imposes an upper bound on the achievable sum-rate of the network under study. The figure further shows that the greater the ground-level shadowing is, the more users the HAPS would serve. This is because when the ground-level shadowing increases, the gain brought by the BS connecting the user to the network decreases, and so the users are more inclined to be served by the HAPS. Likewise, if the HAPS is equipped with more antennas, more users tend to be served by the HAPS, which explains the capacity boost when the HAPS has 40 antennas. Fig. 8 is indeed a crisp illustration of how HAPS helps serving users in both urban and suburban areas; thereby highlighting HAPS roles in connecting the unconnected (through strong HAPS capabilities), and super-connecting the connected (at higher interference levels).

Further, Fig. 9 illustrates the sum-rate of the network versus the maximum power of the HAPS in the low backhaul capacity regime. On the one hand, Fig. 9 first illustrates how the sum rate relying on our proposed solutions (i.e., LBC-IG and LBC-IG-WMMSE, LBC-CD-WMMSE, LBC-DD-WMMSE) remains constant as the maximum power of the HAPS increases, which is due to the fact that the HAPS refrains from serving users when such service is not beneficial to the network-wide sum-rate. On the other hand, the sum-rate resulting from the solutions that rely on channel and distance-based association methods only, i.e., LBC-CD and LBC-DD, decreases as the power of the HAPS increases. This is due to the fact that, unlike our proposed user association algorithm, LBC-CD and LBC-DD employ a channel and distance-based association method irrespective of the FSO link quality. The low quality of the low backhaul capacity regime in this case, therefore, leads to a reduction in the overall network sum-rate as more users served by the HAPS would rather negatively impact the system performance.

### C. Large Networks Simulation

We next simulate a network of large size with a ground footprint of $30 \, \text{km} \times 30 \, \text{km}$ as shown in Fig. 10. In this area, $N_U$ users are distributed in three different subareas. Subarea 1 contains 60 BSs with coordinates: $x$: (0 km to 5 km) and $y$: (0 km to 5 km) and contains 60% of the total number of users. Subarea 2 contains 30 BSs with coordinates: $x$: (25 km to 30 km) and $y$: (25 km to 30 km) and contains 30% of the total number of users. The remaining area is the subarea 3 and contains 10% of the
Fig. 10. Layout of the big network.

Fig. 11. Sum-rate versus users.

Fig. 12. Sum-rate versus maximum power of HAPS.

Fig. 10 shows the layout of the big network. The total number of users and 8 BSs. In this case, subarea 1 can be considered as an urban area. While subarea 2 can be considered as a suburban area, subarea 3 can be considered as a rural area. The geo-satellite and HAPS locations remain at the center of the network as before.

Fig. 11 shows the sum-rate versus the total number of users. It is observed that the proposed solution outperforms all other approaches, especially when the number of users increases. Fig. 11 also shows that joint optimization methods are superior to those algorithms with only user association. Note that the utility of the network in the high backhaul capacity is better than the low backhaul capacity, which further indicates the positive impact of HAPS on large networks throughput. Fig. 12 shows the sum-rate versus the maximum power of HAPS with $N_U = 200$. Similar to Figs. 7, 12 shows that when the maximum power of the HAPS increases, the sum-rate resulting from the joint optimization increases, while the sum-rate resulting from user association only decreases. It is particularly noticeable how the proposed algorithm can bring more significant improvement to large networks than medium networks, mainly due to the higher interference levels. In fact, when the maximum power of HAPS is 30 dBw, Fig. 7 shows that the proposed algorithm can improve the network performance by 22.7%. For large networks, the network performance can be improved by 25.5%, which indicates that the proposed algorithm can improve the network performance, particularly in ultra-dense networks. Fig. 13 shows a graphical illustration of how ground users association changes as the maximum power of the HAPS increases. The figures show that when the power of HAPS is very low, users in urban and suburban areas tend to be served by BSs in Fig. 13(a), while HAPS serves only one user in the rural area. As the power of HAPS increases to 10dBw in Fig. 13(b), HAPS starts serving more users in the rural area, which shows how HAPS help connect the unconnected. As the power increases further in Fig. 13(c), the HAPS starts further serving more users from within the urban area, which is an example of how HAPS ultra-connects the connected.

D. Discussion and Recommendation

Based on the above results, it can be seen that our proposed algorithm provides a superior sum-rate performance as compared to conventional baselines. This is particularly true at the high FSO backhaul capacity regime, and under strong HAPS capabilities (e.g., high transmit power, large number of antennas), where the HAPS offers the prospects of assisting both the unconnected and the connected; thereby improving both rural and metropolitan networks performance. A major sailing outlook of the current paper is that the considered high-speed FSO backhaul...
The adopted FSO path, together with the general advent in HAPS design [11], make the paper results of particular importance in high-demand communication scenarios, e.g., concerts, sports events, etc., as well in the general context of boosting ground level future 6G networks performance.

V. CONCLUSIONS AND FUTURE WORK

Digital inclusion is nowadays celebrated as one of the major drivers towards defining 6G communications networks sustainable architectures. Along this direction, this article proposes an integrated satellite-HAPS-ground network consisting of one geo-satellite, one HAPS, and several ground BSs that collaboratively aim at connecting the unconnected and ultra-connecting the connected. The paper focuses on maximizing the network-wide sum-rate utility, subject to HAPS payload connectivity constraint, HAPS and BSs transmit power constraints, and FSO backhaul constraints, so as to jointly determine the user-association strategy of each user, and the beamforming vectors at the HAPS and BSs. The paper tackles such a complex mixed discrete-continuous optimization problem using an iterative approach, where the user association is determined using a combination of linear integer programming and generalized assignment problems, and where the beamforming strategy is found using a WMMSE approach. The paper results illustrate the appreciable gain of our proposed algorithm, and particularly highlight the numerical prospects of augmenting the ground networks with HAPS for connecting the unconnected (through strong HAPS capabilities), and super-connecting the connected (at the high interference regime), which give promising performance projections about vertical heterogeneous networks prospects. Future research directions in the field include accounting for imperfect channel information at both the HAPS and the ground BSs, evaluating the data-driven optimization approach to solve the problem under consideration in an online fashion, investigating the multi-HAPS scenario, where users are served...
by HAPs directly. Other research extensions of the current work would also include optimizing the end-to-end delay of the system by factoring in the impact of the gateway to geostationary feeder link latency and performance analysis including outage probability, coverage, spot beam, which promises to be a promising future research direction that falls at the intersection of communication, networking, and computing.

APPENDIX A
PROOF OF LEMMA 1

Proof: Similar to the steps of [27], we first rewrite the RF rate (7) as:

\[ R_{ij}^{RF} = \log \det \left( (e_{ij}^{mse})^{-1} \right), \forall i \in I, \forall j \in U_i, \quad (29) \]

where \( e_{ij}^{mse} \) is the associated MSE covariance matrix.

The objective of problem (24) (and equivalently (20)) can therefore be rewritten as follows:

\[ f_1(w_{ij}) = \sum_{i \in I} \sum_{j \in U_i} \lambda_{ij} \det \left( (e_{ij}^{mse})^{-1} \right) + \sum_{j \in U_i} \lambda_{0j} \tau_{0j}. \quad (30) \]

Similarly, the objective function of the equivalent problem (25) can be rewritten as:

\[ f_2(\rho_{ij}, u_{ij}, w_{ij}) = \sum_{i \in I} \sum_{j \in I} \lambda_{ij} \left( \text{Tr}(\rho_{ij} e_{ij}) - \log \det(\rho_{ij}) \right). \quad (31) \]

Since problem (25) is further differentiable, and since its constraints set is separable in the variables \( \rho_{ij}, u_{ij}, w_{ij} \), iteratively solving for one of the variables while fixing the two others, i.e., using a block coordinate descent approach, is guaranteed to converge to a stationary point [27]. To finalize the proof, it suffices to show that the stationary point of (25) is the same as the stationary point of (24) (and equivalently (20)), and that the converse is true.

Since the variables \( \rho_{ij}, u_{ij} \) are unconstrained, their respective first-order optimality conditions yield optimal \( \rho_{ij}^*, u_{ij}^* \) with expressions similar to (27) and (28), i.e.:

\[ u_{ij}^* = u_{ij}^{mse}, \quad \rho_{ij}^* = (e_{ij}^{mse})^{-1}, \forall i \in I, \forall j \in U_i. \quad (32) \]

Let \( w_{ij,r} \) be the \( r \)th entry of the vector \( w_{ij} \), we get:

\[ \frac{\partial f_2(\rho_{ij}^*, u_{ij}^*, w_{ij})}{\partial w_{ij,r}} = \sum_{i \in I} \sum_{j \in U_i} \lambda_{ij} \text{Tr} \left( (e_{ij}^{mse})^{-1} \right) \frac{\partial e_{ij}^{mse}(w_{ij})}{\partial w_{ij,r}} + \sum_{j \in U_i} \lambda_{0j} \sum_{r \in U_r} \frac{\partial \tau_{0j}}{\partial w_{0j,r}} \quad (33) \]

where the second term of the equality (33) can be further developed as:

\[ \lambda_{0j} \frac{\partial \tau_{0j}}{\partial w_{0j,r}} = \begin{cases} \text{Tr} \left( (e_{0j}^{mse})^{-1} \frac{\partial e_{0j}^{mse}(w_{0j})}{\partial w_{0j,r}} \right), & \tau_{0j} = \bar{R}_{0j}^{HAPSF}, \\
0, & \tau_{0j} = \bar{R}_{0j}^{FSO}. \end{cases} \quad (35) \]

The converse follows a reversely traversed equality path, which proves Lemma 1.

REFERENCES

[1] S. Liu, H. Dahrourj, and M.-S. Alouini, “Joint user association and beamforming in integrated Satellite-HAPS-Ground networks,” 2022, arXiv:2204.13257.
[2] N. Saeed, H. Almorad, H. Dahrourj, T. Y. Al-Naffouri, J. S. Shamma, and M.-S. Alouini, “Point-to-point communication in integrated satellite-terrestrial 6G networks: State-of-the-art and future challenges,” IEEE Open J. Commun. Soc., vol. 2, pp. 1505–1525, 2021.
[3] M. S. Alam, G. K. Kurt, H. Yankomeroglu, and P. Zhu, “High altitude platform station based super macro base station constellations,” IEEE Commun. Mag., vol. 59, no. 1, pp. 103–109, Jan. 2021.
[4] A. K. Widjawan and R. Tafazoli, “High altitude platform station (HAPS): A review of new infrastructure development for future wireless communications,” Wireless Pers. Commun., vol. 42, no. 3, pp. 387–404, Aug. 2006.
[5] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, “Space-air-ground integrated network: A survey,” IEEE Commun. Surv. Tuts., vol. 20, no. 4, pp. 2714–2741, Fourthquarter 2018.
[6] S. Chandrasekharan et al., “Designing and implementing future aerial communication networks,” IEEE Commun. Mag., vol. 54, no. 5, pp. 26–34, May 2016.
[7] G. Xiong et al., “A kind of novel ITS based on space-air-ground big-data,” IEEE Intell. Transp. Syst. Mag., vol. 8, no. 1, pp. 10–22, Spring 2016.
[8] M. Casoni, C. A. Grazia, M. Klapcz, N. Patriciello, A. Amditis, and E. Sdongos, “Integration of satellite and LTE for disaster recovery,” IEEE Commun. Mag., vol. 53, no. 3, pp. 47–53, Mar. 2015.
[9] J. Qui, D. Grace, G. Ding, M. D. Zakaria, and Q. Wu, “Air-ground heterogeneous networks for 5G and beyond via integrating high and low altitude platforms,” IEEE Wireless Commun., vol. 26, no. 6, pp. 140–148, Dec. 2019.
[10] S. C. Arum, D. Grace, and P. D. Mitchell, “A review of wireless communication using high-altitude platforms for extended coverage and capacity,” Comput. Commun., vol. 157, pp. 232–250, May 2020.
[11] G. K. Kurt et al., “A vision and framework for the high altitude platform station (HAPS) networks of the future,” IEEE Commun. Surv. Tuts., vol. 23, no. 2, pp. 729–779, Secondquarter 2021.
[12] A. Mohammed, A. Mehmood, F.-N. Pavlidou, and M. Mohoric, “The role of high-altitude platforms (HAPs) in the global wireless connectivity,” Proc. IEEE, vol. 99, no. 11, pp. 1939–1953, Nov. 2011.
[13] S. Karapantazis and F. Pavlidou, “Broadband communications via high-altitude platforms: A survey,” IEEE Commun. Surv. Tuts., vol. 7, no. 1, pp. 2–31, Firstquarter 2005.
[14] O. B. Yahia, E. Erdogan, G. K. Kurt, I. Altunbas, and H. Yankomeroglu, “HAPS selection for hybrid RFFSO satellite networks,” IEEE Trans. Aerosp. Electron. Syst., vol. 50, no. 4, pp. 2855–2867, Aug. 2014.
[15] F. Fidler, M. Knapke, J. Horwath, and W. R. Leeb, “Optical communications for high-altitude platforms,” IEEE J. Sel. Topics Quantum Electron., vol. 16, no. 5, pp. 1058–1070, Sep.–Oct. 2010.
[16] H. Henninger and O. Wilfert, “An introduction to free-space optical communications,” Radioengineering, vol. 19, no. 2, pp. 203–212, Jun. 2010.
[17] L. Liu, R. Zhang, and K.-C. Chua, “Achieving global optimality for weighted sum-rate maximization in the K-user Gaussian interference channel with multiple antennas,” IEEE Trans. Wireless Commun., vol. 11, no. 5, pp. 1933–1945, May 2012.
[18] T. H. Kim and S. Choi, “Interference mitigation via scheduling for the MIMO broadcast channel with limited feedback,” in Proc. IEEE 20th Int. Symp. Pers., Indoor Mobile Radio Commun., 2009, pp. 2035–2039.
[19] J. Wang, D. J. Love, and M. D. Zoltowska, “User selection with zero-forcing beamforming achieves the asymptotically optimal sum rate,” IEEE Trans. Signal Process., vol. 56, no. 8, pp. 3713–3726, Aug. 2008.

[20] R.-J. Reifert et al., “Distributed resource management in downlink cache-enabled multi-cloud radio access networks,” IEEE Trans. Veh. Technol., vol. 71, no. 12, pp. 13120–13136, Dec. 2022.

[21] K. Shen and W. Yu, “Fractional programming for communications systems—Part II: Uplink scheduling via matching,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2631–2644, May 2018.

[22] A. Douik, H. Dahrouj, O. Amin, B. Alosqibi, T. Y. Al-Naffouri, and M.-S. Alouini, “Mode selection and power allocation in multi-level cache-enabled networks,” IEEE Commun. Lett., vol. 24, no. 8, pp. 1789–1793, Aug. 2020.

[23] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, “A tutorial on clique problems in communications and signal processing,” Proc. IEEE, vol. 108, no. 4, pp. 583–608, Apr. 2020.

[24] M. Dörterler, “A new genetic algorithm with agent-based crossover for generalized assignment problem,” Inf. Technol. Control, vol. 48, no. 3, pp. 389–400, Aug. 2019.

[25] H. Dahrouj and W. Yu, “Coordinated beamforming for the multicell multi-antenna wireless system,” IEEE Trans. Wireless Commun., vol. 9, no. 5, pp. 1748–1759, May 2010.

[26] H. Dahrouj and W. Yu, “Multicell interference mitigation with joint beamforming and common message decoding,” IEEE Trans. Commun., vol. 59, no. 8, pp. 2264–2273, Aug. 2011.

[27] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” IEEE Trans. Signal Process., vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[28] B. Dai and W. Yu, “Sparse beamforming and user-centric clustering for downlink cloud radio access network,” IEEE Access, vol. 2, pp. 1326–1339, 2014.

[29] K. Shen and W. Yu, “Fractional programming for communication systems—Part I: Power control and beamforming,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2616–2630, May 2018.

[30] W. Yu, T. Kwon, and C. Shin, “Multicell coordination via joint scheduling, beamforming, and power spectrum adaptation,” IEEE Trans. Wireless Commun., vol. 12, no. 7, pp. 1–14, Jul. 2013.

[31] A. A. Khan, R. Adve, and W. Yu, “Optimizing Multicell Scheduling and Beamforming Via Fractional Programming and Hungarian Algorithm,” in Proc. IEEE Globecom Workshops., 2018, pp. 1–6.

[32] X. Zhu, C. Jiang, L. Yin, L. Kuang, N. Ge, and J. Lu, “Cooperative multi-group multicast transmission in integrated terrestrial-satellite networks,” IEEE J. Sel. Areas Commun., vol. 36, no. 5, pp. 981–992, May 2018.

[33] D. Han, W. Liao, H. Peng, H. Wu, W. Wu, and X. Shen, “Joint cache placement and cooperative multicast beamforming in integrated satellite-terrestrial networks,” IEEE Trans. Veh. Technol., vol. 71, no. 3, pp. 3131–3143, Mar. 2022.

[34] M. Alzenad and H. Yanikomeroglu, “Coverage and rate analysis for vertical heterogeneous networks (VHetNets),” IEEE Trans. Wireless Commun., vol. 18, no. 12, pp. 5643–5657, Dec. 2019.

[35] N. Cherif, M. Alzenad, H. Yanikomeroglu, and A. Yongacoglu, “Downlink coverage and rate analysis of an aerial user in vertical heterogeneous networks (VHetNets),” IEEE Trans. Wireless Commun., vol. 20, no. 3, pp. 1501–1516, Mar. 2021.

[36] H. Jia, Y. Wang, M. Liu, and Y. Chen, “Sum-rate maximization for UAV aided wireless power transfer in space-air-ground networks,” IEEE Access, vol. 8, pp. 216231–216244, 2020.

[37] L. Wang, X. Zhao, C. Wang, and W. Wang, “Resource allocation algorithm based on power control and dynamic transmission protocol configuration for HAPS-IMT integrated system,” Electronics, vol. 20, no. 1, pp. 44–64, Dec. 2021.

[38] A. Alsharoua and M.-S. Alouini, “Improvement of the global connectivity using integrated satellite-airborne-terrestrial networks with resource optimization,” IEEE Trans. Wireless Commun., vol. 19, no. 8, pp. 5088–5100, Aug. 2020.

[39] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M.-S. Alouini, “FSO-based vertical backhaul/fronthaul framework for 5G wireless networks,” IEEE Commun. Mag., vol. 56, no. 1, pp. 218–224, Jan. 2018.

[40] R. Ganian and S. Ordyniak, “Solving integer linear programs by exploiting variable-constraint interactions: A survey,” Algorithms, vol. 12, no. 12, pp. 248–261, Nov. 2019.

[41] G. T. Ross and R. M. Soland, “A branch and bound algorithm for the generalized assignment problem,” Math. Prog., vol. 8, no. 1, pp. 91–103, Dec. 1975.

[42] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[43] H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, “Distributed cloud association in downlink multicloud radio access networks,” in Proc. 49th Annu. Conf. Inf. Sci. Syst., 2015, pp. 1–3.

[44] T. Yoo and A. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,” IEEE J. Sel. Areas Commun., vol. 24, no. 3, pp. 528–541, Mar. 2006.