IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 6, 2020, accepted June 27, 2020, date of publication July 6, 2020, date of current version July 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007501

Resource Allocation and Node Placement in Multi-Hop Heterogeneous Integrated-Access-and-Backhaul Networks

JIUN Y. LAI[1], WU-HSIU WU[2], AND YU T. SU[3], (Life Senior Member, IEEE)
[1]Industrial Technology Research Institute, Hsinchu 31040, Taiwan
[2]MediaTek Inc., Hsinchu 30078, Taiwan
[3]Institute of Communications Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan

Corresponding author: Yu T. Su (ytsu@nctu.edu.tw)

This work was supported in part by the Ministry of Science and Technology (MOST) of Taiwan under Grant MOST 107-2221-E-009-129.

ABSTRACT We consider a heterogeneous MIMO-OFDMA based dense small cell (SC) system in which each macro cell base station (MBS) serves its coverage area with the help of small cell base stations (SBSs) through multi-hop wireless connections. The SBSs act as integrated access and backhaul (IAB) nodes that handle both access and backhaul traffics with wireless links. We first develop an optimal (sum-rate maximization) resource allocation (RA) algorithm which considers subcarriers/spatial subchannels assignment and the associated power allocations. We also present two low-complexity suboptimal RA schemes which, as verified by simulations, incur only minor performance loss in the high SNR region. Our RA algorithms can be applied to other multi-hop networks with general UE association rule and node location distributions. We study the channel aging effect caused by the time lag between the time channel state information (CSI) is measured and that when data transmission occurs. We show the benefit of channel prediction and the limit of a centralized RA approach. The advantages of frequency (channel) reuse and the multi-hop architecture are demonstrated as well. A related but perhaps more important system design issue for an IAB cellular network is the IAB node placement problem. With the given UE association rule and UE location distribution, we present systematic approaches to find the optimal node locations. For two special propagation models, we derive closed-form expressions for the node locations that maximizes a spectral efficiency lower bound. Numerical results validate the accuracy of our estimates based on either numerical evaluations or closed-form solutions.

INDEX TERMS MIMO-OFDMA, miltihop HetNet, imperfect CSI, resource allocation, node placement.

I. INTRODUCTION

Heterogeneous networks (HetNets) that consist of macro and small cells were introduced by 3GPP for LTE advanced (LTE-A) and is also considered for 5G Ultra-Dense Cellular Networks (UDCNs) [1]. Such a heterogeneous network is sometimes known as two-layer heterogeneous network [2]. The structure of a 5G UDCN is similar to a cloud radio access network (C-RAN) in which base stations (BSs) or remote radio heads (RRHs) are connected to a central BS by fibers or by dedicated directional microwave or millimeter wave (mmWave) links. As the required backhaul capacity of an UDCN is often very high and the fiber optic based infrastructure is both costly and time-consuming, the 3GPP has recently completed a study on Integrated Access and Backhaul (IAB). The IAB network relays most backhaul traffic through single or multi-hop mmWave links and allows both UEs and BSs to share the same frequency resources [6]. Those BSs that wirelessly backhauls the access traffic are referred to as IAB-nodes and the associated network shall be called as an IAB network or IAB-UDCN in short. In this paper, whenever we mention UDCN, we are referring to an IAB-UDCN.

To handle the frequent handovers problem in an IAB-UDCN [3], an macro cell BS (MBS) is configured not only to serve nearby user equipments (UEs) but is also responsible for managing resources and controlling the UE handovers among small cell (SC) BS's (SBSs) which act as data plane BSs; not just a complement for the MBS centered network. Energy consumption of such a 5G network

The associate editor coordinating the review of this manuscript and approving it for publication was Ivan Wang-Hei Ho.

is of great concern and has a lot to do with small cell placement and UE association. Reference [4] investigated a multi-hop UDCN with some SBSs act as gateways while the MBS is responsible for control plane. The authors developed backhaul gateways selection and wireless backhaul routing schemes that maximizes the ratio of the network capacity to the consumed energy. A two-tier HetNet with both the MBS and SBSs handling data plane was considered in [5] which has sleeping and range-expansion mechanisms in MBS edges and areas close to the MBS for energy reduction purpose. Since in an IAB-UDCN the SBSs are all IAB nodes, we shall use both IAB nodes and SBSs interchangeably.

The UE association (UA) policy in an UDCN raises the concern of load balancing [7]–[13]. Balancing the workload among MBSs and SBSs can improve system performance and communication qualities [11]. A general principle for load balancing in HetNets is to reduce the load of MBS by pushing some UEs on overloaded MBSs onto the lightly loaded SBSs [7], [8] [11], [12]. For some mmWave based networks, the MBS takes charge of the control plane while SBSs handle data traffic, the MBS may [3] or may not [4] directly service UEs. In [3], the authors consider a cellular network with a single or multiple gateways that is configured by the MBS; the backhaul traffic of a SBS is relayed to an adjacent SBS via a Line-of-Sight (LoS) link, i.e., all backhaul traffics are connected to the MBS by multi-hop mmWave LoS links.

Shown in Fig. 1 is the MIMO-OFDMA based IAB-UDCN we consider in this paper, whose structure is similar to that discussed in [3], [4] and can be regarded as a cloud RAN with a sectorized multi-hop network topology. We assume that all backhaul traffics of SCs are sent to/from the MBS or central BS by either single-hop or two-hop dedicated LoS MIMO links while the MBS also serves neighboring UEs directly. For downlink data, the IAB nodes (SBSs) act as relays to forward the MBS-originated messages to either non-line-of-sight (NLoS) MIMO UEs or to another SBS through an LoS link. The dynamic control messages, including resource allocation (RA) and channel state information (CSI) feedback, are conveyed through lower-band links from/to the MBS. In other words, our network employs a dual connection frequency plan that allocates large spectra at a higher frequency band to regular data links while the sparse control signals are transmitted over a lower frequency band which has a smaller bandwidth but much less path loss. Similar systems which combine MIMO-OFDMA techniques with multiple hop transmission have been considered as a promising architecture for further enhancing wireless capacity and extending the coverage of an MBS [19], [20].

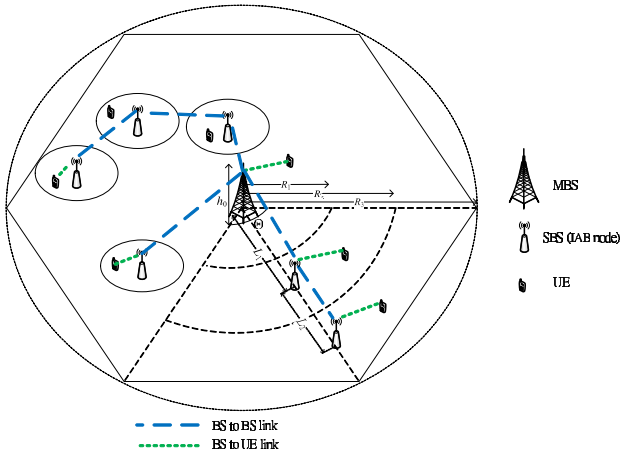

FIGURE 1. A multi-hop heterogeneous sectorized macrocell with multiple small cells and IAB nodes.

As a part of the RA solution, a popular power allocation (PA) scheme that maximizes the sum rate or weighted sum rate under the total transmit power constraint is the classic water-filling approach [14] if the sub-channel (subcarrier) gains are known. A dual problem that minimizes the total transmit power with some rate requirements is also well documented [15]. Numerous variations of these two basic RA metrics have been discussed extensively for OFDMA and other multiuser communication systems [16]–[18]. Some complete or partial analytical solutions were obtained, most were solved by combined analytic/numerical optimization or semi-analytic simulations. Optimal RA for such a multi-hop network is difficult as one has to deal with the allocations of power and frequency-spatial degrees of freedom at all hops simultaneously. An interesting RA metric which maximizes virtual reality users' quality of service (QoS) in an OFDMA based SC network was suggested in [21]. An echo state networks based learning algorithm was developed to solve the corresponding RA problem.

Contributions: The main themes of this paper include RA (joint power-carrier resource assignment) and IAB nodes (SBSs) placement in a dense SC HetNet. The two issues are related but can be separated treated. We develop optimal and suboptimal RA algorithms that try to maximize a weighted sum rate performance with a nominal node location and then find the best node locations using the proposed RA algorithms.

The joint RA solution takes spatial (MIMO) and frequency (subcarriers) and UE degrees of freedom into consideration and makes PA decisions to maximize the weighted downlink sum rate under the total per cell power constraint. An additional per UE subcarrier constraint designed to obtain improved fairness performance is included as an alternative RA metric. The RA optimality is verified by showing that strong duality (zero duality gap) holds (Appendix C) and the constraint relaxed solution turns out to satisfy the original binary constraint (Appendix D). To reduce the complexity of the optimal solution, we propose low-complexity RA algorithms which give suboptimal and near-optimal performance. Our analysis extends to include the channel aging effect caused by the time lag between the time the CSI is measured and that when a RA decision is implemented. A channel predictor is used to compensate for channel aging and the prediction error is considered in computing the

sum rate. We also evaluate the spectral efficiency of a frequency reuse plan with the resulting co-channel interference accounted.

We present a systematic method to numerically find the optimal IAB node locations that achieve the maximum weighted sum rate. For two special cases, we obtain closed-form expressions for near-optimal IAB node locations. The optimal IAB locations predicted by our approaches are proved to be accurate when compared with those found by actually employing the proposed RA schemes and the time-consuming search over all candidate node positions. Although we employ a deterministic UE association rule that guarantees load-balancing in the mean sense for each SC and assume a uniform UE distribution, our approach can be easily modified for solving node placement problems with other UE association rules and different UE distributions.

The rest of this paper is organized as follows. The system and channel models, time-frequency schedule and allocation and related achievable rates are presented in Section II. The channel models we consider include LoS, NLoS (Section II.B.) and aged channels (Section II.C.). The discussion on the distance-dependent mixed channel is deferred until Section VI.D. The RA problem is formulated as a weighted sum rate maximization problem in Section III where we assume that perfect CSI is available. An optimal solution and two suboptimal RA algorithms are developed. In Section IV, we address the IAB node placement issue and present analytic approach to estimate the optimal node locations that give the best spectral efficiency. We derive some closed-form suboptimal IAB node placement solutions which minimize a spectral efficiency lower bound but predict accurate optimal node locations. A deterministic UE association rule is also suggested. In Section V, the channel aging effect and the impact of frequency reuse are discussed; a reuse schedule and a channel prediction method are suggested as well. We provide numerical performance of various proposed RA algorithms and IAB node placement solutions in Section VI to validate our answers, The numerical results and their implications in system design are discussed. Finally, in Section VII we summarize the main results of this paper.

Notations: Unless specially instructed, boldface lowercase letters and boldface uppercase letters are used for vectors and matrices, respectively. $(\cdot)^H$, $(\cdot)^{-1}$, $tr\{\cdot\}$, and rank() denote conjugate transpose, inverse, trace, and rank of a matrix, respectively. diag$\{\cdot\}$ is the diagonalization operator of a vector or matrix. E$\{\cdot\}$ denotes the expectation operator. $\mathbb{C}^{m\times n}$ and I_K stand for space of $m \times n$ matrices and the $K \times K$ identity matrix, respectively. For brevity, we shall denote UE k of SC l by UE(k, l), the BS m to BS $m+1$ link by BBL$(m, m+1)$ and the BS l to UE k link by BUL(k, l). When there is no danger of confusion, we also use UE(k, l) to refer to the multi-hop MBS to UE(k, l) link. The other symbols and abbreviations to be used are listed in Table 1 for convenience of reference.

TABLE 1. Definition of main variables and abbreviations in this paper.

Symbol	Description
$\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m\times n}$	sets of real numbers, n vectors, $m \times n$ matrices
$\mathbb{R}_+, \mathbb{R}_+^n, \mathbb{R}_+^{m\times n}$	sets of non-negative real numbers, related n vectors and $m \times n$ matrices
$\mathbb{C}, \mathbb{C}^n, \mathbb{C}^{m\times n}$	sets of complex numbers, n vectors, $m\times n$ matrices
L_1, L_2	horizontal distances between MBS and SBS 1, and between SBS 1 and SBS 2
h_0, h_1, h_2	heights of MBS, SBS 1 and SBS 2
α_L, α_N	path loss exponents for LoS and NLoS links
N_t, N_r	number of transmit and receive antennas
N_s, N	number of spatial channels of a MIMO link, and subcarriers in each subband
σ_ω^2	white noise power
K_l	number of active users in the SC l
D_m	link distance of BBL$(m, m+1)$
d_{lk}	link distance of BUL(l, k)
ζ_m, ζ_{kl}	shadowing parameters of BBL$(m, m+1)$ and BUL(k, l)
g_L	path loss
ξ_m, ξ_{lk}	large scale fading coefficients including shadowing and path loss of BBL$(m, m+1)$ and BUL(k, l) i.e., $\xi_m = \zeta_m g_L$
$\gamma_m(s,n), \gamma_{lk}(s,n)$	signal to noise ratio of the sth spatial subchannel over the nth vector subcarrier (VC) of BBL$(m, m+1)$ and BUL(k, l).
$p_m(s,n), p_{lk}(s,n)$	power assigned to the sth spatial subchannel over the nth VC of BBL$(m, m+1)$ and BUL(k, l).
$\eta_m(s,n), \eta_{lk}(s,n)$	singular value of the matrix $\mathbf{H}$ corresponding to the sth spatial subchannel over the nth VC of BBL$(m, m+1)$ and BUL(k, l)
$r_m(s,n), r_{lk}(s,n)$	normalized achievable rates of the sth spatial subchannel over the nth VC in the link BBL$(m, m+1)$ and BUL(k, l)
$\rho_k(n)$	nth VC assignment indicator for UE k
$\lambda_l, \varphi_{ln}, \mu_{lk}(n)$	nonnegative Lagrange multipliers

II. SYSTEM AND CHANNEL MODELS

A. A MULTI-HOP HetNet MODEL AND ITS TIME-FREQUENCY PLAN

Consider a MIMO-OFDMA network in which an MBS is equipped with multiple antenna arrays, each is responsible for communicating with UEs and IABs within a sectorized service area. Fig. 1 depicts such a sectorized macrocell in which each sector is served by the MBS and two SBSs that located within the Θ-degree circular sector of a macrocell of radius R_C centered at the MBS and the two SBSs are connected to the MBS via a single-hop or two-hop dedicated LoS link. The two IAB nodes, denoted by SBS 1 and SBS 2, are located on the angle bisector direction and are separated from the MBS by distances L_1 and $L_1 + L_2$. In addition, the height of MBS, SBS 1 and SBS 2 are h_0, h_1 and h_2, respectively. We assume that all UEs are located at height zero.

Each sector is divided into three SCs served by the MBS and two IAB nodes so that the MBS serves UEs in SC 1 and SBS l will serve UEs in SC $l+1$, $l = 1, 2$. We assume that there are K_l UEs in SC l and the UEs are uniformly distributed within the sector. The uniform distribution justifies our initial assumption that the SBSs are located on the angle bisector direction. All BSs (UEs) are equipped with N_t (N_r) antennas that operate in either isotropic or directional mode in disjoint

frequency bands. Every UE in SC ℓ is served by ℓ-hop links and the MBS is in charge of RA for all UEs in the sector.

Shown in Fig. 2 is the downlink time-frequency (T-F) schedule considered in this paper. With this pipeline-like resource utilization schedule, both BSs and UEs, except for few initial time slots, can be active all the time, ensuring maximum T-F resource utility. Reverse and backhaul links' schedules can be similarly planned. The schedule partitions the available T-F resources into frequency bands and time slots so that each SC is given a portion of the transmission resource, ensuring certain degrees of fairness. For if the resource distribution does not take the area-fairness into consideration, most resource will be given to the UEs in SC 1 which are closest to the MBS as such allocation is most sum-rate-efficient. We return to this issue in Sections IV.B and VI.D. Note that this T-F schedule represents the upper layer of our RA plan, i.e., we first partition the available T-F resources into several parts, each for serving a SC. Then the remaining RA tasks are the assignments of subcarriers and spatial subchannels within each designated resource part and the associated power allocation.

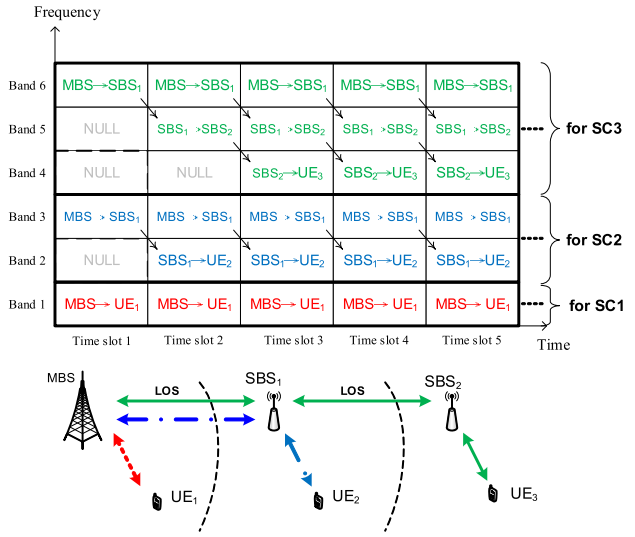

FIGURE 2. A T-F plan/schedule for the multi-hop HetNet shown in Fig. 1. The red dot line denotes a MBS-UE link, the two blue dash-dotted segments denotes a typical two-hop MBS-SBS1-UE2 link while the green solid arrowed lines represent a typical three-hop MBS-SBS1-SBS2-UE link. All those links can be downlink or uplink, UE_i denotes UEs in SC i and SBS i is abbreviated as SBS_i. The arrows in the upper part represent the proposed pipeline-like T-F schedules.

The inter-BS links are LoS links, which can be realized by deploying all BSs at fixed positions with suitably hights [23]. We assume that there are N subcarriers in each subband, and denote the subcarrier spacing by Δf_S. Guard bands of bandwidth Δf_G Hz are inserted into two adjacent subbands. A subcarrier can only be assigned to an UE in an OFDMA system and l frequency subbands are allocated for SC l, each is used to support a hop. For a l-hop link, the vector $(f_1, f_2, \ldots, f_l)$ denotes the subcarrier indexes used in each hop. We refer to such a l-tuple as a vector subcarrier (VC). We index the subcarriers of a subband in ascending carrier frequency order. To reduce the RA complexity, the entries in a VC are assumed to be identical so that there are N (instead of N^l) VCs for SC l to serve UEs in its coverage area. Such a restriction does not loss much generality against the scheme that allows f_i's in a VC to be distinct. We argue as follows. First, the capacity of a multi-hop single-subcarrier link is bounded by the hop with the worst channel state, i.e., the one suffers the most fading loss, which is usually the BS-UE hop (see Fig, 2) as the inter-BS links are LoS links while the BS-UE link is likely to experience much more severe NLoS fading. Second, each LoS BS-BS link has a strong specular component, especially when antenna arrays are used, hence the link gain is dominated by the hop distance and is less sensitive to the subcarrier selected within a subband.

Let $\rho_k(n)$ be the VC assignment indicator, i.e., $\rho_k(n) = 1$ means that the nth VC is allocated to UE k of SC l. This discrete constraint guarantees that an VC is given to one UE only. We define w_{lk} as the weight or priority requirement for UE(k, l) and if UE(k, l) uses VC n, then we denote this UE by UE(k, l, n), the abbreviations, BUL(k, l, n) and BBL$(m, m+1, n)$ are similarly defined. As in the case of UE(k, l), we also use UE(k, l, n) to denote the multi-hop MBS-UE(k, l, n) link.

B. CHANNEL MODEL

We assume a multiple dimensional (time slot, frequency bands, and spatial) transmission with perfect network timing. The CSI measured by every SBS is sent back to the MBS via a separate control channel. The channel remains constant within a time slot but may vary from slot to slot. The signal received by an N_r-antenna receiver is corrupted by an $N_r \times 1$ noise vector whose elements are independent and identically distributed (i.i.d.) zero mean complex Gaussian random variables with variance σ_w^2. Each subcarrier link connecting a pair of BS's suffers from Rician fading while BS-UE links are assumed to suffer from Rayleigh fading.

The mth hop's channel of UE(k, l, n) is described by an $N_r \times N_t$ complex channel matrix $\mathbf{H}$ whose elements are of the form $h = \sqrt{\xi} \times h_{ss}$, where ξ and h_{ss} represent the large-scale fading coefficient (LSFC) and small-scale fading coefficient (SSFC), respectively. The LSFC ξ is the product of the shadowing coefficient ζ and the path loss g_L

$$g_L = \left(\frac{\lambda}{4\pi d_0}\right)^2 \left(\frac{d_0}{d}\right)^{\alpha}, \quad 2 \le \alpha < 6 \tag{1}$$

where λ is the carrier wavelength, d_0 is the closed-in distance, assumed to be 1 meter in subsequent discussions, and d is the link (slant) distance. For BS-BS links, $d = D_1 = \sqrt{L_1^2 + (h_0 - h_1)^2}$ or $d = D_2 = \sqrt{L_2^2 + (h_1 - h_2)^2}$, and for a BS-UE link $d = d_{lk}$. To distinguish LoS channels from NLoS ones, we use α_L and α_N to denote the corresponding path loss exponents.

To separate LSFC from SSFC, we rewrite the channel matrix as $\mathbf{H} = \sqrt{\xi}\mathbf{H}_{SS}$, where the SSFC matrix $\mathbf{H}_{SS}$ is given

by [22]

$$\mathbf{H}_{SS} = \sqrt{\frac{K}{K+1}}\mathbf{H}_{\text{LOS}} + \sqrt{\frac{1}{K+1}}\mathbf{H}_{\text{NLOS}},$$
$$(h_{\text{LOS}})_{pq} = \exp\left(j\frac{2\pi}{\lambda}d_{pq}\right), \tag{2}$$

where $1 \le p \le N_t$, $1 \le q \le N_r$, K is the Rician factor, $\mathbf{H}_{\text{LOS}}$ represent the LoS part of the matrix channel and d_{pq} is the distance between transmit (Tx) antenna p and receive (Rx) antenna q. The NLoS part $\mathbf{H}_{\text{NLOS}}$ is a matrix with i.i.d. zero-mean complex Gaussian random variables with unit variance that varies slot by slot. Unless otherwise specified, we assume that for all BBL$(m, m+1, n)$ links $K = \infty$ and $K = 0$ for all BUL(k, l, n) links. This assumption is based on facts that the locations of both MBS and SBSs are fixed and are so selected such that the BSs on both sides of a relay link are within the LoS of each other; furthermore, each mmWave antenna array used has high directivity and very narrow beamwidth.

Within a slot, each MIMO (subcarrier) channel is frequency-flat faded so that using a proper precoder and receive beamforming vector, we can convert the MIMO channel into a set of parallel independent flat faded sub-channels. The precoding and beamforming vectors, $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$, $\mathbf{U} \in \mathbb{C}^{N_r \times N_r}$, are obtained by performing singular value decomposition (SVD)

$$\mathbf{H} = \mathbf{U}\Xi\mathbf{V}^H, \quad \Xi = \text{diag}[\eta(1), \ldots, \eta(s)] \tag{3}$$

Ξ is diagonal containing the singular values $\eta(1) \ge \eta(2) \ge \ldots \ge \eta(s)$ of $\mathbf{H}$. We assume that the LSFCs remain constant for a period of several time slots.

C. CHANNEL AGING EFFECTS

Transmitting a packet over a l-hop link requires l slots, i.e., if the MBS sends a packet at slot t, it is received by the designated UE at slot $t + l - 1$. It should be noticed that in allocating resource for each hop based on the CSI obtained by the corresponding IAB's, there is a $2(l-1)$-slot lag between the time it is measured by the IAB and the time the resource is used in transmission. We assume that the MBS employs a Wiener-type predictor to predict the channel aging effect and broadcasts the RA information via the lower frequency band downlink control channel. The prediction error is taken into account in evaluating the achievable rate, which is divided into two parts: one accounts for NLoS BS-UE links (MBS-UE or SBS-UE) while the other for LoS BS-BS links (MBS-SBS or SBS-SBS). The latter part is static as the locations of BS's are fixed.

A packet is divided into two parts: the first (preamble) part contains the common pilot sequence and other overhead information, the other (payload) part contains just data. The pilot part, which is usually uncoded, is used for synchronization, channel and signal-to-noise ratio (SNR) estimation. We assume that perfect OFDM carrier and symbol synchronization has been achieved, hence the data part at the nth subcarrier can be expressed as [24]

$$\mathbf{y}(n) = \sqrt{\xi}\mathbf{G}(n)\mathbf{x}(n) + \sqrt{\xi}\big[\mathbf{H}(n) - \mathbf{G}(n)\big]\mathbf{x}(n) + \mathbf{w}(n), \tag{4}$$

where $\mathbf{w}(n) \sim \mathcal{CN}(0, \sigma_w^2\mathbf{I}_{N_t})$ is the spatially white additive Gaussian noise at the receiver side, ξ is the LSFC and

$$\mathbf{G}(n) = \begin{cases} \mathbf{H}(n), & \text{perfect predicted channel} \quad (5) \\ \mathbf{H}_\tau(n), & \text{aged channel} \quad (6) \\ \overline{\mathbf{H}}(n), & \text{predicted channel} \quad (7) \end{cases}$$

with the subscript τ indicates the CSI delay and $\mathbf{G}(n) = \widehat{\mathbf{U}}\widehat{\Xi}\widehat{\mathbf{V}}^H$, $\widehat{\mathbf{U}} \in \mathbb{C}^{N_r \times N_r}$, $\widehat{\Xi} \in \mathbb{C}^{N_r \times N_t}$, and $\widehat{\mathbf{V}} \in \mathbb{C}^{N_t \times N_t}$. MBS uses a precoder $\widehat{\mathbf{V}}(n)$ to map data symbols, $\mathbf{s}(n)$, to transmit antennas. $\mathbf{x}(n) = \widehat{\mathbf{V}}(n)\mathbf{s}(n)$, and $\mathbf{s}(n) \in \mathbb{C}^{N_t \times 1}$ with zero mean $\mathbb{E}[s(n)] = 0$ and covariance matrix $\mathbb{E}[\mathbf{s}(n)\mathbf{s}(n)^H] = \text{diag}[\sqrt{p_1}, \ldots, \sqrt{p_{N_t}}]$, where p_i is allocated power. Denote the detector operation by $\widehat{\mathbf{U}}(n)$, we write detector output $\widetilde{\mathbf{y}}(n) \triangleq \widehat{\mathbf{U}}^H(n)\mathbf{y}(n)$ as

$$\begin{aligned}\widetilde{\mathbf{y}}(n) &= \sqrt{\xi}\widehat{\Xi}(n)\mathbf{s}(n) + \sqrt{\xi}\widehat{\mathbf{U}}^H(n)\big[\mathbf{H}(n) - \mathbf{G}(n)\big]\mathbf{x}(n) \\ &\quad + \widehat{\mathbf{U}}^H(n)\mathbf{w}(n) \\ &\overset{def}{=} \sqrt{\xi}\widehat{\Xi}(n)\mathbf{s}(n) + \Upsilon_{lk}(n) + \widehat{\mathbf{U}}^H(n)\mathbf{w}(n)\end{aligned} \tag{8}$$

As the locations of BSs are fixed, it is reasonable to assume that the LoS channels between BSs are time-invariant and can be perfectly estimated, i.e., $\mathbf{G}(n) = \mathbf{H}(n)$ and $\Upsilon_{lk}(n) = 0$. The instantaneous SNR measured at the receiving end of BBL$(m, m+1)$ is

$$\begin{aligned}\gamma_m(s,n) &= \left(\frac{\xi_m|\eta_m(s,n)|^2}{\sigma_w^2}\right)p_m(s,n), \quad 1 \le m \le l-1 \\ &\overset{def}{=} \varrho_m(s,n)p_m(s,n), \ \xi_m = \zeta_m/D_m^{\alpha_L}\end{aligned} \tag{9}$$

where $\varrho_m(s,n)$ is the channel gain to noise ratio (CNR), other parameters are defined in Table 1. For NLoS BS-UE links the channel is aged and/or cannot be perfectly measured, hence the spatial subchannels formed by combined precoding and receive beamforming are not independent and inter-channel interference does exist, i.e., $\Upsilon_{lk}(n) \ne 0$. We define the instantaneous signal to interference plus noise ratio (SINR) for the NLoS BUL(l, k, n) link as

$$\begin{aligned}\gamma_{lk}(s,n) &= \left(\frac{\xi_{lk}|\eta_{lk}(s,n)|^2}{I_{lk}(s,n) + \sigma_w^2}\right)p_{lk}(s,n) \\ &\overset{def}{=} \varrho_{lk}(s,n)p_{lk}(s,n), \ \xi_{lk} = \zeta_{lk}/d_{lk}^{\alpha_N}\end{aligned} \tag{10}$$

where $p_{lk}(s,n)$ is the transmit power of BUL(l, k, n) and the CSI error induced interference is

$$\begin{aligned}I_{lk}(s,n) &= \Big[\Upsilon_{lk}(n)\Upsilon_{lk}^H(n)\Big]_{ss} = \xi \\ &\times \text{diag}\Big[\sum_{j=1}^{N_t}|b_{1j}|^2p_{lk}(j,n), \ldots, \sum_{j=1}^{N_t}|b_{N_tj}|^2p_{lk}(j,n)\Big]_{ss}\end{aligned} \tag{11}$$

where $\text{diag}[\cdot]_{ss}$ denotes the (s,s)th term of the diagonal matrix $\text{diag}[\cdot]$ and b_{ij} is the (i,j)th element of the matrix

$\widehat{\mathbf{U}}_{lk}^H(n)\big[\mathbf{H}_{lk}(n) - \mathbf{G}_{lk}(n)\big]\widehat{\mathbf{V}}_{lk}(n)$. For UE$(k, l, n)$, the achievable rate of hop m's jth spatial subchannel, normalized by the subcarrier spacing $\triangle f_S$, is

$$\begin{aligned} r_m(j,n) &= \log_2[1+\gamma_m(j,n)],\ 1 \le m < l \\ r_{lk}(j,n) &= \log_2[1+\gamma_{lk}(j,n)] \end{aligned} \tag{12}$$

The corresponding normalized (achievable) rate for the multi-hop link is given by

$$r_{lk}(n) = \min_{m\in\{1,\ldots,l-1\}} \left(\sum_{s=1}^{N_S} r_m(s,n), \sum_{s=1}^{N_S} r_{lk}(s,n) \right) \tag{13}$$

It can be easily verified that $r_{lk}(n)$ is *concave* in $\gamma_m(j,n)$ and $\gamma_{lk}(j,n)$ and is dominated by the worst hop. In the following section, we assume perfect CSI is available so that there is no interference, i.e., $I_{lk}(s,n) = 0$, for all (l,k) and (s,n). The case of nonzero interference caused by channel aging and its remedy are discussed in Section V.

III. RESOURCE ALLOCATION CONSIDERATION

In this section, we investigate the RA problem for multi-hop heterogeneous IAB networks. We employ dual-decomposition to solve the corresponding non-convex optimization problem and develop the optimal joint power and VC allocation policy. We make use of a characteristic of a multi-hop link and that of LoS BS-BS links to simplify the subproblems. To reduce the RA complexity and provide tradeoff between complexity and performance, we provide several low-complexity suboptimal schemes as well.

A. THE OPTIMAL RA PROBLEM

As mentioned before, RA is determined by the MBS using the CSI measured by UEs and SBSs. Given the locations of the MBS and SBSs and the UE association rule, the aim of our RA policy is to maximize the weighted downlink sum capacity of all cells subject to the constraint that the power available for serving UEs in cell l is less than P_l. Formally, the problem is formulated as follow, where, as in Section II, we assume $1 \le m < l$.

$$\textbf{RA-P1:} \max_{\{\rho_k(n),p_m,p_{lk}\}} \sum_{k=1}^{K_l} w_{lk} \sum_{n=1}^{N} \rho_k(n) r_{lk}(n), \tag{14}$$

subject to

$$\sum_{k=1}^{K_l} \rho_k(n) \le 1, \quad \rho_k(n) \in \{0,1\}, \quad \forall\, n \tag{14a}$$

$$\sum_{k=1}^{K_l}\sum_{n=1}^{N} \rho_k(n) \sum_{s=1}^{N_s} \left[\sum_{m=1}^{l-1} p_m(s,n) + p_{lk}(s,n) \right] \le P_l \tag{14b}$$

$$p_m(s,n), p_{lk}(s,n) \ge 0, \quad \forall\,(n,k,m,s) \tag{14c}$$

Note that the UE association rule will determine which UEs are to be included in the first summation of (14) and the locations of the MBS and SBSs affect the achievable rates $r_{lk}(n)$, hence the above formulation is *general enough to accommodate any UE association rule and arbitrary fixed BS location distribution*. As mentioned before, $r_{lk}(n)$ defined by (13) is concave, hence, if we relax the discrete constraint on $\{\rho_k(n)\}$, the above problem can be converted into a equivalent epigraph form. Using the changes of variables $q_m(s,n) = \rho_k(n)p_m(s,n)$ and $q_{lk}(s,n) = \rho_k(n)p_{lk}(s,n)$, we have the equivalent problem

$$\textbf{RA-P2:} \max_{\{q_m,q_{lk}\},\{\rho_k\},\{t_{lk}\}} \sum_{k=1}^{K_l} w_{lk} \sum_{n=1}^{N} t_{lk}(n) \tag{15}$$

subject to

$$t_{lk}(n) \le \rho_k(n) r_{lk}(n), \ \forall\,(n,k) \tag{15a}$$

$$\sum_{k=1}^{K_l} \rho_k(n) \le 1, \quad \forall\, n \tag{15b}$$

$$0 \le \rho_k(n) \le 1, \quad \forall\,(n,k) \tag{15c}$$

$$\sum_{k=1}^{K_l}\sum_{n=1}^{N}\sum_{s=1}^{N_S} \left[\sum_{m=1}^{l-1} q_m(s,n) + q_{lk}(s,n) \right] \le P_l \tag{15d}$$

$$q_m(s,n) \ge 0, \quad q_{lk}(s,n) \ge 0, \ \forall\,(n,k,m,s) \tag{15e}$$

Note that $q_m(s,n)$ and $p_m(s,n)$ are actually functions of k with the relaxed constraint (15c). For notational simplicity we omit their dependence on k and for the convenience of subsequent expressions, we define

$$\{X\} = \{t_{lk}(n)\} \cup \{q_m(s,n), q_{lk}(s,n)\} \cup \{\rho_k(n)\}$$

In Appendix C we prove that **RA-P2** satisfies the strong duality condition, hence, its solution can be obtained by solving the following dual problem. Furthermore, in Appendix D (see also (39)), we show that the resulting solution does satisfy the original discrete constraint (14a), therefore, it is indeed optimal for **RA-P1**.

$$\textbf{RA-P3:} \min_{\lambda_l} \ G_l(\lambda_l) \tag{16}$$

subject to

$$\lambda_l \ge 0, \tag{16a}$$

$$t_{lk}(n) \le \rho_k(n) \sum_{s=1}^{N_S} r_m(s,n), \quad \forall\,(n,k,m) \tag{16b}$$

$$t_{lk}(n) \le \rho_k(n) \sum_{s=1}^{N_S} r_{lk}(s,n), \quad \forall\,(n,k) \tag{16c}$$

$$\sum_{k=1}^{K_l} \rho_k(n) \le 1, \quad \forall\, n \tag{16d}$$

$$0 \le \rho_k(n) \le 1, \quad \forall\,(n,k) \tag{16e}$$

$$q_m(s,n),\ q_{lk}(s,n) \ge 0, \quad \forall\,(n,k,m,s) \tag{16f}$$

where we have explicitly expressed the constraint (15a) as (16b) and (16c) because of (13) and define

$$G_l(\lambda_l) = \max_{\{X\}} \sum_{k=1}^{K_l} \sum_{n=1}^{N} w_{lk} t_{lk}(n) + \lambda_l \times \left\{ P_l - \sum_{k=1}^{K_l} \sum_{n=1}^{N} \sum_{s=1}^{N_S} \left[\sum_{m=1}^{l-1} q_m(s,n) + q_{lk}(s,n) \right] \right\} \quad (17)$$

Since $G_l(\lambda_l)$ can be rewritten as

$$G_l(\lambda_l) = \max_{X} \left[\sum_{n=1}^{N} g_{ln}(\lambda_l) \right] + \lambda_l P_l = \left[\sum_{n=1}^{N} \max_{X} g_{ln}(\lambda_l) \right] + \lambda_l P_l \quad (18)$$

where

$$g_{ln}(\lambda_l) = \sum_{k=1}^{K_l} w_{lk} t_{lk}(n) - \lambda_l \sum_{k=1}^{K_l} \sum_{s=1}^{N_S} \left[\sum_{m=1}^{l-1} q_m(s,n) + q_{lk}(s,n) \right] \quad (19)$$

RA-P3 is equivalent to

$$\textbf{RA-P4:} \min_{\lambda_l} \left\{ \left[\sum_{n=1}^{N} \max_{X} g_{ln}(\lambda_l) \right] + \lambda_l P_l \right\} \quad (20)$$

subject to (16a)–(16e).

Using the Lagrangian (22), as shown at the bottom of the page, which has taken the constraints (16a)–(16e) into account, we obtain $g_{ln}^*(\lambda_l) = \max_X g_{ln}(\lambda)$ by solving

$$\textbf{RA-P4a:} \max_{\substack{X, \varphi_{ln}, \mu_{lk}(n) \\ \{\mu_m(n)\}_{m=1}^{l-1}}} \mathcal{L}_{\text{RA-P4a}} \quad (21)$$

subject to

$$\varphi_{ln} \geq 0, \ \mu_m(n) \geq 0, \ \mu_{lk}(n) \geq 0, \ \forall\, (k,m) \quad (21a)$$

Once we solve **RA-P4a** for each n and obtain the corresponding $g_{ln}^*(\lambda_l)$, the final solution is to be obtained by solving

$$\textbf{RA-P5} \ \min_{\lambda_l} \sum_{n=1}^{N} g_{ln}^*(\lambda_l) + \lambda_l P_l \ \text{ subject to (16d)} \quad (23)$$

B. OPTIMAL RA SOLUTION

To solve **RA-P4a**, we start with the use of the stationarity requirement of the Karush-Kuhn-Tucker (KKT) conditions which implies that, for a fixed n, the optimal PA scheme expressed as a function of $\{\mu_m(n)\}$ and λ_l for $1 \leq m < l$ is given by

$$\frac{q_m(s,n)}{\rho_k(n)} = p_m(s,n) = \left(\frac{\mu_m(n)}{\lambda_l \ln 2} - \frac{1}{\varrho_m(s,n)} \right)^+ \quad (24)$$

where $(x)^+ = \max(0,x)$. This solution is similar to the conventional water-filling solution with the multiple water-levels determined jointly by the spatial subchannel s over VC n in the BUL(l,k) at the mth hop. As for UE(l,k,n), the optimal PA policy is the solution of

$$q_{lk}(s,n) = \frac{\partial \mathcal{L}_{\text{RA-P4a}}}{\partial q_{lk}(s,n)} = 0 \quad (25)$$

When perfect CSI is available, the corresponding optimal PA can be expressed as

$$p_{lk}(s,n) = \left(\frac{\mu_{lk}(n)}{\lambda_l \ln 2} - \frac{1}{\varrho_{lk}(s,n)} \right)^+, \quad (26)$$

which has the same form as (24). The same stationarity requirement also lead to the following lemmas that allow us to determine the Lagrange multipliers $\{\mu_m(n), \mu_{lk}(n)\}$.

Lemma 1: For UE(k,l,n), the corresponding Lagrange multipliers $\mu_m(n), \mu_{lk}(n)$ in determining the optimal PA satisfy

$$\sum_{m=1}^{l-1} \mu_m(n) + \mu_{lk}(n) = w_{lk}. \quad (27)$$

The above equation says that the Lagrange multipliers $\{\mu_m(n), \mu_{lk}(n)\}$ are determined by the corresponding UE weights. With a larger UE weight, the corresponding UE is more likely to be given more power and achieve higher rates. But the optimal solution should satisfy another condition given in

Lemma 2: For UE(k,l,n), the optimal PA scheme satisfies the following equation for all $1 \leq m < l$

$$\sum_{s=1}^{N_S} r_m(s,n) = \sum_{s=1}^{N_S} r_{lk}(s,n) \triangleq \beta_{lk}(n). \quad (28)$$

$$\mathcal{L}_{\text{RA-P4a}}(\{X\}, \lambda_l, \varphi_{ln}, \{\mu_{lk}\}) = \sum_{k=1}^{K_l} w_{lk} t_{lk}(n) - \lambda_l \sum_{k=1}^{K_l} \sum_{s=1}^{N_S} \left[\sum_{m=1}^{l-1} q_m(s,n) + q_{lk}(s,n) \right] + \varphi_{ln} \left(1 - \sum_{k=1}^{K_l} \rho_k(n) \right) + \sum_{k=1}^{K_l} \sum_{m=1}^{l-1} \mu_m(n) \left(\rho_k(n) \sum_{s=1}^{N_S} r_m(s,n) - t_{lk}(n) \right) + \sum_{k=1}^{K_l} \mu_{lk}(n) \left(\rho_k(n) \sum_{s=1}^{N_S} r_{lk}(s,n) - t_{lk}(n) \right) \quad (22)$$

Proof: The complementary slackness requirement (15a) implies that for the BBL(m,$m+1$), we have

$$\mu_m(n)\left(\rho_k(n)\sum_{s=1}^{N_S} r_m(s,n) - t_{lk}(n)\right) = 0 \tag{29}$$

If $\mu_m(n) = 0$, for some m, then according to (24), $p_m(s,n) = 0$ therefore $\mu_m(n) = \mu_{lk}(n) = 0$ for all other m and l,k as any hop in a multi-hop link becomes useless if a single hop is disconnected (zero power). Therefore, $\mu_m(n) > 0,\ \forall\, m$ and (29) implies that $\rho_k(n)\sum_{s=1}^{N_S} r_m(s,n) = t_{lk}(n),\ \ \forall\, m$. A similar argument for BUL(l,k) conclude that

$$\rho_k(n)\sum_{s=1}^{N_S} r_{lk}(s,n) = t_{lk}(n)\ \ \forall\, n \tag{30}$$

which then lead to (28). □

The above lemma implies that for the link UE(l,k,n), the spatial sum rates for all l (hop) are the same whence the link rate is dictated by the hop with the worst CNR. For the multi-hop HetNet we consider, only the last hop is a BS-UE link, the rest are BS-BS links. For a BS-BS LoS MIMO hop, the N_S spatial channels associated with the same subcarrier experience the same large scale fading and the LoS channel matrix defined in (2) is given by

$$\mathbf{H} = \begin{pmatrix} e^{-j\frac{2\pi}{\lambda}d_{11}} & \dots & e^{-j\frac{2\pi}{\lambda}d_{1N_S}} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{\lambda}d_{N_S1}} & \dots & e^{-j\frac{2\pi}{\lambda}d_{N_SN_S}} \end{pmatrix} = \mathbf{U}\Xi\mathbf{V}^H \tag{31}$$

where the second equality is resulted from SVD of $\mathbf{H}$; see (3). As the sum of the squared singular values $\sum_j |\eta(j)|^2$ is equal to the trace of the product matrix $\Xi\Xi^H$, $\mathrm{tr}\{\Xi\Xi^H\}$, we have

$$\mathrm{tr}\{\Xi\Xi^H\} = \mathrm{tr}\{\mathbf{U}^H\mathbf{U}\Xi\mathbf{V}^H\mathbf{V}\Xi^H\} = \mathrm{tr}\{\mathbf{H}\mathbf{H}^H\} \tag{32}$$

The above relation holds for general $\mathbf{H}$ but for LoS channel matrix (31),

$$\mathrm{tr}\{\mathbf{H}\mathbf{H}^H\} = \mathrm{tr}\begin{pmatrix} N_S & \dots & \Delta \\ \vdots & \ddots & \vdots \\ \Delta & \dots & N_S \end{pmatrix} = N_S^2,$$

where Δ denotes the off-diagonal terms. Hence, we conclude that

Lemma 3: For the $N_S \times N_S$ MIMO channel matrix of the form (31), the sum of the square of its singular values is N_S^2.

With this new constraint, we have the RA subproblem of maximizing a LoS link capacity, $r_m(n) = \sum_{s=1}^{N_s}\log_2[1+\gamma_m(s,n)]$, where $\gamma_m(s,n) = \varrho_m(s,n)p_m(s,n)$, and $\varrho_m(s,n) = \xi_m|\eta_m(s,n)|^2/\sigma_w^2$, subject to the total power and squared singular value sum constraint:

$$\textbf{RA-LOS-P1:}\quad \max_{\varrho_m(s,n),p_m(s,n)} \sum_{s=1}^{N_s}\log_2(1+\gamma_m(s,n)) \tag{33}$$

subject to

$$\sum_{s=1}^{N_S}\varrho_m(s,n) = \frac{\xi_m N_S^2}{\sigma_w^2} \tag{33a}$$

$$\sum_{s=1}^{N_S} p_m(s,n) = P(m,n) \tag{33b}$$

$$\varrho_m(s,n) \ge 0, \quad p_m(s,n) \ge 0, \tag{33c}$$

Using the inequality

$$\begin{aligned}
&\frac{1}{N_S}\sum_{s=1}^{N_S}\log_2[1+\varrho_m(s,n)p_m(s,n)] \\
&\le \log_2\left(1+\frac{1}{N_S}\sum_{s=1}^{N_S}\varrho_m(s,n)p_m(s,n)\right) \\
&\le \log_2\left[1+\left(\frac{1}{N_S}\sum_{s=1}^{N_S}\varrho_m(s,n)\right)\left(\frac{1}{N_S}\sum_{s=1}^{N_S}p_m(s,n)\right)\right] \\
&= \log_2\left[1+\frac{\xi_m N_S}{\sigma_w^2}\left(\frac{1}{N_S}\sum_{s=1}^{N_S}p_m(s,n)\right)\right] \qquad (34a) \\
&= \log_2\left(1+\frac{\xi_m P(m,n)}{\sigma_w^2}\right) \qquad (34b)
\end{aligned}$$

where the first inequality follows from Jensen's inequality, the second inequality is resulted from the Chebyshev sum inequality, (34a) is due to (33a) and (34b) is resulted from (33b). The maximum capacity $N_S\log_2(1+\xi_m P(m,n)/\sigma_w^2)$ is achieved by allocating equal power to every spatial subchannel $p_m(s,n) = P(m,n)/N_S$. Note that for a NLoS hop, the equal power for all spatial sunchannels strategy is not optimal. With this result, the optimal PA (24) for $1 \le m < l$ becomes

$$p_m(s,n) = \left(\frac{\mu_m(n)}{\lambda_l \ln 2} - \frac{\sigma_w^2}{N_S\xi_m}\right)^+ = \frac{P(m,n)}{N_s} \tag{35}$$

and the mth LoS hop's sum spatial rate for UE(k,l,n) is

$$\sum_{s=1}^{N_S} r_m(s,n) = N_S\log_2\left(\frac{N_S\xi_m\mu_m(n)}{\lambda_l\sigma_w^2\ln 2}\right) \tag{36}$$

In summary, we have

Lemma 4: For each LoS hop in the link UE(k,l,n), we should allocate the same power to all spatial subchannels and the Lagrange multipliers $\mu_m(n)$ associated with optimal PA of its LoS hops ($1 \le m < l$) have to satisfy

$$\xi_{m'}\mu_{m'}(n) = \xi_m\mu_m(n),\ \ m' \ne m,\ \ 1 \le m' < l \tag{37}$$

and is given by

$$\mu_m(n) = \frac{w_{lk} - \mu_{lk}(n)}{\xi_m\sum_{m'=1}^{l-1}\frac{1}{\xi_{m'}}} \tag{38}$$

(37) is obtained by substituting (36) into (28) while (38) is obtained by using (27) and (37). Based on the above results, a candidate PA method is to use iteratively bisection

approach. We randomly select an initial $\mu_{lk}(n)$ from $[0, w_{lk}]$ to determine all $\mu_m(n)$ by (38). Using those $\mu_m(n)$ to the optimal PA (24), where λ_l is given, then check if the PA values satisfy (28). When the allocated powers fail to meet the condition, we then select a new $\mu_{lk}(n)$ from the other half interval, the bisection based selection process continues until (28) is satisfied. The requirements can be the iteration number and the tolerance of the interval. The detail process is shown in **Algorithm OPA**.

Algorithm OPA An Optimal PA Algorithm for the UE k Over the VC n in the Multi-Hop ($l > 1$) HetNe

1: **Step 1** (Initialization)
2: Given n, k, λ_l, set $\widetilde{\mu}_{lk}(n) \leftarrow w_{lk}$, $\widehat{\mu}_{lk}(n) \leftarrow 0$;
3: $\mu_{lk}(n) \leftarrow (\widetilde{\mu}_{lk}(n) + \widehat{\mu}_{lk}(n))/2$.
4: Select a proper error tolerance ϵ_μ
5: **Step 2** (Compute $\mu_m(n)$ for $1 \le m < l$)
6: Use (27), (38) to compute $\mu_m(n)$ and (35) to compute $\{p_m(s, n)\}$ for $1 \le m < l$.
7: **Step 3** (Update $\mu_{lk}(n)$)
8: **if** $\widetilde{\mu}_{lk}(n) - \widehat{\mu}_{lk}(n) > \epsilon_\mu$ **then**
9: Calculate $\sum_{s=1}^{N_S} r_{lk}(s, n)$ and $\sum_{s=1}^{N_S} r_m(s, n)$.
10: **if** $\sum_{s=1}^{N_S} r_{lk}(s, n) < \sum_{s=1}^{N_S} r_m(s, n)$ **then**
11: $\widehat{\mu}_{lk}(n) \leftarrow \mu_{lk}(n)$
12: **else**
13: $\widetilde{\mu}_{lk}(n) \leftarrow \mu_{lk}(n)$
14: **end if**
15: $\mu_{lk}(n) \leftarrow (\widetilde{\mu}_{lk}(n) + \widehat{\mu}_{lk}(n))/2$.
16: **goto Step 2**.
17: **end if**
18: **Step 4** (Return Values)
19: Compute $p_{lk}(s, n)$ by (26)
20: **return** $\{p_m(s, n)\}$, and $\{p_{lk}(s, n), 1 \le s \le N_s\}$

The remaining undetermined parameters in (22) are $\rho_k(n)$'s, i.e., to which UE VC n should be assigned. We prove in Appendix D that

$$\rho_k(n) = \begin{cases} 1, \ k = \arg\max\limits_{k'} \Big\{ w_{lk'}\beta_{lk'}(n) \\ \qquad -\lambda_l \sum\limits_{s=1}^{N_S} \Big[\sum\limits_{m=1}^{l-1} p_m(s, n) + p_{lk}(s, n) \Big] \Big\} \\ 0, \ \text{otherwise} \end{cases} \tag{39}$$

(39) indicates that although we have relaxed binary constraint on $\{\rho_k(n)\}$, it turns out that the optimal $\{\rho_k(n)\}$ under the relaxed constraint$\{0 \le \rho_k(n) \le 1\}$ is binary after all.

Given the optimal VC indicator $\{\rho_k(n)\}$, we have $1 - \sum_{k=1}^{K_l} \rho_k(n) = 0$. The complementary slackness condition then implies that the corresponding φ_{ln} becomes irrelevant. Furthermore, the same condition requires that the second term on the RHS of (17) must be zero. But if $\lambda_l = 0$, there will be no total power constraint and $G_l(\lambda_l)$ becomes unbounded. Hence, we must have

$$\sum_{k=1}^{K_l} \sum_{n=1}^{N} \sum_{s=1}^{N_S} \Big[\sum_{m=1}^{l-1} q_m(s, n) + q_{lk}(s, n) \Big] = P_l \tag{40}$$

which is equivalent to

$$\sum_{n=1}^{N} \Big\{ \sum_{s=1}^{N_S} \Big[\sum_{m=1}^{l-1} q_m(s, n) + q_{lk^*(n)}(s, n) \Big] \Big\} \stackrel{def}{=} \sum_{n=1}^{N} P(n) = P_l \tag{41}$$

where $k^*(n)$ satisfies (39).

C. COMPLETE OPTIMAL RA ALGORITHM AND FAIRNESS CONCERN

The complete process for finding the optimal RA solution is summarized in **Algorithm 2** which consists of two nested loops. Within the inner loop, we apply the OPA algorithm repeatedly to find $\rho_k(n), \ \forall\, n, k$ and the associated optimal $\mu_m(n), \mu_{lk}(n)$ for a given λ_l. In other words, the purpose of the OPA algorithm is to find the optimum power allocation per hop and spatial subchannel $\{p_m(s, n), 1 \le m < l; p_{lk}(s, n), 1 \le s \le N_s\}$, if VC n is to be assigned to UE k with the Lagrangian coefficient λ_l so that one can compute the corresponding achievable rate for the pair (n, k).

The outer loop is responsible for searching for the λ_l that meets the power constraint (40) and stops when λ_l, which control the provisional total power, reaches a steady state. Both OPA and outer loop use the bisection method to find the corresponding Lagrange coefficients.

To estimate the complexity of the complete algorithm, we have to consider both the μ-iteration of **Algorithm OPA** and the λ-iteration of the outer loop. For the former, the bisection method requires $\log_2(w_{lk}/\epsilon_\mu)$ iterations to achieve the predefined accuracy, ϵ_μ being the stopping threshold. The OPA algorithm is activated for every (n, k) pair in the inner loop, hence NK_l OPA operations are required for each given λ_l. The outer loop needs $\log_2((\widetilde{\lambda}_l - \widehat{\lambda}_l)/\epsilon_\lambda)$ iterations to converge, where $[\widehat{\lambda}_l, \widetilde{\lambda}_l]$ is the initial search interval and ϵ_λ is the outer stopping threshold. The above iteration processes are nested, thus the total complexity of Algorithm 1 is thus given by

$$\mathcal{O}\left(NK_l \log_2\left(\frac{\widetilde{\lambda}_l - \widehat{\lambda}_l}{\epsilon_\lambda} \right) \max_k \log_2\left(\frac{w_{lk}}{\epsilon_\mu} \right) \right).$$

It is well known that an optimal RA scheme tends to giving most resources to the UEs with the best CNR and they are often the UEs closest to the serving BS, resulting in unfair RA decisions. To address the fairness concern without redesigning the RA algorithm, we propose a simple solution which restrict a user from obtaining too much VC resource. This is achieved by imposing a maximum VC resource constraint in the optimal RA algorithm. In Algorithm 2, we add lines 10 to 12 and set the maximum allowed VC number per UE,

$\sum_{n=1}^{N} \rho_k(n)$, to $N - \alpha \times (N/K_l)$, $0 < \alpha < 1$. The effect of this restriction can be measured by Jain's fairness index

$$\mathcal{J} = \frac{\left[\sum_l \sum_n \sum_k r_{lk}(n)\right]^2}{N|\mathcal{K}| \sum_l \sum_n \sum_k [r_{lk}(n)]^2}, \tag{42}$$

where, $|\mathcal{K}|$ is the total number of UEs in the serving area, and $r_{lk}(n) = 0$ if VC n is not assign to UE k of SC l, i.e., if the corresponding $\rho_k(n)$ is zero. To distinguish the optimal RA schemes with and without fairness consideration, we refer to the algorithm without lines 10-12 as the greedy RA scheme while that with the maximum VC resource check as the fair RA scheme. Note that we can also improve the fairness performance by assigning proper weights but it turns out this approach is not as effective. Imposing an upper bound on the power allocated to an UE is not as easy and straightforward, for after converting RA-P2 into its dual **RA-P3**, the total power constraint (15d) has been embedded in the new objective $G_l(\lambda_l)$ and is controlled by the Lagrange multiplier λ_l which is updated in each outer iteration.

D. A SUBOPTIMAL RA SCHEME

A candidate approach for reducing the complexity of the optimal algorithm is simplifying the PA task (**Algorithm OPA** is part (line 7) of **Algorithm 1**). We make two such simplification decisions: (i) assign the same power to every VC used in a SC, i.e., $\sum_{s=1}^{N_S}(\sum_{m=1}^{l-1} p_m(s,n)+p_{lk}(s,n)) = P_l/N$, and (ii) all spatial subchannels in each hop of an MBS-UE link are given the same power, i.e., $p_m(s,n) = \bar{p}_m(n)$ and $p_{lk}(s,n) = \bar{p}_{lk}(n)$ and $\bar{P}_m(n) = N_S\bar{p}_m(n)$. Hence, $\bar{P}_{lk}(n) = N_S\bar{p}_{lk}(n)$ is the power allocated to the nth VC in the jth hop for the kth UE served by BS l. With these two simplifications, the remaining tasks in solving (RA-P2) are (i) VC assignment $\{\rho_k(n)\}$ for all $1 \le n \le N$, and (ii) solving N independent subproblems of optimizing the power distribution $\{\bar{P}_m(n), \bar{P}_{lk}(n)\}$ among hops of the same VC link subject to VC power constraint P_l/N.

For VC n used in an l-hop link, we denote the average CSI error induced interchannel interference by $\bar{I}_{lk}(s,n) = \mathbb{E}[I_{lk}(s,n)]$, then from (11) we obtain

$$\bar{I}_{lk}(s,n) = \xi \times \text{diag}\left[\sum_{j=1}^{N_t} |b_{1j}|^2, \ldots, \sum_{j=1}^{N_t} |b_{N_t j}|^2\right]_{ss} \tag{43}$$

Using the above definition, we express the average SNR per spatial subchannel in a BS-BS hop and a BE-UE hop as

$$\bar{\gamma}_m(s,n) = \varrho_m(s,n)\frac{\bar{P}_m(n)}{N_S}$$
$$\bar{\gamma}_{lk}(s,n) = \frac{\varrho_{lk}(s,n)\bar{P}_{lk}(n)}{\bar{I}_{lk}(s,n)\bar{P}_{lk}(n) + N_S},$$

and the corresponding average spatial subchannel capacities as $\bar{r}_m(s,n) = \log_2(1+\bar{\gamma}_m(s,n))$ and $\bar{r}_{lk}(s,n) = \log_2(1+\bar{\gamma}_{lk}(s,n))$, respectively. As $\bar{I}_{lk}(s,n)$ is now independent of

Algorithm 1 A Bisection Based Iterative RA Algorithm for a Multi-Hop ($l > 1$) IAB HetNet

1: **Step 1** (Initialization)
2: Select proper upper and lower bounds, $\widetilde{\lambda}_l$, $\widehat{\lambda}_l$ and a suitable error tolerance ϵ_λ.
3: **Step 2** (Compute $\{p_m(s,n), p_{lk}(s,n)\}$ and $\{\rho_k(n)\}$ based on λ_l.)
4: Set $\lambda_l \leftarrow (\widetilde{\lambda}_l + \widehat{\lambda}_l)/2$.
5: **for** n=1 to N **do**
6: **for** k=1 to K_l **do**
7: Apply **Algorithm OPA** with (n,k) to find $\{p_m(s,n)\}$ and $p_{lk}(n)$.
8: **end for**
9: **for** k=1 to K_l **do** (lines 10-12 should be bypassed if fairness is of no concern)
10: **if** $\sum_{n=1}^{N} \rho_k(n) > N - \alpha \times (N/K_l)$ **then**
11: continue
12: **end if**
13: Compute $\rho_k(n)$ via (39).
14: **end for**
15: **end for**
16: **Step 3** (Find the optimal λ_l which meets (40))
17: **if** $\widetilde{\lambda}_l - \widehat{\lambda}_l > \epsilon_\lambda$ **then**
18: Calculate $\widetilde{P}_l$ via
19: $\sum_{k=1}^{K_l}\sum_{n=1}^{N}\sum_{s=1}^{N_S}\left[\sum_{m=1}^{l-1} q_m(s,n) + q_{lk}(s,n)\right]$.
20: **if** $\widetilde{P}_l < P_l$ **then**
21: $\widetilde{\lambda}_l \leftarrow \lambda_l$
22: **else**
23: $\widehat{\lambda}_l \leftarrow \lambda_l$
24: **end if**
25: $\lambda_l \leftarrow (\widetilde{\lambda}_l + \widehat{\lambda}_l)/2$
26: **goto Step 2**.
27: **end if**
28: **Step 4** (Return Values)
29: **return** $\{p_m(s,n), p_{lk}(s,n)\}$ and $\{\rho_k(n)\}$.

$p_{lk}(s,n)$, the RA problem is simplified to

$$\textbf{RA-P6:} \max_{\{\bar{P}_m, \bar{P}_{lk}\}, \{\rho_k(n)\}} \sum_{k=1}^{K} \rho_{lk}(n) w_{lk} \bar{r}_{lk}(n) \tag{44}$$

subject to

$$\sum_{k=1}^{K} \rho_k(n) \le 1, \ \rho_k(n) \in \{0,1\}, \ \forall\, n \tag{44a}$$
$$\sum_{m=1}^{l-1} \bar{P}_m(n) + \bar{P}_{lk}(n) = \frac{P_l}{N} \tag{44b}$$

where

$$\bar{r}_{lk}(n) = \min_{1<m\le l} \left\{ \sum_{s=1}^{N_S} \bar{r}_m(s,n), \sum_{s=1}^{N_S} \bar{r}_{lk}(s,n) \right\}. \tag{45}$$

Using the large SINR ($\gamma \gg 1$) approximation $\log_2(1+\gamma) \approx \log_2(\gamma)$, we conclude that

Lemma 5: When the SINRs associated with all N_S spatial subchannles are sufficiently high, the optimal PA solution for RA-P6 is given by

$$\bar{P}_m(n) \approx \frac{\Gamma_m(n)}{\sum_{m=1}^{l-1}\Gamma_m(n)+\Gamma_{lk}(n)} \times \frac{P_l}{N}$$
$$\bar{P}_{lk}(n) \approx \frac{\Gamma_{lk}(n)}{\sum_{m=1}^{l-1}\Gamma_m(n)+\Gamma_{lk}(n)} \times \frac{P_l}{N} \tag{46}$$

where the optimal link power distribution ratio $\Gamma_m(n)$ and $\Gamma_{lk}(n)$ are defined as

$$\Gamma_m(n) = \left(\prod_{s=1}^{N_S} \frac{\varrho_1(s,n)}{\varrho_m(s,n)}\right)^{\frac{1}{N_S}}, \quad \Gamma_{lk}(n) = \frac{-a+\sqrt{a^2+4b}}{2} \tag{47}$$

with (a, b) given by

$$a = \sum_{m=1}^{l-1}\Gamma_m(n) - \left(\frac{\bar{I}_{lk}(s,n)P_T}{NN_S}+1\right)\left(\prod_{s=1}^{N_S}\frac{\varrho_1(s,n)}{\varrho_{lk}(s,n)}\right)^{\frac{1}{N_S}} \tag{48a}$$

$$b = \left(\prod_{s=1}^{N_S}\frac{\varrho_1(s,n)}{\varrho_{lk}(s,n)}\right)^{\frac{1}{N_S}} \sum_{m=1}^{l-1}\Gamma_m(n) \tag{48b}$$

In case perfect CSI ($\sigma^2_{h,l} = 0$) is available, we have $\Gamma_{lk}(n) = \left(\prod_{s=1}^{N_S}\frac{\varrho_1(s,n)}{\varrho_{lk}(s,n)}\right)^{\frac{1}{N_S}}$. Furthermore, the weighted rate maximization VC selection rule is

$$\rho_k(n) = \begin{cases} 1, & k = \arg\max\limits_{k'} w_{lk'}\bar{r}_{lk'}(n) \\ 0, & \text{otherwise} \end{cases} \tag{49}$$

Without the bisection search complexity, this suboptimal algorithm requires $\mathcal{O}(NK_l)$ complexity. We summarize this scheme in Algorithm 3.

Algorithm 2 A Suboptimal Algorithm for a Multi-Hop ($l > 1$) HetNet

Step 1 (Power allocaion)
Constant power allocaion based on (46) and (47).
Step 2 (Find $\rho_k(n)$ according to (49))
for n =1 to N **do**
  for k =1 to K_l **do**
    Compute $\bar{r}_{lk}(n)$ based on (45) and the PA decisions made in **Step 1**
  end for
  for k =1 to K_l **do**
    Compute $\rho_k(n)$ via (49).
  end for
end for

E. AN ALTERNATIVE SUBOPTIMAL RA ALGORITHM

We observe that the complexity of the optimal RA scheme is dominated by that consumed by VC assignment (i.e., computing $\rho_k(n)$). If we first solve the VC-UE association based on the simple equal PA method described in Section III-D and then refine the PA part with this fixed VC assignment by invoking the optimal PA scheme of Section III-A, we can improve the performance of **Algorithm 2**. The resulting hybrid procedure is summarized as **Algorithm** 4 in which the VC assignment decision ($\rho_k(n)$) is based on **Algorithm 2**. As **Step 2** needs $\mathcal{O}\left(N\max_k \log_2\left(\frac{w_{lk}}{\epsilon_\mu}\right)\right)$ OPA operations and $\mathcal{O}\left(\log_2\left(\frac{\widetilde{\lambda}_l-\widehat{\lambda}_l}{\epsilon_\lambda}\right)\right)$ iterations between **Step 2** and **Step 3** are needed, the total complexity of this algorithm becomes $\mathcal{O}\left(N\log_2\left(\frac{\widetilde{\lambda}_l-\widehat{\lambda}_l}{\epsilon_\lambda}\right)\max_k \log_2\left(\frac{w_{lk}}{\epsilon_\mu}\right)+NK_l\right)$.

Algorithm 3 A Hybrid RA Algorithm for a Multi-Hop ($l > 1$) HetNet

Step 1 (Compute $\rho_k(n)$)
for n=1 to N **do**
  for k=1 to K_l **do**
    using (45)-(48b) to compute $\bar{r}_{lk}(n)$
  end for
  for k=1 to K_l **do**
    using (49) to compute $\rho_k(n)$.
  end for
end for
Step 2 (Initialization for power allocation)
Select the upper and lower bounds, $\widetilde{\lambda}_l$, $\widehat{\lambda}_l$, for λ_l and a proper error tolerance ϵ_λ.
Step 3 (Compute $p_m(s,n)$ and $p_{lk}(s,n)$)
Set $\lambda_l \leftarrow (\widetilde{\lambda}_l + \widehat{\lambda}_l)/2$.
for n=1 to N **do**
  apply **Algorithm OPA** to (n, k'), where $k' = \arg\max_k \rho_k(n)$ to obtain $\{p_m(s,n), p_{lk}(s,n)\}$
end for
Step 4 (Find the optimal λ_l that satisfies (40))
if $\widetilde{\lambda}_l - \widehat{\lambda}_l > \epsilon_\lambda$ **then**
  Calculate $\widetilde{P}_T$ via
  $\sum_{k=1}^{K_l}\sum_{n=1}^{N}\sum_{s=1}^{N_S}\left[\sum_{m=1}^{l-1} q_m(s,n) + q_{lk}(s,n)\right]$.
  if $\widetilde{P}_T < P_T$ **then**
    $\widetilde{\lambda}_l \leftarrow \lambda_l$
  else
    $\widehat{\lambda}_l \leftarrow \lambda_l$
  end if
  goto Step 3.
end if
Step 5 (Return Values)
return $\{p_m(s,n), p_{lk}(s,n)\}$ **and** $\{\rho_k(n)\}$.

IV. IAB NODE PLACEMENT AND UE ASSOCIATION

A. LONG TERM AVERAGE RATE OF A MULTIHOP LINK

In contrast to the RA decisions in a mobile environment, which is often made in a frame-by-frame manner, the IAB node placement is a long-term issue in the sense the node locations are to be preselected and will remain fixed for a much longer period. Therefore, node placement decisions should be independent of the instantaneous CNRs upon

which an RA decision relies. That is, if we are seeking for a sum rate maximization placement strategy, the rate should be computed with the long term averaged shadowing and SSFC values and by averaging over the UE locations. Against this backdrop, we define

$$\bar{\varrho}_m(s,n) = \frac{\bar{a}}{D_m^{\alpha_L}}, \quad \bar{\varrho}_{lk}(s,n) = \frac{\bar{b}}{d_{lk}^{\alpha_N}}, \tag{50a}$$

$$\bar{a} = \mathbb{E}\left[\frac{\zeta_m|\eta_m(s,n)|^2}{\sigma_w^2}\right], \quad \bar{b} = \mathbb{E}\left[\frac{\zeta_{lk}|\eta_{lk}(s,n)|^2}{\sigma_w^2}\right] \tag{50b}$$

The long term averaged SNR of BBL$(m, m+1)$ is thus given by

$$\langle\gamma_m\rangle = \bar{\varrho}_m p_m, \quad 1 \le m < l \tag{51}$$

and that for spatial subchannel s in BUL(k, l), assuming a fixed UE(k, l) location, is

$$\langle\gamma_{lk}\rangle = \bar{\varrho}_{lk} p_{lk}, \tag{52}$$

Then, the (average) achievable rate per spatial subchannel for UE(l, k) is given by

$$\langle r_{lk}\rangle = \min_{1\le m<l} \left(\langle r_m\rangle, \langle r_{lk}\rangle\right), \tag{53}$$

where $\langle r_m\rangle = \log_2[1+\langle\gamma_m\rangle]$, and $\langle r_{lk}\rangle = \log_2[1+\langle\gamma_{lk}\rangle]$. To compute this rate we need to know the SBS l to UE k distance d_{lk} and the PA decision $\{p_m, p_{lk}, 1 \le m < l\}$. The distance d_{lk} depends on the node location and affects the path loss of the SBS-UE link, therefore the corresponding link capacity.

Recall that **Lemma 2** (section III) says that the optimal RA solution calls for

$$\sum_{s=1}^{N_S} r_m(s,n) = \sum_{s=1}^{N_S} r_{lk}(s,n) \triangleq \beta_{lk}(n) \tag{54}$$

which means that even with the consideration of all instantaneous channel (LSFC and SSFC) and system (PA) parameters, the optimal strategy is to have the same rate on all hops for each VC.

Based on (51) and (52), the equal hope rate constraint (54) becomes

$$\log_2\left(1+\frac{\bar{a}p_m}{D_m^{\alpha_L}}\right) = \log_2\left(1+\frac{\bar{b}p_{lk}}{d_{lk}^{\alpha_N}}\right) \tag{55}$$

or equivalently,

$$\frac{\bar{a}p_m}{D_m^{\alpha_L}} = \frac{\bar{b}p_{lk}}{d_{lk}^{\alpha_N}} = c_{lk}, \tag{56}$$

Since the power allocated to VC n is limited, we have

$$\sum_{m=1}^{l-1} p_m + p_{lk} = c_{lk}\left(\sum_{m=1}^{l-1}\frac{D_m^{\alpha_L}}{\bar{a}} + \frac{d_{lk}^{\alpha_N}}{\bar{b}}\right) = P \tag{57}$$

and

$$\langle r_{lk}\rangle = \log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}d_{lk}^{\alpha_N} + \bar{b}\sum_{m=1}^{l-1} D_m^{\alpha_L}}\right) \tag{58}$$

which indicates that the long term rate for UE(k, l), if granted a VC and power P, now depends on the SBS-UE distance d_{lk} and the inter-BS distances D_m's. The latter will determine the SBS locations and the SBS-UE distance d_{lk} for a fixed UE position. Note that UE(k, l) can denote any UE using any VC and we are interested in the expected long term rate as a function of the SBS location when the expectation is taken over the random UE positions within a SC of interest.

Let $\rho_l(d_l)$ denotes the probability density function (pdf) which characterizes the probability that a cell-l's UE, d_l meters away from its serving SBS, is granted a VC with power P. It plays a role similar to the indicator function $\rho_k(n)$. For the sector of concern whose service area is a Θ-degree circular sector as shown in Fig. 1, the relation between d_l, the UE location in polar coordinate (z_l, ϕ_l) (assuming all UE heights are at zero) and the location of SBS l $\left(\sum_{m=1}^{l-1} L_m, 0, h_{l-1}\right)$ is given by (59), as shown at the bottom of the next page, and depicted in Fig. 3.

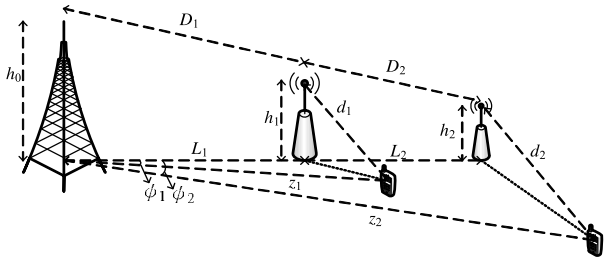

FIGURE 3. The geometry of BS/UE locations and the corresponding BS-UE distances; all UE heights are assumed to be zero.

Assuming a fixed SBS hight h_{l-1}, $\rho(d_l)$ is an implicit function of the SBS and UE locations u_l and (z_l, ϕ_l). Denote the UE location distribution within the circular sector by the pdf $g(z_l, \phi_l)$. The product

$$\rho\left(d_l(u_l; z_l, \phi_l)\right) g_l(z_l, \phi_l) \stackrel{def}{=} p_l\left(d_l(u_l; z_l, \phi_l)\right) \tag{60}$$

is then the pdf characterizing the chance that an UE at (z_l, ϕ_l) can acquire transmission resource under a known RA criterion when the serving SBS is located at u_l. In other words, it is the resource acquisition pdf for the UE at (z_l, ϕ_l) and is function of u_l. When the UEs are uniformly distributed within the sector, ϕ_l will be uniformly distributed over $[-\frac{\Theta}{2}, \frac{\Theta}{2}]$. For UEs in SC l, the conditional location pdf $g(z_l, \phi_l | R_{l-1} \le z_l \le R_l, |\phi| \le \Theta) \stackrel{def}{=} g_l(z_l, \phi_l)$ is given by

$$g_l(z_l, \phi_l) = \frac{2z_l}{\left(R_l^2 - R_{l-1}^2\right)\Theta} \tag{61}$$

The probability of finding an UE d to $d+\triangle$ meters away from the serving BS is proportional to $\pi[(d+\triangle)^2 - d^2]$, i.e., the smaller d is, the less likely we can find an UE. But using the greedy RA approach (**Algorithm 1**), the MBS tends to give resource to the UE with the best BS-UE link CNR. This implies that an UE with a smaller d_l often has a higher probability of being served because of its proximity (smaller path loss) to the serving BS. Taking both factors into account,

we conclude that for a fixed u_l the pdf $P_l(d_l)$ is determined by the distribution of

$$\min_{\substack{R_{l-1}\le x_i\le R_l\\ |\psi_i|\le\Theta/2\\ i=1,2,\cdots,K_l}} d(u_l; x_i, \psi_i), \text{ where } (x_i, \psi_i) \sim g_l(z, \phi) \quad (62)$$

where $\{(x_i, \psi_i), i = 1, 2, \cdots, K_l\}$ are the g_l-distributed locations of K_l UEs in SC l. We depict the resource acquisition pdfs for SC 2 obtained by computing (62) and by simulation for different $u_2 = L_1$ in Fig. 4 (the parameters used in simulation are listed in Table 2 of Section VI). As expected, the pdf is a function of L_1 and for each fixed SBS 1 position, there exists an optimal distance at which the MBS has the highest probability of finding an UE and assigning proper resources. The figure also show that (a) all pdfs can be accurately described by gamma pdfs with proper parameter values and truncations; (b) for a fixed L_1, (62) and the greedy RA simulation predict almost identical pdf peaks, although the latter has heavier tails due to the presence of SSFC and shadowing; (c) when L_1 moves closer to the boundary, the probability of finding an UE at small d's becomes smaller, hence, the pdf peak moves toward a larger d.

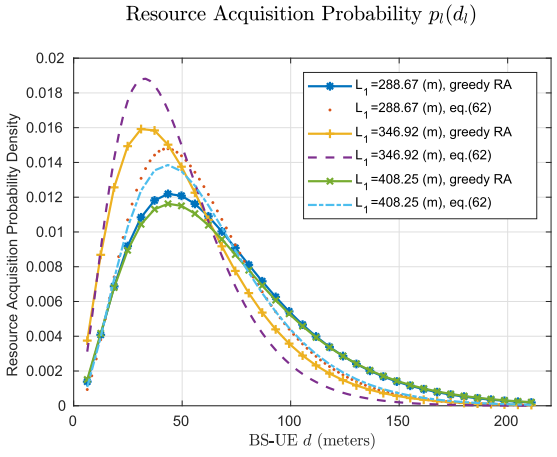

FIGURE 4. Resource acquisition probability as a function of the BS-UE distance. Gamma pdfs with different shape and scale parameters are used to fit the simulated pdfs.

We thus express the average long term achievable single VC link rate as

$$C_l = \mathbb{E}\left[\log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}d_l^{\alpha_N}+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right)\right] \quad (63a)$$

$$= \int_{R_{l-1}}^{R_l}\int_{-\frac{\Theta}{2}}^{\frac{\Theta}{2}} \log_2\left[1+\frac{\bar{a}\bar{b}P}{\bar{a}d_l^{\alpha_N}+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right] \times p_l\left(d_l(u_l; z_l, \phi_l)\right) dz_l d\phi_l, \quad (63b)$$

TABLE 2. Parameter values used in simulations (stdev=standard deviation).

Parameters	Description	Value
R_C	radius of the sector	500 m
Θ	central angle of the sector	60^o
α_L	LoS path-loss exponent	2.2
α_N	NLoS path-loss exponent	3.88
σ_L	stdev of shadowing coefficient (LoS)	3 dB
σ_N	stdev of shadowing coefficient (NLoS)	7 dB
h_B	MBS height	15 m
h_{R1}	SBS 1 height	5 m
h_{R2}	SBS 2 height	5 m
K	number of UEs in the macrocell	18
w_{lk}	UE weighting	1
N_S	number of antennas	2
N	number of subcarriers in a subband	72
Δf_S	subcarrier spacing	60 kHz
Δf_G	bandwidth of the guard band	120 kHz
f_c	central frequency	38 GHz

where we have omitted the notational dependence on k. For the greedy RA scheme, the resource acquisition pdf $p_l(d_l(z_l, \phi_l))$ is a truncated gamma pdf with the truncated domain determined by the SC boundary. The pdf is the same as $g_l(z_l, \phi_l)$ for the fair RA scheme if it assigns all UEs with the same or almost the same amount of resource.

With all channel effects but the path loss have being averaged and system parameters (cell boundaries, SBS height) fixed, the expected rate C_l is a function of the SBS location u_l only. All VCs have the same potential capacity and power constraint. Hence, it is sufficient to consider just a single VC.

B. PLACEMENT OPTIMIZATION

The optimal IAB node locations which maximize (63a) or (63b) can be found by numerical search if $p_l\left(d_l(u_j; z_l, \phi_l\right)$ is known but a general closed-form expression is not available. Nevertheless, as verified in Appendix B, we sill have the following lower bounds

$$C_l \ge \log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}\left[\mathbb{E}(d_l)\right]^{\alpha_N}+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right) \stackrel{def}{=} C_{l1}$$

$$\ge \log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}\left[\mathbb{E}(d_l^2)\right]^{\frac{\alpha_N}{2}}+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right) \stackrel{def}{=} C_{l2}$$

$$\ge \log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}\mathbb{E}[d_l^{\alpha_N}]+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right) \stackrel{def}{=} C_{l3}, \quad (64)$$

This inequality implies that by solving the SBS location parameter u_l that minimizes one of the followings

$$\Omega_{l1} \stackrel{def}{=} \bar{a}\left[\mathbb{E}\left(d_l\right)\right]^{\alpha_N}+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L} \quad (65a)$$

$$d_l(u_l; z_l, \phi_l) = \sqrt{u_l^2+z_l^2-2u_l z_l\cos\phi_l+h_{l-1}^2}, \quad u_l = \sum_{m=1}^{l-1}L_m \quad (59)$$

$$\Omega_{l2} \overset{def}{=} \bar{a}\left[\mathbb{E}(d_l^2)\right]^{\frac{\alpha_N}{2}} + \bar{b}\sum_{m=1}^{l-1} D_m^{\alpha_L} \tag{65b}$$

$$\Omega_{l3} \overset{def}{=} \bar{a}\mathbb{E}\left[d_l^{\alpha_N}\right] + \bar{b}\sum_{m=1}^{l-1} D_m^{\alpha_L} \tag{65c}$$

where

$$\mathbb{E}\left[d_l^m\right] = \int_{R_{l-1}}^{R_l}\int_{-\frac{\Theta}{2}}^{\frac{\Theta}{2}} p_l(d(z_l,\phi_l))[d_l(z_l,\phi_l)]^m \mathrm{d}\phi_l \mathrm{d}z_l \tag{66}$$

we can obtain a series of approximate optimal SBS locations and the approximation accuracies should be consistent with what the inequality has predicted, i.e., the maximizer of (63b), followed by the minimizers of (65a), (65b), and (65c), respectively, with the last one the least accurate. Closed-form expressions of L_m's that minimize Ω_{lk} for arbitrary $\alpha_N \geq 2$ and RA scheme cannot be found. But if $\alpha_N = 2$ or 4, which describes typical LoS and terrestrial NLoS links, we can obtain closed form expressions for the node locations that minimize either Ω_{l2} or Ω_{l3} in case an fair RA scheme is employed. For the 3-SC, 3-hop, system shown in Fig. 1, when $\alpha_N = 2$, we have

$$\Omega_{l2} = \bar{a}\left[\left(u_l^2 + h_{l-1}^2\right) - \frac{8u}{3\Theta}\sin\left(\frac{\Theta}{2}\right)\frac{R_l^3 - R_{l-1}^3}{R_l^2 - R_{l-1}^2} + \frac{\left(R_l^2 + R_{l-1}^2\right)}{2}\right] + \bar{b}\sum_{m=1}^{l-1}\left[L_m^2 + (h_{l-1} - h_l)^2\right] \tag{67}$$

and if $\alpha_N = 4$,

$$\Omega_{l2} = \bar{a}\left[\left(u_l^2 + h_{l-1}^2\right) - \frac{8u}{3\Theta}\sin\left(\frac{\Theta}{2}\right)\frac{R_l^3 - R_{l-1}^3}{R_l^2 - R_{l-1}^2} + \frac{R_l^2 + R_{l-1}^2}{2}\right]^2 + \bar{b}\sum_{m=1}^{l-1}\left[L_m^2 + (h_{l-1} - h_l)^2\right] \tag{68}$$

Similar expressions for Ω_{l3} can be derived.

Hence, we formulate the joint SBSs placement problem as

$$\textbf{PL-P1:} \min_{L_1,L_2} \Omega_{l2},\ l = 2, 3$$
$$\text{subject to } R_1 \leq L_1 \leq R_2, R_2 \leq L_2 \leq R_3 \tag{69}$$

We focus on finding the minimizer of Ω_{l2} instead of (65a) or (65c) as no closed-form expression for Ω_{l1} is available and the lower bound C_{l2} is tighter than C_{l3}. As Ω_{22} is a function of L_1 only, we first find $L_1^* = \arg\min_{L_1} \Omega_{22}$, which is the same as $\arg\max_{L_1} C_{22}$, and then $L_2^* = \arg\min_{L_2} \Omega_{32}(L_1^*) = \arg\max_{L_2} C_{32}(L_1^*)$, where $C_{32}(L_1^*)$ and $\Omega_{32}(L_1^*)$ are computed with $L_1 = L_1^*$. The pair (L_1^*, L_2^*) is then a Pareto-optimal solution of the multi-objective optimization problem (69).

(L_1^*, L_2^*) can be obtained by solving

$$\left.\frac{\partial\Omega_{22}}{\partial L_1}\right|_{L_1=L_1^*} = 0, \quad \left.\frac{\partial\Omega_{32}(L_1^*)}{\partial L_2}\right|_{L_2=L_2^*} = 0, \tag{70}$$

For $\alpha_N = 2$, the above equations lead to

$$L_1^* = \max\left\{\frac{4\bar{a}\sin\frac{\Theta}{2}\left(R_2^3 - R_1^3\right)}{3\Theta\left(\bar{a}+\bar{b}\right)\left(R_2^2 - R_1^2\right)}, R_1\right\} \tag{71a}$$

$$L_1^* + L_2^* = \max\left\{\frac{4\bar{a}\sin\frac{\Theta}{2}\left(R_3^3 - R_2^3\right)}{3\Theta(\bar{a}+\bar{b})\left(R_3^2 - R_2^2\right)} + \frac{\bar{b}L_1^*}{\bar{a}+\bar{b}}, R_2\right\} \tag{71b}$$

where only the lower bounds in the constraints $R_1 \leq L_1 \leq R_2, R_2 \leq L_1 + L_2 \leq R_3$ affect the solution because in our case $\Theta = \pi/3$, $\bar{b}/\bar{a} > 1$, $R_2 = \sqrt{2/3}R_3$ and

$$\frac{4\bar{a}\sin\left(\frac{\Theta}{2}\right)\left(R_2^3 - R_1^3\right)}{3\Theta\left(\bar{a}+\bar{b}\right)\left(R_2^2 - R_1^2\right)} \leq \frac{4\sin\left(\frac{\Theta}{2}\right)(R_2 + R_1)}{3\Theta(1 + \bar{b}/\bar{a})} = \frac{2(R_2 + R_1)}{\pi(1 + \bar{b}/\bar{a})} \leq \frac{2R_2}{\pi} \leq R_2 \tag{72a}$$

$$\frac{4\bar{a}\sin\left(\frac{\Theta}{2}\right)\left(R_3^3 - R_2^3\right)}{3\Theta(\bar{a}+\bar{b})\left(R_3^2 - R_2^2\right)} + \frac{\bar{b}L_1^*}{\bar{a}+\bar{b}} \leq \frac{2(R_3 + R_2) + 2\bar{b}/\bar{a}R_2}{\pi(1 + \bar{b}/\bar{a})} \leq \frac{R_3}{\pi} + \frac{2R_2}{\pi} = \left(\frac{1}{\pi} + \frac{2}{\pi}\sqrt{\frac{2}{3}}\right)R_3 \leq R_3 \tag{72b}$$

When $\alpha_N = 4$, the first part of (70) implies that L_1^* is the solution of $a_3x^3 + a_2x^2 + a_1x + a_0 = 0$, where a_3, a_2, a_1, a_0 are given by

$$a_3 = 1 \tag{73a}$$

$$a_2 = -\frac{3A}{2} \tag{73b}$$

$$a_1 = \frac{1}{2}\left(A^2 + B + 2h_1^2 + \frac{\bar{b}}{\bar{a}}\right) \tag{73c}$$

$$a_0 = -\frac{1}{2}\left(Ah_1^2 + \frac{AB}{2}\right) \tag{73d}$$

and

$$A = \frac{8}{3\Theta}\sin\left(\frac{\Theta}{2}\right)\frac{R_2^3 - R_1^3}{R_2^2 - R_1^2}, \quad B = R_2^2 + R_1^2$$

By using Cardano's formula [27], we obtain

$$L_1^* = S_0 + S_1 - \frac{a_2}{3a_3}, \quad R_1 \leq L_1 \leq R_2 \tag{74}$$

where

$$S_i = \sqrt[3]{T + (-1)^i \times \sqrt{Q^3 + T^2}}, \quad i \in \{0, 1\} \tag{74a}$$

$$Q = \frac{3a_3a_1 - a_2^2}{9a_3^2} \tag{74b}$$

$$T = \frac{9a_3a_2a_1 - 27a_3^2a_0 - 2a_2^3}{54a_3^3} \tag{74c}$$

L_2^* is obtained by solving a similar cubic equation and is derived in Appendix B. The optimal SBS location pair which minimizes Ω_{l3}, i.e., $(L_1^*, L_2^*) = \arg\min_{L_1,L_2}\{\Omega_{23}, \Omega_{33}\}$ can also be similarly derived.

C. UE ASSOCIATION

A conventional approach for UE association, i.e., to which neighboring BS an UE should be linked, is to associate the UE with the BS whose beacon signal strength is the strongest among those broadcasted by neighboring BS's and received

by the UE [12]. This UE association method can lead to serious load-imbalance in the network [25]. For an IAB-UDCN it also causes backhaul and fronthaul overloading. Several central control mechanisms for load-aware energy efficient UE association and PA have been proposed [7]–[9]. They all require that the central control unit has side information such as all the UE and BS locations and instantaneous BS-UE link gains which is the product of the SSFC and LSFC, the latter being the product of the shadowing coefficient and the path loss. Additional system constraints that an UE can be associated with only one BS at a time and each BS can serve a fixed number of UEs with a total transmit power limit are then imposed to avoid load-imbalance.

A simple approach to balance the BS loading is to have the service area partitioned in advance and a BS serves only the UEs located in its designated service area. An admission control mechanism is in place to ensure that the maximum allowed load, i.e., the maximum number of UEs served, is maintained. Like other centralized load-aware UE association schemes, this location based scheme needs the BS and UE location information. The scheme is based on a concept similar to frequency reuse in cellular networks or the fractional frequency reuse scheme for interference control [26]: UEs in different areas are using location-dependent frequency resources. In our case, the resource of concern is the serving BSs. Let $f(x)$ be the pdf of the UE location x over the service area $\mathcal{A}$ and $\{A_j\}_{j=1}^{L}$ is a partition of $\mathcal{A}$ where A_j denote the jth SC. Let K_j is the designed capacity (excluding that reserved for handover UEs) for SC j and there are N UEs within $\mathcal{A}$. Then the probability that a given UE is in SC j is $p_j(x) = \int_{x\in A_j} f(x)dx$. The probability that there are m UEs in cell j follows the binomial law $h_j(m) = \binom{N}{m} p_j^m (1-p_j)^{N-m}$. The probability that there are more UEs in SC j than it can serve is $\sum_{m=K_j+1}^{N} h_j(m) \triangleq F_c(K_j; N, p_j) \triangleq 1 - F(K_j; N, p_j)$, which is called the blocking probability. For a fixed N and p_j, $F_c(n; N, p_j)$ is a function of n. Its inverse function, denoted by $G(\varepsilon, N, p_j)$, represents the integer $n(\varepsilon)$ such that $F_c(n(\varepsilon); N, p_j) = \varepsilon$. Hence, given the UE distribution, and the desired blocking probability, we can easily find the maximum loads K_j's if the SC partition is known or conversely, find the SC partition if K_j's is given.

V. CHANNEL AGING AND FREQUENCY REUSE

The channel aging effects caused by the time lag between the time CSI about a hop is estimated and that when signal is transmitted over that hop is discussed in Section II.C. In this section, we use a model to describe the channel's time-varying behavior and a channel predictor based on that model. To maximize the frequency efficiency, we present a frequency reuse scheme and analyze the resulting co-channel interference.

A. TIME-VARYING CHANNEL AND CHANNEL PREDICTOR

An autoregressive (AR) model is used to approximate Rayleigh fading channel [28] in this paper. A complex AR process of order K can be generated via the time domain recursion

$$h_n = -\sum_{k=1}^{K} a_k h_{n-k} + e_n \tag{75}$$

where e_n is a complex white Gaussian noise process with uncorrelated real and imaginary components. The AR model parpameters consist of the filter coefficients $\{a_1, \ldots, a_K\}$ and the variance σ_e^2 of the driving noise process e_n. Given the desired autocorrelation functions $r_h[k]$ for $k \geq 0$, we can use the Levinson-Durbin recursion to determine $\{a_k\}_{k=1}^{K}$ and σ_e^2, and has a form of

$$\sigma_e^2 = r_h[0] + \sum_{k=1}^{K} a_k r_h[-k] \tag{76}$$

A commonly-used autocorrelation function is the the Jakes model, and in this model, the normalized (unit variance) discrete-time autocorrelation of fading channel coefficients is $r_h[k] = J_0(2\pi f_d T_s |k|)$, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, T_s is the channel sampling duration, and f_d is the maximum Doppler shifts.

Assume we have the past N data samples $\mathbf{h} = [h_{n-N+1}, \ldots, h_n]^T$, the $N \times N$ autocovariance matrix $\mathbf{C}_{hh}$ takes the symmetric Toeplitz form

$$\mathbf{C}_{hh} = \begin{pmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_1 & \phi_0 & \cdots & \phi_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N-1} & \phi_{N-2} & \cdots & \phi_0 \end{pmatrix}, \tag{77}$$

where ϕ_k is the autocorrelation function of the autoregressive process h_n. We wish to predict h_{n+l} for $l \in \mathbb{Z}_+$ based on $\mathbf{h}$. Define the vector $\mathbf{r}_h$ as $\mathbf{r}_h = [\phi_{n+l}, \ldots, \phi_{n+l-N+1}]^T$. The resulting l-step Wiener predictor is given by $\hat{h}_{n+l} = \mathbf{r}_h^T \mathbf{C}_{hh}^{-1} \mathbf{h}$ If we model Jakes' model by the K^{th}-order autoregressive process, the optimal length of data samples for l-step Wiener predictor is equal to K. In other words, the optimal length of data samples N equals the order of the autoregressive process K. Assuming we have the sufficient past data ($N = K$). Therefore, the corresponding l-step Wiener predictor can be written as

$$\hat{h}_{n+l} = \sum_{k=1}^{K} a_n h_{n-k+1}, \tag{78}$$

and α_n is the solution of the equation $\Phi\mathbf{a} = -\mathbf{v}_l$, where Φ is a $K \times K$ autocorrelation function matrix of h_n, $\mathbf{a} = [a_1, \ldots, a_K]^T$, and $\mathbf{v}_l = [\phi_l, \phi_{l+1}, \ldots, \phi_{K-1+l}]^T$. We are interested in the statistical properties of the l-step prediction error $h[n+l] - \hat{h}[n+l]$. Obviously, the prediction error of the l-step Wiener predictor with order K autoregressive process is a zero-mean Gaussian random variable with variance σ_l^2, and this Winener prediction error variance for l-step prediction can be expressed as

$$\sigma_l^2 = \sigma_e^2 + \sum_{k=1}^{l-1} a_k^2 \sigma_{l-k}^2, \quad \sigma_0^2 = 0 \tag{79}$$

B. FREQUENCY REUSE

Frequency reuse is a simple but powerful concept which has been applied in almost all cellular network. The basic idea is to reuse the same frequency band in areas that are geographically separated by a large enough distance such that the resulting co-channel interference (CCI) becomes negligible or tolerable. We consider the simple frequency reuse in the last hop as shown in Fig. 5.

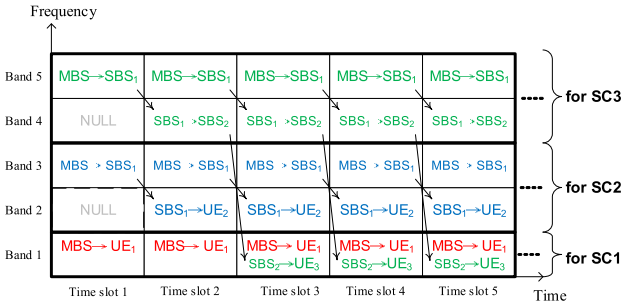

FIGURE 5. A T-F plan/schedule with frequency reuse.

The resulting CCI from MSB to UE$(k, 3, n)$ is given by

$$J_{1k}(s,n) = \rho_k(n) \sum_{k'=1}^{K_3} \rho_{k'}(n) \sum_{s'=1}^{N_S} |\widetilde{\mathbf{H}}_{ss'}|^2 p_{3k'}(s',n) \tag{80}$$

where $\widetilde{\mathbf{H}} = \mathbf{U}_{1k}^H \Xi_{3k} \mathbf{V}_{3k'}$ is the channel response of BUL(k, l, n). The corresponding achievable rate of the sth spatial subchannel for the first hop is

$$r_{1k}(s,n) = \log_2 \left(1 + \frac{\xi_1 |\eta_1(s,n)|^2 p_1(s,n)}{J_{lk}(s,n) + \sigma_w^2} \right) \tag{81}$$

On the other hand, the CCI from SBS 3 to UE$(k, 1, n)$ is given by

$$J_{3k}(s,n) = \rho_k(n) \sum_{k'=1}^{K_1} \rho_{k'}(n) \sum_{s'=1}^{N_S} |\widetilde{\widetilde{\mathbf{H}}}_{ss'}|^2 \tag{82}$$

where $\widetilde{\widetilde{\mathbf{H}}} = \mathbf{U}_{3k}^H \Xi_{1k} \mathbf{V}_{1k'}$ denotes the channel response between the MBS and UE$(k, 3, n)$. The corresponding achievable rate becomes

$$r_{3k}(s,n) = \log_2 \left(1 + \frac{\xi_{3k} |\eta_{3k}(s,n)|^2 p_{3k}(s,n)}{J_{3k}(s,n) + I_{3k}(s,n) + \sigma_w^2} \right) \tag{83}$$

VI. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present the simulated performance to study the effectiveness of our IAB node placement and RA solutions. We consider a MIMO-OFDM based multi-hop heterogeneous IAB network in which both UEs and BSs employ antenna array of size, $N_S = 2$ or 4 with $N = 72$ subcarriers. We have 18 UEs uniformly distributed within the service area–a 60-degree sector of radius $R_C = 500$ meters centered at the MBS with the two IAB nodes (SBS 1, SBS 2) placed on the angle bisector; see Fig. 1. In most cases, we assume that BS-UE (inter-BS) links experience NLoS Rayleigh (LoS Rician) fading with path-loss exponent α_N (α_L) and shadowing parameter $\sigma_{s,N}$ ($\sigma_{s,L}$). In Fig. 12, we examine the scenario when the BS-UE link is described by a distance-dependent mixed NLoS and LoS fading.

The fading process associated with a time-varying channel is modelled as a third-order autoregressive process and the MBS who is in charge of RA applies a Wiener predictor to improve the CSI precision. We define the normalized signal to noise ratio (NSNR) per subcarrier as the averaged SNR measured at the sector edge (MBS-UE distance $= R_c$) when the path loss exponent is $\bar{\alpha} = 2.75$

$$\overline{\text{SNR}} \stackrel{def}{=} \frac{1}{N} \left(\frac{P_l / R_C^{\bar{\alpha}}}{\sigma_w^2} \right) \tag{84}$$

The other parameter values used in our simulation are listed in Table 2.

A. PERFORMANCE OF RA SCHEMES

As an l-hop downlink uses $l\left(N + \Delta f_G / \Delta f_S\right)$ VCs, where $\Delta f_G / \Delta f_S$ is the guard band bandwidth per subband measured in number of VCs, we compute the cell spectrum efficiency (CSE) of SC l by

$$\text{CSE}_l = \frac{1}{l\left(N + \Delta f_G / \Delta f_S\right)} \sum_{k=1}^{K_l} \sum_{n=1}^{N} w_{lk} r_{lk}(n) \tag{85}$$

The sum spectrum efficiency (SSE) of a multi-hop HetNet system is then given by

$$\text{SSE} = \sum_{l=1}^{3} \frac{\text{CSE}_l}{l\left(N + \Delta f_G / \Delta f_S\right)} \tag{86}$$

Fig. 6 compares the SSE performance of the optimal RA scheme (Algorithm 1) and two suboptimal ones (Algorithms 2 and 3). It reveals that Algorithm 3 outperforms Algorithm 2 and yields almost the same SSE as the optimal one. The negligible performance degradation of Algorithm 3 against Algorithm 1 seems to indicate that, although Algorithm 1 may change its VC assignment decision at each iteration, its final decision does not deviate much from that made by Algorithm 2. This is due to the fact that Algorithm 1's PA decisions (26) and (35) depend on the corresponding channels' CNRs, (9)-(10), while Algorithm 2's PA decisions (46)-(48b), although much simpler, are also dependent on the CNRs of the related channels only. Furthermore, both algorithms' VC decisions (39) and (49) are functions of the corresponding PA decisions.

As Algorithm 3 makes the VC assignment decision based on Algorithm 2 and, with this decision, it applies the OPA algorithm iteratively to tune and find the optimal PA. The performance degradation of the latter is a decreasing function of $\overline{\text{SNR}}$ and at $\overline{\text{SNR}} = 10$ dB, it incurs a loss less than 0.3 bps/Hz when used in the 2×2 system. But if we further simplify Algorithm 2 by allocating equal power to each hop so that each hop of an l-hop VC is given transmission power

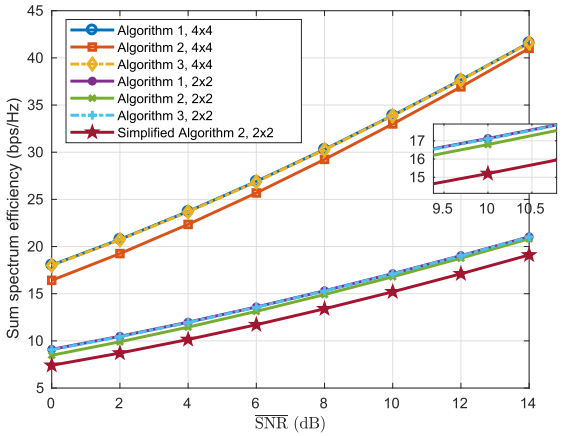

FIGURE 6. SSE (bit/sec/Hz) of a multi-hop HetNet using various RA schemes and different array antenna sizes.

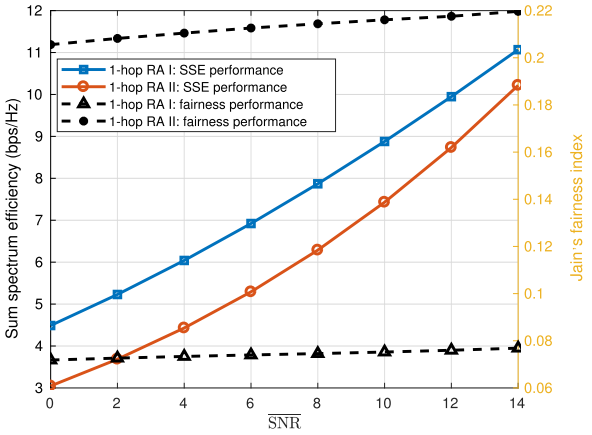

FIGURE 8. SSE and fairness performance of a 2 × 2 MIMO single-hop RA schemes.

of $P_l/(lN)$, the resulting performance (labelled by 'Simplified Algorithm 2) loss increases to more than 2 bps/Hz.

B. ADVANTAGES OF FREQUENCY REUSE AND MULTI-HOP TRANSMISSION

The benefit of frequency reuse is demonstrated in Fig. 7. Obviously, with frequency reuse, all three RA schemes outperform the optimal RA scheme without frequency reuse. The advantage of using multi-hop relays against the single-hop (direct link) scheme can be found in Fig. 8. The single-hop scheme has only the MBS to serve all UEs in the sector covered by three SCs. We consider two single-hop RA schemes: the first scheme, labelled by 1-hop RA-I, has six subbands ($6 \times N$ subcarriers) for serving UEs while the second one, labelled by 1-hop RA-II, allocate two subbands and one third of the total power to serve UEs in each SC. The latter scheme is designed to offer better fair access opportunities but yields poorer SSE performance. As can be seen by comparing these two figures, both single-hop schemes yield much smaller SSEs, indicating that although the single hop system has a larger frequency diversity (more VCs to choose from) and requires only one hop, the improved diversity gain cannot compensate for the much larger path-loss experienced by the BS-UE links in SC 2 and SC 3.

The effect of UE association on the SSE performance can be found by comparing Fig. 8 with Figs. 7-6. To examine the influences of UE association and node locations, we consider an alternate UE association rule that assigns the closest BS to serve the UE of concern and different node placement decisions. In Fig. 9, the SSE performance curve labelled by UEA1 is resulted from our equal serving area UE association rule and the curves labelled by UEA2, UEA3, UEA4 are resulted from the minimum distance rule (MDR) using respectively the optimal node locations, the same node locations as those for UEA1, and the same SBS2 location used by UEA1 but SBS1 is now located at the SC2 right boundary. Both UEA2 and UEA 3 outperform UEA1 by a small margin. This is due to the fact the MDR may obtain

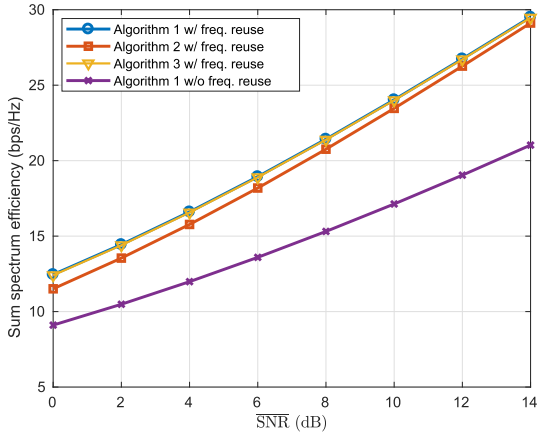

FIGURE 7. SSE performance of a 2 × 2 MIMO multi-hop HetNet with frequency reuse.

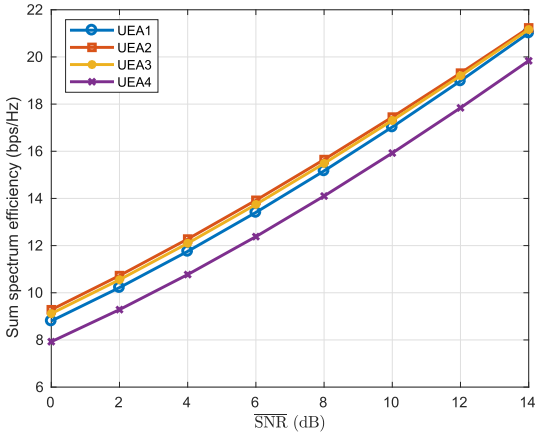

FIGURE 9. The impact of UE association rule on the SSE performance; UEA1 = equal serving area UE association, UEA2, UEA3, UEA4 = MDR with the optimal node locations, MDR using the UEA1 node locations, and MDR using UEA1's SBS2 location while SBS1 is moved to the SC2's right boundary.

better UE-BS link SINRs. But it is also expected that MBS and SBSs will have imbalance loads. The influence of the IAB node locations are to be found by comparing UEA2, UEA3 and UEA4 curves which also indicate that the two different UE association rules have almost the same optimal IAB node locations.

C. CHANNEL AGING AND FAIRNESS ISSUE

Fig. 10 shows the channel aging effect for different time-varying environments parameterized by the normalized Doppler spectrum bandwidth $f_D T$. At 38 GHz carrier frequency and $v = 60$ km/hr vehicle speed, $f_D T = 0.05$ corresponds to a time lag of just 0.024 ms or 1.42 OFDM symbol durations while $f_D T = 0.01, 0.02$ correspond to $v = 12$ and 24 km/hr. We find that significant rate loss incurs at $f_D T = 0.05$ while there is negligible loss at $f_D T = 0.01$ when a proper channel predictor is used. This fact tells us that we must take the channel aging effect into account, especially in a multi-hop network. For multi-hop networks, centralized RA schemes should be employed only to low-mobility UEs if the system is operating at mmWave bands.

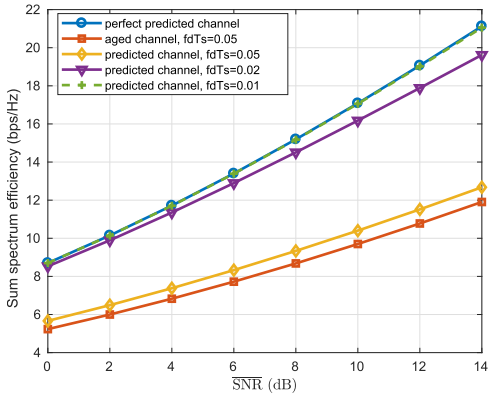

FIGURE 10. The effects of channel aging and channel prediction error on the SSE performance in time-varying environments.

As mentioned in Section III-C, the greedy RA scheme tends to allocate resources to UEs which are closer to the serving BS as the related BS-UE links have better CNRs, resulting highly unfair resource distribution. This behavior is confirmed and the effect of our fair RA scheme is shown in Fig. 11 where the SSE and the corresponding Jain's fairness index performance are plotted as a function of N_u, the maximum number of VCs given to an UE. We see that the larger N_u is (less constraint), the more unfair (smaller $\mathcal{J}$, $\square$ curve) and the better the SSE ($\circ$ curve) becomes. For $N_u = 48$, the Jain's fairness index is about 0.27 and the SSE is close to 15.2 bps/Hz. But if $N_u = 12$, the fairness index is improved to 0.57 while the SSE drops to 7.82 bps/Hz. This is because the RA constraint prohibits an UE with high CNR from obtaining more than its share. The fairness performance of the two direct link RA schemes mentioned before is shown in Fig. 7. Both give much poorer fairness performance against our multi-hop RA schemes and 1-hop RA-I is the worst one,

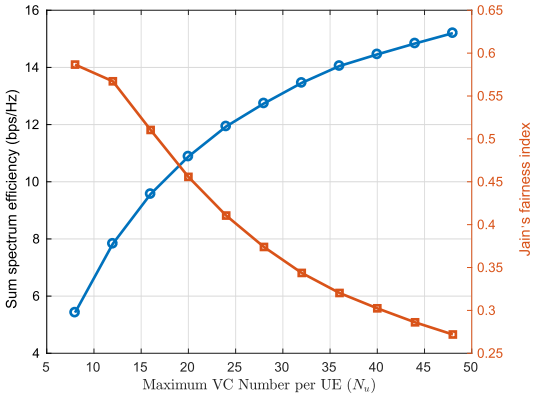

FIGURE 11. SSE (bit/sec/Hz) vs. Jain's fairness index performance of the fair RA scheme at $\overline{\text{SNR}} = 10$ dB.

as expected. The RA-II scheme adopts a RA policy similar to our multi-hop RA schemes which divides the total resources into three equal shares and assign them to three cells thus prevents the MBS from giving most or all resources to the UEs closest to it. In other words, dividing a large service areas into SCs and providing each SC with resources proportional to the average UE numbers each SC is expected to serve has already taken the fairness into account. Within a cell, however, the serving BS still has the tendency to allocate its available resources to close-by UEs. Imposing a maximum per-UE resource constraint is thus a simple and effective solution to achieve the fair access goal.

D. IAB NODE PLACEMENT AND THE EFFECT OF CHANNEL MODELING

As mentioned, for load-balancing, we assign each BS (IAB nodes and MBS) the same service area size which requires $[R_1, R_2, R_3] = [\frac{\sqrt{3}}{3}R_c, \frac{\sqrt{6}}{3}R_c, R_c]$. With a uniform UE distribution, this implies the same average per-cell UE population and hence it is reasonable to allocate the same power to serve each SC.

In Section IV, we have outlined several approaches for deciding the IAB node locations. In Fig. 12, we plot the CSE performance of SC 2 and SC 3 as a function of the serving IAB node locations (L_1 and $L_1 + L_2$), assuming $\alpha_N = 4$. The lower CSE performance curves for both SCs are obtained by using either uniform $p_l(d)$ (for fair access) or the gamma distributed $p_l(d)$ computed by (62) and shown in Fig. 4. We use a predetermined $p_l(d)$ with the node position located in the middle of two boundaries. Since the efficiency computed by (63b) is the average single VC rate instead of sum rate, we normalized the computed CSE curves for both SCs so that they have identical peak values as those CSE obtained by simulation. We found that both approaches give almost the same performance curves for the fair RA scheme. For the greedy RA scheme, discrepancy does exist for SC 1 and SC 2 but both methods predict almost the same L_1^*. For the fair RA scheme, we derive several lower bound based location estimators, we list in Table 3 the optimal SBS 1 and SBS 2 locations L_1^* and $(L_1 + L_2)^*$ estimated by (i) simulating

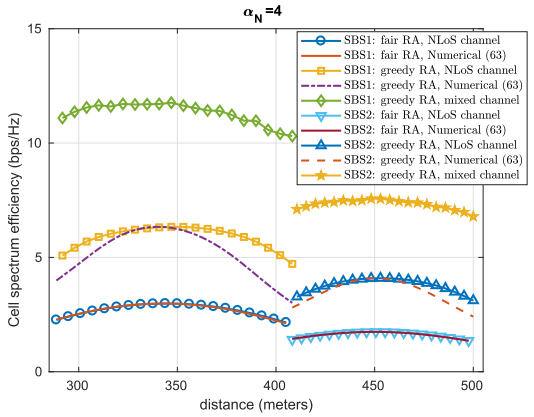

FIGURE 12. CSE vs. IAB node locations measured by their distance to the MBS; $\alpha_N = 4$, SNR = 10 dB. The optimal SBS 1 and SBS 2 locations are at 343.9m and 450.6m for the fair RA scheme and 346.9m and 452.9m for the greedy RA scheme.

TABLE 3. IAB node (SBS) placement results.

Location decision criterion	L_1 (meters) fair RA	$L_1 + L_2$ (meters) fair RA
simulation	343.9	450.6
(63b) numerical	343.9	448.2
(65a) numerical	339	442
(65b) closed-form	336	435
(65c) closed-form	331.2	423.7

the proposed optimal RA schemes, (ii) finding the maximizer of (63b) with uniform or gamma distributed $f(d)$, (iii) computing optimal closed-form locations that minimizes (65b) or (65c), (iv) numerically searching for the minimizer of (65a). Those numbers say that the numerical solution (ii) yields little or no prediction error while those obtained by minimizing (65a)-(65c) are within 2%, 3% and 6% error relative to the true (simulated) one. These performance trends are consistent with what the inequality (64) has predicted. For the greedy RA scheme, the resource acquisition probability $p_l(d)$ is a function of the nod location. Although its peak is not sensitive to the location, its tail part is, therefore, the lower bound based estimators using a pre-determined node location are no longer useful.

Our node placement solutions assume that all BS-UE links are NLoS links. However, measurements on mmWave channels indicate that a BS-UE link may have a distance-dependent LoS probability, P_{LoS} and this probability is a decreasing function of the BS-UE slant distance d. For the urban micro model defined by 3GPP TR 38.901, this probability is given by

$$P_{\text{LoS}}(d) = \begin{cases} 1 & d \le 18 \text{ m} \\ \frac{18}{d} + \left(1 - \frac{18}{d}\right) e^{-d/36} & \text{otherwise} \end{cases} \quad (87)$$

The resulting average normalized rate for the above mixed channel model, which is different from (2), is equal to

$$\mathbb{E}\left\{P_{\text{LoS}}(d) \log_2\left[1 + \gamma_{lk}^{(\text{LoS})}(s,n)\right] + (1 - P_{\text{LoS}})(d) \log_2\left[1 + \gamma_{lk}^{(\text{NLoS})}(s,n)\right]\right\} \quad (88)$$

where the expectation is taken with respect to all fading channel parameters and the BS-UE distance d within the BS service area. $\gamma_{lk}^{(\text{LoS})}(s,n)$, $\gamma_{lk}^{(\text{NLoS})}(s,n)$ have the same expressions as (9) and (10) with the exception that they are now functions of d.

The effect of the mixed fading (87) on the CSE performance and the IAB placement solution is shown in the two upper curves of Fig. 12, with the last hop (the BS-UE link) characterized by the mixed fading model. The figure reveals that the optimal IAB node locations are within 1% of each other for both mixed and pure NLoS environment settings although the CSE for the mixed model is much better, as expected. The same figure also shows the impact of the RA criterion on the system performance and IAB node placement decisions: the greedy RA scheme outperform the fair RA method and places the IAB node a little bit further away from the MBS.

For comparison, we plot the CSE performance when $\alpha_N = 2$ in Fig. 13. With $\alpha_N = 2$, each UE in an SC has LoS link with its serving BS and the channel (link) state is dominated by the link distance. As the length of the first hop is larger than the distance $R_2 - R_1$, to have a minimum worst hop attenuation, SBS 1 is forced to be located at the lower boundary of SC 2. The left CSE performance curve reveals that this is indeed the case for SC 2 and the optimal SBS 1 location is at the lower cell boundary. For SC 3, the corresponding optimal IAB node location is not at its lower boundary but the CSE performance shows no significant variation for locations in between R_2 and the optimal one (427.1 m) as the (performance) dominant (worst) hop is still the first hop.

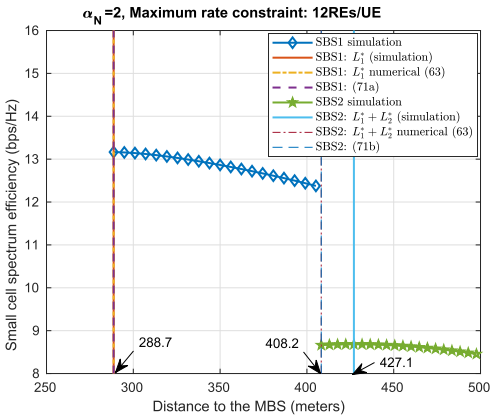

FIGURE 13. CSE performance as a function of the IAB node location (parameterized by its distance to the MBS); fair access RA, $\alpha_N = 2$, SNR = 10 dB. The optimal SBS 1 and SBS 2 locations are at 288.7 m, and 427.1 m (vertical dot-dashed lines) while (71a), (71b) predict that they are at 288.7m and 408.2m (vertical dash line).

VII. CONCLUSION

The issues investigated in this paper have to do with multi-hop heterogeneous IAB networks using MIMO-OFDMA signals, which is a proper candidate model for describing mmWave based UDNs. While our focus is on RA and IAB node placement, various related system design issues are studied as well.

For RA, we develop three RA (VC and power assignments) algorithms for weighted sum rate maximization with and without maximum per UE resource constraint. Two practical low-complexity suboptimal solutions are proposed which explore multi-hop link and LoS channel properties. For IAB node placement, we present a systematic method to solve the IAB node placement problem when the UE association rule is known and apply it specifically to a deterministic UE association rule which guarantees mean load balance. In general, the optimal node locations can be obtained by numerical approach only but for two special propagation loss models, we are able to derive closed-form expressions for the near-optimal (lower bounds-optimized) IAB node locations. We also discuss the effects of channel aging, channel reuse and a proper channel predictor. The impacts of fairness concern on both RA and node placement are examined and the influence of maximum per UE VC constraint and multi-hop connection are discussed by comparing with the single-hop system with and without resource pre-partition.

Numerical results show that although the optimal RA algorithm yields the best performance, the proposed suboptimal RA algorithms suffer small or negligible sum rate performance degradation. Our node placement solutions are verified by simulations and shown to be effective and accurate. The same approach can be applied to different cellular topologies if the UE distribution and association rule are known. We show that our RA algorithms with channel aging consideration can effectively compensate for the performance loss if the time-varying effect is not serious, which means the centralized RA control can be used for low-mobility UEs only. The performance gain of frequency-reuse with proper time-frequency schedule is shown to be significant and so is the multi-hop structure over single-hop systems.

APPENDIX A
THE DERIVATION OF (64)

As C_l defined by (63a) is concave in both d_l and $d_l^{\alpha_N}$, applying Jensen's inequality gives

$$C_l \geq \log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}\mathbb{E}[d_l^{\alpha_N}]+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right) \tag{89a}$$

$$C_l \geq \log_2\left(1+\frac{\bar{a}\bar{b}P}{\bar{a}[\mathbb{E}(d_l)]^{\alpha_N}+\bar{b}\sum_{m=1}^{l-1}D_m^{\alpha_L}}\right) \tag{90a}$$

For $\alpha_N \geq 2$ the power mean inequality

$$\mathbb{E}[x^s]^{1/s} \leq \mathbb{E}[x^t]^{1/t},\ \forall\ -\infty < s \leq t < \infty \tag{91}$$

implies that

$$\mathbb{E}(d_l) \leq \left[\mathbb{E}(d_l^2)\right]^{1/2} \leq \left[\mathbb{E}(d_l^{\alpha_N})\right]^{1/\alpha_N}, \tag{92}$$

and thus

$$[\mathbb{E}(d_l)]^{\alpha_N} \leq \left[\mathbb{E}(d_l^2)\right]^{\frac{\alpha_N}{2}} \leq \mathbb{E}[d_l^{\alpha_N}], \tag{93}$$

which then lead to (64).

APPENDIX B
THE OPTIMAL SBS2 LOCATION

To find the optimal location of SBS2, we begin with the condition $\partial\Omega_3/\partial L_2 = 0$ and, after some algebraic manipulations, find that $L_1+L_2 = u$ must satisfies $b_3u^3+b_2u^2+b_1u+b_0 = 0$, where

$$b_3 = 1 \tag{94a}$$

$$b_2 = -\frac{3A}{2} \tag{94b}$$

$$b_1 = \frac{1}{2}\left(A^2+B+2h_2^2+\frac{\bar{b}}{\bar{a}}\right) \tag{94c}$$

$$b_0 = -\frac{1}{2}\left(Ah_2^2+\frac{AB}{2}+\bar{b}L_1^*\right) \tag{94d}$$

and

$$A = \frac{8}{3\Theta}\sin\left(\frac{\Theta}{2}\right)\frac{R_3^3-R_2^3}{R_3^2-R_2^2}, \quad B = R_3^2+R_2^2 \tag{95}$$

The unique $R_2 \leq u = L_1+L_2 \leq R_3$ that solves (94a) is

$$u^* = (L_1+L_2)^* = s_0+s_1-\frac{b_2}{3b_3} \tag{96}$$

where

$$s_i = \sqrt[3]{t+(-1)^i\times\sqrt{q^3+t^2}}, \quad i\in\{0,1\} \tag{96a}$$

$$q = \frac{3b_3b_1-b_2^2}{9b_3^2} \tag{96b}$$

$$t = \frac{9b_3b_2b_1-27b_3^2b_0-2b_2^3}{54b_3^3} \tag{96c}$$

Hence, $L_2^* = (L_1+L_2)^* - L_1^*$.

APPENDIX C
STRONG DUALITY OF (RA-P2)

We first convert **RA-P2** into the equivalent standard optimization form

$$\textbf{RA-P2a:}\ \max_{\{q_m,q_{lk}\},\{\rho_k\},\{t_{lk}\}} F_0 \overset{def}{=} \sum_{k=1}^{K_l}\sum_{n=1}^{N} w_{lk}t_{lk}(n) \tag{97}$$

subject to

$$t_{lk}(n)-\rho_k(n)r_{lk}(n) \leq 0, \quad \forall\ (n,k,m) \tag{97a}$$

$$\sum_{k=1}^{K_l}\rho_k(n)-1 \leq 0, \quad \forall\ n \tag{97b}$$

$$-\rho_k(n) \leq 0, \quad \forall\ (n,k) \tag{97c}$$

$$\rho_k(n)-1 \leq 0, \quad \forall\ (n,k) \tag{97d}$$

$$\sum_{k=1}^{K_l}\sum_{n=1}^{N}\sum_{s=1}^{N_S}\left[\sum_{m=1}^{l-1}q_m(s,n)+q_{lk}(s,n)\right]-P_l \leq 0 \tag{97e}$$

$$-q_m(s,n) \leq 0, \quad \forall\ (n,k,m,s) \tag{97f}$$

$$-q_{lk}(s,n) \leq 0, \quad \forall\ (n,k,s) \tag{97g}$$

We denote the inequality constraints (97a)-(97g) by $F_i \leq 0, i = 1,\cdots,M$, and $M = 1+N+2NK_l+NK_ll+NK_lN_Sl$.

Then, the feasible domain of **RA-P2a** can be expressed as $\mathcal{D}_{\text{RA-P2a}} = \bigcap_{i=0}^{M} \mathbf{dom} F_i$, where the set $\mathbf{dom} F_i$ represents the domain of F_i. We use **relint** $\mathcal{D}_{\text{RA-P2a}}$ to denote the relative interior of the domain $\mathcal{D}_{\text{RA-P2a}}$ and **int** $\mathcal{D}_{\text{RA-P2a}}$ to denote the interior of $\mathcal{D}_{\text{RA-P2a}}$.

Consider the point

$$\begin{aligned} Z^{\diamond} &= \left(t_{lk}(n), q_m(s,n), \rho_k(n)\right) \\ &= \left(0, \frac{\psi P_l}{K_l N N_S l}, \frac{\psi}{K_l}\right), \quad \forall\, (n,k,m,s) \end{aligned} \tag{98}$$

where $\psi \in \mathbb{R}_+$ and $\psi < 1$. It is easy to see that $F_i(Z^{\diamond}) < 0,\ i = 1, \cdots, M$. Thus, $Z^{\diamond} \in \mathcal{D}_{\text{RA-P2}}$ and the Slater's condition is satisfied. To prove that $Z^{\diamond} \in$ **int** $\mathcal{D}_{\text{RA-P2}}$ whence $\mathcal{D}_{\text{RA-P2}}$ has nonempty interior, we notice that, for a given $\psi \in \mathbb{R}_+$ and $\psi < 1$, $\exists$ a two-norm ball $\mathcal{B}\left(Z^{\diamond}, \varepsilon\right) \subseteq \mathcal{D}_{\text{RA-P2}}$ and the two-norm ball radius $\varepsilon = \min\{\psi P_T / K N_S N_A l, \psi/K\} > 0$. A two-norm ball $\mathcal{B}\left(Z^{\diamond}, \varepsilon\right) \subseteq \mathcal{D}_{\text{RA-P2}}$ means $Z^{\diamond} \in$ **int** $\mathcal{D}_{\text{RA-P2}}$. In this case, it also implies that there exist a $Z^{\diamond} \in$ **relint** $\mathcal{D}_{\text{RA-P2}}$ with $F_i(Z^{\diamond}) < 0,\ i = 1, \cdots, M$. The strong duality holds for **RA-P2a** and **RA-P2**.

APPENDIX D
OPTIMAL VC SOLUTIONS

Using (28) and the complementary slackness requirements of (15a), we rewrite the Lagrangian $\mathcal{L}_{\text{RA-P4}}$ as

$$\begin{aligned} \mathcal{L}_{\text{RA-P4}} &= \sum_{k=1}^{K_l} w_{lk} \rho_k(n) \beta_{lk}(n) \\ &\quad - \lambda_l \sum_{k=1}^{K_l} \sum_{s=1}^{N_S} \left[\sum_{m=1}^{l-1} \rho_k(n) p_m(s,n) + \rho_k(n) p_{lk}(s,n) \right] \\ &\quad + \varphi_{ln} \left(1 - \sum_{k=1}^{K_l} \rho_k(n)\right), \quad \lambda_l \geq 0,\ \varphi_{ln} \geq 0 \\ &= \sum_{k=1}^{K_l} \rho_k(n) \mathcal{L}'_{\text{RA-P4}}(k) + \varphi_{ln}, \end{aligned} \tag{99}$$

where $\mathcal{L}'_{\text{RA-P4}}(k) \overset{def}{=} \partial \mathcal{L}_{\text{RA-P4}} / \partial \rho_k(n)$. Clearly, $\mathcal{L}_{\text{RA-P4}}$ is linear in $\rho_k(n)$ as

$$\begin{aligned} \mathcal{L}'_{\text{RA-P4}}(k) &= w_{lk} \beta_{lk}(n) - \varphi_{ln} \\ &\quad - \lambda_l \sum_{s=1}^{N_S} \left(\sum_{m=1}^{l-1} p_m(s,n) + p_{lk}(s,n) \right) \end{aligned} \tag{100}$$

is independent of $\rho_k(n)$. Hence, unless $\mathcal{L}'_{\text{RA-P4}}(k) = 0$, the extreme values of $\mathcal{L}_{\text{RA-P4}}$ occur at the boundary points, i.e., either at $\rho_k(n) = 1$ or $\rho_k(n) = 0$. But the complementary slackness condition on $\mathcal{L}_{\text{RA-P4}}$ requires that the solution

$$\{\rho_k^*(n)\} = \arg \max_{\{\rho_k(n)\}} \mathcal{L}_{\text{RA-P4}} \tag{101}$$

satisfies $\varphi_{ln} \left(1 - \sum_{k=1}^{K_l} \rho_k^*(n)\right) = 0$ which implies that either $\varphi_{ln} > 0$ and $\left(1 - \sum_{k=1}^{K_l} \rho_k^*(n)\right) = 0$ or $\varphi_{ln} = 0$ and $\left(1 - \sum_{k=1}^{K_l} \rho_k^*(n)\right) > 0$. The later case is obviously not feasible as it says that $\rho_k(n) = 0,\ \forall\, k$, i.e., VC n is not used and $\mathcal{L}_{\text{RA-P4}} = 0$. Therefore, we have $\sum_{k=1}^{K_l} \rho_k^*(n) = 1$ and $\rho_k^*(n) = 1$ for a unique $1 \leq k \leq K_l$. It is easy to see the unique k, denoted by $k^\star$ is given by

$$\begin{aligned} k^\star = \arg \max_k \Bigg\{ & w_{lk} \beta_{lk}(n) \\ & - \lambda_l \sum_{s=1}^{N_S} \left(\sum_{m=1}^{l-1} p_m(s,n) + p_{lk}(s,n) \right) \Bigg\} \end{aligned} \tag{102}$$

and the corresponding optimal Lagrangian is

$$\begin{aligned} \mathcal{L}^{\star}_{\text{RA-P4}} &= w_{lk^\star} \beta_{lk^\star}(n) \\ &\quad - \lambda_l \sum_{s=1}^{N_S} \left(\sum_{m=1}^{l-1} p_m(s,n) + p_{lk^\star}(s,n) \right) \\ &\geq \varphi_{ln} > 0 \end{aligned} \tag{103}$$

As for the case when $\mathcal{L}'_{\text{RA-P4}}(k) \overset{def}{=} \partial \mathcal{L}_{\text{RA-P4}} / \partial \rho_k(n) = 0$, for some k or all k, then $\mathcal{L}_{\text{RA-P4}}(k)$ is constant for those k's and the corresponding optimal $\rho_k(n)$ can assume any value in the interval [0, 1], 0 and 1 included. The rule (102) thus is valid without loss of generality.

REFERENCES

[1] M. Ding, D. Lopez-Perez, H. Claussen, and M. A. Kaafar, "On the fundamental characteristics of ultra-dense small cell networks," *IEEE Netw.*, vol. 32, no. 3, pp. 92–100, May 2018.

[2] S. Alireza Banani, A. W. Eckford, and R. S. Adve, "Analyzing dependent placements of small cells in a two-layer heterogeneous network with a rate coverage constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9801–9816, Dec. 2016.

[3] X. Ge, S. Tu, G. Mao, and C. X. Wang, "5g ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.

[4] X. Ge, S. Tu, G. Mao, V. K. N. Lau, and L. Pan, "Cost efficiency optimization of 5G wireless backhaul networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 12, pp. 2796–2810, Dec. 2019.

[5] R. Tao, W. Liu, X. Chu, and J. Zhang, "An energy saving small cell sleeping mechanism with cell range expansion in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2451–2463, May 2019.

[6] *Study on Integrated Access and Backhaul*, Standard TR 38.874 (Rel. 15), 3GPP, 2018.

[7] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. M. Leung, and H. V. Poor, "Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1936–1947, Sep. 2017.

[8] Y. L. Lee, T. C. Chuah, A. A. El-Saleh, and J. Loo, "User association for backhaul load balancing with quality of service provisioning for heterogeneous networks," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2338–2341, Nov. 2018.

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

[10] T. Zhou, Z. Liu, J. Zhao, C. Li, and L. Yang, "Joint user association and power control for load balancing in downlink heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2582–2593, Mar. 2018.

[11] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2713, Jun. 2013.

[12] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in hetnets: Old myths and open problems," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[13] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5G: Performance analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2669–2684, Dec. 2019.

[14] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 1394–1398.
[15] C. Yui Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
[16] X. Xiao, X. Tao, and J. Lu, "QoS-aware energy-efficient radio resource scheduling in multi-user OFDMA systems," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 75–78, Jan. 2013.
[17] S. Gortzen and A. Schmeink, "Optimality of dual methods for discrete multiuser multicarrier resource allocation problems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3810–3817, Oct. 2012.
[18] M. Tao, Y.-C. Liang, and F. Zhang, "Resource allocation for delay differentiated traffic in multiuser OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2190–2201, Jun. 2008.
[19] T. Li, H. Yu, and S. Liu, "Adaptive power allocation for decode-and-forward MIMO-OFDM relay system," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2009, pp. 1–4.
[20] Y.-B. Lin, W.-H. Wu, and Y. T. Su, "Optimal and suboptimal resource allocations for multi-hop MIMO-OFDMA networks," in *Proc. IEEE 23rd Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 502–506.
[21] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-Service model and learning-based resource management," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5621–5635, Nov. 2018.
[22] F. Bohagen, P. Orten, and G. E. Oien, "Design of capacity-optimal highrank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1420–1425, Apr. 2007.
[23] G. Liu, F. R. Yu, H. Ji, and V. C. M. Leung, "Energy-efficient resource allocation in cellular networks with shared full-duplex relaying," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3711–3724, Aug. 2015.
[24] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.
[25] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.
[26] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4294–4305, Dec. 2011.
[27] S. Gottwald, W. Gellert, and M. Hellwich, *The VNR Concise Encyclopedia of Mathematics*, vol. 1, 2nd ed. Springer, Feb. 1990.
[28] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.

JIUN Y. LAI received the B.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2001, and the M.S. degree in communications engineering from National Chiao Tung University, Hsinchu, in 2003, where he is currently pursuing the Ph.D. degree with the Institute of Communications Engineering. He joined Information and Communications Research Laboratories, Industrial Technology Research Institute, in 2003, where he has been working in the area of emerging wireless communication system development. He has published several international conference papers and holds several patents in U.S., Chinese, and Taiwan. His research interests include novel digital signal processing in high-performance base station, radio resource allocation in wireless networks, and emerging wireless communication technology.

WU-HSIU WU received the B.S. and M.S. degrees in communications engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2012 and 2013, respectively. He joined MediaTek Inc., Hsinchu, in 2014, and has participated in 5G NR chipset development. His research interests include signal processing for communication systems, radio resource allocation in wireless networks, and architectures for new wireless communication systems.

YU T. SU (Life Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Southern California, USA, in 1983. From 1983 to 1989, he was with LinCom Corporation, Los Angeles, USA, where he was involved in various satellite communication and measurement systems design. He was a Corporate Scientist when he left LinCom. Since 1989, he has been with National Chiao Tung University, Hsinchu, Taiwan, where he was an Associative Dean of the College of Electrical and Computer Engineering, from 2004 to 2007, was the Head of the Communications Engineering Department, from 2001 to 2003, and served as a Deputy Director, from 1997 to 2000, and is currently a Professor with the Department of Electrical and Computer Engineering and also affiliated with the Microelectronic and Information Systems Research Center. From 2005 to 2008, he was the Area Coordinator of Taiwan National Science Council's Telecommunications Program. His main research interests include communication systems, networking, and statistical signal processing.