

Power-Domain Interference Graph Estimation for Full-Duplex Millimeter-Wave Backhauling

Haorui Li^{ID}, *Student Member, IEEE*, Daqian Ding^{ID}, *Student Member, IEEE*, Yibo Pi^{ID}, *Member, IEEE*,
and Xudong Wang^{ID}, *Fellow, IEEE*

Abstract—Traditional wisdom for network resource management allocates separate frequency-time resources for measurement and data transmission tasks. As a result, the two types of tasks have to compete for resources, and a heavy measurement task inevitably reduces available resources for data transmission. This prevents interference graph estimation (IGE), a heavy yet important measurement task, from being widely used in practice. To resolve this issue, we propose to use power as a new dimension for interference measurement in full-duplex millimeter-wave backhaul networks, such that data transmission and measurement can be done simultaneously using the same frequency-time resources. Our core insight is to consider the mmWave network as a linear system, where the received power of a node is a linear combination of the channel gains. By controlling the powers of transmitters, we can find unique solutions for the channel gains of interference links and use them to estimate the interference. To accomplish resource allocation and IGE simultaneously, we jointly optimize resource allocation and IGE with power control. Extensive simulations show that significant links in the interference graph can be accurately estimated with minimal extra power consumption, independent of the time and carrier frequency offsets between nodes.

Index Terms—mmWave backhauling, interference graph estimation, resource allocation.

I. INTRODUCTION

NETWORK densification is a key mechanism to increase spectrum reuse such that the network capacity can be greatly boosted for future explosive growth of user demand. However, denser small cells will inevitably increase the deployment cost if small cells are connected with a fiber backhaul [2]. For fast and cost-effective deployment of dense small cells, millimeter-wave (mmWave) backhauling is proposed as a promising backhaul alternative. In recent years, mmWave backhauling has been adopted in the integrated access and backhaul (IAB) networks for the deployment of 5G and beyond systems [3], [4], [5], which jointly optimizes the access

and backhaul networks for the overall network performance. With the self-interference cancellation (SIC) techniques [6], [7], full-duplex capability can be integrated for mmWave backhauling. Full-duplex mmWave backhauling can alleviate the network bottleneck, which greatly improves the latency and throughput of IAB networks [8].

In mmWave communications, to combat large path loss in high-frequency bands, the transmitters generally employ beamforming techniques to direct their beams towards the receivers, resulting in less interference between links. Despite that, under dense deployment, mmWave networks are shown to still operate in the interference-limited regime: the network capacity is significantly affected by the interference among nodes [9]. Full-duplex communication allows more concurrent transmissions in the backhaul networks but also incurs more severe and complicated interference [10]. It is thus critical to effectively estimate and manage interference for resource allocation in full-duplex mmWave backhaul networks. However, compared to extensive research on interference management, interference estimation is less explored. Most of the existing works focus on resource allocation under the assumption that interference can be effectively and accurately estimated [11], [12]. Our work complements existing efforts on resource allocation by estimating the interference graph that depicts the interference between any pair of mmWave links in the network. Efficient interference graph estimation (IGE) is crucial for the practical use of resource allocation algorithms in full-duplex mmWave backhauling.

In dense mmWave networks, interference is depicted in problem formulations in different forms. The simplest form is a conflict graph (or contention graph) dictating if each pair of links can be active at the same time [13], [14]. Based on the conflict graph, scheduling plans of concurrent transmissions can be determined [15], [16]. In a conflict graph, links not allowed for concurrent transmissions significantly interfere with each other. To construct a conflict graph, we typically need to estimate the interference between each pair of links and compare it against a threshold [17]. However, as discussed in [18], it incurs high overheads to directly measure inter-link interference: a network with n links includes $O(n^2)$ pairs of links to be measured. It is possible to reduce the measurement overheads by measuring only a subset of links and inferring the rest using matrix completion [19] or graph embedding techniques [20]. However, these techniques leverage node or link similarities to predict interference distribution, incapable of guaranteeing link-level interference estimation accuracy for

Manuscript received 31 October 2023; revised 4 March 2024 and 3 May 2024; accepted 3 May 2024. Date of publication 29 May 2024; date of current version 11 October 2024. This work was supported by the National Natural Science Foundation of China under Grant 62201346. An earlier version of this paper was presented at the IEEE GLOBECOM 2023 in Malaysia in December, 2023 [DOI: 10.1109/GLOBECOM54140.2023.10437959]. The associate editor coordinating the review of this article and approving it for publication was C.-K. Wen. (*Corresponding author: Yibo Pi.*)

The authors are with the UM-SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: haorui.li@sjtu.edu.cn; daqian.ding@sjtu.edu.cn; yibo.pi@sjtu.edu.cn; wxudong@sjtu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2024.3403150>.

Digital Object Identifier 10.1109/TWC.2024.3403150

1536-1276 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

the construction of conflict graphs. Moreover, conflict graphs depict the binary relations between links, which cannot capture the additive nature of interference. As a result, conflict graphs cannot effectively avoid significant interference when multiple concurrent links exist.

The second form to characterize interference in mmWave communications is modelling, which can be used to estimate the magnitude of interference and to capture its additive nature. With modelling, the channel and antenna gains for each mmWave link can be easily obtained to estimate the interference. Modelling the channel and antenna patterns requires choosing proper model parameters based on the propagation environment and there exist a variety of choices. A mmWave channel could consist of either a single line of sight (LOS) component, purely scattering components, or a mix of both the LOS and scattering components [21], [22]. Each component could further experience different degrees of path loss and fading [23], [24]. In outdoor spaces, path loss is typically assumed to be inversely proportional to the square of the distance and frequency, whereas the path loss exponent becomes larger in environment with rich obstacles [23]. Similarly, the fading channel gain is assumed following different distributions depending on the propagation environment [24]. It is, therefore, rather difficult to model a complex propagation environment, which could be best depicted with a combination of dynamic channel models. Moreover, estimating the directional antenna gain often requires approximating the actual antenna patterns, which cannot be simply approximated by a sectorized antenna model due to irregular radiation shapes caused by hardware imperfections, as demonstrated in real-world experiments [25]. Both the complexities and errors in modelling the channel and antenna patterns make it impractical to estimate the interference accurately. This leads to poor resource allocation decisions and, consequently, degraded network performance.

More recently, considering the complexities in both interference modelling and measurements, machine learning techniques have been used to learn a direct mapping from network information to resource allocation decisions without directly estimating the interference. In [18], a deep learning approach is proposed to use the geographic locations of nodes as input and learn link scheduling decisions from a large scale of network layouts to achieve the maximum sum rate in device-to-device networks. Further improvement using graph embedding and deep neural networks reduces the scale of training data from hundreds of thousands to hundreds of network layouts [26]. Similarly, deep learning models have also been used to improve other performance metrics, e.g., age of information (AoI) [27]. These machine learning techniques typically learn the mapping from static network information to link scheduling decisions, not adaptive to fast changing network conditions, e.g., fading.

In cellular networks, to support coordinated multi-point transmission, 4G Long Term Evolution Advanced (LTE-A) standards introduce channel state information interference measurement (CSI-IM) reference signals to measure the interference between BSs and UEs. Specifically, when an UE reports its received signal strength (RSS) on a resource

element (RE) where no signal is transmitted by its serving BS, the reported RSS is then the interference from other BSs on the subcarrier of the RE [28]. For scheduling purposes, interference has to be measured separately for each interfering BS, which requires operations telling UEs which RBs to measure and which RBs to be used for reporting back measurements. This incurs heavy measurement overheads when the number of interfering BSs is large, especially in dense networks. More importantly, this process requires symbol-level synchronization between BSs and accurate estimation of carrier frequency offsets (CFOs) between UEs and their interfering BSs [29]. However, strict time synchronization is difficult in a fiber backhaul [30] and becomes even harder in a multi-hop wireless backhaul [31].

Motivated by these challenges, we want to 1) efficiently estimate the interference graph, depicting the channel gains of interference links, for full-duplex mmWave backhaul networks under time offsets (TO) and CFO, and 2) use the interference graph to improve the efficiency of resource allocation. We propose to use power as a new dimension for interference graph estimation such that data transmission and measurement can be done simultaneously with the same frequency-time resources. Our core insight is that the received power of a node can be expressed as a linear combination of the channel gains. In a time-slotted network, manipulating the powers of transmitters enables each receiving node to have different received powers across time slots, i.e., a group of linear equations for each node. If the powers of transmitters form a full-rank matrix, the channel gains have a unique solution. Compared to the reference signal (RS)-based IGE, sensitive to TO and CFO, our approach achieves IGE in the power domain, making its performance robust to TO and CFO.

In summary, our contributions in this paper are as follows.

- We show that the mmWave network can be considered as a linear system and propose to estimate the interference graph by manipulating the transmit powers of nodes, which is robust to TO and CFO. We further derive the error bounds of the estimated channel gains for both communication and interference links.
- We formulate a multi-objective optimization problem to jointly optimize the energy efficiency in resource allocation and the condition number in IGE, such that IGE can be conducted accurately and simultaneously with resource allocation using the same frequency-time resources. We design iterative algorithms to solve the problem and prove the convergence of these algorithms.
- We evaluate our approach with extensive simulations under various network settings and show that our approach can accomplish resource allocation and IGE simultaneously with minimal overhead on power consumption.

Note that compared to our previous conference paper [1], we have added substantial additional material in this paper, including the practical power control scheme for achieving power linearity, the bounds of channel gain estimation errors, and a new multi-objective formulation and solution of the joint optimization problem for resource allocation and IGE. The rest of this paper is organized as follows. Section II presents the

Notation	Definition
$x_s[i]$	The i -th transmitted sample of node s
$y_s[i]$	The i -th received sample of node s
$X_s[i]$	The modulated symbol on the i -th subcarrier of node s
$X_s^{I,\max}$	The maximum value of $X_s^I[i]$'s of node s
$X_s^{Q,\max}$	The maximum value of $X_s^Q[i]$'s of node s
$g_{(s,d),(k,z)}$	The equivalent channel gain from link (s,d) to (k,z)
$\mathbf{g}_{(k,z)}$	Vector of equiv. channel gains from other links to (k,z)
$v[i]$	The i -th sample of noise at a single antenna
$p_s^{tx}[i]$	The expected transmit power of node s in the i -th block
$\bar{p}_s^{tx}[i]$	The average transmit power of node s in the i -th block
$p_s^{rx}[i]$	The expected received power of node s in the i -th block
$\bar{p}_s^{rx}[i]$	The average received power of node s in the i -th block
N_c	Number of subcarriers
N_g	Length of the cyclic prefix
N_s	Number of samples in an OFDM symbol
$\mathbf{f}_s$	Errors in the received powers of node s across blocks

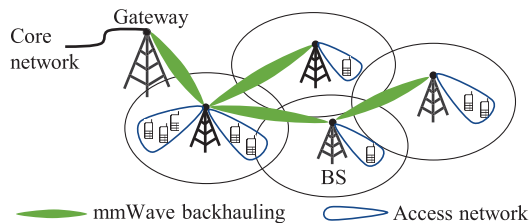


Fig. 1. Multi-hop mmWave backhaul network.

II. SYSTEM MODEL

A. Millimeter-Wave Backhauling

We consider a multi-hop mmWave backhaul network with K BSs, one of which serves as a gateway to the core network, as shown in Figure 1. Each BS is equipped with a full-duplex radio such that it can transmit data to one node and receive data from another node simultaneously. Self-interference cancellation (SIC) techniques are employed at each BS to cancel self-interference. All BSs are assumed to be time-synchronized and scheduled to transmit and receive data at specified time slots. The scheduling decision is made at the gateway in a centralized way, which requires each BS to report local power measurements of received signals for IGE. After a scheduling plan is made, it is disseminated from the gateway to all BSs. Both the data collection and dissemination for network management purposes are done through the backhaul network

We consider the CFO and TO between BSs in the wireless backhaul network. Since BSs are not mobile, the Doppler frequency shift is negligible and CFO is mainly caused by the oscillator mismatch in frequency [29]. The TO arises from the synchronization errors among BSs as well as the difference in the propagation delays when considering the signals from multiple transmitters to the same receiver [32]. We assume that no cooperative transmission is used and that each receiver is assumed to receive data from a single transmitter. For demodulation, each receiver only needs to estimate the CFO of the transmitter and will not compensate the CFOs of other transmitters. Moreover, per the 5G standards [33], the subcarrier spacing for mmWave communications could be as large as 960kHz, equating to approximately $1\mu\text{s}$ symbol time and $0.07\mu\text{s}$ CP length. Since nanosecond-level time synchronization is hard to achieve even in wired networks [30], let alone in wireless backhauling networks [31], we assume that TO could be larger than CP length.

B. Channel Model

Let (s, z) be the directional link from BS s to z , and $\mathcal{E}$ be the set of all directional links. For simplicity, each BS is assumed to have M transmit antennas and M receive antennas. The channel from BS s to z is denoted as $\mathbf{H}_{(s,z)} \in \mathbb{C}^{M \times M}$. When BS s is transmitting to BS z , the transmit beamforming and receive combining vectors of BSs s and z are $\tilde{\gamma}_s \in \mathbb{C}^{M \times 1}$ and $\tilde{\omega}_z \in \mathbb{C}^{M \times 1}$, respectively. Since BSs are typically deployed at high places with fixed locations, the channel coherence time between BSs is as large as hundreds of milliseconds for mmWave communications, even in windy days [34]. We thus can assume that $\mathbf{H}_{(s,z)}$ is approximately constant during the process of IGE, which takes much less time than the coherence time.¹ We denote the residual CFO of BS s with respect to BS z as $\phi_{(s,z)}$, normalized to the subcarrier spacing with a range between -0.5 and 0.5 , and the TO of link (s, d) as $\mu_{(s,d)}$. Let $x_z[i]$ and $y_z[i]$ be the i -th transmitted and received samples of nodes z , respectively. The i -th received sample of node z with beamforming towards node k can be expressed as

$$y_z[i] = \sum_{(s,d) \in \mathcal{E} \setminus \{(z,k)\}} e^{j2\pi i \phi_{(s,z)}} \tilde{\omega}_z^T \mathbf{H}_{(s,z)} \tilde{\gamma}_s x_s[i - \mu_{(s,d)}] + \underbrace{\sqrt{\eta} \tilde{\omega}_z^T \mathbf{H}_z^{SI} \tilde{\gamma}_z x_z[i]}_{r_z^{SI}[i]} + \tilde{\omega}_z^T \mathbf{v}[i], \quad (1)$$

where $\mathcal{E}$ is the set of all links, (z, k) is the directional link originating from node z , $\mathbf{H}_z^{SI}$ is the self-interference channel, η is the SIC capability provided by the SIC techniques,

¹Our goal is not to estimate $\mathbf{H}_{(s,z)}$ for precoding between the transmitters and receivers, but rather to estimate the gains of interference channels for resource allocation, as detailed in Section II-C.

$r_z^{SI}[i]$ is the residual self-interference (SI), and $v[i] \sim \mathcal{CN}(\mathbf{0}, \sigma_v^2 \mathbf{I})$ is the noise variance at each antenna of the receiving BS. Let $h_{(s,d),(k,z)}^{\text{eq}} = e^{j2\pi i \phi(s,z)} \tilde{\omega}_z^T \mathbf{H}_{(s,z)} \tilde{\gamma}_s$ be the equivalent channel from link (s,d) to (k,z) considering transmit beamforming and receive combining. Similarly, let $h_z^{SI} = \sqrt{\eta} \tilde{\omega}_z^T \mathbf{H}_z^{SI} \tilde{\gamma}_z$ be the equivalent residual SI channel. We can rewrite Eq. (1) as

$$y_z[i] = \sum_{(s,d) \in \mathcal{E}} h_{(s,d),(k,z)}^{\text{eq}} x_s[i - \mu_{(s,d)}] + n_z[i], \quad (2)$$

where $h_{(z,d),(k,z)}^{\text{eq}} = h_z^{SI}$ is the equivalent residual SI channel and $n_z[i] = \tilde{\omega}_z^T v[i]$. Since beam coherence time is longer than channel coherence time [35], we can consider that $h_{(s,d),(k,z)}^{\text{eq}}$ and h_z^{SI} are approximately constant during the channel coherence time. We want to estimate the magnitudes of these equivalent channels and subsequently use them for resource allocation. The time-domain sample $x_s[i]$ is the inverse discrete Fourier transform (IDFT) of the modulated symbols, $\{X_s[r]\}_{r=0}^{N_c-1}$, i.e.,

$$x_s[i] = \frac{1}{\sqrt{N_c}} \sum_{r=0}^{N_c-1} X_s[r] e^{j2\pi ir/N_c}, 0 \leq i \leq N_c + N_g - 1, \quad (3)$$

where $X_s[r]$ is the modulated symbol on subcarrier r transmitted by BS s , and N_g is the length of the cyclic prefix (CP).

C. Interference Model

Resource allocation problems involving interference need to adopt an interference model, where interference is typically assumed to be additive. We present the additive interference model for full-duplex mmWave networks and relate the interference model with the above channel model, such that our IGE approach can be easily integrated with the existing resource allocation problems. In the mmWave network, each node employs the beamforming techniques to generate a radiation pattern with a main beam and multiple side lobes. Due to hardware constraints, the mmWave nodes could have limited codebook size for beamforming and irregular radiation patterns [25], which makes interference nulling difficult. We thus consider the interference from both the main beam and the side lobes of transmitters. In mmWave communications, interference depends on the directions of both transmission and reception. Suppose that nodes s and k are transmitting to nodes d and z , respectively. Let $g_{(s,d),(k,z)}$ denotes the equivalent channel gain from link (s,d) to link (k,z) , where $g_{(z,d),(k,z)}$ is the equivalent residual SI channel gain of node z . The equivalent channel $g_{(s,d),(k,z)}$ is a product of the transmission gain at node s , the channel gain of the interference link from node s to z , and the reception gain at node z , denoted as $g_{(s,d),(k,z)}^t, g_{(s,d),(k,z)}^c$, and $g_{(s,d),(k,z)}^r$, respectively. Let p_s^{tx} be the transmit power of node s to node d , and p_z^{rx} be the received power of node z from node k . The received power of node z can be expressed as

$$p_z^{rx} = \sum_{(s,d) \in \mathcal{E}} \underbrace{g_{(s,d),(k,z)}^t g_{(s,d),(k,z)}^c g_{(s,d),(k,z)}^r}_{g_{(s,d),(k,z)}} p_s^{tx} + W_z,$$

where W_z is the noise power of node z . Suppose that node k is transmitting to node z , we can express the signal-to-interference-plus-noise ratio (SINR) at node z as

$$\text{SINR}_{(k,z)} = \frac{g_{(k,z),(k,z)} p_k^{tx}}{W_z + \sum_{(s,d) \in \mathcal{E} \setminus (k,z)} g_{(s,d),(k,z)} p_s^{tx}}, \quad (4)$$

where $g_{(s,d),(k,z)} = |h_{(s,d),(k,z)}^{\text{eq}}|^2$.

III. INTERFERENCE GRAPH ESTIMATION: A POWER-DOMAIN APPROACH

In this section, we present the feasibility of power-domain IGE and design a practical power control scheme. After that, we analyze and bound the channel gain estimation error.

A. Estimating Interference Graph With Power Control

Lemma 1: In full-duplex mmWave networks, if $\mathbb{E}[|X_k[i]|^2] = \mathbb{E}[|X_k[j]|^2]$ for $i \neq j$ and $\mathbb{E}[X_k[i]] = 0$ for all i 's and k 's, the expected receive power of node z with beamforming towards node k is a linear combination of the equivalent channel gains, independent of TOs and CFOs, i.e.,

$$\mathbb{E}[|y_z[i]|^2] = \sum_{(s,d) \in \mathcal{E}} g_{(s,d),(k,z)} \mathbb{E}[|x_s[i]|^2] + W_z, \quad (5)$$

where W_z is the expected thermal noise power.

Proof: Please refer to Appendix A. ■

Let s_i and d_i be the senders and receiver of the i -th directional link, respectively, and $p_{s_i}^{tx}[t]$ and $p_{d_i}^{rx}[t]$ be the transmit and receive powers of nodes s_i and d_i at time t .

Theorem 1: The equivalent channel gains between links can have a unique solution, if the transmit powers of senders of different links over time satisfy

$$\text{rank}(\mathbf{P}) = |\mathcal{E}|, \quad (6)$$

where $\mathbf{P} = [\mathbf{p}_{s_1}^{tx}, \dots, \mathbf{p}_{s_{|\mathcal{E}|}}^{tx}]$ and $\mathbf{p}_{s_i}^{tx} = [p_{s_i}^{tx}[j], \dots, p_{s_i}^{tx}[j + n - 1]]^T$ includes the transmit powers of the sender of link (s_i, d_i) from time j to $j + n - 1$.

Proof: Suppose that node z directs beamforming towards node k . Let $\mathbf{p}_z^{rx} = [p_z^{rx}[j], \dots, p_z^{rx}[j + n - 1]]^T$ be the receive powers of node z including the self-interference at different times, and $\mathbf{w} = [W_z[j], \dots, W_z[j + n - 1]]^T$ be the vector of noise power. According to Eq. (5), we can have

$$\mathbf{p}_z^{rx} = \mathbf{P} \mathbf{g}_{(k,z)} + \mathbf{w}, \quad (7)$$

where $\mathbf{g}_{(k,z)} = [g_{(s_1, d_1), (k, z)}, \dots, g_{(s_{|\mathcal{E}|}, d_{|\mathcal{E}|}), (k, z)}]^T$, $(s_i, d_i) \in \mathcal{E}$, and $g_{(s_i, d_i), (k, z)}$ is the residual SI channel when $s_i = z$. Since $\mathbf{P} \in \mathbb{R}^{n \times |\mathcal{E}|}$ and $n \geq |\mathcal{E}|$, $\mathbf{g}_{(k,z)}$ has a unique solution if $\text{rank}(\mathbf{P}) = |\mathcal{E}|$. ■

Theorem 1 shows that it is feasible to estimate channel gains if the actual transmit powers can approximate the expected transmit powers and form a full-rank matrix. We can also see that compared to the RS-based approach, sensitive to TO and CFO, our approach conducts IGE in the power domain, making its performance robust to TO and CFO. The estimated channel gains can be then used in Eq. (4) to estimate the SINR for resource allocation problems.

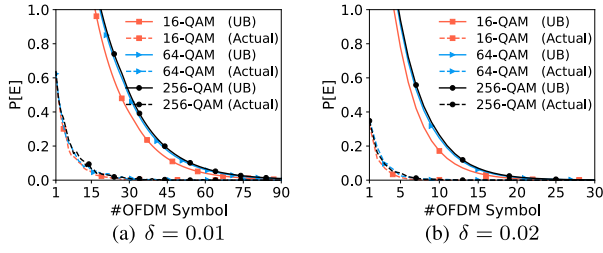


Fig. 2. Error bounds under square M -QAM modulations, where $\mathbb{P}[E]$ denotes the probability of the approximation error falling beyond the tolerable range.

B. Approximating the Expected Transmit Power

Suppose that the node k has N_k modulated symbols to send on each subcarrier, equating to N_k OFDM symbols including $N_k N_s$ time-domain samples, where $N_s = N_c + N_g$ is the number of time-domain samples within an OFDM symbol. Let $X_k^I[i]$ and $X_k^R[i]$ be the imaginary and real parts of $X_k[i]$, respectively. The maximum values of $X_k^I[i]$ and $X_k^R[i]$ are denoted as $X_k^{I,\max}$ and $X_k^{Q,\max}$. We further denote $x_k^I[i]$ and $x_k^Q[i]$ as the I/Q components of $x_k[i]$ and the average transmit power over N_k OFDM symbols is $\bar{p}_k^{tx} = \frac{1}{N_k N_s} \sum_{j=1}^{N_k N_s} |x_k[j]|^2$ and the expectation of $\bar{p}_k^{tx}$ is denoted as $\mu_k = \mathbb{E}[\bar{p}_k^{tx}]$. We want to approximate μ_k with $\bar{p}_k^{tx}$ and derive the condition for the difference $|\bar{p}_k^{tx} - \mu_k|$ to be less than $\mu_k \delta$, where $0 \leq \delta \leq 1$ is the tolerable approximation error.

Theorem 2: Assuming that CP-OFDM is used for modulation and that modulated symbols are chosen with equal probability from the constellation such that $\mathbb{E}[X_k[i]] = 0$ and $\text{Var}(X_k^I[i]) = \text{Var}(X_k^R[i]) = \sigma_k^2$ for all i 's and k 's, the average transmit power can be bounded as

$$\mathbb{P}[|\bar{p}_k^{tx} - \mu_k| \geq \mu_k \delta] \leq 2 \min_{\Delta} \left\{ e^{f_1(\Delta)} + e^{f_2(\Delta)} \right\},$$

with

$$f_1(\Delta) = -\frac{N_k N_g (N_c^2 - N_c) \sigma_k^4}{4(X_k^{I,\max})^4} Q\left(\frac{(X_k^{I,\max})^2 N_s \Delta}{N_g (N_c - 1) \sigma_k^4}\right),$$

and

$$f_2(\Delta) = -\frac{N_k N_c \eta_k^2}{b_k^2} Q\left(\frac{b_k(\mu_k \delta - \Delta)}{\eta_k^2}\right),$$

where Δ is chosen to minimize the sum of $e^{f_1(\Delta)}$ and $e^{f_2(\Delta)}$, $Q(t) = (1+t) \log(1+t)$, $b_k = 2(X_k^{I,\max})^2 - 2\sigma_k^2$, and $\eta_k^2 = \mathbb{E}[|X_k[i]|^4] - 4\sigma_k^4$.

Proof: Please refer to the Appendix B. ■

Remark 1: For modulations with symbols of equal energy, e.g., quadrature phase shift keying (QPSK) and binary phase-shift keying (BPSK), the average power $\bar{p}_k^{tx}$ is equal to μ_k , indicating that $\mathbb{P}[|\bar{p}_k^{tx} - \mu_k| \geq \mu_k \delta] = 0$ for $\delta > 0$.

Using Theorem 2, we show that it is feasible to approximate the expected transmit power with a small number of OFDM symbols. Let E denote the event that $|\bar{p}_k^{tx} - \mu_k|$ is greater than $\mu_k \delta$. We want $\mathbb{P}[E]$ to be controlled within a tolerable range. Fig. 2 shows the simulation results for $\mathbb{P}[E]$'s under different square M -quadrature amplitude modulation

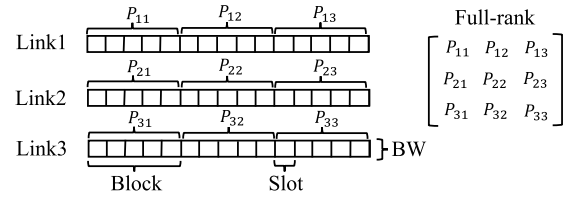


Fig. 3. Power control scheme for interference graph estimation.

(M -QAM) modulations, where $N_c = 1024$ and $N_g = 72$. When $\delta = 0.01$, the solid lines in Fig. 2(a) show that more than 78 OFDM symbols are needed to reduce the upper bound (UB) of $\mathbb{P}[E]$ to 0.01 or lower for 16-QAM modulation. For 64-QAM modulation, this number increases to 86 OFDM symbols. From 64-QAM to 256-QAM modulations, the UBs are almost the same. To understand $\mathbb{P}[E]$ in random cases, we randomly select modulated symbols with equal probability from the constellation and check if event E occurs. This process is repeated 10,000 times under each symbol size (i.e., the number of OFDM symbols used) to calculate $\mathbb{P}[E]$. The dashed lines show that the probability of event E in random cases is much less than the upper bound. Using only about 25 OFDM symbols can control $\mathbb{P}[E]$ to 0.01 or lower. Relaxing the approximation error (δ) to 0.02 can further reduce the required number of OFDM symbols to 5 for achieving the same $\mathbb{P}[E]$. From the above results, we can conclude that it is possible to approximate the expected transmit power with a reasonable number of OFDM symbols.

C. Power Control Scheme for Interference Graph Estimation

Theorems 1 and 2 together indicate that we can design a simple power control scheme for IGE in a time-slotted mmWave multi-hop network. As shown in Fig. 3, to approximate the expected transmit power, each link has to transmit with the same power over several consecutive time slots, referred to as a *block*. The number of consecutive time slots (denoted as n_s) in a block is determined based on Theorem 2. Since links may differ in modulation schemes, blocks of different lengths are needed for each link. For design simplicity, we enforce all links to use the same block length. Since this example includes three links, a 3-by-3 full-rank transmit power matrix needs to be constructed. Suppose that a network includes n_l links. A total number of $n_s n_l$ slots is required for the IGE process. Reducing n_s helps accelerate the IGE process at the cost of an increasing approximation error to the expected transmit power. The trade-off between the time consumption and accuracy of IGE will be further discussed in Section V-B.

From Theorem 1, we know that the expected received power is a linear combination of the equivalent channel gains and the expected transmit powers, i.e., $\mathbf{p}_z^{rx} = \mathbf{P} \mathbf{g}_{(k,z)} + \mathbf{w}$, which can be rewritten as

$$\mathbf{P} \mathbf{g}_{(k,z)} = \mathbf{b}_z, \quad (8)$$

where $\mathbf{b}_z = \mathbf{p}_z^{rx} - \mathbf{w}$. However, as the number of OFDM symbols used for transmission is limited, $\mathbf{b}_z$ is not directly measurable. The receiver can only measure the actual received

power, denoted as $\bar{b}_z$, by averaging over the received symbols in each block. We estimate the channel gains as

$$\mathbf{P}\hat{\mathbf{g}}_{(k,z)} = \bar{b}_z, \quad (9)$$

where $\hat{\mathbf{g}}_{(k,z)} = [\hat{g}_{(s_1,d_1),(k,z)}, \dots, \hat{g}_{(s_{|\mathcal{E}|},d_{|\mathcal{E}|}),(k,z)}]^T$ denotes the estimated equivalent channel gains from other links to link (k, z) .

D. Error Analysis

This section presents two types of channel estimation errors and their bounds. The *per-link error* is the difference between $\mathbf{g}_{(k,z)}$ and $\hat{\mathbf{g}}_{(k,z)}$, which describes the channel estimation errors between link (k, z) and each of the rest. The *per-pair error* is the channel estimation error between two specific links, which is of more granularity. Both the estimation errors will be considered in the formulation of the resource allocation problems in Section IV.

Suppose that n_b blocks are used for the IGE process, where $n_b \geq |\mathcal{E}|$. Let $\mathbf{f}_z \in \mathbb{R}^{n_b \times 1}$ be the difference between the actual and the expected received powers of node z , i.e., $\bar{b}_z = \mathbf{b}_z + \mathbf{f}_z$. Let $\epsilon_b = \|\mathbf{f}_z\|/\|\mathbf{b}_z\|$ be the relative perturbation for the received power.

Theorem 3: (Per-link error) *The channel estimation error for link (k, z) can be bound as*

$$\|\hat{\mathbf{g}}_{(k,z)} - \mathbf{g}_{(k,z)}\| \leq \kappa(\mathbf{P})\|\mathbf{g}_{(k,z)}\|\epsilon_b$$

where $\kappa(\mathbf{P})$ is the condition number of $\mathbf{P}$.

Proof: Recall that $\mathbf{P} \in \mathbb{R}^{n_b \times |\mathcal{E}|}$ and $n_b \geq |\mathcal{E}|$, where $\mathbf{P}$ is a column full-rank matrix with the left inverse $\mathbf{P}^\dagger$. Combining Eq. (8) and (9), we have $\hat{\mathbf{g}}_{(k,z)} - \mathbf{g}_{(k,z)} = \mathbf{P}^\dagger \mathbf{f}_{(k,z)}$. Hence,

$$\begin{aligned} \frac{\|\hat{\mathbf{g}}_{(k,z)} - \mathbf{g}_{(k,z)}\|}{\|\mathbf{g}_{(k,z)}\|} &\leq \frac{\|\mathbf{P}^\dagger\|\|\mathbf{f}_z\|}{\|\mathbf{g}_{(k,z)}\|} \leq \|\mathbf{P}^\dagger\|\|\mathbf{P}\| \frac{\|\mathbf{f}_z\|}{\|\mathbf{P}\|\|\mathbf{g}_{(k,z)}\|} \\ &\leq \|\mathbf{P}^\dagger\|\|\mathbf{P}\| \frac{\|\mathbf{f}_z\|}{\|\mathbf{b}_z\|} = \kappa(\mathbf{P})\epsilon_b. \end{aligned}$$

Theorem 3 indicates that we can lower the channel estimation bound for link (k, z) by reducing either $\kappa(\mathbf{P})$ or ϵ_b . We can minimize $\kappa(\mathbf{P})$ by carefully choosing the transmit powers of nodes, as will be further discussed in Section IV. We can further decrease $\kappa(\mathbf{P})$ by using a transmit power matrix of a larger size.

Theorem 4: *If the transmit powers of nodes can be controlled, for $\mathbf{P} \in \mathbb{R}^{n_b \times |\mathcal{E}|}$, the condition number of $\mathbf{P}$ can be reduced by using a larger n_b .*

Proof: Let $\mathbf{P}_i = [\mathbf{p}^{tx}[1], \dots, \mathbf{p}^{tx}[j], \dots, \mathbf{p}^{tx}[i]]^T$ be a submatrix consisting of the first i rows of $\mathbf{P}$, where $\mathbf{p}^{tx}[j] = [p_{s_1}^{tx}[j], \dots, p_{s_{|\mathcal{E}|}}^{tx}[j]]$ and $i = |\mathcal{E}|, \dots, n_b$. Let α_i and β_i be the largest and smallest singular values of $\mathbf{P}_i$, respectively, and $k_i = \alpha_i/\beta_i$ be the condition number. It has been shown in [36] that $\alpha_{i+1}^2 \leq \alpha_i^2 + (\mathbf{u}_{i+1}^T \mathbf{p}[i+1])^2$ and $\beta_{i+1}^2 \geq \beta_i^2 + (\mathbf{v}_{i+1}^T \mathbf{p}[i+1])^2$, where $\mathbf{u}_{i+1}$ and $\mathbf{v}_{i+1}$ are the eigenvectors of $\mathbf{P}_{i+1}^T \mathbf{P}_{i+1}$ corresponding to α_i^2 and β_i^2 , respectively. Hence,

$$k_{i+1}^2 \leq k_i^2 \times \frac{1 + (\mathbf{u}_{i+1}^T \mathbf{p}^{tx}[i+1])^2/\alpha_i^2}{1 + (\mathbf{v}_{i+1}^T \mathbf{p}^{tx}[i+1])^2/\beta_i^2}.$$

Since $\alpha_i^2 \geq \beta_i^2$, if $\mathbf{v}_{i+1}^T \mathbf{p}^{tx}[i+1] > \mathbf{u}_{i+1}^T \mathbf{p}^{tx}[i+1]$, we can conclude $k_{i+1} < k_i$. The condition can be satisfied if we let the transmit power matrix $\mathbf{P}_{i+1} = [\mathbf{P}_i; \mathbf{p}^{tx}[i+1]]$, where $\mathbf{p}^{tx}[i+1] = \mathbf{v}_{i+1}$. This means that using a larger transmit matrix helps reduce the condition number of $\mathbf{P}$. ■

We will show how the choice of n_b decreases the condition number of $\mathbf{P}$ with experiments in Section V-B.

Lemma 2: *Assuming that $\mathbf{f}_z$ is known, the estimation error for the equivalent channel gain between a pair of links can be bounded as*

$$|\hat{g}_{(s,d),(k,z)} - g_{(s,d),(k,z)}| \leq \|\mathbf{r}_i\| \|\mathbf{f}_z\|, \quad (11)$$

where i is the index of $g_{(s,d),(k,z)}$ in $\mathbf{g}_{(k,z)}$ and $\mathbf{r}_i$ is the i -th row of $\mathbf{P}^\dagger$, the left inverse of $\mathbf{P}$.

Proof: Recall that we want to estimate $\hat{\mathbf{g}}_{(k,z)}$ from the linear system, $\mathbf{P}\hat{\mathbf{g}}_{(k,z)} = \mathbf{b}_z + \mathbf{f}_z$. Based on the perturbation theory in [37], the component-wise error can be written as $|\hat{g}_{(s,d),(k,z)} - g_{(s,d),(k,z)}| = \|\mathbf{r}_i\| \|\mathbf{b}\| \epsilon_b \cos \psi_{\mathbf{f},i}$, where $\cos \psi_{\mathbf{f},i}$ is the angle between $\mathbf{r}_i$ and $\mathbf{f}_z$. Hence,

$$\begin{aligned} |\hat{g}_{(s,d),(k,z)} - g_{(s,d),(k,z)}| &= \|\mathbf{r}_i\| \|\mathbf{f}_z\| |\cos \psi_{\mathbf{f},i}| \\ &\leq \|\mathbf{r}_i\| \|\mathbf{f}_z\|. \end{aligned}$$

Lemma 2 tells us that the upper bound of the component-wise error depends on $\|\mathbf{f}_z\|$. By further bounding $\|\mathbf{f}_z\|$, we can derive a bound for the estimation error. However, bounding $\|\mathbf{f}_z\|$ requires knowing $\mathbf{g}_{(k,z)}$, which cannot be obtained. Instead, we want to bound the expected estimation error for the equivalent channel gain, denoted as $\Delta_{(s,d),(k,z)} = \mathbb{E}_g[|\hat{g}_{(s,d),(k,z)} - g_{(s,d),(k,z)}|]$.

To bound $\Delta_{(s,d),(k,z)}$, we write $f_z[i]$ as in Eq. (10), shown at the bottom of the next page, where

$$\begin{aligned} F_{(k,z)}[r] &= \sum_{(s,d) \in \mathcal{E}} \sum_{\substack{(l,v) \in \mathcal{E} \\ [-3pt] (l,v) \neq (s,d)}} q_{(s,d),(k,z)}[r] q_{(l,v),(k,z)}^*[r], \\ V_{(k,z)}[r] &= n_z[r] \sum_{(s,d) \in \mathcal{E}} q_{(s,d),(k,z)}[r], \\ s[r] &= \sum_{(s,d) \in \mathcal{E}} g_{(s,d),(k,z)} (\mathbb{E}[|x_s[r]|^2] - |x_s[r]|^2), \\ q_{(s,d),(k,z)}[r] &= h_{(s,d),(k,z)}^{\text{eq}} x_{(s,d)}[r - \mu_{(s,d)}]. \end{aligned}$$

We want to bound the three items in Eq. (10) separately. Suppose that m_i samples are needed for the probability of items (i) being greater than A_i to be less than β_i , i.e., $\mathbb{P}[(i) > A_i] \leq \beta_i$, for $i = 1, 2, 3$. We can calculate A_i 's as follows.

Lemma 3: *For edge $(k, z) \in \mathcal{E}$, given β and the number of samples, m_i 's, A_i 's can be calculated as*

$$\begin{aligned} A_1^2 &= \frac{4}{m_1 \beta_1} \sum_{(s,d) \in \mathcal{E}} \sum_{\substack{(l,v) \in \mathcal{E} \\ (l,v) \neq (s,d)}} \sigma_s^2 \sigma_l^2 \mathbb{E}[g_{(s,d),(k,z)}] \mathbb{E}[g_{(l,v),(k,z)}], \\ A_2^2 &= \frac{1}{m_2 \beta_2} \sum_{(s,d) \in \mathcal{E}} \mathbb{E}[g_{(s,d),(k,z)}^2] (4\sigma_s^4 - \frac{4}{N_c} \sigma_s^4), \\ A_3^2 &= \frac{1}{m_3 \beta_3} \sigma_v^4. \end{aligned}$$

Proof: Please refer to Appendix C. ■

With the probabilistic bound for each item in Eq. (10), the per-pair estimation error can then be bounded as follows. The per-pair error will be used in the optimization problem (P1) to incorporate the impact of channel gain estimation error.

Theorem 5: (Per-pair error) For edge $(k, z) \in \mathcal{E}$, the expected channel gain estimation error, $\Delta_{(s,d),(k,z)}$, can be bounded with probabilistic guarantee as

$$\mathbb{P} \left[\Delta_{(s,d),(k,z)} \leq \|\mathbf{r}_i\| \sum_{j=1}^3 A_j \right] \geq 1 - \beta,$$

where the number of samples used is the maximum of m_1 , m_2 , and m_3 , and $\beta = 1 - \prod_{j=1}^3 (1 - \beta_j)$.

Proof: We know that $\mathbb{P}[|\odot_j| \leq A_j] \geq 1 - \beta_j$. Hence,

$$\begin{aligned} \mathbb{P} \left[\Delta_{(s,d),(k,z)} \leq \|\mathbf{r}_i\| \sum_{j=1}^3 A_j \right] &\geq \prod_{j=1}^3 \mathbb{P}[|\odot_j| \leq A_j] \\ &\geq \prod_{j=1}^3 (1 - \beta_j) \end{aligned}$$

Given the probabilistic guarantee, $1 - \beta$, we can calculate the β_j 's that must be satisfied by each item. ■

IV. JOINT OPTIMIZATION FOR INTERFERENCE GRAPH ESTIMATION AND RESOURCE ALLOCATION

As our IGE approach is in the power domain, we can jointly optimize the accuracy of IGE and the network performance with power control. We consider the impact of estimation errors on resource allocation in the problem formulation, such that our power allocation scheme is robust to estimation errors.

A. Problem Formulation

Since each node is equipped with a single full-duplex radio, each node has a single path to the gateway, determined by the routing algorithm. Based on the buffer status reports of UEs, we can infer the traffic demand for each link. Our goal is to 1) maximize the energy efficiency of the mmWave backhaul network in satisfying the traffic demand within the required time and 2) minimize estimation errors. As indicated by Theorem 3, the per-link error can be reduced by decreasing the condition number of the transmit power matrix. Let $\mathcal{I}$ be the set of indices for blocks and $\mathcal{N}$ be the set of indices for nodes. We can formulate a multi-objective joint optimization problem as follows.

$$\begin{aligned} (\mathcal{P1}) \quad &\min_{\mathbf{P}, \delta} \left[\sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} p_k^{tx}[i] \tau, \kappa(\mathbf{P}) \right] \\ \text{s.t.} \quad &(C_1) \sum_{k \in \mathcal{N}} \delta_{(k,z)}[i] \leq 1, \forall i \in \mathcal{I}, \forall z \in \mathcal{N} \end{aligned}$$

$$\begin{aligned} (C_2) \quad &\sum_{z \in \mathcal{N}} \delta_{(k,z)}[i] \leq 1, \forall i \in \mathcal{I}, \forall k \in \mathcal{N} \\ (C_3) \quad &\delta_{(k,z)}[i] \in \{0, 1\}, \forall (k, z) \in \mathcal{E}, \forall i \in \mathcal{I} \\ (C_4) \quad &\sum_{i=1}^{|\mathcal{I}|} \delta_{(k,z)}[i] \geq \left\lceil \frac{D_{(k,z)}}{R_{(k,z)} \tau} \right\rceil, \forall (k, z) \in \mathcal{E}, \\ (C_5) \quad &\widehat{\text{SINR}}_{(k,z)} \geq \gamma_{(k,z)} \delta_{(k,z)}[i], \forall (k, z) \in \mathcal{E}, \\ (C_6) \quad &p_k^{tx}[i] \geq \delta_{(k,z)}[i] P_{min}, \forall (k, z) \in \mathcal{E}, \\ (C_7) \quad &p_k^{tx}[i] \leq \delta_{(k,z)}[i] P_{max}, \forall (k, z) \in \mathcal{E}, \\ (C_8) \quad &\text{rank}(\mathbf{P}) = |\mathcal{E}|, \end{aligned}$$

where τ is the block length, $\delta_{(k,z)}[i]$ indicates if node k is sending to node z at the i -th block, $D_{(k,z)}$ is the traffic demand for link (k, z) , $R_{(k,z)}$ is the transmission rate under the selected modulation and coding scheme (MCS), $\gamma_{(k,z)}$ is the required SINR by link (k, z) to support the transmission rate $R_{(k,z)}$, P_{max} and P_{min} are the boundary powers, and $\widehat{\text{SINR}}_{(k,z)}$ is the SINR calculated using the conservative channel gains that consider the estimation errors for link (k, z) . We define $\widehat{\text{SINR}}_{(k,z)}$ as

$$\widehat{\text{SINR}}_{(k,z)} = \frac{g_{(k,z),(k,z)} p_k^{tx}}{W_z + \sum_{(s,d) \in \mathcal{E} \setminus (k,z)} g_{(s,d),(k,z)}^{ub} p_s^{tx}}, \quad (12)$$

where $g_{(k,z),(k,z)}$ is the equivalent channel gain for the communication link from node k to z , $g_{(s,d),(k,z)}^{ub} = \hat{g}_{(s,d),(k,z)} + \Delta_{(s,d),(k,z)}$, and $\hat{g}_{(s,d),(k,z)}$ is the estimated equivalent channel gain for the interference channel from link (s, d) to (k, z) . For each communication link, the channel gain will be estimated for the demodulation of received symbols. We simply assume that $g_{(k,z),(k,z)}$ be accurately estimated for the communication links, which can be easily extended to account for the estimation errors. The equivalent channel gain, $\hat{g}_{(s,d),(k,z)}$, is estimated based on Eq. (9) with the received powers measured in the previous round of scheduling.

B. Proposed Solution

P1 is a multi-objective mixed integer nonlinear programming (MINLP) problem with a nonlinear objective function and a rank constraint. When the decision space is large (with hundreds of binary and continuous variables), it is challenging to find the optimal solution for P1 due to the high computational complexity. We thus propose a heuristic approach with the workflow shown in Fig. 4. Specifically, we decompose P1 into two subproblems P2 and P3. P2 is an MINLP problem aiming to maximize the energy efficiency while satisfying user traffic demand, i.e.,

$$\begin{aligned} (\mathcal{P2}) \quad &\min_{\mathbf{P}, \delta} \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} p_k^{tx}[i] \\ \text{s.t.} \quad &C_1 - C_7, \end{aligned}$$

$$f_z[i] = \underbrace{\frac{1}{m} \sum_{r=(m-1)i+1}^{mi} (F_{(k,z)}[r] + V_{(k,z)}[r])}_{\textcircled{1}} + \underbrace{\frac{1}{m} \sum_{r=(m-1)i+1}^{mi} s[r]}_{\textcircled{2}} + \underbrace{\frac{1}{m} \sum_{r=(m-1)i+1}^{mi} |n_z[r]|^2}_{\textcircled{3}}, i \in \{1, \dots, n_b\} \quad (10)$$

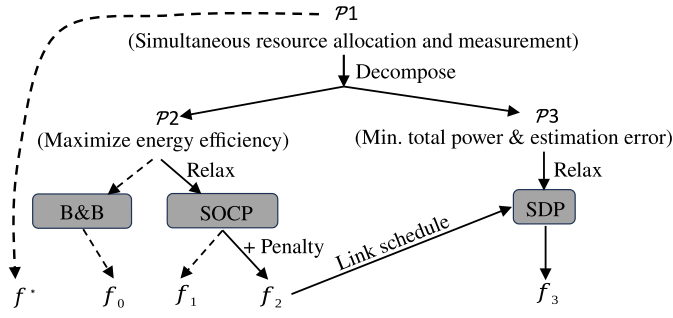


Fig. 4. Workflow of our proposed solution.

where the link schedule δ including the on/off states of links in each block is determined. $\mathcal{P}_3$ takes as input the link schedule δ from $\mathcal{P}_2$ to minimize two objectives, the total power consumption and estimation error. $\mathcal{P}_3$ can be written as a linear scalarization problem to combine the two objectives using weights, i.e.,

$$(\mathcal{P}_3) \min_{\mathbf{P}} \alpha_1 \frac{1}{\beta_1} \kappa(\mathbf{P}) + \alpha_2 \frac{1}{\beta_2} \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} p_k^{tx}[i]$$

s.t. C_5 - C_8 ,

where variable $\alpha_i > 0$, $\sum_i \alpha_i = 1$, specifies the priority of the i -th objective, and β_1 and β_2 are the normalization factors for the condition number and power consumption. Varying α_i 's allows us to obtain a set of Pareto optimal solutions on the Pareto front. Based on δ , $\mathcal{P}_3$ obtains the power allocation $\mathbf{P}$ that satisfies the rank constraint and minimize the weighted condition number. $\mathcal{P}_3$ differs from $\mathcal{P}_2$ in that as δ is given in $\mathcal{P}_3$, its variable list does not include δ , making it a problem with only continuous variables.

$\mathcal{P}_2$ will be relaxed into a second-order cone programming (SOCP) problem with items penalizing the binary decision variables deviating from zero or one, and $\mathcal{P}_3$ will be relaxed into a semi-definite programming (SDP) problem, where f_2 is the objective value for solving the relaxed $\mathcal{P}_2$ with penalty items and f_3 is the objective value of our heuristic solution for $\mathcal{P}_1$. For optimality loss, we want to evaluate the performance gap between the optimal and the heuristic solutions. In Fig. 4, f^* is the optimal objective value for $\mathcal{P}_1$, f_0 is the optimal objective value for $\mathcal{P}_2$, and f_1 is the objective value for relaxed $\mathcal{P}_2$ with no penalty terms. We know that these values can be ordered as $f_1 \leq f_0 \leq f^* \leq f_3$.² For small-sized problems, f_0 can be obtained with the branch-and-bound (B&B) approach. In Section V, we will use the performance gap between f_0 and f_3 to evaluate the performance gap between the optimal and our heuristic solution.

C. Solution to $\mathcal{P}_2$

We want to approximately solve $\mathcal{P}_2$ with continuous and convex relaxations. We first relax the binary variables $\delta_{(k,z)}[i]$'s in $\mathcal{P}_2$ into continuous variables between 0 and 1.

²Since f_0 are f_1 are the objective values for the original and relaxed $\mathcal{P}_2$, respectively, we have $f_1 \leq f_0$. Since $\mathcal{P}_1$ has one more constraint (i.e., the rank constraint) than $\mathcal{P}_2$, we expect that $f_0 \leq f^*$. Moreover, as f_3 is the objective value of the heuristic solution for $\mathcal{P}_1$, we know that $f^* \leq f_3$.

As a result, the problem becomes a nonconvex nonlinear problem. After the problem is solved, we want to convert relaxed continuous $\delta_{(k,z)}[i]$'s back to binary and thus introduce a penalty term in the objective function to penalize $\delta_{(k,z)}[i]$'s for deviating from 0 or 1. The penalty term is $f(\delta) = g_1(\delta) + g_2(\delta)$, where $g_1(\delta) = \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} \delta_{(k,z)}[i]$ and $g_2(\delta) = -\sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} (\delta_{(k,z)}[i])^2$. Suppose $\delta^{(t-1)}$ is a feasible solution at the $(t-1)$ -th iteration. $g_2(\delta)$ can be linearized by its first-order Taylor approximation near $\delta^{(t-1)}$ as

$$g_2(\delta) \leq \tilde{g}_2(\delta) \triangleq g_2(\delta^{(t-1)}) + \nabla g_2^T(\delta^{(t-1)})(\delta - \delta^{(t-1)}).$$

Next, we want to relax the nonconvex constraint C_5 into a convex one and use the successive convex approximation (SCA) technique to iteratively solve the convex problem until convergence. To relax the constraint C_5 , we divide it into two constraints, i.e.,

$$(C_{5.1}) a_{(k,z)}[i] = BN_0 + \sum_{(l,d) \in \mathcal{E} \setminus (k,z)} g_{(l,d),(k,z)} p_l^{tx}[i], \quad (13)$$

and

$$(C_{5.2}) \frac{p_{(k,z)}^{tx}[i]}{\gamma_{(k,z)}} g_{(k,z),(k,z)} \geq \delta_{(k,z)}[i] a_{(k,z)}[i]. \quad (14)$$

Following the principle of SCA, we need to find a convex surrogate function for the nonconvex term in $C_{5.2}$ as

$$\delta_{(k,z)}[i] a_{(k,z)}[i] \leq \frac{\phi_{(k,z)}[i]}{2} (\delta_{(k,z)}[i])^2 + \frac{(a_{(k,z)}[i])^2}{2\phi_{(k,z)}[i]} \quad (15)$$

for any constant $\phi_{(k,z)}[i] > 0$, where the equality holds when $\phi_{(k,z)}[i] = a_{(k,z)}[i]/\delta_{(k,z)}[i]$. Then, constraint $C_{5.2}$ becomes

$$(C_{5.3}) \frac{g_{(k,z),(k,z)}}{\gamma_{(k,z)}} p_{(k,z)}^{tx}[i] \geq \frac{\phi_{(k,z)}[i]}{2} (\delta_{(k,z)}[i])^2 + \frac{(a_{(k,z)}[i])^2}{2\phi_{(k,z)}[i]},$$

which is a convex constraint.

Let $\mathbf{A}$ be the matrix for $a_{(k,z)}[i]$'s. At the t -th iteration, $\mathcal{P}_2$ can then be relaxed as

$$(\mathcal{P}_2^*) \min_{\mathbf{P}, \delta, \mathbf{A}} \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} p_k^{tx}[i] + \lambda g_1(\delta) + \lambda \tilde{g}_2(\delta)$$

s.t. $C_1, C_2, C_4, C_{5.1}, C_{5.3}, C_6, C_7$,
 $\delta_{(k,z)}[i] \in [0, 1], \forall (k,z) \in \mathcal{E}, \forall i \in \mathcal{I}$,

which is a SOCP problem. The solution to $\mathcal{P}_2$ can be obtained by iteratively solving $\mathcal{P}_2^*$ until convergence.

D. Solution to $\mathcal{P}_3$

When δ is given, constraint C_5 becomes linear. As shown in [38], rank constraints are discontinuous and nonconvex, which needs to be gradually approximated with a sequence of semidefinite programming (SDP) problems. To approximate constraint C_8 , we reformulate it with the rank constraints of two semidefinite matrices. Specifically, we have $\text{rank}(\mathbf{P}) = |\mathcal{E}|$ if and only if there exists a $\mathbf{Z} \in \mathbb{S}^{n_b}$ such that

$$\text{rank}(\mathbf{Z}) = |\mathcal{E}|,$$

and

$$\text{rank}(\mathbf{U}) \leq |\mathcal{E}|, \quad \mathbf{U} = \begin{bmatrix} \mathbf{I}_{|\mathcal{E}|} & \mathbf{P}^T \\ \mathbf{P} & \mathbf{Z} \end{bmatrix}, \quad (16)$$

where $\mathbb{S}^{n_b}$ is the set of symmetric $n_b \times n_b$ matrices. Based on the theorem in [38], we have that when $e = 0$ and $\mathbf{U}$ is a positive semidefinite matrix, $\text{rank}(\mathbf{Z}) = |\mathcal{E}|$ and $\text{rank}(\mathbf{U}) \leq |\mathcal{E}|$ are equivalent to

$$V^T \mathbf{Z} V > 0, \text{ and } e \mathbf{I}_{n_b} - \mathbf{W}^T \mathbf{U} \mathbf{W} \succeq 0$$

where the vector $V \in \mathbb{R}^{n_b \times 1}$ is the eigenvector corresponding to the $n_b - |\mathcal{E}| + 1$ smallest eigenvalue of $\mathbf{Z}$, and the matrices $\mathbf{W} \in \mathbb{R}^{(n_b + |\mathcal{E}|) \times n_b}$ are the eigenvectors corresponding to the n_b smallest eigenvalues of $\mathbf{U}$.

Replacing the rank constraints with the SDP ones, we can solve $\mathcal{P}3$ by iteratively solving the following problem:

$$(\mathcal{P}3-1) \min_{\mathbf{P}^{(t)}, \mathbf{Z}^t, e^t} \alpha \kappa(\mathbf{P}) + \frac{(1-\alpha)}{P_{\text{SOCP}}^*} \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} p_k^t[i] + w^t e^t \quad (17a)$$

$$\text{s.t.} \quad (V^{t-1})^T \mathbf{Z}^t V^{t-1} > 0, \quad (17b)$$

$$e^t \mathbf{I}_{n_b} - (\mathbf{W}^{t-1})^T \mathbf{U}^t \mathbf{W}^{t-1} \succeq 0, \quad (17c)$$

$$0 \leq e^t \leq e^{t-1}, \quad (17d)$$

$$C_5, C_6, C_7, \quad (17e)$$

$$\mathbf{Z}^t \succeq 0, \mathbf{U}^t \succeq 0, \quad (17f)$$

where $p_k^t[i]$ is the transmit power of node k at the t -th iteration, $\mathbf{W}^{t-1}$ includes the eigenvectors corresponding to the n_b smallest eigenvalues of $\mathbf{U}^{t-1}$, w^t is a weighting factor increasing with the iteration count t , and P_{SOCP}^* is the power consumption of $\mathcal{P}2^*$ used to normalize the power consumption. The starting point at the 0-th iteration can be obtained by solving Eq. (17) without constraints (17b)-(17d). At the 0-th iteration, e^0 is set to be the n_b -th smallest eigenvalue of $\mathbf{U}^0$.

To handle the nonconvexity of $\kappa(\mathbf{P})$, we express the condition number in terms of the singular values. The condition number of $\mathbf{P}$ is the ratio of its largest and smallest singular values, i.e., $\kappa(\mathbf{P}) \triangleq \left(\frac{\lambda_{\max}(\mathbf{P}^T \mathbf{P})}{\lambda_{\min}(\mathbf{P}^T \mathbf{P})} \right)^{\frac{1}{2}}$. We can rewrite $\mathcal{P}3-1$ as an SDP problem, i.e.,

$$(\mathcal{P}3-2) \min_{\mathbf{P}, \gamma} \alpha \gamma + \frac{(1-\alpha)}{P_{\text{SOCP}}^*} \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} p_k^t[i]$$

$$\text{s.t. } C_5-C_8, C_9,$$

where C_9 constrains the smallest and largest singular values of $\mathbf{P}$ as

$$(C_9) \quad \mu \mathbf{I}_{|\mathcal{E}|} \prec \mathbf{P}^T \mathbf{P} \prec \gamma \mu \mathbf{I}_{|\mathcal{E}|}.$$

It is clear that this problem is nonlinear and nonconvex, which needs to be further simplified. We know from Eq. (17c) that when the iterative approach to $\mathcal{P}3$ converges, $e = 0$ and $\text{rank}(\mathbf{U}) \leq |\mathcal{E}|$. It has been shown in [38] that $\text{rank}(\mathbf{U}) \leq |\mathcal{E}|$ indicates $\text{rank}(\mathbf{Z} - \mathbf{P} \mathbf{P}^T) = 0$, i.e., $\mathbf{Z} = \mathbf{P} \mathbf{P}^T$. Since it can be easily proved that the non-zero eigenvalues of $\mathbf{P}^T \mathbf{P}$ are also the eigenvalues of $\mathbf{P} \mathbf{P}^T$, we can replace $\mathbf{P}^T \mathbf{P}$ with $\mathbf{Z}$ in

constraint C_{10} . Defining new variables $\nu = 1/\mu$, $\tilde{Z}_{ij} = Z_{ij}/\mu$, $\tilde{P}_{ij} = P_{ij}/\mu$ and $\tilde{e}^t = e^t/\mu$, we can rewrite C_{10} as

$$(C_{10}^*) \quad \begin{bmatrix} \mathbf{I}_{|\mathcal{E}|} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \prec \tilde{\mathbf{Z}}^t \prec \gamma \begin{bmatrix} \mathbf{I}_{|\mathcal{E}|} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\tilde{\mathbf{Z}}^t = [\tilde{Z}_{ij}]$. Following the substitution, we can rewrite C_5-C_7 as $C_5^*-C_7^*$. Then, the problem $\mathcal{P}3-2$ can be solved by solving the following problem iteratively until convergence:

$$\min_{\tilde{\mathbf{P}}, \tilde{\mathbf{Z}}, \tilde{e}^t, \gamma, \nu} \alpha \gamma + \frac{(1-\alpha)}{P_{\text{SOCP}}^*} \sum_{i \in \mathcal{I}} \sum_{(k,z) \in \mathcal{E}} \tilde{p}_k^t[i] + w^t \tilde{e}^t \quad (18a)$$

$$\text{s.t.} \quad C_5^*-C_7^*, C_9^* \quad (18b)$$

$$(\tilde{V}^{t-1})^T \tilde{\mathbf{Z}}^t \tilde{V}^{t-1} > 0, \quad (18c)$$

$$\tilde{e}^t \mathbf{I}_{n_b} - (\tilde{\mathbf{W}}^{t-1})^T \tilde{\mathbf{U}}^t \tilde{\mathbf{W}}^{t-1} \succeq 0, \quad (18d)$$

$$0 \leq \tilde{e}^t \leq \tilde{e}^{t-1}, \quad (18e)$$

$$\tilde{\mathbf{Z}}^t \succeq 0, \tilde{\mathbf{U}}^t \succeq 0, \nu > 0, \quad (18f)$$

where $\tilde{\mathbf{U}} = \begin{bmatrix} \nu \mathbf{I}_{|\mathcal{E}|} & \tilde{\mathbf{P}}^T \\ \tilde{\mathbf{P}} & \tilde{\mathbf{Z}} \end{bmatrix}$, $\tilde{\mathbf{P}} = [\tilde{P}_{ij}]$. $\tilde{V}^{t-1}$ and $\tilde{\mathbf{W}}^{t-1}$ are computed by $\tilde{\mathbf{Z}}^{t-1}$ and $\tilde{\mathbf{U}}^{t-1}$ as what V^{t-1} and $\mathbf{W}^{t-1}$ to $\mathbf{Z}^{t-1}$ and $\mathbf{U}^{t-1}$, respectively.

E. Convergence Analysis

We first prove that the iterative approach for $\mathcal{P}2^*$ converges. Let $(\mathbf{P}^t, \delta^t, \mathbf{A}^t)$ be the solution of $\mathcal{P}2^*$ at the t -th iteration. At the $(t+1)$ -th iteration, if we replace $(\mathbf{P}, \delta, \mathbf{A})$ by $(\mathbf{P}^t, \delta^t, \mathbf{A}^t)$, all the constraints in $\mathcal{P}2^*$ are still satisfied. This means that the solution at the t -th iteration is a feasible solution at the $(t+1)$ -th iteration. As a result, the solution obtained at the $(t+1)$ -th iteration is less or equal to that at the t -th objective. In addition that the solution of $\mathcal{P}2^*$ is bounded due to the power constraints, we have that $\mathcal{P}2^*$ is convergent. Since $\mathcal{P}3$ is solved iteratively, we want to prove the convergence of the iterative approaches. Specifically, we first want to prove that if there exists a solution at the $(i-1)$ -th iteration, a solution will also exist at the i -th iteration, i.e., the problem is always solvable during the iterative process. Let $(\mathbf{U}^t, e^t)$ be the solution to Eq. (17) at the t -th iteration. Then, $(\mathbf{U}^t, e^t)$ satisfies $e^t \mathbf{I}_{n_b} - (\mathbf{W}^{t-1})^T \mathbf{U}^t \mathbf{W}^{t-1} \succeq 0$, i.e., $e^t \geq \lambda_1 = \lambda_{\max} \left((\mathbf{W}^{t-1})^T \mathbf{U}^t \mathbf{W}^{t-1} \right)$. We want to prove that $e^t \mathbf{I}_{n_b} - (\mathbf{W}^t)^T \mathbf{U}^t \mathbf{W}^t \succeq 0$ is true, i.e., $e^t \geq \lambda_{\max} \left((\mathbf{W}^t)^T \mathbf{U}^t \mathbf{W}^t \right)$, such that Eq. (17) has at least a solution $(\mathbf{U}^{(t)}, e^t)$ at the $(t+1)$ -th iteration. By definition, we know the n_b -th smallest eigenvalue $\lambda_2 = \lambda_{\max} \left((\mathbf{W}^t)^T \mathbf{U}^t \mathbf{W}^t \right)$. We next want to prove that $\lambda_1 \geq \lambda_2$. Recall that $\mathbf{W}^t$ includes the eigenvectors corresponding to the n_b smallest eigenvalues of $\mathbf{U}^t$. It has been proved in [39] that for $\mathbf{W}^T \mathbf{W} = \mathbf{I}_{n_b}$, $\lambda_{\max} \left(\mathbf{W}^T \mathbf{U}^t \mathbf{W} \right) \geq \lambda_2$. Since $(\mathbf{W}^{t-1})^T \mathbf{W}^{t-1} = \mathbf{I}_{n_b}$, we have $\lambda_1 \geq \lambda_2$. For the first iteration ($t = 1$), the pair $(\mathbf{U}^0, e^0)$ is a feasible solution for the iterative problem at the first iteration, where e^0 is the $(|\mathcal{E}| + 1)$ -th largest eigenvalue of $\mathbf{U}^0$. Since Eq. (17) has a solution at each iteration and $e^t \leq e^{t-1}$, the iterative approach converges.

V. PERFORMANCE EVALUATION

A. Simulation Setup

The performance of our proposed joint optimization for IGE and resource allocation is evaluated via extensive simulations. A dense multi-hop mmWave backhaul network is simulated, where BSs are randomly located in a 0.01km^2 square field with density more than 700BSs/km^2 . To prevent BSs from being either too close or distant, we control the distances between BSs within the range of 15m to 100m . To align with the 3GPP standards [4], we divide time into slots of equal length and each BS transmits data at the beginning of each time slot. Each time slot includes 14 OFDM symbols and the length of a time slot depends on the subcarrier spacing, ranging from 1ms down to $62.6\mu\text{s}$. BSs are synchronized with TOs and experience CFOs with the fractional component between -0.5 and 0.5 . Among all BSs, one is selected to serve as the gateway, while all other BSs backhaul traffic to/from the gateway. Between each BS and the gateway exists only a single path, determined by the Dijkstra algorithm based on the distances between BSs. The links between BSs are directional, allowing both upstream and downstream traffic, where the total number of links is twice the number of BSs minus one. The mmWave channel $\mathbf{H}_{(k,z)}$ is modelled as a Rician channel, with a LOS component $\mathbf{H}_{(k,z)}^L$ and a scattering component $\mathbf{H}_{(k,z)}^S$, i.e., $\mathbf{H}_{(k,z)} = \sqrt{\frac{\varpi(k,z)\zeta(k,z)}{\zeta(k,z)+1}}\mathbf{H}_{(k,z)}^L + \sqrt{\frac{\varpi(k,z)}{\zeta(k,z)+1}}\mathbf{H}_{(k,z)}^S$. The Rician K-factor $\zeta(k,z)$ is set to 2 and $\varpi(k,z) = 10^{-(10\rho_1\log_{10}d(k,z) + \rho_2\log_{10}(4\pi/\lambda))}/10$, where $d(k,z)$ is the distance of link (k,z) , λ is the carrier wavelength, and ρ_1, ρ_2 are empirical propagation parameters as in [40]. We simulate the randomness in the Rician channel by changing the scattering component $\mathbf{H}_{(k,z)}^S$, consisting of independent and identically distributed entries following Gaussian distribution as in [40] and [41].

Each BS has access to a uniform linear array (ULA) with 100 antennas and the antenna spacing is equal to half of the wavelength. As a BS's transmitter is relatively close to its own receiver compared with the distance away from other nodes, we assume that SI channel suffers from Rayleigh flat fading without any effect of path loss, i.e., $(\mathbf{H}_z^{SI})_{i,j} \sim \mathcal{CN}(0,1)$. Unless specified otherwise, we assume that the system has sufficient SI suppression with a SIC capability $\eta = -100\text{dB}$ according to [42] and [43]. The angle-of-arrival (AoA) at each BS is the direction from which the strongest signal strength is detected, which are used to estimate the transmit and receive beamforming vectors of BSs. We allow each BS to control its transmit power between 800mW and 1200mW and use a different transmit power for each block. Each scheduling period is pre-determined by the number of BSs and the block length. We will discuss the choice of a proper block length in our experiments. The MCS for each link is determined based on the channel quality. The demand for each link takes a random number of time slots, less than the scheduling period, to send with the chosen modulation scheme. We use the square M -QAM modulation in our experiments and $M \in \{4, 16, 64, 256\}$. Each transmitter randomly selects modulated symbols from the constellation to

TABLE II
SIMULATION PARAMETERS

#Antenna per BS	100	BS distance	$\geq 15\text{m}$
OFDM symbols per slot	14	Tx power range	$800\text{--}1200\text{mW}$
Carrier frequency f_c	28GHz	Noise spectrum	-174dBm/Hz
Total bandwidth	250MHz	BS density	$\geq 700\text{BSs/km}^2$

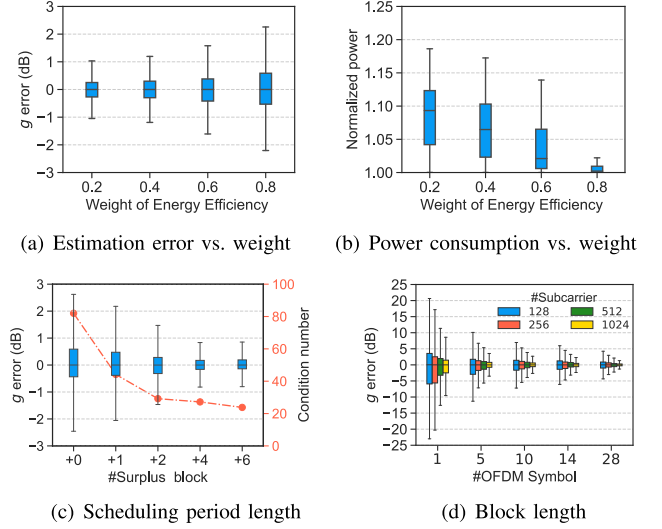


Fig. 5. Design parameter selection.

send on each subcarrier. The simulation parameters are listed in Table II.

B. Design Parameter Selection and Convergence

Weight Choice for Multi-Objective Functions.: To strike a balance between the channel gain estimation error and power consumption, we evaluate the two metrics under different weight choices. Since the experiments are conducted on random topologies, for better comparison, we normalize the total power consumption as in the objective function in Eq. (17a). From Fig. 5(a) and 5(b), we can see that as the weight of power consumption increases, the channel gain estimation error increases and the normalized power consumption decreases, as expected. Proper weights can be chosen based on the specific requirements on channel gain estimation and power consumption. In our experiments, we choose the power consumption weight of 0.8 to control the extra power overhead attributed to IGE.

Scheduling Period Length: As indicated by Theorem 4, a larger scheduling period helps reduce the condition number of the transmit power matrix. However, a larger scheduling period means a slower update of the channel estimation. We know that the minimum required number of blocks is equal to the number of BSs. The blocks exceeding the minimum required ones are referred to as *surplus blocks*. Fig. 5(c) shows the estimation errors under different numbers of surplus blocks, where the network topologies are random with 8, 12, or 16 links. It can be seen that the median condition number decreases as the number of surplus blocks increases, which aligns with Theorem 4. As a result of the decreasing condition numbers, the estimation errors decrease sharply with

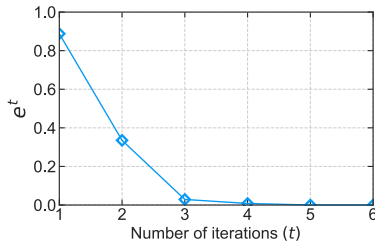


Fig. 6. Convergence of our approach under a random 16-link topology.

several surplus blocks. We thus use two surplus blocks in our experiments.

Block Length: As mentioned, we want to use the average transmit power to approximate the expected one with a sufficient number of samples. Fig. 5(d) shows the distribution of the channel gain estimation errors under different numbers of subcarriers and OFDM symbols, where more than 100,000 channel gains are estimated under 16-link networks with random topologies for each configuration. We can see that the estimation error decreases as more subcarriers and OFDM symbols are used, equating to using more samples in the time domain. When the number of subcarriers is 1024, the average absolute estimation error is as low as 0.6dB, even when a single OFDM symbol is used. There is no significant difference in estimation error when the number of OFDM symbols increases from 14 to 28. We thus set blocks to include 14 OFDM symbols, i.e., one time slot, in our experiments.

Convergence: After all design parameters are determined, we evaluate the convergence performance of our approach under random topologies. As we discussed in Section IV-E, when e^t approaches to zero, the iterative approach to Eq. (18) converges. Fig. 6 shows the convergence process under a random topology of 16 links. As the number of iterations t increases, e^t decreases to 0. It can be seen that our iterative approach takes only 6 iterations to converge.

C. Performance Evaluation

Channel Gain Estimation: We measure the estimation error of the equivalent channel gain from link (s, d) to (k, z) as $10 \log_{10} (\hat{g}(s, d, (k, z)) / g(s, d, (k, z)))$ in dB, which is the log ratio of the estimated and actual equivalent channel gains. The ratio is equal to zero when the estimated and actual channel gains are equal and increases in its absolute value as the measured one deviates away from the actual one. Fig. 7(a) shows the boxplot of the estimation errors using our approach under different network sizes, where the network size is measured by the number of links in the backhaul network. We can see that our approach achieves estimation errors less than 2dB for almost all the links, with half of the errors less than 0.8dB. The estimation errors slightly increase with the network size. As the network size increases, the condition number of the transmit power matrix tends to increase due to a larger dimension and results in larger estimation errors.

Fig. 7(b) shows the distribution of the channel gain estimation errors for 16-link networks with random topologies, where the relative magnitude of a link is calculated with respect

to g_{max} , the maximum channel gain in the network. We can see that smaller channel gains tend to have larger estimation errors. For extremely small channel gains, the estimated values may fall below zero due to estimation errors. We choose to set such channel gains to the minimum channel gain in the network. Fortunately, such links with small equivalent channel gains are typically weak interference links, with signal strength much less than the communication links. Fig. 7(c) shows the estimation errors for communication and interference links, respectively. It can be seen that the interference links generally experience larger estimation errors than the communication links. As a result, the estimated SINR is close to the actual one, ensuring that the joint optimization can perform as expected under estimation errors.

Robustness to TO and CFO: Fig. 8(a) shows the estimation errors under different TOs, where the TO is normalized with respect to CP length. Since our approach estimates channel gains in the power domain, it is robust to TO and achieves very small estimation errors under all TOs. Fig. 8(b) shows the estimation errors under different ranges of CFOs, where CFOs are randomly selected for each range. We can see that our approach achieves almost the same distribution of estimation errors as when no CFO exists (i.e., CFO = 0). This indicates that our approach is robust to CFOs. The channel gain estimation errors are mainly due to the randomness in received powers, not CFOs.

Comparison With the Existing Methods: We compare our approach with two existing methods for channel gain estimation: reference signal (RS)-based and model-based. The RS-based approach sends known reference symbols for measurement and the model-based approach is assumed using the same Rician channel model as in our experiments, where the K-factor is set to 2. We compare our approach and RS-based approach under different TOs. As shown in Fig. 8(a), the RS-based approach is accurate when TO is within CP and degrades rapidly when TO becomes greater than CP, while our approach is robust to TO. We compare our approach and the model-based approach under different Rician K-factors. Fig. 9 shows that the model-based approach suffers from the randomness in the scattering component of the mmWave channel. Only when the LOS component is dominant (for large K-factors), the model-based approach can have small estimation errors as our approach does.

Optimizing the Condition Number: In our proposed solution, the second step is to minimize the condition number of the transmit power matrix. Fig. 10 shows the distributions of estimation errors before and after minimizing the condition number. We can see that reducing condition number helps reduce the estimation errors, consistent with what Lemma 2 implies. Looking into how much condition numbers are reduced, we find that the reduction is significant for condition numbers of different magnitudes and that large condition numbers generally experience larger reductions than small ones.

Power Overhead: Our approach uses power as a new dimension for interference measurement and thus may have extra power overhead. We compare our approach with the pure resource allocation scheme not considering IGE (i.e., $\mathcal{P}2^*$) in

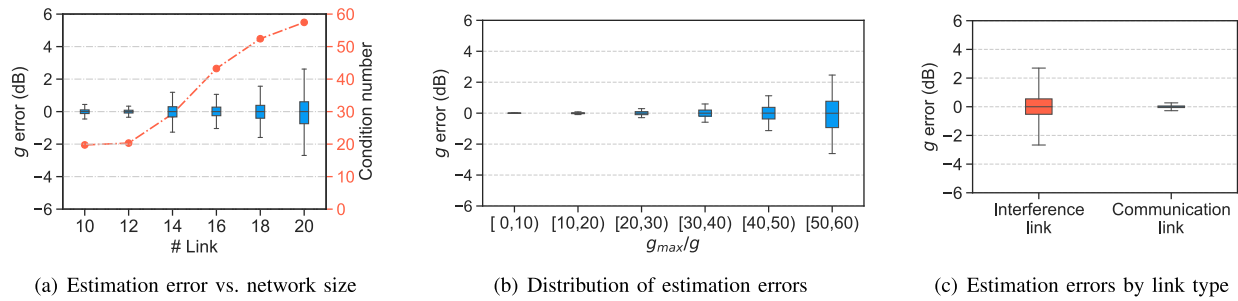


Fig. 7. The accuracy of equivalent channel gains is estimated with respect to network size, channel gain magnitude, and link type. In summary, both the interference and communication channel gains can be accurately estimated under different network sizes.

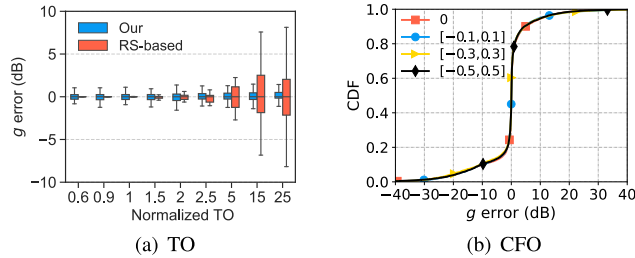


Fig. 8. Equivalent channel gains are estimated under TO and CFO. Our approach is robust to TO and CFO.

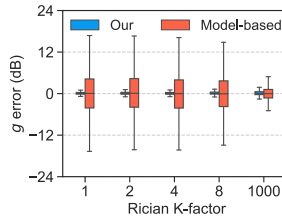


Fig. 9. When the scattering component of the mmWave channel is large (small Rician K-factors), the model-based approach fails to accurately estimate the channel gains, while our approach performs well regardless of the K-factor.

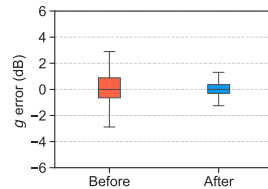


Fig. 10. Impact of condition number.

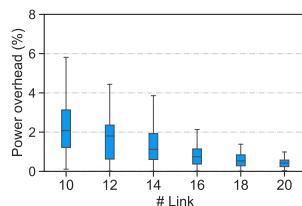


Fig. 11. Power overhead.

terms of power consumption. Let P_1 and P_0 denote the power consumption in our approach and the pure power resource allocation scheme, respectively. We define *power overhead* as $(P_1 - P_0)/P_0$. Fig. 11 shows the boxplot for power overheads under different network sizes, where 100 random layouts are used for each network size. The positive power overheads indicate that our approach estimates the interference graph at

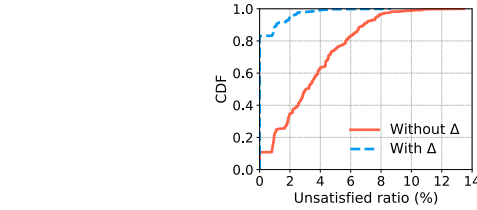


Fig. 12. Robustness to channel gain estimation errors.

the cost of power consumption. When the network includes 10 links, the 99-percentile power overhead is below 6% and the average is about 2%. The power overhead in general decreases as the network size increases because the extra power consumption needed for power control increases slower than the total power consumption of the network.

Robustness to Channel Estimation Errors: To combat channel estimation errors, we have incorporated the error bounds of estimated channel gains in our problem formulation ($\mathcal{P}1$). We define the unsatisfied ratio as the percentage of links not satisfying the SINR constraints due to estimation errors. Fig. 12 shows that the distributions of unsatisfied ratios with and without considering estimation errors (Δ) under 700 random topologies. Compared with the power allocation without considering Δ , our approach significantly reduces the unsatisfied ratios, with only 2% of links slightly violating the SINR constraints for 5% of random topologies. In contrast, about 60% of topologies have an unsatisfied ratio greater than 2% when Δ is not considered. The violation is because we use the bounds of the expected channel estimation errors for power allocation. A more conservative bound can be used to further reduce the unsatisfied ratios.

Impact of SIC Capability: The above experiments are conducted under a fixed SIC capability of 100 dB. We re-do the above experiments for channel gain estimation errors under different SIC capabilities. Fig. 13 shows that the distribution of the estimation errors does not have significant changes, even when the SIC capability is as low as 70 dB. This is because we consider the equivalent residual SI channels in the estimation and can accurately estimate their channel gains. Nonetheless, when a residual SI channel is overly strong and significantly affects the overall magnitude of the received signal, it may affect the estimation accuracy of weak channels as indicated by the error analysis in Section III-D.

Optimality Loss: We compare our approach with the B&B approach that aims to find the optimal solution for resource

TABLE III
COMPARISON WITH THE B&B APPROACH

	4 links	8 links	12 links	16 links
Power Ratio	1.03	1.03	1.02	1.02
Run Time Ratio	0.07	0.05	0.02	0.006

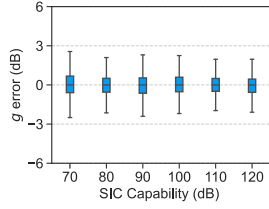


Fig. 13. Channel gain estimation errors under different SIC capabilities.

allocation in terms of power consumption and run time. Table III shows the experimental results under four different network sizes, where the power and run time ratios are the power consumption and run time of our approach divided by those of the B&B approach, respectively. It can be seen that the power ratio is close to one, indicating that our approach experiences no significant optimality loss for power consumption. Meanwhile, our approach takes only a few percent of the run time of the B&B approach and the gap in run time continues to grow as the network size increases.

VI. CONCLUSION

In this paper, we proposed to use power as a new dimension for interference graph estimation, such that it can be done simultaneously with resource allocation using the same frequency-time resources. To this end, we proved that the mmWave backhaul network can be considered as a linear system such that the interference graph can be measured with power control. We derived bounds for the estimation errors of channel gains and found that minimizing condition number of the transmit power matrix is effective in reducing the estimation errors. We formed a multi-objective joint optimization of the resource allocation and IGE, and showed that our approach is robust to timing and carrier frequency offsets. Moreover, the overhead of our approach was shown to be minimal by experiments. As for future work, we will apply the joint optimization framework to other wireless networks.

APPENDIX A PROOF OF LEMMA 1

Since $\mathbf{v}[i] \sim \mathcal{CN}(\mathbf{0}, \sigma_v^2 \mathbf{I})$, $\mathbb{E}[n_z[i]] = \mathbb{E}[\tilde{\omega}_z^T \mathbf{v}[i]] = 0$ and $\mathbb{E}[|n_z[i]|^2] = \tilde{\omega}_z^T \mathbb{E}[\mathbf{v}[i]\mathbf{v}[i]^H] \tilde{\omega}_z^* = \sigma_v^2$. Let $F_{(s,d),(k,z)}[i] = h_{(s,d),(k,z)}^{\text{eq}} x_s[i - \mu_{(s,d)}]$. Based on Eq. (2), the expected receive power of node z can be calculated by omitting the items multiplying $n_z[i]$ as

$$\begin{aligned} & \mathbb{E}[|y_z[i]|^2] \\ &= \sum_{(s,d) \in \mathcal{E}} |h_{(s,d),(k,z)}^{\text{eq}}|^2 \mathbb{E}[|x_s[i - \mu_{(s,d)}]|^2] \\ &+ \sum_{(s,d) \in \mathcal{E}} \sum_{\substack{(l,v) \in \mathcal{E} \\ (l,v) \neq (s,d)}} \mathbb{E}[F_{(s,d),(k,z)}[i] F_{(l,v),(k,z)}^*[i]] + \sigma_v^2. \end{aligned}$$

Since $x_s[i]$ and $x_l[i]$ are independent for $s \neq l$, we have $\mathbb{E}[x_s[i] x_l[i]^*] = \mathbb{E}[x_s[i]] \mathbb{E}[x_l[i]^*] = 0$, where $\mathbb{E}[x_s[i]] = 0$ for

all s 's, because $x_s[i] = \frac{1}{\sqrt{N_c}} \sum_{r=0}^{N_c-1} X_s[r] e^{j2\pi i r / N_c}$ and $\mathbb{E}[X_s[r]] = 0$ for $1 \leq m \leq N_c$. This implies that when $(l, v) \neq (s, d)$,

$$\begin{aligned} & \mathbb{E}[F_{(s,d),(k,z)}[i] F_{(l,v),(k,z)}^*[i]] \\ &= q \mathbb{E}[x_s[i - \mu_{(s,d)}]] \mathbb{E}[x_l^*[i - \mu_{(l,v)}]] = 0, \end{aligned}$$

where $q = h_{(s,d),(k,z)}^{\text{eq}} \left(h_{(l,v),(k,z)}^{\text{eq}} \right)^* \mathbb{E}_{\Delta\phi} [e^{j2\pi i \Delta\phi}]$ and $\Delta\phi = \phi_{(s,z)} - \phi_{(l,z)}$.

APPENDIX B PROOF OF THEOREM 2

Since the I/Q components of $x_k[i]$ are assumed to be independent and that the modulated symbols for each component are selected from the constellation equi-probably, we have that $\mathbb{E}[X_k^I[i]] = \mathbb{E}[X_k^Q[i]] = 0$ and $\sigma_k^2 = \text{Var}(X_k^I[i]) = \text{Var}(X_k^Q[i])$. Suppose we want to calculate the average power of received signals over N_k OFDM symbols. As each OFDM symbol includes $N_s = N_c + N_g$ time-domain samples, the average signal power is $\bar{p}_k^{\text{tx}} = \frac{1}{N_k N_s} \sum_{j=1}^{N_k N_s} |x_k[j]|^2$. We want to apply the concentration inequalities to bound the difference between $\bar{p}_k^{\text{tx}}$ and $\mathbb{E}[|x_k[i]|^2]$. However, the concentration inequalities generally assume the independence between variables, while the transmitted signals, $x_k[i]$'s, are the IDFTs of all modulated symbols on N_c subcarriers and are thus dependent. We thus want to relate $x_k[i]$ with the modulated symbols, $X_k[z]$'s, that are independently chosen.

Based on the Parseval's Theorem for discrete signals, the sum of transmitted signals within an OFDM symbol is

$$\sum_{i=0}^{N_s-1} |x_k[i]|^2 = \sum_{z=0}^{N_c-1} |X_k[z]|^2 + \sum_{i=0}^{N_g-1} |x_k[i]|^2. \quad (19)$$

As $x_k[i]$ is the IDFT of $X_k[z]$'s, it can be written as

$$\begin{aligned} |x_k[i]|^2 &= \frac{1}{N_c} \sum_{d=0}^{N_c-1} \sum_{z=0}^{N_c-1} X_k[d] X_k[z]^* e^{j2\pi(d-z)i/N}, \\ &= \frac{1}{N_c} \sum_{d=0}^{N_c-1} |X_k[d]|^2 \\ &+ \frac{2}{N_c} \sum_{d=1}^{N_c-1} \sum_{z=0}^{d-1} \text{Re}\{X_k[d] X_k[z]^* e^{j\theta}\}, \end{aligned}$$

where $\theta = 2\pi(d-z)i/N$. Then, this implies that

$$\begin{aligned} \mathbb{E}[|x_k[i]|^2] &= \mathbb{E}[|X_k[z]|^2] = \mathbb{E}[(X_k^I[z])^2] + \mathbb{E}[(X_k^Q[z])^2] \\ &= 2\sigma_k^2, \end{aligned}$$

and Eq. (19) can be rewritten as

$$\sum_{i=0}^{N_s-1} |x_k[i]|^2 = \frac{N_s}{N_c} \sum_{z=0}^{N_c-1} |X_k[z]|^2 + \frac{2}{N_c} \sum_{i=0}^{N_g-1} \sum_{d=1}^{N_c-1} \sum_{z=0}^{d-1} f_k(X),$$

where $f_k(X) = \text{Re}\{X_k[d] X_k[z]^* e^{j\theta}\}$.

As $X_k[d] = X_k^I[d] + jX_k^Q[d]$, we can rewrite $f_k(X)$ as

$$\begin{aligned} f(X_k) &= X_k^I[d] X_k^I[z] \cos(\theta) + X_k^I[d] X_k^Q[z] \sin(\theta) \\ &- X_k^Q[d] X_k^I[z] \sin(\theta) + X_k^Q[d] X_k^Q[z] \cos(\theta), \end{aligned}$$

where $d \neq z$. For zero-mean independent variables x_i 's, we have

$$\begin{aligned} \text{Var}(x_1 \dots x_d) &= \prod_{i=1}^d (\text{Var}(x_i) + \mathbb{E}[x_i]) - \prod_{i=1}^d \mathbb{E}[x_i] \\ &= \prod_{i=1}^d \text{Var}(x_i), \end{aligned}$$

which implies $\text{Var}(X_k^I[z]X_k^I[d]) = \text{Var}(X_k^I[z])\text{Var}(X_k^I[d])$. In addition that $\text{Var}(X_k^I[z]\cos(\theta)) = \mathbb{E}[(X_k^I[z])^2]\cos^2(\theta)$ and $\text{Var}(X_k^I[d]) = \text{Var}(X_k^Q[d])$, we can have $\text{Var}(f(X_k)) = 2\sigma_k^4$. To use the Bennett's inequality, we further need to bound $f_k(X)$ as

$$|f_k(X)| \leq |X_k[d]| |X_k[z]| |e^{j\theta}| \leq B_{k,1} = 2 \left(X_k^{I,max} \right)^2.$$

Combining the above facts, $f(X_k)$ can be bounded as

$$\mathbb{P} \left[\frac{1}{n} \sum_{i=0}^{n-1} f(X_k) \geq \Delta \right] \leq \exp \left(-\frac{n\sigma_f^2}{B_{k,1}^2} h \left(\frac{B_{k,1}\Delta}{\sigma_f^2} \right) \right),$$

where n is the number of samples and $\sigma_f^2 = \text{Var}(f(X_k))$.

Let $\mu_k = \mathbb{E}[|x_k[i]|^2]$ and $0 \leq \delta \leq 1$ denote the approximation error between $\bar{p}_k^{tx}$ and μ_k . The probability of $\bar{p}_k^{tx}$ being greater than $\mu_k(1 + \delta)$ is

$$\begin{aligned} &\mathbb{P}[\bar{p}_k^{tx} - \mu_k \geq \mu_k \delta] \\ &= \mathbb{P} \left[\frac{1}{N_k N_c} \sum_{i=0}^{N_k-1} \sum_{z=0}^{N_c-1} |X_k[z + N_c i]|^2 - \mu_k \geq \mu_k \delta - f(X) \right] \\ &\leq \mathbb{P}[f(X) \leq \Delta] \mathbb{P}[E] + \mathbb{P}[f(X) \geq \Delta], \end{aligned}$$

where

$$f(X) = \frac{2}{N_k N_s N_c} \sum_{i=0}^{N_k-1} \sum_{l=0}^{N_g-1} \sum_{d=1}^{N_c-1} \sum_{z=0}^{d-1} f_k(X),$$

and

$$\mathbb{P}[E] = \mathbb{P} \left[\frac{1}{N_k N_c} \sum_{i=0}^{N_k-1} \sum_{z=0}^{N_c-1} |X_k[z + N_c i]|^2 - \mu_k \geq \mu_k \delta - \Delta \right].$$

Based on the Bennett's inequality, we have

$$\mathbb{P}[E] \leq \exp \left(-\frac{N_k N_c \sigma_S^2}{B_{k,2}^2} h \left(\frac{B_{k,2}(\mu_k \delta - \Delta)}{\eta_k^2} \right) \right).$$

where B_k and η_k^2 are the upper bound and variance for $|X_k[z]|^2 - \mu_k$, respectively. Since $\mu_k = 2\sigma_k^2$ and $X_k^{I,max} \geq 2\sigma_k^2$, $B_k = 2 \left(X_k^{I,max} \right)^2 - 2\sigma_k^2$, and $\eta_k^2 = \mathbb{E}[|X_k[z]|^4] - 4\sigma_k^4$. Moreover, we know that $\bar{p}_k^{tx} - \mu_k \leq -\mu_k \delta$ if and only if $-\bar{p}_k^{tx} + \mu_k \geq \mu_k \delta$, which implies that $\mathbb{P}[\bar{p}_k^{tx} - \mu_k \geq \mu_k \delta] \leq 2\mathbb{P}[\bar{p}_k^{tx} - \mu_k \geq \mu_k \delta]$.

APPENDIX C PROOF OF THEOREM 3

We want to compute a bound for each item using the Chebyshev's inequality, which requires knowing the variance of the variable. It has been shown in [44] that $x_s[i]$ approximately follows the complex Gaussian distribution when the number

of subcarriers is large based on the central limit theorem. Since $x_s[i - \mu_{(s,d)}]$ have same distribution with $x_s[i]$ and recall Eq.(3), [44] has shown it become complex gaussian with variance $2\sigma_s^2$ (σ_s^2 is the variance of modulated symbol X_s 's real/imaginary part) by the central limit theorem. As a result, $\mathbb{E}[|x_s[r]|^2] = 2\sigma_s^2$.

Let $D_{(k,z)}[r] = F_{(k,z)}[r] + V_{(k,z)}[r]$, which is a real-valued variable. We first calculate $\mathbb{E}[D_{(k,z)}^2[r]]$ for item ①. As $\mathbb{E}_x[x_s[r]] = 0$, $\mathbb{E}_x[q_{(s,d),(k,z)}[r]] = 0$ and $\mathbb{E}_x[q_{(s,d),(k,z)}^*[r]] = 0$. We thus have that

$$\begin{aligned} \text{Var}_x(\textcircled{1}) &= \mathbb{E}_x[\textcircled{1}^2] - (\mathbb{E}_x[\textcircled{1}])^2 \\ &= \frac{1}{m_1^2} \mathbb{E}_x \left[\left(\sum_{r=(m_1-1)i+1}^{m_1 i} D_{(k,z)}[r] \right)^2 \right] \\ [3pt] &= \frac{1}{m_1} \sum_{e_1 \in \mathcal{E}} \sum_{\substack{e_2 \in \mathcal{E} \\ e_2 \neq e_1}} \mathbb{E}_x[|q_{e_1,(k,z)}[r]|^2] \mathbb{E}_x[|q_{e_2,(k,z)}^*[r]|^2] \\ &= \frac{4}{m_1} \sum_{e_1 \in \mathcal{E}} \sum_{\substack{e_2 \in \mathcal{E} \\ e_2 \neq e_1}} |h_{e_1,(k,z)}^{\text{eq}}|^2 |h_{e_2,(k,z)}^{\text{eq}}|^2 \sigma_s^2 \sigma_l^2, \end{aligned}$$

where $e_1 = (s, d)$ and $e_2 = (l, v)$. Using the Chebyshev's inequality, we have $\mathbb{P}[\textcircled{1} > A_1] \leq \frac{\text{Var}(\textcircled{1})}{A_1^2}$.

For item ②, we know $\mathbb{E}[\textcircled{2}] = 0$ and

$$\begin{aligned} \text{Var}_x(\textcircled{2}) &= \mathbb{E}_x[\textcircled{2}^2] - (\mathbb{E}_x[\textcircled{2}])^2 \\ &= -4 \left(\sum_{(s,d) \in \mathcal{E}} |h_{(s,d),(k,z)}^{\text{eq}}|^2 \sigma_s^2 \right)^2 \\ &\quad + \frac{1}{m_2^2} \mathbb{E} \left[\left(\sum_{r=(m_2-1)i+1}^{m_2 i} \sum_{(s,d) \in \mathcal{E}} |h_{(s,d),(k,z)}^{\text{eq}}|^2 |x_s[r]|^2 \right)^2 \right] \\ &= \frac{1}{m_2} \sum_{(s,d) \in \mathcal{E}} |h_{(s,d),(k,z)}^{\text{eq}}|^4 (\mathbb{E}_x[|x_s[r]|^4] - 4\sigma_s^4) \end{aligned}$$

Using Eq. (3), we can compute

$$\begin{aligned} \mathbb{E}_x[|x_s[r]|^4] &= \frac{1}{N_c^2} \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} |X_s[i]|^2 |X_s[j]|^2 \\ &\quad + \frac{1}{N_c^2} \sum_{i=0}^{N_c-1} \sum_{\substack{j=0 \\ j \neq i}}^{N_c-1} |X_s[i]|^2 |X_s[j]|^2 \\ &= 8\sigma_s^4 - \frac{4}{N_c} \sigma_s^4 \end{aligned}$$

Using the Chebyshev's inequality, we have $\mathbb{P}[\textcircled{2} > A_2] \leq \frac{\text{Var}(\textcircled{2})}{A_2^2}$.

For item ③, since $v[i] \sim \mathcal{CN}(\mathbf{0}, \sigma_v^2 \mathbf{1})$, and $n_z[i] = \tilde{\omega}_{(k,z)}^T v[i]$, we have $n_z[i] \sim \mathcal{CN}(0, \sigma_v^2)$. Thus, $\mathbb{E}_v[\textcircled{3}] = \sigma_v^2$ and

$$\text{Var}(\textcircled{3}) = \mathbb{E}_n[\textcircled{3}^2] - (\mathbb{E}_n[\textcircled{3}])^2$$

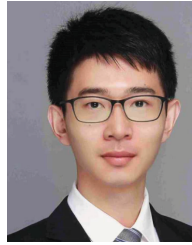
$$\begin{aligned}
&= \frac{1}{m_3^2} \mathbb{E}_n \left[\left(\sum_{r=(m_3-1)i+1}^{m_3 i} (n_z^I[r])^2 + (n_z^Q[r])^2 \right) \right] - \sigma_v^4 \\
&= \frac{1}{m_3} \sigma_v^4.
\end{aligned}$$

Using the Chebyshev's inequality, we have $\mathbb{P}[|\textcircled{3} - \sigma_v^2| > A_3] \leq \frac{\text{Var}(\textcircled{3})}{A_3^2} = \frac{\sigma_v^4}{m_3 A_3^2}$.

REFERENCES

- [1] H. Li, D. Ding, and Y. Pi, "Interference graph estimation for full-duplex mm wave backhauling: A power control approach," in *Proc. GLOBECOM*, pp. 6273–6278, Dec. 2023.
- [2] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 4th Quart., 2016.
- [3] M. Polese et al., "Integrated access and backhaul in 5G mmWave networks: Potential and challenges," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 62–68, Mar. 2020.
- [4] *Technical Specification Group Radio Access Network; NR; Study on Integrated Access and Backhaul (Release 16)*, document TR 38.874, Dec. 2018.
- [5] S. Ranjan, P. Jha, A. Karandikar, and P. Chaporkar, "A flexible IAB architecture for beyond 5G network," *IEEE Commun. Standards Mag.*, vol. 7, no. 3, pp. 64–71, Sep. 2023.
- [6] I. P. Roberts, J. G. Andrews, H. B. Jain, and S. Vishwanath, "Millimeter-wave full duplex radios: New challenges and techniques," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 36–43, Feb. 2021.
- [7] J. Zhang, H. Luo, N. Garg, A. Bishnu, M. Holm, and T. Ratnarajah, "Design and analysis of wideband in-band-full-duplex FR2-IAB networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4183–4196, Jun. 2022.
- [8] M. Gupta, I. P. Roberts, and J. G. Andrews, "System-level analysis of full-duplex self-backhauled millimeter wave networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1130–1144, Feb. 2023.
- [9] B. Yang, G. Mao, M. Ding, X. Ge, and X. Tao, "Dense small cell networks: From noise-limited to dense interference-limited," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4262–4277, May 2018.
- [10] S. Goyal, P. Liu, S. S. Panwar, R. A. Difazio, R. Yang, and E. Bala, "Full duplex cellular systems: Will doubling interference prevent doubling capacity?" *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121–127, May 2015.
- [11] J.-H. Kwon, B. Lim, and Y.-C. Ko, "Resource allocation and system design of out-band based integrated access and backhaul network at mmWave band," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6503–6517, Jun. 2022.
- [12] M. Pagin, T. Zugno, M. Polese, and M. Zorzi, "Resource management for 5G NR integrated access and backhaul: A semi-centralized approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 753–767, Feb. 2022.
- [13] Y. Li, J. Luo, W. Xu, N. Vucic, E. Pateromichelakis, and G. Caire, "A joint scheduling and resource allocation scheme for millimeter wave heterogeneous networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
- [14] Z. Ma, B. Li, Z. Yan, and M. Yang, "QoS-oriented joint optimization of resource allocation and concurrent scheduling in 5G millimeter-wave network," *Comput. Netw.*, vol. 166, Jan. 2020, Art. no. 106979.
- [15] S. C. Ergen and P. Varaiya, "TDMA scheduling algorithms for wireless sensor networks," *Wireless Netw.*, vol. 16, no. 4, pp. 985–997, May 2010.
- [16] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless networks: The joint routing and scheduling problem," in *Proc. 9th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2003, pp. 42–54.
- [17] M. Peltomaki, J.-M. Koljonen, O. Tirkkonen, and M. Alava, "Algorithms for self-organized resource allocation in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 346–359, Jan. 2012.
- [18] W. Cui, K. Shen, and W. Yu, "Spatial deep learning for wireless scheduling," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1248–1261, Jun. 2019.
- [19] Y. Zhao, W. Li, J. Wu, and S. Lu, "A joint scheduling and resource allocation scheme for millimeter wave heterogeneous networks," in *Proc. IEEE INFOCOM*, Mar. 2015, pp. 2218–2226.
- [20] W. Li, J. Zhang, and Y. Zhao, "Conflict graph embedding for wireless network optimization," in *Proc. IEEE INFOCOM*, 2017, pp. 1–9.
- [21] Q. Xue, X. Fang, and C.-X. Wang, "Beamspace SU-MIMO for future millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1564–1575, Jul. 2017.
- [22] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.
- [23] *Study on Channel Model for Frequencies From 0.5 To 100 GHz (Release 16)*, document TR 38.901, 3GPP, V16.1.0, 2019.
- [24] O. Semiari, W. Saad, M. Bennis, and Z. Dawy, "Inter-operator resource management for millimeter wave multi-hop backhaul networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5258–5272, Aug. 2017.
- [25] A. Ortiz, A. Asadi, G. H. Sim, D. Steinmetzer, and M. Hollick, "SCAROS: A scalable and robust self-backhauling solution for highly dynamic millimeter-wave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 12, pp. 2685–2698, Dec. 2019.
- [26] M. Lee, G. Yu, and G. Y. Li, "Graph embedding-based wireless link scheduling with few training samples," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2282–2294, Apr. 2021.
- [27] L. Luo, Z. Liu, Z. Chen, M. Hua, W. Li, and B. Xia, "Age of information-based scheduling for wireless D2D systems with a deep learning approach," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1875–1888, Sep. 2022.
- [28] H. Elgendi, M. Mäenpää, T. Levanen, T. Ihalainen, S. Nielsen, and M. Valkama, "Interference measurement methods in 5G NR: Principles and performance," in *Proc. 16th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2019, pp. 233–238.
- [29] Y.-R. Tsai, H.-Y. Huang, Y.-C. Chen, and K.-J. Yang, "Simultaneous multiple carrier frequency offsets estimation for coordinated multi-point transmission in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4558–4568, Sep. 2013.
- [30] H. Li, L. Han, R. Duan, and G. M. Garner, "Analysis of the synchronization requirements of 5G and corresponding solutions," *IEEE Commun. Standards Mag.*, vol. 1, no. 1, pp. 52–58, Mar. 2017.
- [31] P. Chen and Z. Yang, "Understanding precision time protocol in today's Wi-Fi networks: a measurement study," in *Proc. USENIX Annu. Tech. Conf.*, 2021, pp. 597–610.
- [32] Y. Medjahdi, M. Terre, D. L. Ruyet, D. Roviras, and A. Dziri, "Performance analysis in the downlink of asynchronous OFDM/FBMC based multi-cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2630–2639, Aug. 2011.
- [33] *Technical Specification Group Radio Access Network; NR; Physical Channels and Modulation (Release 15)*, document TS 38.211, V16.6.0, 3GPP, 2021.
- [34] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [35] V. Va, J. Choi, and R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5014–5029, Jun. 2017.
- [36] A. Dax, "Kaczmarz's anomaly: A surprising feature of Kaczmarz's method," *Linear Algebra Appl.*, vol. 662, pp. 136–162, Apr. 2023.
- [37] S. Chandrasekaran and I. Ipsen, "Perturbation theory for the solution of systems of linear equations," Dept. Comput. Sci., Yale Univ., New Haven, CT, USA, Tech. Rep. YALEU/DCS/RR-886, 1991.
- [38] C. Sun and R. Dai, "Rank-constrained optimization and its applications," *Automatica*, vol. 82, pp. 128–136, Aug. 2017.
- [39] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [40] L. Zhao, Z. Wei, D. W. K. Ng, J. Yuan, and M. C. Reed, "Multi-cell hybrid millimeter wave systems: Pilot contamination and interference mitigation," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5740–5755, Nov. 2018.
- [41] L. Zhao, D. W. Kwan Ng, and J. Yuan, "Multiuser precoding and channel estimation for hybrid millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
- [42] D. W. K. Ng, Y. Wu, and R. Schober, "Power efficient resource allocation for full-duplex radio distributed antenna networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2896–2911, Apr. 2016.

- [43] B. Li, L.-L. Yang, R. G. Maunder, and S. Sun, "Resource allocation in millimeter-wave multicarrier-division duplex systems with hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 7921–7935, Aug. 2021.
- [44] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 282–289, Jul. 2001.



Yibo Pi (Member, IEEE) received the B.S. and M.S. degrees in electrical and computer engineering from Shanghai Jiao Tong University in 2012 and 2015, respectively, and the Ph.D. degree in computer science from the University of Michigan in 2021. He is currently an Assistant Professor with the UM-SJTU Joint Institute, Shanghai Jiao Tong University. His current research interests include network measurement, wireless networking, and network telemetry.



Haorui Li (Student Member, IEEE) received the bachelor's degree in electronic information and electrical engineering from Dalian University of Technology. He is currently pursuing the Ph.D. degree with the Network Measurement and Systems Laboratory, UM-SJTU Joint Institute, Shanghai Jiao Tong University. His current research interests include mmWave backhauling networks, interference graph estimation, and network telemetry.



Xudong Wang (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology in 2003. He was a Senior Research Engineer, a Senior Network Architect, and a Research and Development Manager with several companies. He is currently the John Wu and Jane Sun Professor with the UM-SJTU Joint Institute, Shanghai Jiao Tong University. He is also an Affiliate Professor with the Department of Electrical and Computer Engineering, University of Washington. His research interests include wireless



Daqian Ding (Student Member, IEEE) received the bachelor's degree in information engineering from Beijing University of Posts and Telecommunications. He is currently pursuing the master's degree with the Network Measurement and Systems Laboratory, UM-SJTU Joint Institute, Shanghai Jiao Tong University. His current research interests include 5G networks and network resource scheduling.

communication networks, joint communications and sensing, edge computing, and federated learning. He was a Voting Member of IEEE 802.11 and 802.15 Standard Committees. He was the General Chair of the 2017 IEEE 5G Summit in Shanghai and the TPC Co-Chair of the 32nd International Conference on Information Networking. He was the Demo Co-Chair of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC 2006), the Technical Program Co-Chair of the Wireless Internet Conference (WICON) 2007, and the General Co-Chair of WICON 2008. He was an Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Ad Hoc Networks* (Elsevier), and *China Communications*.