Predictive Delay-Aware Scheduling with Receiver Rotation Detection and mmWave Channel Learning

Yifei Sun, Bojie Lv, Rui Wang, Haisheng Tan, Francis C.M. Lau

Abstract—In this paper, the joint downlink delay-aware scheduling in a large time span, where the rotation of User Equipments (UEs) may lead to significant channel variation, is investigated via a novel approximate Markov Decision Process (MDP) method. Specifically, we consider the joint downlink power allocation and receiving UE selection of a number of successive frames in a millimeter Wave (mmWave) system with quasi-static scattering clusters in the channel and rotating UEs. The propagation statistics of scattering clusters can be tracked via a learning method. Since the rotation of UEs can be detected, future channel statistics can be forecast via embedded motion sensors. Hence, the overall scheduling is formulated as a finite-horizon MDP with non-stationary predictable state transition probabilities, where the average queuing delay and probability of transmission buffer overflow are considered in the objective of scheduling optimization. A novel low-complexity solution framework with an analytical performance bound is proposed to save the efforts of value iteration. Benefiting from the forecast of system statistics, superior performance to the benchmarks is shown by numerical simulations, particularly in the suppression of buffer overflow rate. Preliminary experiments via an mmWave testbed are conducted to demonstrate the feasibility of the sensor-assisted mmWave beam alignment.

I. INTRODUCTION

Millimeter Wave (mmWave) communication is a key technique for future wireless communication systems due to its ultra-wide spectrum [2], [3]. In the mmWave band, because of directional patterns of antenna arrays and a very limited number of propagation paths between the Base Station (BS) and User Equipments (UEs), UEs with motions (e.g., rotation) may suffer from severe Signal-to-Noise Ratio (SNR) fluctuation, which may lead to large transmission queuing delay and buffer overflow rate. We show in this paper that exploiting motion sensors at UEs, the SNR fluctuation becomes predictable. Hence, large-time-scale joint scheduling could effectively prepare the transmitter against packet drop and Quality-of-Service (QoS) loss. This raises a new stochastic optimization issue of joint transmission scheduling in a large time scale with non-stationary but predictable channel statistics and random packet arrivals.

A. Related Work

In mmWave systems, beam alignment is essential for high data rate communications. There have been a number of works on mmWave beam training algorithm design. For example in [4], the Rosenbrock algorithm was proposed for searching a good transceiver beam pair from all possible beams to establish a communication link with a high data rate. In [5], a multi-beam search approach with predesigned analog hierarchical codebooks was proposed to achieve a high success rate of beam alignment. Exploiting the sparsity of mmWave Multiple-Input Multiple-Output (MIMO) channel matrix, compressive sensing has been widely considered for low-complexity channel estimation and beam alignment [6]–[8]. However, the above beam training methods may lead to significant overhead when the UEs are in motion. Particularly, rapid changes in Angle of Arrival (AoA) or Angle of Departure (AoD) induced by UE rotation may evoke more stringent requirements than translational mobility [9], [10]. In fact, embedded motion sensors at UEs can be exploited to assist beam tracking when UEs are rotating. For instance, in [9], [11], the motion sensors were utilized to predict rotation angles, such that the beam directions could be compensated to maintain the alignment. In practice, due to the limitation of antenna Field-of-View[1] (FoV), phased arrays of mobile UEs operating on mmWave may not be able to steer the beams towards all directions within $360°$ (typically, FoV is less than $120°$ [16]). Hence, UE rotation may make Line-of-Sight (LoS) path out of the FoV and thus result in link disconnection. Therefore, a robust beam alignment design should jointly exploit the motion sensors and geometric information on scattering clusters, such that the best propagation path within the antenna FoV can be found quickly.

In addition to the frequent beam alignment in the physical layer, UE rotation and limited antenna FoV may lead to large packet delivery latency or even packet drop rate in the Medium Access Control (MAC) layer due to link quality degradation. Notice that delay-aware scheduling in a large time-scale has been investigated in different mmWave communication systems. For example in [17], a delay-aware admission control and beam allocation problem in a fronthaul downlink network was investigated via the Lyapunov method. In [18], delay-aware scheduling in an integrated mmWave and

Yifei Sun is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China, and also the Department of Computer Science, the University of Hong Kong, Hong Kong, China. Bojie Lv and Rui Wang are with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China. Haisheng Tan is with LINKE Lab, University of Science and Technology of China, Hefei, China. Francis C.M. Lau is with the Department of Computer Science, the University of Hong Kong, Hong Kong, China.

[1]This may be because of the inter-element spacing, sidelobes [12], [13], intrinsic properties of patch antenna elements [14] or self-induced blockage by users' hands and bodies [15].

sub-6G communication system was studied via the tabular Q-learning approach. In [19], a number of throughput-optimal delay-aware policies were proposed for joint dynamic routing and link scheduling design in mmWave-based cellular backhaul networks. However, these works did not consider the impact of UE mobility on the transmission queuing delay and transmitter buffer overflow. In fact, the state transition of mmWave communication systems would become non-stationary when the UEs are rotating. This is because of the variation of channel statistics. Hence, the scheduling designs of the above works, which are based on stationary system statistics, cannot be applied in the rotation scenario. To the best of our knowledge, delay-aware scheduling in a non-stationary mmWave communication system has not been well addressed.

B. Our Contributions

In this paper, we shed light on the above issue by proposing a predictive transmission scheduling framework for non-stationary mmWave systems due to UE rotation and limited antenna FoV. Specifically, we consider the joint downlink transceiver beam design, power allocation and user selection in a number of frames for an mmWave system with UE rotation, where the average queuing delay and buffer overflow rate are considered in the optimization objective. Due to the randomness of channel gain and packet arrival, the overall problem is a finite-horizon Markov Decision Process (MDP). The main contributions of this paper are summarized below.

- We propose a sensor-assisted learning method to predict the non-stationary transition kernel of the system versus time. Note that the above finite-horizon MDP cannot be solved unless its time-varying transition kernel, particularly the channel statistics, is known. We show that the prediction is feasible by tracking the scattering clusters in the channel and sensing the orientation of phased arrays at the UEs.
- We propose a low-complexity semi-distributed method to solve the above finite-horizon MDP with a non-trivial analytical upper bound on the achievable average system cost. The proposed algorithm achieves a good balance between complexity and performance via one-step policy iteration over analytical value functions. As a comparison, the optimal policy of finite-horizon MDP is intractable; the neural network-based methods require computation-intensive training before scheduling, and the performance can hardly be analyzed.
- An experiment based on an mmWave testbed is demonstrated to verify the feasibility of UE orientation sensing and the adopted beam alignment scheme, which exploits both motion sensors and channel statistics.

With numerical simulations, we verify the performance gain of the proposed algorithms and offer novel insights on how UE mobility affects the downlink scheduling in a large time scale, shedding light on the predictive delay-aware scheduling. For example, the BS tends to schedule more transmission resources to the UEs which will suffer from worse channels due to rotation.

The rest of this paper is organized as below. The system model is introduced in Section II. In Section III, we formulate the scheduling of downlink transmission as a dynamic programming problem. A low-complexity suboptimal solution is proposed and analyzed in Section IV. Efficient learning algorithms are developed in Section V to address the issue of unknown system statistics. The experiments, numerical simulations and discussions are provided in Section VI, and the conclusion is drawn in Section VII.

Notation: Bold lowercase $\mathbf{a}$ denotes a column vector, bold uppercase $\mathbf{A}$ denotes a matrix, non-bold letters a, A denote scalar values, and letter $\mathcal{A}$ denotes sets. Using this notation, $|a|$ is the magnitude of a scalar, $(a)^+$ denotes $\max(0, a)$. $[\mathbf{A}]_{i,j}$, $\mathbf{A}^\mathsf{T}$, and $\mathbf{A}^\mathsf{H}$ denote the (i, j)-th element, transpose, and conjugate transpose of $\mathbf{A}$, respectively. $\mathbf{1}_i$ denotes the column vector whose i-th element is 1 and other elements are 0. $\mathcal{CN}(m, R)$ denotes complex Gaussian distribution with mean m and variance R. $\mathbb{E}[\cdot]$ and $\mathrm{Var}[\cdot]$ denote an expectation and a variance respectively. $\mathbb{I}[\cdot]$ denotes an indicator function, whose value is 1 when the event is true and 0 otherwise. $\mathbb{C}^{M \times N}$ and $\mathbb{R}^{M \times N}$ denote spaces of $M \times N$ matrices with complex and real entries, respectively.

II. SYSTEM MODEL

In this section, we first introduce the overview of an mmWave system with UE rotation as well as the channel model. Then a beam alignment scheme based on channel statistical parameters is proposed. Finally, we define the system queue dynamics. The main notations used are listed in Table I.

A. mmWave System with UE Rotation

The downlink mmWave transmission scheduling from one BS to K UEs in a two-dimensional plane $\mathfrak{P}$ is considered[2], where the set of UEs is denoted by $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$. An example of the system is illustrated in Fig. 1(a). The analog MIMO transceiver with a single Radio Frequency (RF) chain and a half-wavelength Uniform Linear phased Array (ULA) is adopted at both the BS and UEs as in [3], where the ULAs at the BS and UEs are with N_T and N_R antenna elements, respectively[3]. Hence, analog precoders and combiners can be adopted at the BS and UEs respectively to enhance the receiving SNR [3], [22]. Some UEs may be rotating in the plane $\mathfrak{P}$ during downlink transmission. This might happen when playing games with Virtual Reality (VR) headsets or watching videos via handsets. Due to UE rotation and limited antenna FoV, the downlink channel, particularly the number of scattering clusters in the FoV and their arrival (or departure) angles with respect to the receiving (or transmission) phased arrays, may change significantly. Thus, the downlink channel statistics are not temporally homogeneous.

---

[2]For elaboration convenience, we consider the mmWave communication in a two-dimensional plane. However, the same principles can be extended to transmission scheduling with three-dimensional rotation, as long as the planar phased arrays and three-dimensional angular acceleration detection via gyroscopes are considered. In fact, a two-dimensional assumption is commonly made to simplify the algorithm description, e.g., [6], [20], [21].

[3]We assume that all ULAs at the UEs have the same number of antenna elements for elaboration convenience. However, the proposed scheme in this paper can be easily extended to UEs with different numbers of antenna elements by defining different combiner codebooks.
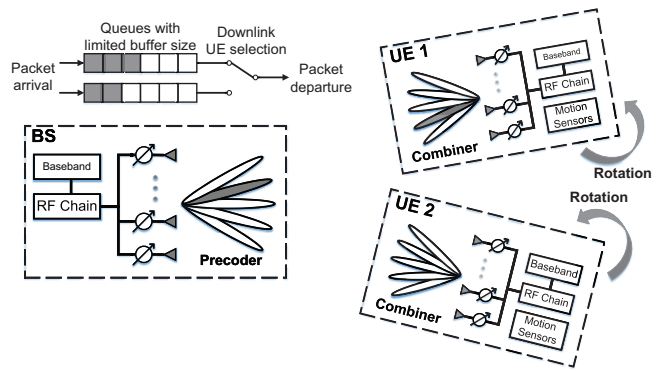
TABLE I. Main notations.

Symbol	Description
Network	
$K/\mathcal{K}$	Number/Set of UEs
$N_\mathrm{R}/N_\mathrm{T}$	Number of antenna elements at the UEs/BS
$T/\mathcal{T}$	Number/Set of frames in a scheduling period
$\mathcal{I}_t^\mathrm{SCSI}$	SCSI in the t-th frame
$\mathcal{I}_\mathrm{sta}^\mathrm{SCSI}$	Static SCSI in the scheduling period
Channel	
$\mathbf{H}_{t,k}$	Channel matrix from the BS to the k-th UE
N_k^cl	Number of scattering clusters in the k-th channel
$N_{k,i}^\mathrm{ray}$	Number of propagation paths in the (k,i)-th cluster
$\alpha_{t,k,i,\ell}$	Instantaneous complex gain
$\phi_{t,k,i,\ell}/\theta_{t,k,i,\ell}$	AoA/AoD of the (k,i,ℓ)-th path in the t-th frame
$\Lambda_\mathrm{R}(\phi)/\Lambda_\mathrm{T}(\theta)$	Gains of antenna elements at the UEs/BS
$\mathbf{a}_\mathrm{R}(\phi)/\mathbf{a}_\mathrm{T}(\theta)$	Array response vectors of the ULAs at the UEs/BS
$\mathbf{w}_{t,k}/\mathbf{f}_{t,k}$	Combiner/Precoder at the k-th UE/BS
$\mathcal{W}/\mathcal{F}$	Combiner/Precoder codebook
d_t	UE selection of downlink transmission
P_t	Downlink transmission power
$P_{t,k}$	Transmission power if the k-th UE is selected
$R_{t,k}$	Spectral efficiency
$Y_{t,k}$	Baseband channel power gain
Queuing	
$Q_{t,k}/Q_{t,k}^\mathrm{D}$	Pre-decision/Post-decision queue length
$A_{t,k}/D_{t,k}$	Number of arrival/departure packets
Problem	
$\mathcal{S}_t/\mathcal{S}_t^\mathrm{D}$	Aggregation of pre-decision/post-decision states
$\mathcal{Q}_t/\mathcal{Q}_t^\mathrm{D}$	Aggregation of pre-decision/post-decision QSI
$\mathcal{Y}_t$	Aggregation of baseband channel power gains
Ω_t	Scheduling policy in the t-th frame
$\Omega_t^\star$	Optimal scheduling policy in the t-th frame
Π	Base policy
Ψ_t	Proposed policy in the t-th frame
g_t	System cost in the t-th frame
G	Overall cost in a scheduling period

(a) System model.

(b) Channel model.

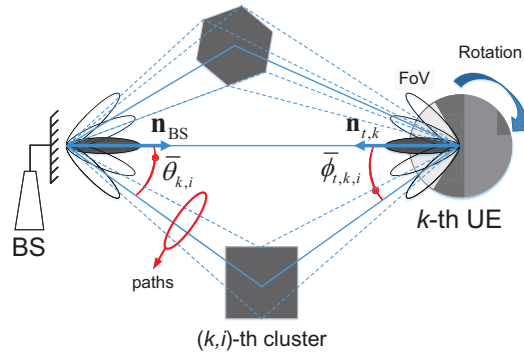Fig. 1. Illustration of the system model and the channel model.

Specifically, the UE rotation is modeled as follows. The transmission time is organized by frames[4], and the wireless channel is assumed to be quasi-static within one frame. UEs are either quasi-static (with zero angular velocity) or rotating with constant angular velocities in a number of frames. Similar to [23], [24], the period during which all the UEs are rotating with constant angular velocities is referred to as one *scheduling period*[5]. Each scheduling period consists of T frames and the set of frame indices in one scheduling period is denoted by $\mathcal{T} \triangleq \{1, 2, \ldots, T\}$. We shall focus on the joint UE selection and power allocation within one scheduling period, where the angular velocity of the k-th UE is denoted as ω_k, its boresight direction at the 1-st frame of the scheduling period is denoted as $\mathbf{n}_{1,k} \in \mathfrak{P}$, and the boresight direction of the BS's array is denoted as $\mathbf{n}_\mathrm{BS} \in \mathfrak{P}$. Then for the k-th UE, the rotation angle during the first $(t-1)$ frames ($\forall t \in \mathcal{T}$) and the boresight

direction in the t-th frame can be written as

$$\Delta\phi_{t,k} \triangleq (t-1)\omega_k T_\mathrm{F} \tag{1}$$

and

$$\mathbf{n}_{t,k} = \begin{bmatrix} \cos(\Delta\phi_{t,k}) & -\sin(\Delta\phi_{t,k}) \\ \sin(\Delta\phi_{t,k}) & \cos(\Delta\phi_{t,k}) \end{bmatrix} \mathbf{n}_{1,k}, \tag{2}$$

respectively, where T_F denotes the frame duration.

The motion sensors embedded in the UEs are able to detect the rotation. For instance, the magnetometers and gyroscopes are able to detect the orientations and angular velocities of UEs, respectively [9]. Hence, we assume the orientations of each UE in all the frames can be predicted at the beginning of one scheduling period. Moreover, we consider the scenario where the locations of scattering clusters are static. Therefore, with the knowledge of $\{\mathbf{n}_{t,k}|\forall t, k\}$, $\mathbf{n}_\mathrm{BS}$ and scattering clusters, the distribution of Channel State Information (CSI) in every frame of the scheduling period can be predicted at the very beginning, which is elaborated in the following part.

B. Channel Model

As in [3], [25], [26], the geometric channel model illustrated in Fig. 1(b) is adopted. Specifically, there are N_k^cl quasi-static scattering clusters in the channel from the BS to the k-th UE[6], and $N_{k,i}^\mathrm{ray}$ propagation paths in the i-th cluster. The LoS path can be treated as a special cluster with a single strong path. For convenience of exposition, the i-th cluster in the channel

---

[4]In fact, the scheduling unit in this work, i.e., the frame, can be not only a physical frame but other resource units in the time domain, such as a slot, a mini-slot or a symbol. The proposed scheduling scheme can be extended to these finer-granularity timescales as long as the orientations of UEs can be predicted.

[5]In this paper, to simplify the elaboration of the algorithms later on, we consider the transmission optimization within one scheduling period, where the angular velocities of UEs are constant. In fact, the scheduling methods proposed in this work are feasible as long as the UE orientations can be predicted deterministically or statistically, e.g., the angular velocities or accelerations are constant or their distributions can be predicted.

[6]We consider quasi-static scattering clusters like walls, monitors, etc. Their locations can be sensed before the scheduling. In the case of mobile scattering clusters (like walking persons), wireless sensing can be exploited to track their trajectories [27]–[29]. As a result, it might be feasible to treat them as quasi-static in a scheduling period.

$$\textbf{PDF of AoAs: } f_{\phi;t,k,i}(\phi) \triangleq \begin{cases} \frac{\beta_{\phi;k,i}}{\sqrt{2}\sigma_{\phi;k,i}} \exp\left(-\frac{\sqrt{2}|\phi-\bar{\phi}_{t,k,i}|}{\sigma_{\phi;k,i}}\right) & |\phi-\bar{\phi}_{t,k,i}| \leq \delta_{\phi;k,i}, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$\textbf{PDF of AoDs: } f_{\theta;k,i}(\theta) \triangleq \begin{cases} \frac{\beta_{\theta;k,i}}{\sqrt{2}\sigma_{\theta;k,i}} \exp\left(-\frac{\sqrt{2}|\theta-\bar{\theta}_{k,i}|}{\sigma_{\theta;k,i}}\right) & |\theta-\bar{\theta}_{k,i}| \leq \delta_{\theta;k,i}, \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

between the BS and the k-th UE is referred to as the (k,i)-th cluster, and the ℓ-th path of the (k,i)-th cluster is referred to as the (k,i,ℓ)-th path. Similar to [30], the channel matrix from the BS to the k-th UE in the t-th frame $\mathbf{H}_{t,k} \in \mathbb{C}^{N_R \times N_T}$ can be represented by

$$\mathbf{H}_{t,k} = \sum_{i=1}^{N_k^{cl}} \sum_{\ell=1}^{N_{k,i}^{ray}} \alpha_{t,k,i,\ell} \mathbf{a}_R(\phi_{t,k,i,\ell}) \mathbf{a}_T^H(\theta_{t,k,i,\ell}) \\ \times \Lambda_R(\phi_{t,k,i,\ell}) \Lambda_T(\theta_{t,k,i,\ell}), \tag{3}$$

where $\alpha_{t,k,i,\ell}$, $\phi_{t,k,i,\ell}$ and $\theta_{t,k,i,\ell}$ are the instantaneous complex gain, AoA and AoD of the (k,i,ℓ)-th path in the t-th frame, respectively, $\Lambda_R(\phi)$ and $\Lambda_T(\theta)$ refer to the receiving and transmission antenna gains respectively, $\mathbf{a}_R(\phi)$ and $\mathbf{a}_T(\theta)$ represent the normalized array response vectors of the ULAs at the UEs and the BS respectively. Thus,

$$\mathbf{a}_R(\phi) \triangleq \frac{1}{\sqrt{N_R}} \left[1, e^{-j\pi \sin\phi}, \ldots, e^{-j\pi(N_R-1)\sin\phi}\right]^T, \tag{4}$$

$$\mathbf{a}_T(\theta) \triangleq \frac{1}{\sqrt{N_T}} \left[1, e^{-j\pi \sin\theta}, \ldots, e^{-j\pi(N_T-1)\sin\theta}\right]^T. \tag{5}$$

The patterns of antenna elements are modeled as identical and ideal-sectored as in [26], i.e.,

$$\Lambda_R(\phi) \triangleq \begin{cases} 1 & \phi \in [\phi_{\min}, \phi_{\max}], \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

$$\Lambda_T(\theta) \triangleq \begin{cases} 1 & \theta \in [\theta_{\min}, \theta_{\max}], \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Remark 1 (Effect of UE Rotation). *Due to the limited FoV of the receiving antenna elements (i.e., Λ_R), the phased array cannot capture all propagation rays in $360°$ azimuth. Thus, the effective scattering clusters in the channel may vary during the UE rotation. For example, the paths via the scattering cluster represented by the black square in Fig. 1(b) may be out of the FoV after a few frames when the k-th UE is rotating clockwise.*

Similar to the channel model in [3], [25], [26], the instantaneous gains $\{\alpha_{t,k,i,\ell}\}$, AoAs $\{\phi_{t,k,i,\ell}\}$ and AoDs $\{\theta_{t,k,i,\ell}\}$ are drawn from independent distributions in each frame. Specifically, $\alpha_{t,k,i,\ell} \sim \mathcal{CN}(0, \sigma_{\alpha;k,i}^2)$ follows the complex Gaussian distribution with zero mean and variance $\sigma_{\alpha;k,i}^2$. Moreover, we model the angular distribution of propagation paths as truncated Laplacian as in [31]. Thus, the Probability Density Functions (PDFs) of $\phi_{t,k,i,\ell}$ and $\theta_{t,k,i,\ell}$ ($\forall t,k,i,\ell$) are given by (8) and (9), where $\sigma_{\phi;k,i}$ and $\sigma_{\theta;k,i}$ denote the standard deviations of the AoA and AoD distributions of the (k,i)-th cluster respectively, $\bar{\phi}_{t,k,i}$ and $\bar{\theta}_{k,i}$ represent the mean AoA and AoD in the t-th frame respectively, $\delta_{\phi;k,i}$ and $\delta_{\theta;k,i}$ denote the truncation limits of the AoA and

AoD distributions respectively, the normalization factors are $\beta_{\phi;k,i} = \frac{1}{1-e^{-\sqrt{2}\delta_{\phi;k,i}/\sigma_{\phi;k,i}}}$ and $\beta_{\theta;k,i} = \frac{1}{1-e^{-\sqrt{2}\delta_{\theta;k,i}/\sigma_{\theta;k,i}}}$, respectively. Due to the rotation of UEs and the quasi-static scattering clusters, the mean AoD of the (k,i)-th cluster is a constant, and the mean AoAs of the (k,i)-th cluster in one scheduling period satisfy

$$\bar{\phi}_{t,k,i} = \bar{\phi}_{1,k,i} + \Delta\phi_{t,k} = \bar{\phi}_{1,k,i} + (t-1)\omega_k T_F, \ \forall t,k,i. \tag{10}$$

For elaboration convenience, we define the Statistical Channel State Information (SCSI) as the tuple of parameters sufficiently characterizing the distribution of channel matrix in (3) as follows.

Definition 1 (SCSI). *The SCSI in the t-th frame is defined by*

$$\mathcal{I}_t^{SCSI} \triangleq \left(\mathcal{I}_{sta}^{SCSI}, \{\bar{\phi}_{t,k,i} | \forall k, i\}\right), \tag{11}$$

where $\mathcal{I}_{sta}^{SCSI} \triangleq \left(\{N_k^{cl} | \forall k\}, \{N_{k,i}^{ray} | \forall k,i\}, \{\sigma_{\alpha;k,i}^2 | \forall k,i\}, \{\bar{\theta}_{k,i} | \forall k,i\}, \{\sigma_{\theta;k,i}^2 | \forall k,i\}, \{\sigma_{\phi;k,i}^2 | \forall k,i\}\right)$ is the tuple of quasi-static channel statistical parameters.

Note that the component $\mathcal{I}_{sta}^{SCSI}$ of SCSI is constant which can be learned before the scheduling. Moreover, with the knowledge of phased array orientations $\{\mathbf{n}_{t,k} | \forall t,k\}$, $\{\bar{\phi}_{t,k,i} | \forall t,k,i\}$ can also be estimated. As a result, the distribution of channel matrices $\{\mathbf{H}_{t,k} | \forall t,k\}$ can be predicted at the beginning of the scheduling period. In the following elaboration, we first assume the knowledge of $\{\mathcal{I}_t^{SCSI} | \forall t\}$ is available, and discuss the learning algorithms for $\mathcal{I}_{sta}^{SCSI}$ in Section V.

C. SCSI-based Beam Alignment

Due to the single RF chain at the BS, one UE is selected for downlink transmission in each frame. Let $\mathbf{w}_{t,k} \in \mathbb{C}^{N_R \times 1}$ and $\mathbf{f}_{t,k} \in \mathbb{C}^{N_T \times 1}$ be the analog combiner at the k-th UE and the analog precoder at the BS in the t-th frame respectively, if the k-th UE is selected. Due to hardware limitation, the beam directions of $\mathbf{w}_{t,k}$ and $\mathbf{f}_{t,k}$ are not continuously adjustable, but selected from the following finite codebooks, respectively,

$$\mathbf{w}_{t,k} \in \mathcal{W} \triangleq \{\mathbf{a}_R(\phi_q) | q \in \mathcal{N}_R\}, \tag{12}$$

$$\mathbf{f}_{t,k} \in \mathcal{F} \triangleq \{\mathbf{a}_T(\theta_p) | p \in \mathcal{N}_T\}, \tag{13}$$

where $\phi_q \triangleq \arcsin\left(\frac{2(q-1)}{N_R}-1\right)$, $\theta_p \triangleq \arcsin\left(\frac{2(p-1)}{N_T}-1\right)$, $\mathcal{N}_R \triangleq \{1,2,\ldots,N_R\}$ and $\mathcal{N}_T \triangleq \{1,2,\ldots,N_T\}$. Hence, the spectral efficiency of the k-th UE in the t-th frame is given by

$$R_{t,k} = \log_2\left(1 + \frac{P_{t,k}Y_{t,k}}{N_0 W}\right), \tag{14}$$

where

$$Y_{t,k} \triangleq |\mathbf{w}_{t,k}^H \mathbf{H}_{t,k} \mathbf{f}_{t,k}|^2 \tag{15}$$

$$(q_{t,k}^\dagger, p_{t,k}^\dagger) = \underset{q \in \mathcal{N}_{\mathrm{R}}, p \in \mathcal{N}_{\mathrm{T}}}{\arg\max} \; \mathbb{E}_{\mathbf{H}_{t,k}} \left[\left| \mathbf{a}_{\mathrm{R}}^{\mathsf{H}}(\phi_q) \mathbf{H}_{t,k} \mathbf{a}_{\mathrm{T}}(\theta_p) \right|^2 \right] \tag{17}$$

$$= \underset{q \in \mathcal{N}_{\mathrm{R}}, p \in \mathcal{N}_{\mathrm{T}}}{\arg\max} \; \sum_{i=1}^{N_k^{\mathrm{cl}}} N_{k,i}^{\mathrm{ray}} \sigma_{\alpha;k,i}^2 \int_{\phi_{t,k,i}^-}^{\phi_{t,k,i}^+} \left| f_{\mathrm{R}}(\phi_q, \phi_{t,k,i,1}) \right|^2 f_{\phi;k,i}(\phi_{t,k,i,1}) \mathrm{d}\phi_{t,k,i,1}$$

$$\times \int_{\theta_{k,i}^-}^{\theta_{k,i}^+} \left| f_{\mathrm{T}}(\theta_p, \theta_{t,k,i,1}) \right|^2 f_{\theta;k,i}(\theta_{t,k,i,1}) \mathrm{d}\theta_{t,k,i,1}, \tag{18}$$

denotes the channel power gain in baseband, $P_{t,k}$ is the transmission power of the BS, N_0 is the noise power spectral density, and W is the bandwidth.

The analog precoder $\mathbf{f}_{t,k}$ and combiner $\mathbf{w}_{t,k}$ in each frame can be chosen to maximize the instantaneous channel power gain $Y_{t,k}$ via compressive-sensing-based channel estimation [6], [7] or exhaustive beam search [4], [5]. However, both approaches increase the signaling overhead. Instead, we exploit the SCSI prediction and adopt the following statistical beam alignment scheme, which maximizes the average baseband SNR without instantaneous CSI estimation.

Scheme 1 (SCSI-Based Beam Alignment). *The analog combiner and precoder for the k-th UE in the t-th frame are selected according to*

$$\mathbf{w}_{t,k} = \mathbf{a}_{\mathrm{R}}(\phi_{q_{t,k}^\dagger}), \; \mathbf{f}_{t,k} = \mathbf{a}_{\mathrm{T}}(\theta_{p_{t,k}^\dagger}), \tag{16}$$

where $(q_{t,k}^\dagger, p_{t,k}^\dagger)$ is given by (18), $\phi_{t,k,i}^+ = \max(\bar{\phi}_{t,k,i} + \delta_{\phi;k,i}, \phi_{\max})$, $\phi_{t,k,i}^- = \min(\bar{\phi}_{t,k,i} - \delta_{\phi;k,i}, \phi_{\min})$, $\theta_{k,i}^+ = \max(\bar{\theta}_{k,i} + \delta_{\theta;k,i}, \theta_{\max})$, $\theta_{k,i}^- = \min(\bar{\theta}_{k,i} - \delta_{\theta;k,i}, \theta_{\min})$, $f_{\mathrm{R}}(\phi_q, \phi) \triangleq \mathbf{a}_{\mathrm{R}}^{\mathsf{H}}(\phi_q) \mathbf{a}_{\mathrm{R}}(\phi) = \exp(-j\pi(\sin\phi - \sin\phi_q)(N_{\mathrm{R}} - 1)/2) \frac{\sin(\pi N_{\mathrm{R}}(\sin\phi - \sin\phi_q)/2)}{N_{\mathrm{R}} \sin(\pi(\sin\phi - \sin\phi_q)/2)}$ *and* $f_{\mathrm{T}}(\theta_p, \theta) \triangleq \mathbf{a}_{\mathrm{T}}^{\mathsf{H}}(\theta_p) \mathbf{a}_{\mathrm{T}}(\theta) = \exp(-j\pi(\sin\theta - \sin\theta_p)(N_{\mathrm{T}} - 1)/2) \frac{\sin(\pi N_{\mathrm{T}}(\sin\theta - \sin\theta_p)/2)}{N_{\mathrm{T}} \sin(\pi(\sin\theta - \sin\theta_p)/2)}$.

Since the integrals in (18) depend on SCSI in the t-th frame $\mathcal{I}_t^{\mathrm{SCSI}}$ which can be predicted from $\mathcal{I}_1^{\mathrm{SCSI}}$ and angular velocities $\{\omega_k | \forall k \in \mathcal{K}\}$, the precoders and combiners for all the frames within one scheduling period can be pre-designed at the beginning of the scheduling period[7]. Given Scheme 1, the Cumulative Distribution Function (CDF) of the baseband channel power gain can be derived as follows.

Lemma 1 (CDF of $Y_{t,k}$). *Denote the variances of* $f_{\mathrm{R}}(\phi_{q_{t,k}^\dagger}, \phi_{t,k,i,1}) \Lambda_{\mathrm{R}}(\phi_{t,k,i,1})$ *and* $f_{\mathrm{T}}(\theta_{p_{t,k}^\dagger}, \theta_{t,k,i,1}) \Lambda_{\mathrm{T}}(\theta_{t,k,i,1})$ *as* $\sigma_{\mathrm{R};t,k,i}^2$ *and* $\sigma_{\mathrm{T};t,k,i}^2$, *respectively, the expectations of* $|f_{\mathrm{R}}(\phi_{q_{t,k}^\dagger}, \phi_{t,k,i,1}) \Lambda_{\mathrm{R}}(\phi_{t,k,i,1})|$, $|f_{\mathrm{T}}(\theta_{p_{t,k}^\dagger}, \theta_{t,k,i,1}) \Lambda_{\mathrm{T}}(\theta_{t,k,i,1})|$ *and* $|f_{\mathrm{R}}(\phi_{q_{t,k}^\dagger}, \phi_{t,k,i,1}) f_{\mathrm{T}}(\theta_{t,k,i,1}, \theta_{p_{t,k}^\dagger}) \Lambda_{\mathrm{R}}(\phi_{t,k,i,1}) \Lambda_{\mathrm{T}}(\theta_{t,k,i,1})|$ *as* $\mu_{\mathrm{R};t,k,i}$, $\mu_{\mathrm{T};t,k,i}$ *and* $\mu_{t,k,i}$, *respectively. With Scheme 1 and sufficiently large*

$N_{k,i}^{\mathrm{ray}}(\forall i)$, *the CDF of $Y_{t,k}$ is given by*

$$F_{Y_{t,k}}(x) \triangleq \Pr\left[Y_{t,k} \le x\right] = 1 - \exp\left(-x \Big/ \sum_{i=1}^{N_k^{\mathrm{cl}}} N_{k,i}^{\mathrm{ray}} \sigma_{\rho;t,k,i,1}^2\right),$$

where $x > 0$ and

$$\sigma_{\rho;t,k,i,1}^2 = \frac{\pi \sigma_{\alpha;k,i}^2 + 4\sigma_{\alpha;k,i}}{4} (\sigma_{\mathrm{R};t,k,i}^2 \sigma_{\mathrm{T};t,k,i}^2 + \sigma_{\mathrm{R};t,k,i}^2 \mu_{\mathrm{T};t,k,i}^2 + \sigma_{\mathrm{T};t,k,i}^2 \mu_{\mathrm{R};t,k,i}^2) + \sigma_{\alpha;k,i} \mu_{t,k,i}^2. \tag{19}$$

Proof. Please refer to Appendix A. $\square$

D. System Queue Dynamics

There are K downlink transmission queues at the BS, each for one UE. The arrival data of each queue is organized by packets, each with B information bits. It is assumed that the number of arrival packets at the k-th UE in the t-th frame, denoted as $A_{t,k}$, follows an independent Poisson distribution with expectation λ_k, i.e.,

$$\Pr[A_{t,k} = n] = \frac{\lambda_k^n e^{-\lambda_k}}{n!}. \tag{20}$$

Let $\mathcal{A}_t \triangleq \{A_{t,k} | \forall k \in \mathcal{K}\}$ represent the aggregate packet arrivals in the t-th frame. Without loss of generality, it is assumed that all packets arrive at the end of each frame.

Supposing that the d_t-th UE is selected in the t-th frame, the number of departure packets from the queue for the selected UE in the t-th frame is given by

$$D_{t,d_t} = \left\lfloor \frac{W R_{t,d_t} T_{\mathrm{F}}}{B} \right\rfloor. \tag{21}$$

Denote $Q_{t,k}$ as the queue length (number of packets) of the k-th queue at the beginning of t-th frame and $Q_{\max}$ as the buffer size for each queue. The queue dynamics can be expressed as

$$Q_{t+1,k} = \min\left(Q_{t,k}^{\mathrm{D}} + A_{t,k}, Q_{\max}\right), \tag{22}$$

where the arrival packets will be discarded if the buffer is full (i.e., buffer overflow), and the post-decision queue length $Q_{t,k}^{\mathrm{D}}$ is defined as[8]

$$Q_{t,k}^{\mathrm{D}} = \begin{cases} (Q_{t,k} - D_{t,k})^+ & k = d_t, \\ Q_{t,k} & k \ne d_t. \end{cases} \tag{23}$$

The allocation of d_t and P_{t,d_t} in the t-th frame will affect the queuing dynamics of future frames. For instance, suppose one UE's LoS path would vanish in future frames due to limited FoV and UE rotation. The BS may raise power and schedule more transmission time for this UE, such that the BS can deliver as many packets as possible to this UE and the chance of buffer overflow could be suppressed. On the

---

[7]On one hand, if there is one scattering cluster, Scheme 1 may adjust the receiving beam to compensate for the rotation angle based on the knowledge of ω_k, which is similar to the sensor-assisted beam alignment in [9]. On the other hand, if there are multiple scattering clusters, Scheme 1 may steer both transmission and receiving beams to the best scattering cluster (with maximum average baseband channel power gain) based on the knowledge of $\mathcal{I}_t^{\mathrm{SCSI}}$ duration rotation.

[8]In the following section, we shall define Bellman's equations with respect to the post-decision queue length.

$$\Pr[\mathcal{S}_{t+1}|\mathcal{S}_t, \Omega_t] = \mathbb{E}_{\mathcal{A}_t} \Pr[\mathcal{Q}_{t+1}|\mathcal{Q}_t, \mathcal{Y}_t, \Omega_t, \mathcal{A}_t] \Pr[\mathcal{Y}_{t+1}] \tag{25}$$

$$= \mathbb{E}_{\mathcal{A}_t} \mathbb{I}\Big[Q_{t+1,d_t} = \min\big\{(Q_{t,d_t} - D_{t,d_t})^+ + A_{t,d_t}, Q_{\max}\big\}\Big|P_{t,d_t}, Y_{t,d_t}, A_{t,d_t}\Big]$$

$$\times \prod_{k \neq d_t} \mathbb{I}[Q_{t+1,k} = \min(Q_{t,k} + A_{t,k}, Q_{\max})|A_{t,k}] \prod_k \Pr[Y_{t+1,k}]. \tag{26}$$

---

other hand, other UEs may suffer from degradation of queuing performance. Hence, transmission scheduling of all the frames in one scheduling period should be optimized jointly with a global objective. Due to the randomness of channel and packet arrival, the joint scheduling design is a stochastic optimization problem, which is formulated as a finite-horizon MDP problem in the following section.

III. PROBLEM FORMULATION

In order to facilitate the MDP formulation, the system state, scheduling action and policy, state transition probability, and system cost are first defined as follows.

Definition 2 (System State). *At the beginning of the t-th frame, the system state is represented by $\mathcal{S}_t \triangleq (\mathcal{Q}_t, \mathcal{Y}_t)$, consisting of Queuing State Information (QSI) of all the UEs $\mathcal{Q}_t \triangleq \{Q_{t,k}|\forall k \in \mathcal{K}\}$, and baseband channel power gains to all the UEs $\mathcal{Y}_t \triangleq \{Y_{t,k}|\forall k \in \mathcal{K}\}$.*

Definition 3 (Scheduling Action and Policy). *At the beginning of the t-th frame, the scheduling actions include the selection of downlink transmission UE $d_t \in \mathcal{K}$ and the downlink transmission power $P_t \triangleq P_{t,d_t}$, where the following instantaneous power constraint at the BS should be satisfied,*

$$P_t \leq P_{\max}, \ \forall t \in \mathcal{T}. \tag{24}$$

Hence, the scheduling policy of the BS, denoted as Ω_t, is a mapping from the system state $\mathcal{S}_t$ to the scheduling actions, i.e., $\Omega_t(\mathcal{S}_t) = (d_t, P_t)$.

Given the scheduling policy of the t-th frame, the transition probabilities of the system state from the t-th frame ($\forall t \in \mathcal{T}$) to the $(t+1)$-th frame are given by (26).

In this paper, the scheduling policies $\{\Omega_t|\forall t \in \mathcal{T}\}$ are designed to minimize the average system queuing and power costs, including the transmission power, the average packet transmission delay and the penalty of the packet drop. According to Little's law, the average delay of packet transmission is proportional to the average number of packets in the system [32]. Since we consider the transmission scheduling in a finite scheduling period, the summation of queuing packet numbers of all the frames in the scheduling period can be used as an equivalent measurement of average transmission delay [33]–[35]. Hence, we define the following weighted sum of the transmission power consumption, the number of queuing packets and full buffer penalty as the system cost in the t-th frame ($\forall t \in \mathcal{T}$),

$$g_t(\mathcal{S}_t, d_t, P_t) \triangleq w_{\mathrm{P}} P_t + \sum_{k \in \mathcal{K}} (Q_{t,k} + w_{\mathrm{Q}} \mathbb{I}[Q_{t,k} = Q_{\max}]),$$

where w_{P} and w_{Q} are the weights of the power consumption and full-buffer penalty respectively, and $\mathbb{I}[Q_{t,k} = Q_{\max}]$ is 1

when the buffer is full and 0 otherwise. Note that packet drop will occur when packets arrive at full buffers.

The overall minimization objective of one scheduling period with the initial system state $\mathcal{S}_1$ is then given by

$$G(\mathcal{S}_1, \Omega) \triangleq \mathbb{E}_{\mathcal{A}, \mathcal{Y}}^{\Omega} \left[\sum_{t=1}^T g_t(\mathcal{S}_t, \Omega_t(\mathcal{S}_t)) + \varrho(\mathcal{Q}_{T+1}) \middle| \mathcal{S}_1 \right],$$

where $\mathcal{A} \triangleq \{\mathcal{A}_t|\forall t \in \mathcal{T}\}$, $\mathcal{Y} \triangleq \{\mathcal{Y}_t|\forall t \in \mathcal{T}\}$, $\Omega \triangleq \{\Omega_t|\forall t \in \mathcal{T}\}$, and $\varrho(\mathcal{Q}_{T+1}) \triangleq \sum_{k \in \mathcal{K}} Q_{T+1,k}$ accounts for the remaining packet number at the end of one scheduling period. The expectation is taken over the randomness of packet arrivals $\mathcal{A}$ and baseband channel power gains $\mathcal{Y}$, and the system state transition follows (26). Let $\Omega_t^\star$ and $\Omega^\star \triangleq \{\Omega_t^\star|\forall t \in \mathcal{T}\}$ be the optimal scheduling policy of the t-th frame and the optimal policy aggregation respectively, the transmission design can be formulated as the following dynamic programming problem.

P1 : $\Omega^\star = \underset{\Omega}{\arg\min} \ G(\mathcal{S}_1, \Omega) \tag{27}$

$$\text{s.t.} \quad 0 \leq P_t \leq P_{\max}, \ \forall d_t \in \mathcal{K}, \forall t \in \mathcal{T}$$

P1 is a finite-horizon MDP with T stages (frames). Due to the limited antenna FoV and UE rotation, the distributions of baseband channel power gain are non-stationary. Hence, the optimal policies of different frames could be different. This is different from the infinite-horizon MDP formulation in [35] or the MDP formulation for Stochastic Shortest Path (SSP) in [36], where the optimal policies of different control stages (e.g., frames) are homogeneous. As a result, instead of finding an optimal policy for all frames, we should optimize the scheduling policies for all the frames respectively.

Generally, P1 can be solved via Bellman's equations, which are different for different frames. In order to avoid complicated calculation of transition matrix and expectation of future cost in the phase of online scheduling [37], we introduce the following post-decision system state and adopt the Bellman's equations with *post-decision value functions* in the solution.

Definition 4 (Post-Decision State). *At the beginning of the t-th frame, the post-decision system state is defined by $\mathcal{S}_t^{\mathrm{D}} \triangleq (\mathcal{Q}_t^{\mathrm{D}}, \mathcal{Y}_t)$, consisting of post-decision QSI (defined in (23)) of all the UEs $\mathcal{Q}_t^{\mathrm{D}} \triangleq \{Q_{t,k}^{\mathrm{D}}|\forall k \in \mathcal{K}\}$ and baseband channel power gains to all the UEs $\mathcal{Y}_t$.*

Note that $\mathcal{Y}_t$ is independently distributed in different frames, it can be averaged out from the value functions as in [38]. Denote the post-decision value function of the optimal policy (optimal value function for short) as $W_t(\mathcal{Q}_t^{\mathrm{D}})$; the Bellman's equation with respect to the post-decision QSI only can be

$$\Pr[\mathcal{Q}^{\mathrm{D}}_{t+1}|\mathcal{Q}^{\mathrm{D}}_t, \Omega_{t+1}] = \mathbb{E}_{\mathcal{A}_t, \mathcal{Y}_{t+1}} \Pr\left[\mathcal{Q}^{\mathrm{D}}_{t+1}\big|\mathcal{Q}^{\mathrm{D}}_t, \mathcal{Y}_{t+1}, \Omega_{t+1}, \mathcal{A}_t\right] \tag{29}$$

$$= \mathbb{E}_{\mathcal{A}_t, \mathcal{Y}_{t+1}} \mathbb{I}\Big[Q^{\mathrm{D}}_{t+1,d_{t+1}} = \big\{ \min(Q^{\mathrm{D}}_{t,d_{t+1}} + A_{t,d_{t+1}}, Q_{\max}) - D_{t+1,d_{t+1}} \big\}^+ \Big| P_{t+1,d_{t+1}}, Y_{t+1,d_{t+1}}, A_{t,d_{t+1}}\Big]$$

$$\times \prod_{k \neq d_{t+1}} \mathbb{I}\left[Q^{\mathrm{D}}_{t+1,k} = \min\left(Q^{\mathrm{D}}_{t,k} + A_{t,k}, Q_{\max}\right)\big| A_{t,k}\right], \ \forall t \in \mathcal{T}. \tag{30}$$

written as

$$W_t(\mathcal{Q}^{\mathrm{D}}_t) = \min_{\Omega_{t+1}(\mathcal{S}_{t+1})} \mathbb{E}_{\mathcal{A}_t, \mathcal{Y}_{t+1}}[g_{t+1}(\mathcal{S}_{t+1}, \Omega_{t+1}(\mathcal{S}_{t+1}))$$
$$+ W_{t+1}(\mathcal{Q}^{\mathrm{D}}_{t+1})|\mathcal{Q}^{\mathrm{D}}_t], \ \forall t \in \mathcal{T}, \tag{28}$$

where $W_T(\mathcal{Q}^{\mathrm{D}}_T) \triangleq \mathbb{E}_{\mathcal{A}_T}\left[\varrho(\mathcal{Q}_{T+1})\big|\mathcal{Q}^{\mathrm{D}}_T\right]$ for notation convenience. Based on the system transition probabilities in (26), the transition probabilities of post-decision QSI in (28) are given by (30).

With the above optimal value function, the optimal scheduling policy of the t-th frame ($\forall t \in \mathcal{T}$) for P1 can be obtained by

$$\Omega^\star_t(\mathcal{S}_t) = \arg\min_{\Omega_t(\mathcal{S}_t)} \left\{ g_t\big(\mathcal{S}_t, \Omega_t(\mathcal{S}_t)\big) + W_t\left(\mathcal{Q}^{\mathrm{D}}_t\right) \right\} \tag{31}$$
$$\text{s.t.} \quad 0 \leq P_t \leq P_{\max}.$$

It can be observed from (31) that the optimal value function of the t-th frame $W_t(\mathcal{Q}^{\mathrm{D}}_t)$ should be calculated before the derivation of the optimal scheduling policy for the t-th frame $\Omega^\star_t(\mathcal{S}_t)$. However, because of the minimization in (28), it is difficult to derive the closed-form expression of the optimal value function in each frame $\{W_t(\mathcal{Q}^{\mathrm{D}}_t)|\forall t \in \mathcal{T}\}$. It is also difficult to calculate the optimal value functions for all possible system states numerically due to the huge system state space, as there are $(Q_{\max} + 1)^K$ possible queuing states for K UEs. Conventional low-complexity solutions addressing such curse of dimensionality rely on the approximation of value function via neural networks [39], linear structure [40], etc. However, these general methods may not fully exploit the analytical communication models established in Section II, thus posing significant challenges on performance analysis. In the following section, a novel low-complexity solution is proposed with a non-trivial analytical performance bound.

IV. Low-Complexity Scheduling

The low-complexity solution framework is illustrated in Fig. 2. We first adopt the backpressure algorithm [41] as the base policy, denoted as Π, and derive the closed-form expressions of its value functions at all frames in Section IV-B, denoted as $\{W^{\Pi}_t(\mathcal{Q}^{\mathrm{D}}_t)|\forall t, \mathcal{Q}^{\mathrm{D}}_t\}$. Then in Section IV-C, the derived value functions are used to approximate the optimal value functions, i.e., $W_t(\mathcal{Q}^{\mathrm{D}}_t) \approx W^{\Pi}_t(\mathcal{Q}^{\mathrm{D}}_t)$, and calculate the scheduling policies by solving the Right-Hand Side (RHS) of (31) in each frame (i.e., policy iteration). Finally, an analytical performance lower-bound is introduced for the obtained new policies.

A. Base Policy

In our proposed solution framework, the base policy provides an approximation of average future cost in the optimization of current scheduling actions in (31). Given the system QSI at the beginning of the scheduling period $\mathcal{Q}_1$, the base policy uses the average downlink throughput with constant power P_Π and packet arrival rate to approximate the queue dynamics and adopt the *backpressure algorithm* [41], which has been widely used in delay-aware scheduling to select the downlink UE. Particularly, the predicted average spectral efficiency of the k-th UE in the t-th frame ($\forall t \in \mathcal{T}, \forall k \in \mathcal{K}$) is given by

$$R^{\Pi}_{t,k}(P_\Pi) \triangleq \int \log_2\left(1 + \frac{P_\Pi x}{N_0 W}\right) \frac{\mathrm{d}F_{Y_{t,k}}(x)}{\mathrm{d}x}\mathrm{d}x, \tag{32}$$

where $F_{Y_{t,k}}(x)$ is the CDF of the baseband channel power gain of the k-th UE in the t-th frame. Then in the base policy, the index of the selected UE is given by

$$d^{\Pi}_t = \arg\max_{k \in \mathcal{K}} \left[R^{\Pi}_{t,k}(P_\Pi) \times Q^{\Pi}_{t,k}\right], \ \forall t \in \mathcal{T}. \tag{33}$$

Moreover, the QSI in the $(t+1)$-th frame is approximated as

$$Q^{\Pi}_{t+1,k} = \begin{cases} \min\left\{(Q^{\Pi}_{t,k} - D^{\Pi}_{t,k})^+ + \lambda_k, Q_{\max}\right\} & k = d^{\Pi}_t, \\ \min(Q^{\Pi}_{t,k} + \lambda_k, Q_{\max}) & k \neq d^{\Pi}_t, \end{cases} \tag{34}$$

where $Q^{\Pi}_{1,k} = Q_{1,k}$, $D^{\Pi}_{t,k} \triangleq \left\lfloor \frac{W R^{\Pi}_{t,k}(P_\Pi) N_{\mathrm{F}}}{B} \right\rfloor, \forall t \in \mathcal{T}, \forall k \in \mathcal{K}^9$. Applying (33) and (34) iteratively, the UE selections of the base policy, denoted as $\{d^{\Pi}_1, d^{\Pi}_2, \ldots, d^{\Pi}_T\}$, can be determined. Hence, the base policy can be summarized as follows.

Policy 1 (Base Policy Π). *The transmission power to each selected UE is fixed to $P_t = P_\Pi$, $\forall t$. Moreover, the UE selection is determined by applying* (33) *and* (34) *iteratively, which is denoted as $\{d^{\Pi}_1, d^{\Pi}_2, \ldots, d^{\Pi}_T\}$.*

B. Value Function of the Base Policy

The value function of the base policy Π, which measures the average system cost from the $(t+1)$-th frame with the base policy, can be written as

$$W^{\Pi}_t(\mathcal{Q}^{\mathrm{D}}_t) = (T - t)w_{\mathrm{P}} P_\Pi + \sum_{k \in \mathcal{K}} W^{\Pi}_{t,k}(Q^{\mathrm{D}}_{t,k}), \tag{35}$$

where the *local value function* of the base policy $W^{\Pi}_{t,k}(Q^{\mathrm{D}}_{t,k})$ is the average queuing cost raised by the k-th UE since the $(t+1)$-th frame given the base policy Π and post-decision QSI $\mathcal{Q}^{\mathrm{D}}_t$. Thus,

$$W^{\Pi}_{t,k}(Q^{\mathrm{D}}_{t,k}) \triangleq \mathbb{E}^{\Pi}_{\mathcal{A}, \mathcal{Y}}\Big[\sum_{\tau = t+1}^{T} \big(Q_{\tau,k} + w_{\mathrm{Q}}\mathbb{I}[Q_{\tau,k} = Q_{\max}]\big)$$
$$+ Q_{T+1,k}\Big|Q^{\mathrm{D}}_{t,k}\Big]. \tag{36}$$

Notice that given the base policy and initial system state of the 1-st frame, the QSI for each UE evolves as a Markov chain. Thus, the QSI distribution of each downlink transmission

---

[9] λ_k in (34) can be learned in an online manner as in [42].

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2024.3400377
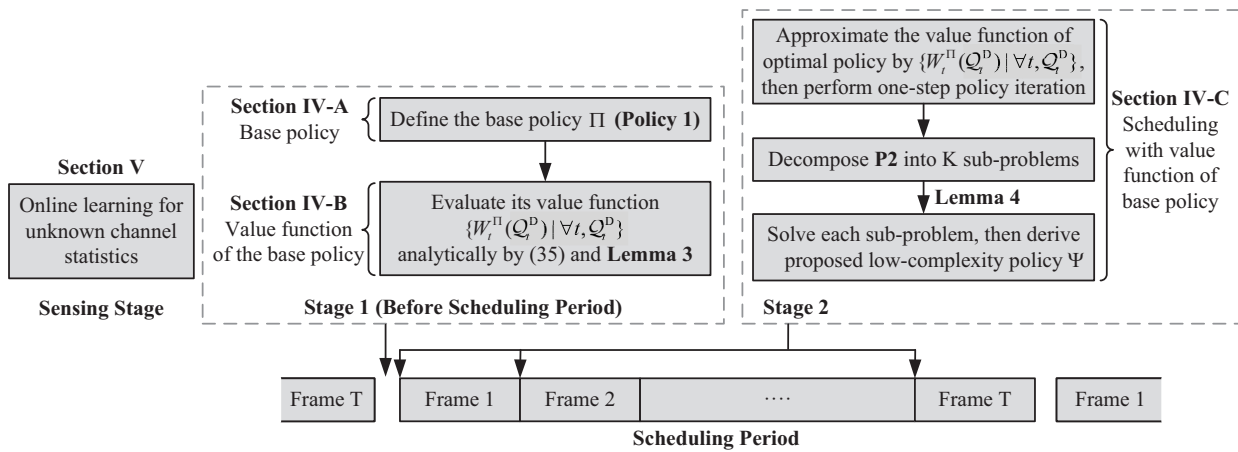
8

Fig. 2. Proposed low-complexity solution framework.

queue, as well as the average queuing cost for each UE $W_{t,k}^{\Pi}(Q_{t,k}^{D})$, can be derived with QSI transition probabilities. The QSI transition depends on the statistics of both packet arrival and departure. Since the former is known, we first introduce the following conclusion on the distribution of the departure packet number under the base policy.

Lemma 2. *Let $D_{t,k}^{\Pi}$ be the number of packets delivered to the k-th UE in the t-th frame under the base policy (if the k-th UE is selected). According to the beam selection scheme (Scheme 1), given $\{\omega_k|\forall k \in \mathcal{K}\}$ and $\mathcal{I}_1^{SCSI}$, the Probability Mass Function (PMF) of $D_{t,k}^{\Pi}$ is then given by*

$$\Pr\left[D_{t,k}^{\Pi}=n\right]=F_{Y_{t,k}}\left[\left(2^{\frac{(n+1)B}{WT_F}}-1\right)\frac{N_0W}{P_{\Pi}}\right]$$
$$-F_{Y_{t,k}}\left[\left(2^{\frac{nB}{WT_F}}-1\right)\frac{N_0W}{P_{\Pi}}\right], \quad (37)$$

where $F_{Y_{t,k}}$ is provided in Lemma 1.

Proof. According to the definition of $D_{t,k}$ in (21), $\Pr[D_{t,k}^{\Pi}=n]=\Pr[nB \leq WR_{t,k}T_F < (n+1)B|\Pi]$. Substituting $R_{t,k}$ with its definition in (14), the proof is straightforward. $\square$

Hence, the local value functions $W_{t,k}^{\Pi}(Q_{t,k}^{D})(\forall t, k, Q_{t,k}^{D})$ are derived in the following lemma.

Lemma 3 (Value Function of Base Policy). *Let $\mathbf{c}^{(1)}, \mathbf{c}^{(2)} \in \mathbb{R}^{(Q_{max}+1)\times 1}$ where $[\mathbf{c}^{(1)}]_i \triangleq (i-1) + w_Q\mathbb{I}[i-1 = Q_{max}]$ and $[\mathbf{c}^{(2)}]_i \triangleq i-1$. The entries of matrices $\mathbf{P}_k, \mathbf{M}_{t,k} \in \mathbb{R}^{(Q_{max}+1)\times(Q_{max}+1)}$ are specified in Table II and Table III, respectively, where the expression of $D_{t,k}^{\Pi}$ in Table III is given by Lemma 2. Then $W_{t,k}^{\Pi}(Q_{t,k}^{D})$ can be represented as*

$$W_{t,k}^{\Pi}(Q_{t,k}^{D}) = \mathbf{1}_{Q_{t,k}^{D}+1}^{T}\mathbf{v}_{t,k}, \ \forall t, k, Q_{t,k}^{D}, \quad (38)$$

where

$$\mathbf{v}_{t,k} \triangleq \sum_{\tau=t+1}^{T}\mathbf{P}_k\mathbf{X}_k(t,\tau)\mathbf{c}^{(1)} + \mathbf{P}_k\mathbf{X}_k(t,T+1)\mathbf{c}^{(2)}, \quad (39)$$

$$\mathbf{X}_k(t,\tau) \triangleq \prod_{n=t+1}^{\tau-1}\mathbf{M}_{n,k}^{\mathbb{I}(d_n^{\Pi}=k)}\mathbf{P}_k^{\mathbb{I}(d_n^{\Pi}\neq k)}. \quad (40)$$

TABLE II. Non-zero entries of matrix $\mathbf{P}_k$

i	j	$[\mathbf{P}_k]_{i,j}$
$1, ..., Q_{max}$	$i, ..., Q_{max}$	$\Pr[A_{t,k} = j-i]$
$Q_{max}+1$	$Q_{max}+1$	1
$1, ..., Q_{max}$	$Q_{max}+1$	$\Pr[A_{t,k} \geq Q_{max}+1-i]$

TABLE III. Non-zero entries of matrix $\mathbf{M}_{t,k}$

i	j	$[\mathbf{M}_{t,k}]_{i,j}$
$1, ..., Q_{max}+1$	$Q_{max}+1$	$\Pr[A_{t,k}-\min(D_{t,k}^{\Pi},i-1) \geq Q_{max}+1-i]$
$1, ..., Q_{max}+1$	1	$\Pr[D_{t,k}^{\Pi} \geq i-1]\Pr[A_{t,k}=0]$
$1, ..., Q_{max}+1$	$2, ..., Q_{max}$	$\Pr[A_{t,k}-\min(D_{t,k}^{\Pi},i-1) = j-i]$

Proof. Please refer to Appendix B. $\square$

As a result, the value functions of the base policy for all the frames $\{W_t^{\Pi}(\mathcal{Q}_t^{D})|\forall t, \mathcal{Q}_t^{D}\}$ can be derived by (35) at the beginning of scheduling period. The procedure is summarized in Algorithm 1.

C. Scheduling with Approximate Value Function

In this part, we use the value functions of the base policy $\{W_t^{\Pi}(\mathcal{Q}_t^{D})|\forall t, \mathcal{Q}_t^{D}\}$ to approximate the optimal value func-

---

Algorithm 1: Evaluation of local value functions of base policy

Input :
$\mathcal{Q}_1$: QSI in the 1-st frame
$\{\mathcal{I}_t^{SCSI}|\forall t \in \mathcal{T}\}$: SCSI in each frame
Output:
$\{W_{t,k}^{\Pi}(Q_{t,k}^{D})|\forall t, k, Q_{t,k}^{D}\}$: Local value functions of the base policy

1 The BS calculates $\{d_1^{\Pi}, d_2^{\Pi}, \ldots, d_T^{\Pi}\}$ by applying (33) and (34) iteratively and broadcasts the results to UEs

2 **for** $k = 1$ *to* K **do in parallel**

3   The k-th UE calculates $\{W_{t,k}^{\Pi}(Q_{t,k}^{D})|\forall t, Q_{t,k}^{D}\}$ locally based on SCSI $\{\mathcal{I}_t^{SCSI}\}$ via Lemma 3

$$\textbf{P2 [One-Step Policy Iteration]: } \Psi_t(\mathcal{S}_t) \triangleq \left(d_t^\Psi, P_t^\Psi\right) = \arg\min_{d_t \in \mathcal{K}, P_t} \left\{ g_t(\mathcal{S}_t, d_t, P_t) + W_t^\Pi\left(\mathcal{Q}_t^{\mathrm{D}}(\mathcal{S}_t, \Omega_t)\right) \right\} \tag{41}$$
$$\text{s.t.} \quad 0 \le P_t \le P_{\max}.$$

$$\textbf{P2(k) : } P_{t,k}^\psi = \arg\min_{P_{t,k}} \left\{ w_{\mathrm{P}} P_{t,k} + W_{t,k}^\Pi\left(Q_{t,k}^{\mathrm{D}}\left(Q_{t,k}, Y_{t,k}, P_{t,k}\right)\right) \right\}, \tag{42}$$
$$\text{s.t.} \quad 0 \le P_{t,k} \le P_{\max}.$$

---

tions $\{W_t(\mathcal{Q}_t^{\mathrm{D}})|\forall t, \mathcal{Q}_t^{\mathrm{D}}\}$, and derive the proposed scheduling policies $\{\Psi_t(\mathcal{S}_t)|\forall t \in \mathcal{T}\}$. Because the expressions are provided, conventional value iteration to evaluate the value functions can be avoided, which significantly reduces the computation complexity. The proposed scheduling policy in the t-th ($\forall t$) frame, denoted by $\Psi_t(\mathcal{S}_t)$, can be obtained by solving P2.

Since both $g_t(\mathcal{S}_t, \Omega_t(\mathcal{S}_t))$ and $W_t^\Pi(\mathcal{Q}_t^{\mathrm{D}}(\mathcal{S}_t, \Omega_t))$ in (41) can be decoupled to each UE, the power allocation in P2 can be decomposed into K sub-problems each for a UE assuming it is selected. The k-th sub-problem ($\forall k \in \mathcal{K}$) is given by P2(k). The optimal power allocation for the k-th UE in P2(k) is provided in the following lemma. Let $G_{t,k}^\psi$ be the minimized objective of P2(k), the user selection and corresponding power allocation can be obtained via

$$\begin{cases} d_t^\Psi = \arg\min_{k \in \mathcal{K}} G_{t,k}^\psi, & (43) \\ P_t^\Psi = P_{t,d_t^\Psi}^\psi. & (44) \end{cases}$$

Lemma 4 (Local Power Optimization). *The optimal transmission power in* P2(k) *is given by*

$$P_{t,k}^\psi = \arg\min_{P_{t,k} \in \mathscr{P}_{t,k}} \left\{ w_{\mathrm{P}} P_{t,k} + \Delta \mathbf{z}_{t,k}^{\mathsf{T}}(P_{t,k}) \mathbf{v}_{t,k} \right\}, \tag{45}$$

where $\mathscr{P}_{t,k} \triangleq \left\{ 0, \frac{2^{\frac{B}{W N_{\mathrm{F}}}}-1}{Y_{t,k}}, \frac{2^{\frac{2B}{W N_{\mathrm{F}}}}-1}{Y_{t,k}}, \dots, \min\left(\frac{2^{\frac{Q_{\max}B}{W N_{\mathrm{F}}}}-1}{Y_{t,k}}, P_{\max}\right) \right\}$ *is the feasible power set,* $\Delta \mathbf{z}_{t,k}(P_{t,k}) = \mathbf{1}_{Q_{t,k}^{\mathrm{D}}(P_{t,k})+1} - \mathbf{1}_{Q_{t,k}+1}$, *and* $Q_{t,k}^{\mathrm{D}}(P_{t,k}) = \left(Q_{t,k} - \left\lfloor \frac{W R_{t,k}(P_{t,k}) T_{\mathrm{F}}}{B} \right\rfloor\right)^+$.

Proof. Please refer to Appendix C. □

As a summary, the proposed solution framework can be implemented in a semi-distributed manner, which is elaborated in Algorithm 2. Because the decoupled value functions of each UE, i.e., $\{W_{t,k}^\Pi(Q_{t,k}^{\mathrm{D}})|\forall t, Q_{t,k}^{\mathrm{D}}\}$, are evaluated at the beginning of a scheduling period and stored locally for future calculation in each frame, both the computation complexity and signaling overhead are significantly reduced.

D. Performance Bound and Complexity Analysis

Furthermore, we have the following bounds on the value function of the proposed scheduling policy.

Lemma 5 (Performance Bound of Ψ). *Let*

$$W_t^\Psi\left(\mathcal{Q}_t^{\mathrm{D}}\right) \triangleq \mathbb{E}_{\mathcal{A}, \mathcal{Y}}^\Psi \left[\sum_{\tau=t+1}^T g_\tau\left(\mathcal{S}_\tau, \Psi_\tau(\mathcal{S}_\tau)\right) + \varrho(\mathcal{Q}_{T+1}) \middle| \mathcal{Q}_t^{\mathrm{D}} \right],$$

($\forall t \in \mathcal{T}$) be the value function of the proposed policies $\Psi \triangleq \{\Psi_\tau | \forall \tau = t+1, \dots, T\}$, *then*

$$W_t\left(\mathcal{Q}_t^{\mathrm{D}}\right) \le W_t^\Psi\left(\mathcal{Q}_t^{\mathrm{D}}\right) \le W_t^\Pi\left(\mathcal{Q}_t^{\mathrm{D}}\right), \ \forall t \in \mathcal{T}. \tag{46}$$

Algorithm 2: The proposed solution framework

Input :
$\mathcal{Q}_1$: QSI in the 1-st frame
$\{\mathcal{I}_t^{\mathrm{SCSI}}|\forall t \in \mathcal{T}\}$: SCSI in each frame
Output:
Ψ_t: Proposed scheduling actions $\{\Psi_t(\mathcal{S}_t)|\forall t \in \mathcal{T}\}$

1 **for** *each scheduling period* **do**
2   Conduct Algorithm 1
3   **for** $t = 1$ *to* T **do**
4   **for** $k = 1$ *to* K **do in parallel**
5   Given the analog precoder and combiner, the k-th UE obtain the baseband channel power gain $Y_{t,k}$ via channel estimation
6   The k-th UE calculates the optimal power $P_{t,k}^\psi$ via (45), and reports $G_{t,k}^\psi$ and $P_{t,k}^\psi$ to the BS via uplink signaling channel
7   *The BS determines* Ψ_t *via* (43) *and* (44)*, and starts downlink transmission of the t-th frame*

Proof. Because Ψ is the suboptimal scheduling policy based on the value function of the base policy, $W_t(\mathcal{Q}_t^{\mathrm{D}}) \le W_t^\Psi(\mathcal{Q}_t^{\mathrm{D}})$ holds. Moreover, due to the *policy improvement property* [36], $W_t^\Psi(\mathcal{Q}_t^{\mathrm{D}}) \le W_t^\Pi(\mathcal{Q}_t^{\mathrm{D}})$ holds. □

The overall computation complexity of Algorithm 2 consists of Algorithm 1 and per-frame scheduling. In Algorithm 1, the overall time complexity is mainly contributed by line 3, which is $O(T^2 Q_{\max}^3)$ for each UE in the parallel calculation. The time complexity for online scheduling in each frame (line 4–7 of Algorithm 2) is mainly contributed by line 7, which is $O(K)$. As a comparison, the time complexity of conventional value iteration for optimal scheduling policy is $O(TKQ_{\max}^{4K+1})$, which grows exponentially with respect to the number of UEs.

V. ONLINE LEARNING FOR UNKNOWN CHANNEL STATISTICS

In the previous section, the value function of the base policy $W_t^\Pi(\mathcal{Q}_t^{\mathrm{D}})$ is derived analytically by taking the SCSI $\mathcal{I}_t^{\mathrm{SCSI}}$ as apriori knowledge. However, the SCSI may not be available in practice at the very beginning. Therefore, online learning algorithms tracking the SCSI in the 1-st frame become necessary. Note that the SCSI of the following frames can be derived from that of the 1-st frame via (10). Since the scattering clusters are quasi-static, it is assumed that the number of clusters N_k^{cl}, and the cluster mean AoD and AoA of each cluster, i.e., $\bar{\theta}_{k,i}$ and $\bar{\phi}_{1,k,i}$, have already been detected

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2024.3400377
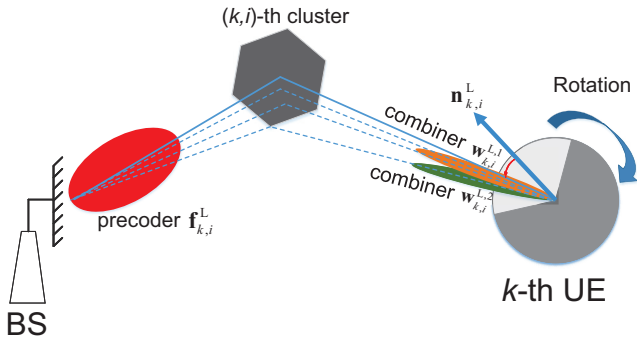
10

Fig. 3. Online learning for unknown channel statistics.

via the approaches in the existing literature [43], [44]. For elaboration convenience, it is also assumed that the AoA distribution supports of clusters do not overlap. Hence, the statistical learning algorithms for the ray number $N_{k,i}^{\mathrm{ray}}$, the AoD standard deviation $\sigma_{\theta;k,i}$, AoA standard deviation $\sigma_{\phi;k,i}$ and standard deviation of the complex gain $\sigma_{\alpha;k,i}$ of the (k,i)-th cluster ($\forall k, i$) are elaborated in this section.

Without loss of generality, we elaborate the learning algorithm for the AoA standard deviation, while the AoD standard deviation can be tracked similarly by the uplink according to the channel reciprocity. As illustrated in Fig. 3, online learning of the AoA standard deviation $\sigma_{\phi;k,i}$ can be conducted when the k-th UE's phased array is at a particular direction $\mathbf{n}_{k,i}^{\mathrm{L}}$ (or equivalently, the mean AoA of the (k,i)-th cluster is at the direction $\bar{\phi}_{k,i}^{\mathrm{L}}$). This is feasible because the UE's rotation and orientation can be detected.

Particularly, a wide beam is used as $\mathbf{f}_{k,i}^{\mathrm{L}} \triangleq \mathbf{a}_{\mathrm{T}}\big(\theta_{p_{k,i}^{\mathrm{L}}}\big) \in \mathcal{F}$ such that the (k,i)-th cluster is within the beam coverage. The BS transmits pilot signal s_k with power P_{L} and probing precoder $\mathbf{f}_{k,i}^{\mathrm{L}}$ twice, and the k-th UE applies the two distinct probing combiners $\mathbf{w}_{k,i}^{\mathrm{L},1} \triangleq \mathbf{a}_{\mathrm{R}}\big(\phi_{q_{k,i}^{\mathrm{L},1}}\big) \in \mathcal{W}$ and $\mathbf{w}_{k,i}^{\mathrm{L},2} \triangleq \mathbf{a}_{\mathrm{R}}\big(\phi_{q_{k,i}^{\mathrm{L},2}}\big) \in \mathcal{W}$, $(q_{k,i}^{\mathrm{L},1} \neq q_{k,i}^{\mathrm{L},2})$, for pilot receiving, respectively. The above baseband channel estimation is repeated for Ξ times when the boresight of the k-th UE's phased array is at the direction $\mathbf{n}_{k,i}^{\mathrm{L}}$[10]. Denote $y_{k,i}^{\xi,j}$ as the received signal with probing combiner $\mathbf{w}_{k,i}^{\mathrm{L},j}$ ($j=1,2$) in the ξ-th channel estimation, then

$$y_{k,i}^{\xi,j} = \sqrt{P_{\mathrm{L}}}[\mathbf{w}_{k,i}^{\mathrm{L},j}]^{\mathsf{H}}\mathbf{H}_{k,i}^{\xi}\mathbf{f}_{k,i}^{\mathrm{L}}s_k + [\mathbf{w}_{k,i}^{\mathrm{L},j}]^{\mathsf{H}}\mathbf{z}^{\xi}, \ j=1,2 \quad (47)$$

where $\mathbf{H}_{k,i}^{\xi}$ and $\mathbf{z}^{\xi}$ denote the channel matrix and the noise vector with i.i.d. $\mathcal{CN}(0, \sigma_z^2)$ noise, respectively. Denote $\eta_{k,i}^{\xi,j} \triangleq |y_{k,i}^{\xi,j}|^2$ as the received signal power. After Ξ channel estimations, the average received signal power can be written as

$$\frac{1}{\Xi}\sum_{\xi=1}^{\Xi}\eta_{k,i}^{\xi,j} \xrightarrow{\Xi\to\infty} \mathbb{E}\big[\eta_{k,i}^{\xi,j}\big]$$

$$= P_{\mathrm{L}}\mathbb{E}_{\mathbf{H}_{k,i}^{\xi}}\left[\left|[\mathbf{w}_{k,i}^{\mathrm{L},j}]^{\mathsf{H}}\mathbf{H}_{k,i}^{\xi}\mathbf{f}_{k,i}^{\mathrm{L}}\right|^2\right] + \sigma_z^2 \quad (48)$$

$$\approx P_{\mathrm{L}}\mathbb{E}_{\mathbf{H}_{k,i}^{\xi}}\left[\left|[\mathbf{w}_{k,i}^{\mathrm{L},j}]^{\mathsf{H}}\mathbf{H}_{k,i}^{\xi}\mathbf{f}_{k,i}^{\mathrm{L}}\right|^2\right], j=1,2, \quad (49)$$

[10]Note that as Ξ channel estimations are not in the same scheduling period due to the constraint on the particular boresight direction $\mathbf{n}_{k,i}^{\mathrm{L}}$, the online learning algorithm may last for a number of scheduling periods.

where the approximation is for the high SNR regime. Hence, we have the following lemma.

Lemma 6 (Ratio of Average Received Signal Power)**.** *For sufficiently large* N_{R}, N_{T} *and* Ξ,

$$\frac{\sum_{\xi=1}^{\Xi}\eta_{k,i}^{\xi,1}}{\sum_{\xi=1}^{\Xi}\eta_{k,i}^{\xi,2}} \longrightarrow \frac{\int_{\mathcal{P}_{\mathrm{R};q_{k,i}^{\mathrm{L},1}}} f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\mathrm{L}})\mathrm{d}\phi}{\int_{\mathcal{P}_{\mathrm{R};q_{k,i}^{\mathrm{L},2}}} f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\mathrm{L}})\mathrm{d}\phi}$$

$$= \frac{g_+(\phi_{q_{k,i}^{\mathrm{L},1}}, \sigma_{\phi;k,i}) - g_-(\phi_{q_{k,i}^{\mathrm{L},1}}, \sigma_{\phi;k,i})}{g_+(\phi_{q_{k,i}^{\mathrm{L},2}}, \sigma_{\phi;k,i}) - g_-(\phi_{q_{k,i}^{\mathrm{L},2}}, \sigma_{\phi;k,i})}, \quad (50)$$

where $\mathcal{P}_{\mathrm{R};q} \triangleq \left\{\phi \,\big|\, |\sin\phi - \sin\phi_q| \leq \frac{1}{N_{\mathrm{R}}}\right\}$, $\mathcal{P}_{\mathrm{T};p} \triangleq \left\{\theta \,\big|\, |\sin\theta - \sin\theta_p| \leq \frac{1}{N_{\mathrm{T}}}\right\}$, $g_+(\phi_{q_{k,i}^{\mathrm{L},j}}, \sigma_{\phi;k,i})$ *and* $g_-(\phi_{q_{k,i}^{\mathrm{L},j}}, \sigma_{\phi;k,i})$ *are given by* (51) *and* (52), *respectively.*

Proof. Please refer to Appendix D. $\qquad\square$

Then the estimation of $\sigma_{\phi;k,i}$ after Ξ channel estimations, denoted as $\hat{\sigma}_{\phi;k,i}^{\Xi}$, is given by

$$\hat{\sigma}_{\phi;k,i}^{\Xi}$$
$$= \arg\min_{x\in\mathbb{R}^+}\left|\frac{\sum_{\xi=1}^{\Xi}\eta_{k,i}^{\xi,1}}{\sum_{\xi=1}^{\Xi}\eta_{k,i}^{\xi,1}} - \frac{\int_{\mathcal{P}_{\mathrm{R};q_{k,i}^{\mathrm{L},1}}} f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\mathrm{L}})\mathrm{d}\phi}{\int_{\mathcal{P}_{\mathrm{R};q_{k,i}^{\mathrm{L},2}}} f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\mathrm{L}})\mathrm{d}\phi}\right|.$$

It is clear that $\hat{\sigma}_{\phi;k,i}^{\Xi} \xrightarrow{\Xi\to\infty} \sigma_{\phi;k,i}$. Based on the observations $\{\eta_{k,i}^{\xi,j}|\forall \xi = 1, 2, \ldots, \Xi, \forall j = 1, 2\}$, $N_{k,i}^{\mathrm{ray}}$ and $\sigma_{\alpha;k,i}^2$ can also be estimated in the following lemma.

Lemma 7. *Denote* $\zeta_{k,i} = N_{k,i}^{\mathrm{ray}}\sigma_{\alpha;k,i}^2$ *as the average channel power gain of the* (k,i)-th *cluster. Let* (53), $\hat{N}_{k,i}^{\Xi} = \ln\left(\frac{1}{\Xi}\sum_{\xi=1}^{\Xi}\mathbb{I}\big[\eta_{k,i}^{\xi,1} \leq \tau_{k,i}^{\mathrm{L}}\big]\right)\big/\ln(\nu)$ *and* $\hat{\sigma}_{\alpha;k,i}^{\Xi} = \sqrt{\hat{\zeta}_{k,i}^{\Xi}/\hat{N}_{k,i}^{\Xi}}$ *be the estimation of* $\zeta_{k,i}$, $N_{k,i}^{\mathrm{ray}}$ *and* $\sigma_{\alpha;k,i}$ *after* Ξ *baseband channel estimations, where* $\tau_{k,i}^{\mathrm{L}}$ *is a threshold satisfying* $\sigma_z^2 \ll \tau_{k,i}^{\mathrm{L}} \ll P_{\mathrm{L}}\sigma_{\alpha;k,i}^2$, *and* ν *is given by*

$$\nu = 1 - \int_{\mathcal{P}_{\mathrm{R};q_{k,i}^{\mathrm{L},j}}} f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\mathrm{L}})\mathrm{d}\phi$$
$$\times \int_{\mathcal{P}_{\mathrm{T};p_{k,i}^{\mathrm{L}}}} f_{\theta;k,i}(\theta; \sigma_{\theta;k,i}, \bar{\theta}_{k,i}^{\mathrm{L}})\mathrm{d}\theta. \quad (54)$$

Then we have $\hat{\zeta}_{k,i}^{\Xi} \xrightarrow{\Xi\to\infty} \zeta_{k,i}$, $\hat{N}_{k,i}^{\Xi} \xrightarrow{\Xi\to\infty} N_{k,i}^{\mathrm{ray}}$, *and* $\hat{\sigma}_{\alpha;k,i}^{\Xi} \xrightarrow{\Xi\to\infty} \sigma_{\alpha;k,i}$.

Proof. Please refer to Appendix D. $\qquad\square$

VI. Experiments and Simulations
A. Sensing-Based Beam Alignment

In this part, the feasibility and performance of the SCSI-based beam alignment scheme (Scheme 1) are demonstrated. The platform and environment of the experiment are shown in Fig. 4. There is one transmitter and one receiver, working with 60.48 GHz carrier frequency. Both the transmitter and the receiver are developed with one NI USRP-X310s [45] and one Sivers IMA development kit [46]. The former is used to generate and capture baseband signals. The latter is equipped with a 16-antenna phased array, serving as the RF front-end.

$$g_+\big(\phi_{q_{k,i}^{\mathrm{L},j}}, \sigma_{\phi;k,i}\big) = \exp\Big(\sqrt{2}\big(\arcsin\big(\sin\phi_{q_{k,i}^{\mathrm{L},j}} + \frac{1}{N_{\mathrm{R}}}\big) - \bar{\phi}_{k,i}^{\mathrm{L}}\big)/\sigma_{\phi;k,i}\Big), \ j = 1, 2. \tag{51}$$

$$g_-\big(\phi_{q_{k,i}^{\mathrm{L},j}}, \sigma_{\phi;k,i}\big) = \exp\Big(\sqrt{2}\big(\arcsin\big(\sin\phi_{q_{k,i}^{\mathrm{L},j}} - \frac{1}{N_{\mathrm{R}}}\big) - \bar{\phi}_{k,i}^{\mathrm{L}}\big)/\sigma_{\phi;k,i}\Big), \ j = 1, 2. \tag{52}$$

$$\hat{\zeta}_{k,i}^{\Xi} = \frac{\frac{1}{\Xi}\sum_{\xi=1}^{\Xi}\eta_{k,i}^{\xi,1}}{\int_{\mathcal{P}_{\mathrm{R};q_{k,i}^{\mathrm{L},j}}} f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\mathrm{L}})\mathrm{d}\phi \int_{\mathcal{P}_{\mathrm{T};p_{k,i}^{\mathrm{L}}}} f_{\theta;k,i}(\theta; \sigma_{\theta;k,i}, \bar{\theta}_{k,i}^{\mathrm{L}})\mathrm{d}\theta} \tag{53}$$
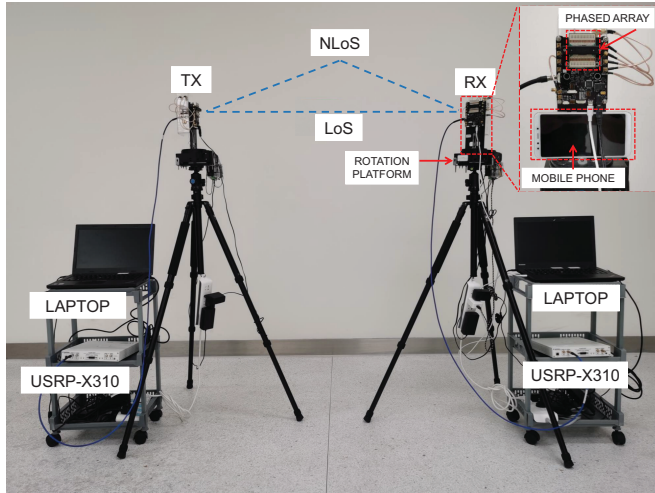
Fig. 4. Front view of the experimental platform and environment. In real measurements, the transmitter and receiver are placed 10 meters away.

(a) Initial antenna orientations.

(b) Final antenna orientations after UE rotating 140° clockwise.

Fig. 5. Experimental deployment at the beginning and the end of the rotation.

Two laptops are used to control the beam patterns of the phased arrays at the transmitter and the receiver respectively. There is one quasi-omnidirectional pattern and 63 directional patterns in the beam codebook of the phased array. Main lobe directions of the directional beam patterns range from $-45°$ to $45°$, leading to a limited FoV.

The receiving phased array is mounted on a rotating platform, which is to emulate the rotation of the UE on the horizontal plane. A mobile phone is tied to the receiving phased array, which can periodically measure the rotation angles by the embedded magnetometer (or gyroscope) and report its reads to the laptops. As a result, the laptops can adapt the transmission and receiving beams according to the rotation angle.

The experiment scenario, when the receiving phased array is rotating clockwise, is illustrated in Fig. 5. In our experimental environment, there is one LoS path and one Non-Line-of-Sight (NLoS) path (via a wall) between the transmitter and the receiver. Thus, the number of clusters is $N_k^{\mathrm{cl}} = 2$ in (3). As shown in our previous experiment [27], the specular reflection is dominant in the mmWave signals' scattering off a flat wall. Hence, for the NLoS path via the wall, the angles of incidence and reflection are identical as illustrated in Fig. 5.

Before the rotation, the channel statistics, including the AoAs, AoDs and average channel power gains of the LoS and NLoS paths, have been estimated via the spatial smoothing MUltiple SIgnal Classification (MUSIC) algorithm as in [27]. Due to the limited antenna FoV, only the LoS path is available at the beginning stage as in Fig. 5 (a). Particularly, when the
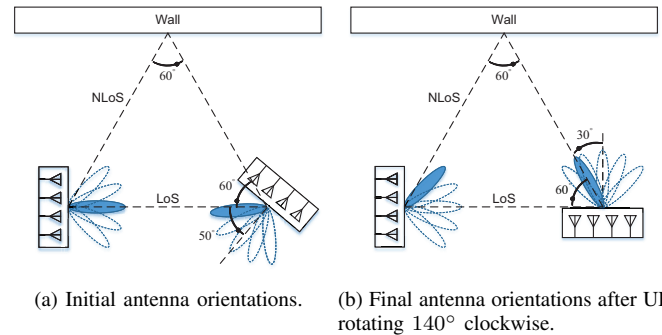
rotation angle is less than around $55°$, only the LoS path is within the FoV; when the rotation angle is within $[55°, 120°]$, both the LoS and NLoS paths are within the FoV; and finally, when the rotation angle is greater than $120°$, only the NLoS path is within the FoV as in Fig. 5 (b).

During the rotation of the receiver platform, the laptops periodically adjust the transmission and receiving beams according to channel statistics and the rotation angle as in Scheme 1. The average SNR and beam selection versus the rotation angle are shown in Fig. 6(a) and Fig. 6(b), respectively. In both figures, the following three benchmarks are compared with our proposed beam alignment scheme: (1) both transmission and receiving beams are exhaustively searched to maximize the channel power gain in each frame (*Optimal*); (2) the transmission and receiving beams are aligned to LoS direction initially; the receiving beams are then adjusted to compensate the rotation angle according to the reads of magnetometer, such that the alignment along the LoS direction is maintained (*Sensor-Assisted Compensation* [9]); (3) the transmission and receiving beams are aligned to the LoS direction initially, and no beam adjustment with rotation (*No Compensation*).

It can be observed that the proposed Scheme 1 achieves similar SNR performance to the optimal one without the overhead of exhaustive beam search. Both the optimal and proposed schemes select the LoS path when the rotation angle is less than $120°$, and the NLoS path thereafter. As a comparison, the sensor-assisted compensation scheme sticks to the LoS path and suffers from severe SNR degradation when the LoS is out of the FoV (the rotation angle is greater than $120°$). This demonstrates the effectiveness of the proposed beam alignment scheme in the scenario with multiple scattering clusters: with rotation sensing, the transmitter and the receiver can maintain the beam alignment; when the aligned path is out of array FoV, the location knowledge of the scattering clusters can help to
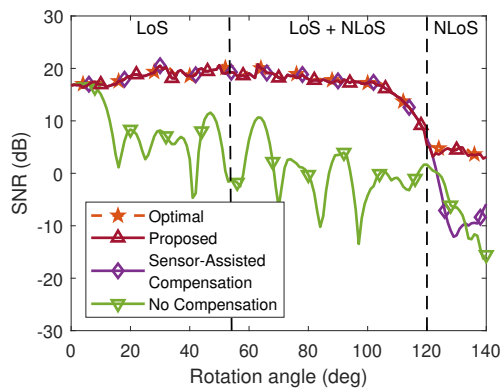
(a) SNR versus rotation angle.
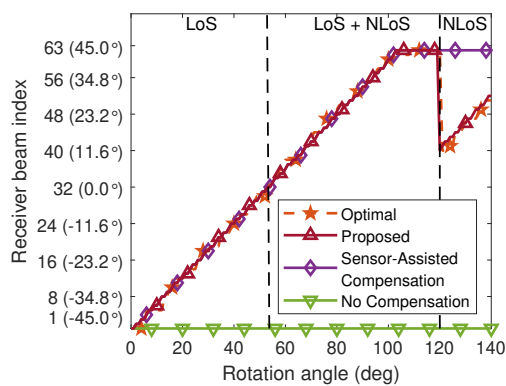
(b) Beam index versus rotation angle.

Fig. 6. Performance and beam switch of different beam alignment schemes during rotation.

establish a new beam alignment.

B. Numerical Simulations

In this part, the performance of the proposed downlink scheduling algorithm is demonstrated via numerical simulations. In the simulation, there are eight UEs in the system, four of which are static (indexed with $1 \sim 4$) and the others are rotating with angular velocity 2 rad/s (indexed with $5 \sim 8$). There are 3 scattering clusters in the channel, and the simulation parameters are summarized in Table IV.

We compare the proposed algorithms with the following three benchmarks. For each benchmark, the transmission power P_{BM} is either 27 dBm or 30 dBm. For a fair comparison, the base policy and benchmarks all adopt the proposed SCSI-based beam alignment and the same transmission power, i.e., $P_{\Pi} = P_{\mathrm{BM}}$.

Dynamic Backpressure (DBP) [41]: The downlink transmission UE selection is based on the backpressure algorithm according to the real-time data rate and queue length, i.e., $d_t = \arg\max_{k \in \mathcal{K}} [R_{t,k}(P_{\mathrm{BM}}) \times Q_{t,k}]$.

Largest-Rate First (LRF): In each frame, the UE with the largest data rate is selected, i.e., $d_t = \arg\max_{k \in \mathcal{K}} R_{t,k}(P_{\mathrm{BM}})$.

Longest-Queue First (LQF): In each frame, the UE with the longest queue is selected, i.e., $d_t = \arg\max_{k \in \mathcal{K}} Q_{t,k}$.

The instantaneous SNRs and queue lengths of a static UE (indexed by $k = 1$) and a rotating UE (indexed by $k = 5$) in one trial of the scheduling period are illustrated in Fig. 7,

TABLE IV. Parameter configuration of simulations.

| Parameter	Symbol	Value
# of UEs	K	8
# of frames per scheduling period	T	100
Buffer size	$Q_{\max}$	200
Packet arrival rate	λ_k	$\mathcal{U}(2, 6)$
Frame duration	N_{F}	10 ms
Packet size	B	400 bytes
# of antenna elements	$N_{\mathrm{R}}, N_{\mathrm{T}}$	16, 32
Antenna FoV of the UEs/BS	$[\phi_{\min}, \phi_{\max}],$ $[\theta_{\min}, \theta_{\max}]$	$[-30°, 30°],$ $[-90°, 90°]$
Maximum transmission power	$P_{\max}$	30 dBm
# of clusters and rays	$N_k^{\mathrm{cl}}, N_{k,i}^{\mathrm{ray}}$	$3, \mathcal{U}(20, 40)$
Variance of complex gain	$\sigma_{\alpha;k,i}^2$	$\mathcal{U}(4 \times 10^{-15},$ $4 \times 10^{-14})$
Cluster mean AoAs/AoDs	$\bar{\phi}_{1,k,i}, \bar{\theta}_{k,i}$	$\mathcal{U}(-180°, 180°)$
Angular spread of AoAs/AoDs	$\sigma_{\phi;k,i}, \sigma_{\theta;k,i}$	$\mathcal{U}(5°, 10°)$
Angular velocity	ω_k	0 for $k=1 \sim 4$; 2 rad/s for $k=5 \sim 8$
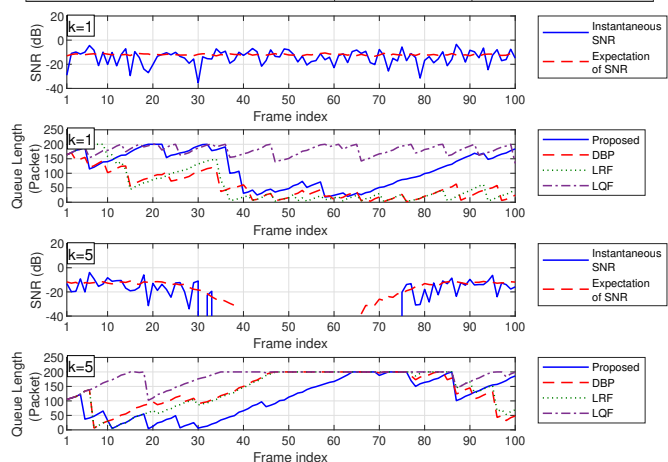
Fig. 7. Dynamics of SNRs and queue lengths.

where the transmission power is 27 dBm. There is no strong scattering cluster for the 5-th UE due to rotation in the middle of the scheduling period, leading to weak SNRs. It can be observed that the proposed scheme can predict the low SNR period of the rotating UEs and schedule more transmission opportunities for them before the low SNR period so that the packet drop rate can be reduced. As a comparison, the benchmarks suffer from high packet drop rate during the low SNR period. This corroborates our motivation and demonstrates the performance gain of the sensing-based channel prediction and the scheduling framework introduced in this paper.

While Fig. 7 shows the system performance in a trial, Fig. 8(a) shows the CDF of the system cost of the proposed scheme as well as the three benchmarks. It can be observed that the proposed algorithm has significantly better CDF curves than the benchmarks. More insights can be obtained from Fig. 8(b), where average transmission power, average queuing delay, and average number of packet-drop UEs are illustrated respectively. DBP has the medium cost of delay and packet drop penalty. This is because it makes UE selection according to both the queue length and data rate. LRF has the least cost of delay because it attempts to decrease the queue length as much as possible in every single frame. LQF only takes the QSI into account but neglects the CSI, which results in the worst performance. The proposed scheme manages to achieve the minimum packet drop rate, while keeping the average queuing
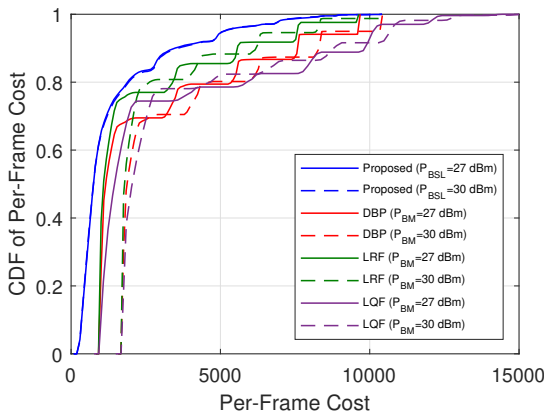
(a) CDF of the per-frame cost.
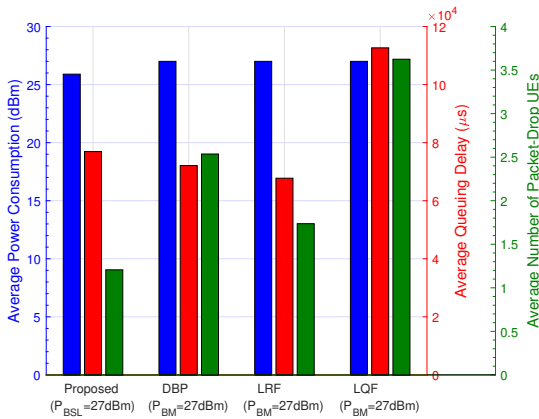
(b) Average transmission power, queuing delay and number of packet-drop UEs.

Fig. 8. Numerical analysis of system cost.

delay at a low level. This demonstrates the benefits of channel prediction of the proposed scheme in suppressing the packet drop rate with non-stationary and intermittent mmWave link status.

The average overall cost versus the number of UEs is shown in Fig. 9, where the maximum number of UEs is increased to 16. It can be observed that the average overall cost of the proposed scheme is always lower than the benchmarks in the different number of UEs, which implies the better performance of the proposed scheme. Moreover, the performance gain is not sensitive to the number of UEs.
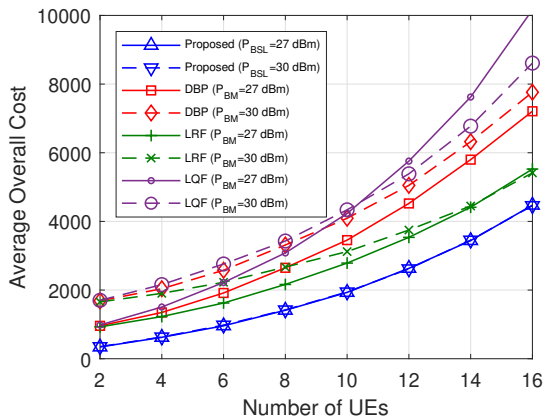

Fig. 9. Average overall cost versus the number of UEs.

The convergence of the channel statistics is shown in Fig.

10. Due to the page limitation, we only show the convergence of channel statistics of the $(1, 1)$-th cluster. The tracking of AoA standard deviation $\hat{\sigma}_{\phi;1,1}^{\Xi}$, the number of paths $\hat{N}_{1,1}^{\Xi}$, and the standard deviation of the complex gain $\hat{\zeta}_{1,1}^{\Xi}$ converge after around 100 channel estimations.
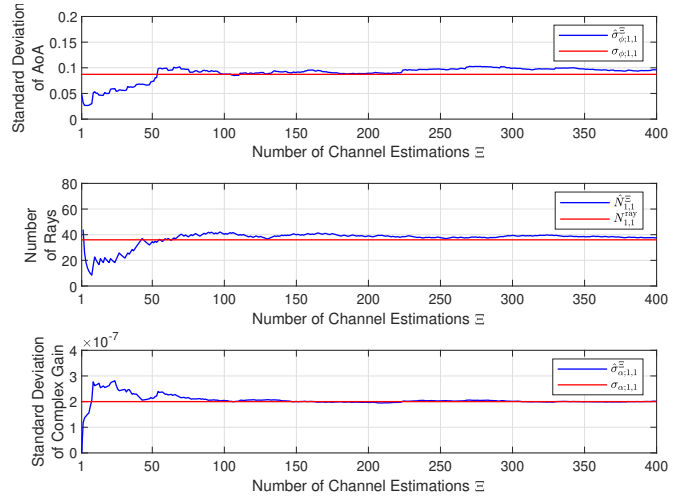

Fig. 10. Convergence of the channel statistics.

VII. CONCLUSION

In this paper, we consider the downlink transmission scheduling in an mmWave cell with one BS and multiple UEs. Each UE is either static or rotating with a predictable angular velocity for a number of frames, where the angular velocity of rotation can be measured by the motion sensors in the UEs and reported to the BS. We first propose a SCSI-based beam alignment scheme, where the precoder and combiner can be selected in each frame without any estimation of the instantaneous channel. Then, we formulate the joint optimization of the downlink UE selection and power allocation as a finite-horizon MDP in a scheduling period. To address the *curse of dimensionality*, we also propose a novel approximate MDP approach via one-step policy iteration over a base policy, where the analytical performance bound can be obtained. Finally, we propose efficient learning algorithms when some system statistics are unknown. It is shown by simulations that the proposed scheduling framework can effectively exploit the motion sensors to predict future performance, resulting in better performance than the benchmarks.

APPENDIX A
PROOF OF LEMMA 1

By substituting $\mathbf{H}_{t,k}$ with (3), $Y_{t,k}$ can be represented by (55), where $\rho_{t,k,i,\ell} \sim \mathcal{CN}(0, \sigma_{\rho;t,k,i,1}^2)$ $(\forall \ell)$, $\sigma_{\rho;t,k,i,1}^2$ is given by (56).

Since $f_{\mathrm{R}}(\phi_{q_{t,k}^{\dagger}}, \phi_{t,k,i,\ell})\Lambda_{\mathrm{R}}(\phi_{t,k,i,\ell})$ and $f_{\mathrm{T}}(\theta_{t,k,i,\ell}, \theta_{p_{t,k}^{\dagger}})$ $\Lambda_{\mathrm{T}}(\theta_{t,k,i,\ell})$ are independent, we can derive (19). According to the central limit theorem, with sufficiently large $N_{k,i}^{\mathrm{ray}}$ $(\forall i)$, $\sum_{\ell=1}^{N_{k,i}^{\mathrm{ray}}} \rho_{t,k,i,\ell} \sim \mathcal{CN}(0, N_{k,i}^{\mathrm{ray}} \sigma_{\rho;t,k,i,1}^2)$ and thus $\sum_{i=1}^{N_k^{\mathrm{cl}}} \sum_{\ell=1}^{N_{k,i}^{\mathrm{ray}}} \rho_{t,k,i,\ell} \sim \mathcal{CN}(0, \sum_{i=1}^{N_k^{\mathrm{cl}}} N_{k,i}^{\mathrm{ray}} \sigma_{\rho;t,k,i,1}^2)$. $Y_{t,k}$ follows an exponential distribution, i.e., $Y_{t,k} \sim \mathrm{Exp}(1/\sum_{i=1}^{N_k^{\mathrm{cl}}} N_{k,i}^{\mathrm{ray}} \sigma_{\rho;t,k,i,1}^2)$.

$$Y_{t,k} = \left| \sum_{i=1}^{N_k^{\text{cl}}} \sum_{\ell=1}^{N_{k,i}^{\text{ray}}} \underbrace{\alpha_{t,k,i,\ell} f_{\text{R}}\left(\phi_{q_{t,k}^\dagger}, \phi_{t,k,i,\ell}\right) f_{\text{T}}\left(\theta_{t,k,i,\ell}, \theta_{p_{t,k}^\dagger}\right) \Lambda_{\text{R}}(\phi_{t,k,i,\ell}) \Lambda_{\text{T}}(\theta_{t,k,i,\ell})}_{\rho_{t,k,i,\ell}} \right|^2 \tag{55}$$

$$\begin{aligned} \sigma_{\rho;t,k,i,1}^2 =& \frac{\pi \sigma_{\alpha;k,i}^2 + 4\sigma_{\alpha;k,i}}{4} \text{Var}\left[f_{\text{R}}\left(\phi_{q_{t,k}^\dagger}, \phi_{t,k,i,1}\right) f_{\text{T}}\left(\theta_{t,k,i,1}, \theta_{p_{t,k}^\dagger}\right) \Lambda_{\text{R}}(\phi_{t,k,i,1}) \Lambda_{\text{T}}(\theta_{t,k,i,1}) \right] \\ &+ \sigma_{\alpha;k,i} \mathbb{E}^2 \left[\left| f_{\text{R}}\left(\phi_{q_{t,k}^\dagger}, \phi_{t,k,i,1}\right) f_{\text{T}}\left(\theta_{t,k,i,1}, \theta_{p_{t,k}^\dagger}\right) \Lambda_{\text{R}}(\phi_{t,k,i,1}) \Lambda_{\text{T}}(\theta_{t,k,i,1}) \right| \right] \end{aligned} \tag{56}$$

$$\begin{aligned} \frac{1}{\Xi} \sum_{\xi=1}^\Xi \eta_{k,i}^{\xi,j} \xrightarrow{\Xi \to \infty} & P_{\text{L}} \mathbb{E}_{\mathbf{H}_{k,i}^\xi} \left[|[\mathbf{w}_{k,i}^{\text{L},j}]^{\text{H}} \mathbf{H}_{k,i}^\xi \mathbf{f}_{k,i}^{\text{L}}|^2 \right] \\ =& P_{\text{L}} N_{k,i}^{\text{ray}} \sigma_{\alpha;k,i}^2 \int_{\phi_{\min}}^{\phi_{\max}} \left| f_{\text{R}}(\phi_{q_{k,i}^{\text{L},j}}, \phi) \right|^2 f_{\phi;k,i}(\phi; \sigma_{\phi;k,i}, \bar{\phi}_{k,i}^{\text{L}}) \text{d}\phi \times \int_{\theta_{\min}}^{\theta_{\max}} \left| f_{\text{T}}(\theta_{p_{k,i}^{\text{L}}}, \theta) \right|^2 f_{\theta;k,i}(\theta; \sigma_{\theta;k,i}, \bar{\theta}_{k,i}^{\text{L}}) \text{d}\theta \\ \xrightarrow{N_{\text{R}}, N_{\text{T}} \to \infty} & P_{\text{L}} N_{k,i}^{\text{ray}} \sigma_{\alpha;k,i}^2 \Pr\left[\phi \in \mathcal{P}_{\text{R};q_{k,i}^{\text{L},j}}\right] \Pr\left[\theta \in \mathcal{P}_{\text{T};p_{k,i}^{\text{L}}}\right], \; j = 1,2 \end{aligned} \tag{57}$$

APPENDIX B
PROOF OF LEMMA 3

Denote $\mathbf{s}_{t,\tau,k} \in \mathbb{R}^{(Q_{\max}+1)\times 1}$ as the pre-decision probability vector for the k-th queue in the τ-th frame given $Q_{t,k}^{\text{D}}$. $\mathbf{1}_{Q_{t,k}^{\text{D}}+1}$ represents the pre-decision probability vector for the k-th queue in the t-th frame. $\mathbf{M}_{t,k}$ and $\mathbf{P}_k$ are transition probability matrices for the k-th queue considering both packet departure and arrival and only the packet arrivals, respectively. $[\mathbf{c}^{(1)}]_i$ and $[\mathbf{c}^{(2)}]_i$ represent the per-frame queuing and packet-drop cost for the k-th UE in the τ-th frame for cases $t+1 \le \tau \le T$ and $\tau = T+1$, respectively. Then (36) can be written as

$$W_{t,k}^\Pi(Q_{t,k}^{\text{D}}) = \sum_{\tau=t+1}^T \mathbf{s}_{t,\tau,k}^{\text{T}} \mathbf{c}^{(1)} + \mathbf{s}_{t,T+1,k}^{\text{T}} \mathbf{c}^{(2)}, \tag{58}$$

where

$$\mathbf{s}_{t,\tau,k} = \mathbf{X}_k^{\text{T}}(t,\tau) \mathbf{P}_k^{\text{T}} \mathbf{1}_{Q_{t,k}^{\text{D}}+1}, \; \tau = t+1,\dots,T. \tag{59}$$

In (58), $\mathbf{s}_{t,\tau,k}^{\text{T}} \mathbf{c}^{(1)}$ and $\mathbf{s}_{t,T+1,k}^{\text{T}} \mathbf{c}^{(2)}$ counts for the average queuing and packet-drop cost in the τ-th frame for cases $t+1 \le \tau \le T$ and $\tau = T+1$, respectively. Hence, Lemma 3 is straightforward.

APPENDIX C
PROOF OF LEMMA 4

The feasible power set $\mathscr{P}_{t,k}$ contains the minimum required transmission powers to transmit all possible integer numbers of packets. Then P2(k) can be solved by one-dimensional search in $\mathscr{P}_{t,k}$. Adding the objective function in P2(k) with the constant $\sum_{k' \ne k} W_{t,k'}^\Pi\left(Q_{t,k'}^{\text{D}}(Q_{t,k'}, Y_{t,k'})\right) - W_t^\Pi(\mathcal{Q}_t)$, we have $W_t^\Pi\left(\mathcal{Q}_t^{\text{D}}(\mathcal{S}_t, \psi_{t,k})\right) - W_t^\Pi(\mathcal{Q}_t) = W_{t,k}^\Pi\left(Q_{t,k}^{\text{D}}(P_{t,k})\right) - W_{t,k}^\Pi(Q_{t,k}) = \Delta \mathbf{z}_{t,k}^{\text{T}}(P_{t,k}) \mathbf{v}_{t,k}$.

APPENDIX D
PROOF OF LEMMA 6 AND LEMMA 7

According to [47], with sufficiently large N_{R} and N_{T}, we have $|f_{\text{R}}(\phi_q, \phi)| \to 1$ for $\phi \in \mathcal{P}_{\text{R};q}$ and $|f_{\text{R}}(\phi_q, \phi)| \to 0$ otherwise, and $|f_{\text{T}}(\theta_p, \theta)| \to 1$ for $\theta \in \mathcal{P}_{\text{T};p}$ and $|f_{\text{T}}(\theta_p, \theta)| \to 0$ otherwise. Then (49) can be represented by (57).

With sufficiently large Ξ, the ratio of the average received

signal powers under the two combiners is given by

$$\frac{\sum_{\xi=1}^\Xi \eta_{k,i}^{\xi,1}}{\sum_{\xi=1}^\Xi \eta_{k,i}^{\xi,2}} \xrightarrow{\Xi \to \infty} \frac{\Pr\left[\phi \in \mathcal{P}_{\text{R};q_{k,i}^{\text{L},1}}\right]}{\Pr\left[\phi \in \mathcal{P}_{\text{R};q_{k,i}^{\text{L},2}}\right]}. \tag{60}$$

Hence, Lemma 6 is straightforward.

With sufficiently large Ξ, we have

$$\frac{1}{\Xi} \sum_{\xi=1}^\Xi \mathbb{I}[\eta_{k,i}^{\xi,1} \le \tau_{k,i}^{\text{L}}] \xrightarrow{\Xi \to \infty} \Pr[\eta_{k,i}^{\xi,1} \le \tau_{k,i}^{\text{L}}] = \Pr[\Phi_{k,i} = 0],$$

where $\Phi_{k,i}$ follows a binomial distribution, i.e., $\Phi_{k,i} \sim \mathbb{B}(N_{k,i}^{\text{ray}}, \Pr[\phi \in \mathcal{P}_{\text{R};q_{k,i}^{\text{L},1}}] \Pr[\theta \in \mathcal{P}_{\text{T};p_{k,i}^{\text{L}}}])$. Hence, Lemma 7 is straightforward.

REFERENCES

[1] Y. Sun, B. Lv, R. Wang, H. Tan, and F. C. M. Lau, "Predictive resource allocation in mmWave systems with rotation detection," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 2753–2759.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, 2016.

[4] B. Li, Z. Zhou, W. Zou, X. Sun, and G. Du, "On the efficient beamforming training for 60GHz wireless personal area networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 504–515, 2013.

[5] Z. Xiao, P. Xia, and X.-G. Xia, "Hierarchical multi-beam search for millimeter-wave MIMO systems," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5.

[6] J. Choi, "Beam selection in mm-Wave multiuser MIMO systems using compressive sensing," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2936–2947, 2015.

[7] N. J. Myers, A. Mezghani, and R. W. Heath, "Swift-link: A compressive beam alignment algorithm for practical mmWave radios," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1104–1119, 2019.

[8] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2909–2913.

[9] J. Bao, D. Sun, and H. Li, "Motion sensor aided beam tracking in mobile devices of millimeter-wave communications," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–7.

[10] M. Park and H. K. Pan, "Effect of device mobility and phased array antennas on 60 GHz wireless networks," in *Proceedings of the 2010 ACM international workshop on mmWave communications: from circuits to networks*, 2010, pp. 51–56.

[11] Y. Lin, C. Shen, and Z. Zhong, "Sensor-aided predictive beam tracking for mmWave phased array antennas," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–5.

[12] V. Dham. (2017) Programming chirp parameters in TI radar devices. [Online]. Available: bit.ly/3MVCSSl

[13] M. N. Hamdy. (2020) Beamformers explained. [Online]. Available: bit.ly/40jIk4z

[14] S. Wang, J. Huang, X. Zhang, H. Kim, and S. Dey, "X-array: Approximating omnidirectional millimeter-wave coverage using an array of phased arrays," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.

[15] V. Raghavan, L. Akhoondzadeh-Asl, V. Podshivalov, J. Hulten, M. A. Tassoudji, O. H. Koymen, A. Sampath, and J. Li, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3010–3024, 2019.

[16] E. Öjefors, M. Andreasson, T. Kjellberg, H. Berg, L. Aspemyr, R. Nilsson, K. Brink, R. Dahlbäck, D. Wu, K. Sjögren, and M. Carlsson, "A 57-71 GHz beamforming SiGe transceiver for 802.11 ad-based fixed wireless access," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2018, pp. 276–279.

[17] C.-H. Fang, L.-H. Shen, T.-P. Huang, and K.-T. Feng, "Delay-aware admission control and beam allocation for 5G functional split enhanced millimeter wave wireless fronthaul networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2430–2444, 2022.

[18] Y. Cao, B. Sun, and D. H. K. Tsang, "Delay-aware scheduling over mmWave/sub-6 dual interfaces: A reinforcement learning approach," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[19] J. García-Rois, R. Banirazi, F. J. González-Castaño, B. Lorenzo, and J. C. Burguillo, "Delay-aware optimization framework for proportional flow delay differentiation in millimeter-wave backhaul cellular networks," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2037–2051, 2018.

[20] Y. Heng and J. G. Andrews, "Machine learning-assisted beam alignment for mmWave systems," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[21] W. Zhang, Y. Wei, S. Wu, W. Meng, and W. Xiang, "Joint beam and resource allocation in 5G mmWave small cell systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10 272–10 277, 2019.

[22] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, 2014.

[23] W. Zhang, W. Zhang, and J. Wu, "UAV beam alignment for highly mobile millimeter wave communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8577–8585, 2020.

[24] I. Mavromatis, A. Tassi, R. J. Piechocki, and A. Nix, "Beam alignment for millimetre wave links with motion prediction of autonomous vehicles," in *Antennas, Propagation RF Technology for Transport and Autonomous Platforms 2017*, 2017, pp. 1–8.

[25] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, 2018.

[26] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[27] Y. Sun, J. Li, T. Zhang, R. Wang, X. Peng, X. Han, and H. Tan, "An indoor environment sensing and localization system via mmWave phased array," *Journal of Communications and Information Networks*, vol. 7, no. 4, pp. 383–393, 2022.

[28] C. Yu, Y. Sun, Y. Luo, and R. Wang, "mmAlert: mmwave link blockage prediction via passive sensing," *IEEE Wireless Communications Letters*, pp. 1–1, 2023.

[29] C. Yu, Y. Luo, R. Chen, and R. Wang, "Passive handwriting tracking via weak mmWave communication signals," *IEEE Wireless Communications Letters*, vol. 13, no. 3, pp. 874–878, 2024.

[30] H. Seleem, A. I. Sulyman, and A. Alsanie, "Hybrid precoding-beamforming design with hadamard RF codebook for mmWave large-scale MIMO systems," *IEEE Access*, vol. 5, pp. 6813–6823, 2017.

[31] A. Forenza, D. J. Love, and R. W. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, 2007.

[32] J. D. Little and S. C. Graves, "Little's law," in *Building intuition*. Springer, 2008, pp. 81–100.

[33] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6284–6297, 2016.

[34] B. Lyu, Y. Hong, H. Tan, Z. Han, and R. Wang, "Cooperative jobs dispatching in edge computing network with unpredictable uploading delay," *Journal of Communications and Information Networks*, vol. 5, no. 1, pp. 75–85, 2020.

[35] R. Wang and V. K. N. Lau, "Delay-aware two-hop cooperative relay communications via approximate MDP and stochastic learning," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7645–7670, 2013.

[36] D. Bertsekas, *Dynamic programming and optimal control*, 4th ed. Belmont, MA, USA: Athena scientific, 2012, vol. 2.

[37] W. B. Powell, *Approximate Dynamic Programming: Solving the curses of dimensionality*. John Wiley & Sons, 2007, vol. 703.

[38] B. Lv, L. Huang, and R. Wang, "Joint downlink scheduling for file placement and delivery in cache-assisted wireless networks with finite file lifetime," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4177–4192, 2019.

[39] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[40] A. Geramifard, T. J. Walsh, S. Tellex, G. Chowdhary, N. Roy, and J. P. How, "A tutorial on linear function approximators for dynamic programming and reinforcement learning," *Foundations and Trends® in Machine Learning*, vol. 6, no. 4, pp. 375–451, 2013.

[41] L. Georgiadis, M. J. Neely, and L. Tassiulas, *Resource allocation and cross-layer control in wireless networks*. Now Publishers Inc, 2006.

[42] S. Huang, B. Lv, R. Wang, and K. Huang, "Scheduling for mobile edge computing with random user arrivals—an approximate MDP and reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7735–7750, 2020.

[43] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[44] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[45] USRP X310. [Online]. Available: https://bit.ly/3JRWEMh

[46] EVK 06002/00. [Online]. Available: https://bit.ly/3pD96Zx

[47] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.
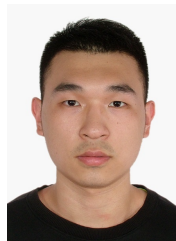
Yifei Sun (Graduate Student Member, IEEE) received the B.E. degree in communication engineering from Southern University of Science and Technology (SUSTech), Shenzhen, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Computer Science, The University of Hong Kong (HKU) and Department of Electrical and Electronic Engineering, SUSTech. His research interests include wireless sensing and sensing-assisted wireless resource allocation.
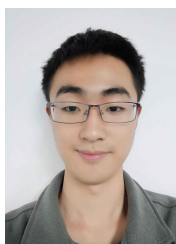
Bojie Lv received the B.E. degree in communication engineering from Southern University of Science and Technology (SUSTech), Shenzhen, China, in 2018, and the M.Phil. degree (ranked top 2) in information and communication engineering from Harbin Institute of Technology, Harbin, China, in 2020. He is currently pursuing the Ph.D. degree in mathematics with SUSTech. His research interests are in the area of stochastic optimization and resource allocation in the wireless networks.

Rui Wang (Member, IEEE) received the B.S. degree from the University of Science and Technology of China in 2004, and the Ph.D. degree in wireless communications from The Hong Kong University of Science and Technology in 2008.

From 2009 to 2012, he was a Senior Research Engineer with Huawei Technologies, Co., Ltd. Since 2012, he has joined the Southern University of Science and Technology of China as an Associate Professor. He has research experience in both academia and industry. He has authored over 100 papers in top-level IEEE journals and flagship international conferences, especially in the area of wireless radio resource optimization and interference management. He has contributed over 20 U.S. patent applications and over 100 Chinese patent applications (50 of them have been granted).

Haisheng Tan (Senior Member, IEEE) received the B.E. degree (Hons.) in software engineering and the B.S. degree (Hons.) in management from the University of Science and Technology of China (USTC), and the Ph.D. degree in computer science from The University of Hong Kong (HKU). He is currently a Professor with USTC. His research interests lie primarily in networking algorithm design and system implementation, where he has published over 100 papers in prestigious journals and conferences, including IEEE/ACM ToN, NSDI, INFOCOM, and EruoSys. He recently received the Best Paper Award in WASA'19, CWSN'20, PDCAT'20, and ICPADS'21.

Francis C.M. Lau received the Ph.D. degree from the Department of Computer Science, University of Waterloo. He was formerly Head of Department of Computer Science at The University of Hong Kong, China, where he now serves as an honorary professor. His research interests include computer systems, networks, AI, and application of computing in arts and music. He was the Editor-in-Chief of the Journal of Interconnection Networks from 2011 to 2020.