Towards Optimal Pilot Spacing and Power Control in Multi-Antenna Systems Operating Over Non-Stationary Rician Aging Channels

Sajad Daei[†], Gábor Fodor[⋆†], Mikael Skoglund[†], Miklós Telek[‡♯]

[†]KTH Royal Institute of Technology, Stockholm, Sweden.
[⋆]Ericsson Research, Stockholm, Sweden.
[‡]Budapest University of Technology and Economics, Budapest, Hungary.
[♯]MTA-BME Information Systems Research Group, Budapest, Hungary.

***Abstract*—Several previous works have addressed the inherent trade-off between allocating resources in the power and time domains to pilot and data signals in multiple input multiple output systems over block-fading channels. In particular, when the channel changes rapidly in time, channel aging degrades the performance in terms of spectral efficiency without proper pilot spacing and power control. Despite recognizing non-stationary stochastic processes as more accurate models for time-varying wireless channels, the problem of pilot spacing and power control in multi-antenna systems operating over non-stationary channels is not addressed in the literature. In this paper, we address this gap by introducing a refined first-order autoregressive model that exploits the inherent temporal correlations over non-stationary Rician aging channels. We design a multi-frame structure for data transmission that better reflects the non-stationary fading environment than previously developed single-frame structures. Subsequently, to determine the optimal pilot spacing and power control within this multi-frame structure, we develop an optimization framework and an efficient algorithm based on maximizing a deterministic equivalent expression for the spectral efficiency, demonstrating its generality by encompassing previous channel aging results. Our numerical results indicate the efficacy of the proposed method in terms of spectral efficiency gains over the single frame structure.**

***Index terms*—** Channel aging, frame design, multi-antenna systems, Rician non-stationary channels, spectral efficiency

I. Introduction

As several previous works have pointed out, the performance of the uplink of multiple input multiple output (MIMO) systems depends critically on the quality of the available channel state information at the receiver (CSIR) [1]–[3]. Therefore, assuming a block fading reciprocal channel, a finite number of symbols in the time and frequency domains are typically made available for CSIR acquisition, while the remaining symbols are used for data transmission [4], [5]. Also, under a fixed power budget, pilot symbols reduce the transmitted energy for data symbols, as it has been shown in [6], [7], where the near-optimal pilot-to-data power ratio (PDPR) for various pilot patterns and receiver structures have been derived. Subsequently, reference [8] has optimized the pilot overhead for single-user wireless fading channels, and studied how various system parameters of interest (e.g. fading rate, signal-to-noise ratio (SNR)) depend on this overhead. Reference [9] has derived analytical expressions for the achievable spectral efficiency (SE) as a function of the PDPR in the presence of correlated channels. Note that the above papers have considered block-fading channels, where the subsequent channel realizations are assumed to be independent and drawn from the same distribution. For block-fading channel models, important prior works have developed the conceptually elegant and practically useful concept of the deterministic equivalent signal-to-interference-plus-noise ratio (SINR) that helps to derive analytical expressions for the achievable SINR and spectral efficiency of multi-antenna systems [10]–[14] .

In contrast, some previous works have suggested that in the case when the subsequent channel realizations are correlated, the temporal correlation structure can be exploited to improve the quality of the channel estimates [15], [16].

Another set of important related works have considered the problem of acquiring CSIR in the presence of channel aging [17]–[23]. Channel aging refers to the evolution of the channel between subsequent estimation instances, and can be conveniently modelled as a first-order or higher-order stationary autoregressive (AR) process, whose states can be estimated and predicted using Kalman or Wiener filters [19], [22].

Specifically, the work in [19] explored the effects of channel aging on the SINR performance in single input multiple output uplink systems and multiple input single output downlink systems, and derived closed-form expressions for the deterministic equivalent SINR under channel aging. Extending that work, reference [23] considered a more general AR model, assuming a specific state transition matrix in the form of exponential decaying matrices to formulate a closed-form expression for the deterministic equivalent SINR in AR aging channels.

As it has been shown in, for example, [23]–[25], channel aging gives rise to the inherent problem of pilot spacing or frame size dimensioning. Frame size dimensioning refers to determining how densely pilot symbols should be inserted in the flow of data symbols such that the available CSIR is properly updated at the expense of some pilot overhead. A key insight provided by these papers is that the performance of the uplink of MIMO systems depends both on the PDPR and the frame size, and both should be properly set such that a good balance between CSIR acquisition and communication is achieved in the presence of some power budget and channel

aging.

Along a related research line, several recent works have shown that in many wireless communication scenarios, the evolution of the wireless channel is advantageously modeled as a non-stationary stochastic process, whose first and second order statistics – including the mean and covariance matrices of the channels – evolve in time [26]–[33]. Specifically, the measurement results obtained in a MIMO system and reported in [29], [34] suggest that the propagation conditions and thereby the wireless channels are non-stationary in space and time even under low mobility due to varying mobile user positions and changes in the propagation environment. Non-stationary channel aging is also considered in [28], where a convergent minimum mean squared error (MMSE) beamforming algorithm is proposed, which outperforms state-of-the-art beamforming methods in the downlink of MIMO systems. However, these papers do not treat the inherent trade-offs discussed above.

In the light of the above results on the PDPR, frame size dimensioning (pilot spacing) and the importance of modelling the evolution of the wireless channels as non-stationary processes, in this paper we argue that CSIR acquisition should be designed for MIMO systems operating over non-stationary wireless systems. Specifically, in this paper, we consider the uplink of a multiuser multiple input multiple output (MU-MIMO) system operating over non-stationary aging Rician channels. In these channels, both the mean and covariance matrices of the channel age in time, which can lead to a degradation in performance in terms of SE.

Therefore, in this paper, we first introduce a refined first-order AR model, meticulously tailored to capture the intricacies of time-varying Rician non-stationary aging channels. This model harnesses the underlying temporal correlations of the channel by allowing the channel covariance matrices to change in time.

Next, we introduce a multi-frame structure, where each frame consists of one pilot time slot, while the remaining slots are allocated for data transmission, as illustrated in Figure 3. The frame sizes (pilot spacing), the number of frames, and the PDPR are not predetermined. We propose an analytical optimization framework to maximize the deterministic equivalent SE under specific power constraints. This framework identifies critical parameters, such as the frame size, the number of frames, and the optimal power allocation between data and pilot symbols. Significantly, this optimization relies solely on the temporal dynamics of the channel, without dependence on measurements or channel and data estimates. Consequently, the proposed design can be executed at the transmitter side. Central to our work is the pivotal question of determining the optimal rate for updating the CSIR, that is the optimal pilot spacing. Our proposed optimization framework explicitly addresses this fundamental question.

Furthermore, the outcomes of extensive numerical experiments validate the efficacy of our methodology. They underscore its profound impact on optimizing pilot power, data power, frame sizes, and the number of frames, which are all critical elements in the quest for enhanced SE.

A. Contributions and Key Differences Compared with Prior Works

In this subsection, we highlight the main contributions of this paper and discuss the key differences compared with prior works.

1) **Refined AR Channel Modeling**: We propose a refined representation of the time-varying channel by a refined AR channel model in Proposition 1, which not only exploits temporal channel correlations with previous time instances, but it also takes into account the innovative channel information of the current time. The state transition matrix depends explicitly on the so-called correlation matrix, which is equivalent to the auto-correlation function of the normalized centered channel.
   The proposed AR model addresses the Rician non-stationary aging channel, which is a more general case than the Rayleigh stationary scenario discussed in [19], [23]–[25]. Furthermore, in the specific case of Rayleigh stationary fading examined in [23]–[25], determining the optimal AR evolution matrix and the covariance matrix of the AR model's error remains unclear. In contrast, our model establishes explicit connections between the AR evolution matrix and the correlation matrix, which can be practically obtained.
2) **Deterministic Equivalent Spectral Efficiency**: We propose an analytical framework in Theorems 1 and 2 for Rician aging non-stationary fast-fading channels, utilizing concentration inequalities derived from random matrix theory tools. This framework provides an exact deterministic approximation of the average achievable SE, termed the deterministic equivalent SE, as the number of base station (BS) antennas grows significantly. The outcome is expressed as a function of various parameters, including frame sizes, the number of frames, pilot and data powers, the number of receiver antennas, and the Rician K factor.
   In the context of stationary Rayleigh block fading channels, prior works such as [10] and [35] obtained approximations for the deterministic equivalent SINR, while [19], [23] established an upper bound for the average achievable SE in stationary Rayleigh aging channels. Notably, in our SE derivations, we employ the MMSE estimator for data symbol estimation and incorporate an aging-aware MMSE receiver combiner. This approach contrasts with [23], which relies on least square estimation for the optimal receiver combiner. The distinction underscores that our method effectively utilizes the pre-existing aging information of the channels to determine the optimal receiver combiner for data estimation, providing a crucial advantage over the approach in [23].
3) **Optimal Frame Design and Power Allocation in Multi-frame Data Transmission**: Our approach provided in Section III introduces a multi-frame framework, as illustrated in Figure 3, wherein each frame comprises one pilot time slot and the remaining slots are allocated for data transmission. This flexible framework accommodates an arbitrary number of frames and data time slots within each frame. Subsequently, we formulate an

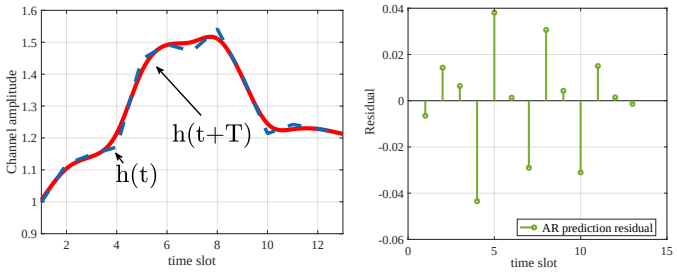
optimization problem with the objective of maximizing the deterministic equivalent SE averaged across all data time slots and frames, while satisfying some power constraints. This optimization task entails determining optimal values for various parameters, including the number of frames, frame sizes, as well as pilot and data powers. The proposed problem inherently falls under mixed-integer programming, a category generally recognized as NP-hard. To address this complexity, we provide a novel, efficient algorithm in Algorithm 1 named OptResource designed to tackle the intricacies of solving the optimization problem.

While [23] exclusively optimizes pilot spacing (frame size) in Rayleigh stationary fading scenarios, making the implicit assumption of equal frame sizes, predetermined knowledge of the number of frames (number of pilots), pilot and data powers, our method goes beyond by optimizing the number of frames, diverse frame sizes, and the allocation of pilot and power in the broader and more complex context of Rician aging non-stationary environments. Notably, even in stationary environments, our proposed multi-frame optimization outperforms the single-frame optimization presented in [23] in terms of SE. Furthermore, it is crucial to highlight that [23] seeks to maximize an upper-bound on the average SE when determining optimal pilot spacing, while our proposed framework exactly maximizes the average achievable SE. This distinction is significant, as maximizing upper-bounds does not guarantee the maximization of the average SE.

4) **Optimization at the Transmitter or Receiver**: Our findings provided in Section III and V indicate that interference components such as path loss, pilot and data powers, and Doppler frequencies impact the achieved deterministic spectral efficiency but are irrelevant in determining optimal frame design. This suggests that all optimization tasks regarding frame design can be carried out at the transmitter side while optimal power-domain resources can be calculated at the receiver and the receiver can inform the transmitter about them by some control or feedback loops. To the best of our knowledge, prior works such as [23]–[25] assert the achievement of optimal pilot spacing at the receiver side.
5) **Accessible Correlation Information in Non-stationary Environments**: We derive explicit formulas in Proposition 2 for the required correlation matrix regarding the proposed framework in Section III, and show that they are accessible beforehand in practical non-stationary Rician time-varying environments. Notably, in the special scenario of stationary Rayleigh aging channels, our findings align with the well-known Jakes model, as previously derived in [36].

B. Outline

The paper is organized as follows. Section II proposes a channel model that exploits temporal correlations, and presents the measurement model at the receiver. Section III specifies the road to obtain the instantaneous (random) SINR and SE and proposes a deterministic expression for the achievable SE based on random matrix theory tools that concentrates around the random instantaneous SE. In Section IV, we propose an optimization problem that takes into account this deterministic equivalent SE and finds the optimal values of frame sizes, number of frames and pilot and data powers. We also propose a heuristic algorithm to find the optimal values of frame size, number of frames and pilot and data powers. The paper is concluded in Section VI.



(a) (b)

Figure 1. The channel at time $t+T$ is described based on the channel at time t using the proposed AR model. (a): Time-varying actual channel and its AR prediction. (b): AR prediction residual.

C. Notation

Scalars are denoted by italic letters, while matrices and vectors are shown by bold upright letters. $\mathbf{I}_N$ denotes the identity matrix of size N. **vec** stands for the column stacking vector operator that transforms a matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$ into its vectorized version $\mathbf{x} \triangleq \mathbf{vec}(\mathbf{X}) \in \mathbb{C}^{MN\times 1}$. For a random process $\mathbf{h}(t) \in \mathbb{C}^{N_r\times 1}$ that shows the channel at the time slot $t \in \mathbb{N}^+$, $\overline{\mathbf{h}}(t) \triangleq \mathbb{E}[\mathbf{h}(t)]$ denotes the channel mean and $\widetilde{\mathbf{h}}(t) \triangleq \mathbf{h}(t) - \overline{\mathbf{h}}(t)$ denotes the centered channel. Furthermore, the autocovariance and autocorrelation are given by $\mathbf{C}_{\mathbf{h}}(t) \triangleq \mathbb{E}\left[\widetilde{\mathbf{h}}(t)\widetilde{\mathbf{h}}(t)^{\mathrm{H}}\right]$ and $\mathbf{R}_{\mathbf{h}}(t) \triangleq \mathbb{E}\left[\mathbf{h}(t)\mathbf{h}(t)^{\mathrm{H}}\right]$. The cross-covariance of the channel between t_1 and t_2 is shown by $\mathbf{C}_{\mathbf{h}}(t_1,t_2) \triangleq \mathbb{E}\left[\widetilde{\mathbf{h}}(t_1)\widetilde{\mathbf{h}}^{\mathrm{H}}(t_2)\right]$. The normalized centered channel is shown by $\mathbf{h}'(t) \triangleq \mathbf{C}_{\mathbf{h}}^{-\frac{1}{2}}(t)\widetilde{\mathbf{h}}(t)$. To avoid complexity, we occasionally eliminate the time dependency when referring to autocovariance or autocorrelation matrices. $\mathbf{e}_k \in \mathbb{R}^{N\times 1}$ refers to the k-th unit vector, with components equal to zero except for the k-th component that is one. $\mathbf{1}_N \in \mathbb{R}^{N\times 1}$ is an all-one vector of size N. $j \triangleq \sqrt{-1}$ is the complex imaginary unit. $\lambda_{\max}(\mathbf{A})$ stands for the spectral radius of matrix $\mathbf{A}$. x_l stands for the l-th element of $\mathbf{x} \in \mathbb{C}^{N\times 1}$. $[\mathbf{A}]_{k,l}$ denotes the (k,l)-th element of the matrix $\mathbf{A}$. $J_0(\cdot)$ is the Bessel function of zero kind and is defined as $J_0(x) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-jx\sin(t)}\mathrm{d}t$. The inner product between two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as $\langle \mathbf{A},\mathbf{B}\rangle \triangleq \mathrm{trace}(\mathbf{A}^{\mathrm{H}}\mathbf{B})$. For a vector $\mathbf{x} \in \mathbb{C}^{N\times 1}$, the ℓ_2 norm is defined as $\|\mathbf{x}\|_2 \triangleq \sqrt{\sum_{l=1}^{N}|x_l|^2}$. The Kronecker product between two vectors $\mathbf{x} \in \mathbb{C}^{N\times 1}$ and $\mathbf{y} \in \mathbb{C}^{M\times 1}$ is shown by $\mathbf{x}\otimes\mathbf{y} \in \mathbb{C}^{MN\times 1}$. $\mathbb{E}_{x,y}$ denotes the expectation operator over the joint distribution of random variables x and y. To simplify notation, throughout the paper, we tag User-1 as the intended user, and will sometimes drop index $k=1$ when referring to the tagged user.

II. System Model

A. Channel Model

Consider a time $t = nT$, where $n \in \mathbb{N}^+$ represented as an integer multiple of the symbol duration T. The wireless communication link between a user with a single antenna and a BS equipped with N_r antennas at time slot t, represented by $\mathbf{h}(t) \in \mathbb{C}^{N_r \times 1}$, is typically characterized as a non-stationary complex stochastic process. Specifically, the mean ($\overline{\mathbf{h}}(t)$) and the covariance matrix ($\mathbf{C}_{\mathbf{h}}(t)$) of $\mathbf{h}(t)$ change with time t, and the cross-covariance matrix $\mathbf{C}_{\mathbf{h}}(t_1, t_2)$ between time slots t_1 and t_2 depends not only on the time lag (i.e., $t_2 - t_1$) but also on the reference time t_1. In the following proposition, we state how the time-varying channel at time $t + T$ can be modeled as an AR process exploiting the temporal correlations with previous time t.

Proposition 1. *Let $t = nT$ be any time slot and $\mathbf{h}(t) \in \mathbb{C}^{N_r \times 1}$ be a random channel at time t. The channel at time $t+T$ can be represented by the channel at time t as follows:*

$$\widetilde{\mathbf{h}}(t+T) = \underbrace{\mathbf{A}(t)\widetilde{\mathbf{h}}(t)}_{\text{Correlation information}} + \underbrace{\boldsymbol{\xi}(t+T)}_{\text{Innovative information}}, \quad (1)$$

where the state transition matrix $\mathbf{A}(t)$ is defined as

$$\mathbf{A}(t) \triangleq \mathbf{C}_{\mathbf{h}}(t+T, t)\mathbf{C}_{\mathbf{h}}^{-1}(t), \quad (2)$$

and the AR noise $\boldsymbol{\xi}(t+T)$ is distributed as $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Theta}(t+T))$ with

$$\boldsymbol{\Theta}(t+T) = \mathbf{C}_{\mathbf{h}}(t+T) - \mathbf{C}_{\mathbf{h}}(t+T, t)\mathbf{C}_{\mathbf{h}}^{-1}(t)\mathbf{C}_{\mathbf{h}}(t, t+T). \quad (3)$$

Proof. See Appendix A.

Remark 1. (Channel modeling intuition) As shown in Figure 1, the channel at time $t+T$ in (1) consists of two terms, which are independent from each other. The first term specifies the temporal correlation of the channel at time $t+T$ with that of time t, while the second term provides the innovative channel information of the current time. However, the covariance of the second term is related to the temporal correlation, as it specifies the error caused by representing the channel vector at time $t+T$, based on the channel at time t. For a given integer n, the AR model in (1) implies that the spectral radius of the correlation matrix between $t+nT$ and t, that is the cross-covariance of the normalized centered channel, and is denoted by

$$\mathbf{P}_{\mathbf{h}}(t+nT, t) \triangleq \mathbb{E}[\mathbf{h}'(t+nT)\mathbf{h}'^{\mathrm{H}}(t)] = \mathbf{C}_{\mathbf{h}}^{-\frac{1}{2}}(t+nT)\mathbf{C}_{\mathbf{h}}(t+nT, t)\mathbf{C}_{\mathbf{h}}^{-\frac{1}{2}}(t), \quad (4)$$

must be less than one. This makes sure that the correlation between the channel values decreases as they become further apart in time. This makes sense, because as time progresses, environmental conditions, user mobility, and other factors introduce variations in the channel, leading to decreasing correlation between channel observations made at different time instances. It is worth noting that the lag-n correlation matrix between $t+nT$ and t can be derived from the lag-1 correlation matrices between $t+T$ and t. Specifically, the lag-1 correlation matrix between $t+T$ and t allows us to express the n-step correlation matrix as follows:

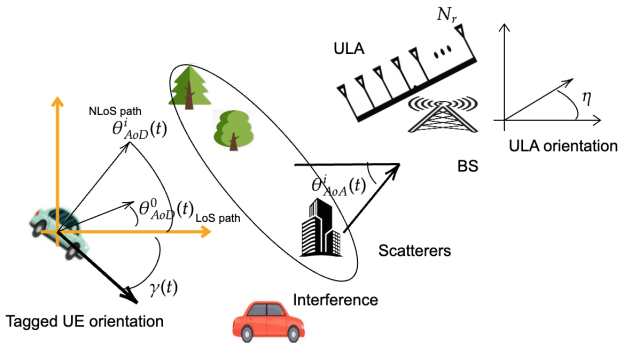


Figure 2. A typical example of Rician time-varying channel in vehicular networks.

$$\mathbf{P}_{\mathbf{h}}(t+nT, t) = \prod_{i=t}^{t+(n-1)T} \mathbf{P}_{\mathbf{h}}(i+T, i). \quad (5)$$

This relation holds true only when the AR model (1) is applicable, which is generally satisfied when the channel follows a Gaussian distribution and the symbol duration T is sufficiently small.

B. Correlation Matrix in Time-Varying Environments

The correlation matrix of the channel is typically influenced by several factors, including the propagation geometry, user's velocity, and antenna characteristics. Consider a tagged vehicle moving towards direction $\gamma(t)$ at time t with the speed $\nu(t)$ as shown in Figure 2.

Suppose that the BS is equipped with an uniform linear array (ULA) of N_r number of antennas that is aligned in direction η (see Figure 2). The distance between antenna elements is denoted by d. The environment between the tagged user equipment (UE) and the BS is composed of $L(t)$ number of scatterers and the Doppler frequencies of the line of sight (LoS) and non-line of sight (NLoS) paths are obtained by

$$f_d^0(t) = \frac{\nu(t)\cos(\gamma(t) - \theta^0_{\mathrm{AoD}}(t))}{\lambda}, f_d^i(t) = \frac{\nu(t)\cos(\gamma(t) - \theta^i_{\mathrm{AoD}}(t))}{\lambda}, \quad (6)$$

where $i = 1, ..., L(t)$, $\lambda \triangleq \frac{c}{f_c}$ is the wavelength of the source, $c = 3 \times 10^8$ m/s, f_c is the carrier frequency, $\theta^0_{\mathrm{AoD}}(t))$ is the Angle of Departure (AoD) for the LoS angle and $\theta^i_{\mathrm{AoD}}(t)), i = 1, ..., L(t)$ are the AoDs for NLoS paths. For this case, we state the following proposition, which will be useful in the sequel.

Proposition 2. *Assume that the number of scatterers is sufficiently large ($L(t) \to \infty$) and the AoD and Angle of Arrivals (AoAs) are distributed according to some known distributions. Then, the correlation matrix $\mathbf{P}_{\mathbf{h}}(t_1, t_2)$ can be obtained as*

$$\mathbf{P}_{\mathbf{h}}(t_1, t_2) = \rho_{\mathrm{temporal}}(t_1, t_2)\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_1, t_1)\mathbf{P}_{\mathrm{spatial}}(t_1, t_2)\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_2, t_2) \quad (7)$$

where

$$\rho_{\mathrm{temporal}}(t_1, t_2) \triangleq \mathbb{E}_{\theta_{\mathrm{AoD}}(t_1)), \theta_{\mathrm{AoD}}(t_2))}\left[e^{j\frac{2\pi}{\lambda}\left[t_1\nu(t_1)\cos(\gamma(t_1) - \theta_{\mathrm{AoD}}(t_1)) - t_2\nu(t_2)\cos(\gamma(t_2) - \theta_{\mathrm{AoD}}(t_2))\right]}\right]$$
$$\rho_{\mathrm{spatial}}(t_1, t_2, k, l) \triangleq [\mathbf{P}_{\mathrm{spatial}}(t_1, t_2)]_{k,l} =$$
$$\mathbb{E}_{\theta_{\mathrm{AoA}}(t_1)), \theta_{\mathrm{AoA}}(t_2))}\left[e^{j\frac{2\pi d}{\lambda}\left[(k-1)\cos(\eta - \theta_{\mathrm{AoA}}(t_1)) - (l-1)\cos(\eta - \theta_{\mathrm{AoA}}(t_2))\right]}\right] \quad (8)$$

Proof. See Appendix B.

This model aligns with the widely recognized U-shaped band-limited power spectral density, resulting from the assumption that the propagation environment is two-dimensional (2-D) with non-isotropic scattering (more detailed can be found in [26], [36], [37]). If the distributions of AoAs and AoD are known in advance, then the correlation matrix can be fully calculated by numerical integration. It is crucial to emphasize that having a predictive understanding of the evolving behavior of the channel covariance over time, as explored in studies such as [28], [34], [38], [39], can significantly benefit our approach. Utilizing the direct relationship provided by Equation (4), we can access the time-varying cross-covariances. These cross-covariances play a critical role in our channel estimation procedure.

Remark 2. (Special cases: stationary environments) Define $\tau \triangleq t_2 - t_1$. If the tagged user and the environment are stationary, i.e., the parameters $\nu(t) = \nu, \gamma(t) = \gamma, \theta_{\mathrm{AoA}}(t) = \theta_{\mathrm{AoA}}, \theta_{\mathrm{AoD}}(t) = \theta_{\mathrm{AoD}}$ do not depend on the time and the AoDs follow non-uniform von Mises distribution with probability density function (PDF)

$$f(\theta_{\mathrm{AoD}}) = \frac{\mathrm{e}^{\kappa_{\mathrm{AoD}} \cos(\theta_{\mathrm{AoD}} - \theta^c_{\mathrm{AoD}})}}{2\pi J_0(j\kappa_{\mathrm{AoD}})}, \tag{9}$$

then, we have a simple closed-form formula for the correlation matrix as follows:

$$\mathbf{P_h}(\tau) = \frac{J_0\left(j\sqrt{\kappa^2_{AoD} - \frac{4\pi^2}{\lambda^2}\tau^2\nu^2 - j\frac{4\pi}{\lambda}\tau\nu\kappa_{AoD}\cos(\gamma - \theta^c_{AoD})}\right)}{J_0(j\kappa_{AoD})}\mathbf{I}_{N_r} \tag{10}$$

Here, θ^c_{AoD} denotes the central AoD and κ_{AoD} measures the extent of concentration around this central value. When $\kappa_{\mathrm{AoD}} = 0$, we have uniform distribution for AoD which leads to

$$\mathbf{P_h}(\tau) = J_0\left(-\tfrac{2\pi}{\lambda}\tau\nu\right)\mathbf{I}_{N_r}, \tag{11}$$

and aligns with the well-known Jakes model in [40].

Remark 3. (AR model limitation) There are two ways to obtain the correlation matrix between $t + nT$ and t where T is the symbol duration. The first method relies on relation (5) and the assumption that the centered channel follows the AR model (1). The second way is to directly use (7). As an example, consider $\kappa_{\mathrm{AoD}} = 50, T = 0.001$, $\nu = 1m/s$, $f_c = 1000Hz$, $N_r = 2$, $n = 2$. The correlation matrix obtained from one-step correlations is

$$\mathbf{P_h}(t + 2T, t) = \mathbf{P_h}(t + 2T, t + T)\mathbf{P_h}(t + T, t) = (0.9843 - 0.1243j)\mathbf{I}_2 \tag{12}$$

and the correlation matrix given by (11) is

$$\mathbf{P_h}(t + 2T, t) = (0.9771 - 0.1215j)\mathbf{I}_2. \tag{13}$$

The above example shows that for a non-uniform distribution with a high value of κ_{AoD} (indicating low variance), the correlation matrix of the AR model can be approximately described by (10). However, in the case of a uniform AoD distribution, this approximation is valid only for very small symbol duration T.

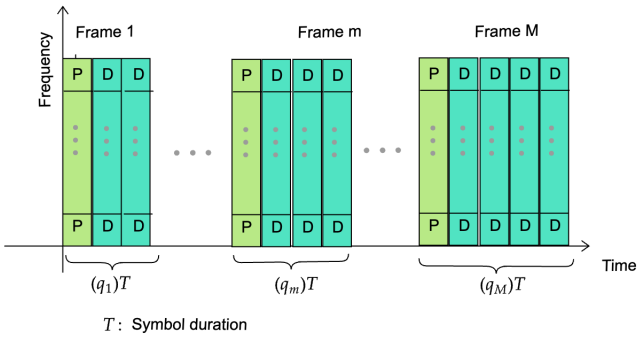


Figure 3. A schematic figure for multi-frame data transmission. $q_1, ..., q_M$ specifies the length of frames $1, ..., M$. Each time slot contains F symbols in frequency domain. The first time slot of each frame is considered for pilot transmission (shown by P) and the rest is for data transmission (shown by D).

Remark 4. (Availability of channel correlation and covariance matrices in non-stationary environments)

Predicting the correlation matrix outlined in (4) becomes feasible when the future velocities of users and the distribution of antennas at the BS are known. In stationary scenarios, this prediction relies on the velocities of users and antenna spacing, particularly in the case of ULA. Various methods exist to track and predict the variations of the channel covariance matrices. Kalman filtering [41] and machine learning techniques [42] offer real-time predictions of forthcoming changes in non-stationary environments. Kalman filtering is particularly advantageous when the underlying dynamics of the system is well-understood or can be accurately modeled [41]. For instance, in [41], a Kalman filtering approach is employed to predict and track the time-varying channel covariance matrix in specific environments. Alternatively, neural networks can be trained on historical data and relevant features to predict future channel covariance matrices. By learning the complex relationships between input features and covariance matrix variations, machine learning models, as demonstrated in [42], can deliver precise predictions of the covariance matrices in non-stationary environments.

C. Uplink Pilot Signal Model

We consider M frames $m = 1, ..., M$ in which frame m consists of q_m time slots. The frame sizes are represented by vector $\mathbf{q} \triangleq [q_1, \ldots, q_M]^{\mathrm{T}}$. The first time slot of each frame is devoted to pilot transmission and the rest is employed to transmit data as shown in Figure 3. Within frame m out of the total of M frames, each user sends τ_p pilot symbols in the first time slot, followed by $q_m - 1$ data time slots, each of which contains F data symbols according to Figure 3. From this point on, each symbol is transmitted within a duration of T which is assumed to be one without loss of generality. By defining $\delta_m \triangleq \sum_{l=1}^{m} q_l + 1$, the total duration is $\delta_M - 1 = \sum_{l=1}^{M} q_l$. The times slots of frame m are from δ_{m-1} to $\delta_m - 1$. We consider a scenario involving K single-antenna UEs and a BS equipped with an ULA consisting of N_r antennas. User-k transmits each of the F pilot symbols with transmit power $P_{\mathrm{p},k}$, and each data symbol in slot-i with transmit power $P_{\mathrm{d},k}$, $k = 1 \ldots K$. Assuming that the coherence bandwidth accommodates at least τ_p pilot symbols in the frequency domain as follows:

$$\mathbf{s} \triangleq [s(1), ..., s(\tau_{\rm p})]^{\rm T} \in \mathbb{C}^{\tau_{\rm p}\times 1}, \tag{14}$$

in which $\|\mathbf{s}\|_2^2 = \tau_p$. By assuming that the pilot sequences of users are orthogonal to each other, the received signal from the tagged user at pilot time slot, takes the following form:

$$\mathbf{Y}_p(i) = \alpha\sqrt{P_{\rm p}}\mathbf{h}(i)\mathbf{s}^{\rm T} + \mathbf{N}_{\rm p}(i) \quad \in \mathbb{C}^{N_r\times\tau_{\rm p}}, \tag{15}$$

where $\mathbf{h}(i) \in \mathbb{C}^{N_r\times 1}$ is the channel vector at time slot i with mean $\overline{\mathbf{h}}(i)$ and covariance matrix $\mathbf{C}_{\mathbf{h}}(i) \in \mathbb{C}^{N_r\times N_r}$. Furthermore, α_k denotes the large scale fading of k-user where user $k=1$ is the tagged one, $P_{\rm p}$ denotes the pilot power of the tagged user at each pilot time slot, and $\mathbf{N}_{\rm p}(i) \in \mathbb{C}^{N_r\times\tau_{\rm p}}$ is the additive white Gaussian noise (AWGN) with element-wise variance $\sigma_{\rm p}^2$. For notational convenience, we denote the large scale fading of the tagged user by $\alpha \triangleq \alpha_1$. It is beneficial to write equation (15) in a matrix-vector form as follows:

$$\mathbf{y}_p(i) \triangleq \mathbf{vec}\left(\mathbf{Y}_p(i)\right) = \alpha\sqrt{P_{\rm p}}\overline{\mathbf{S}}\mathbf{h}(i) + \mathbf{n}_{\rm p}(i) \ \in \mathbb{C}^{\tau_{\rm p}N_r\times 1}, \tag{16}$$

where $\mathbf{y}_p(i)$, $\mathbf{n}_{\rm p}(i) \triangleq \text{vec}(\mathbf{N}_p(i)) \in \mathbb{C}^{\tau_{\rm p}N_r\times 1}$ and $\overline{\mathbf{S}} \triangleq \mathbf{I}_{N_r} \otimes \mathbf{s} \in \mathbb{C}^{\tau_{\rm p}N_r\times N_r}$ is such that $\overline{\mathbf{S}}^{\rm H}\overline{\mathbf{S}} = \tau_{\rm p}\mathbf{I}_{N_r}$. We also assume that the pilot power corresponding to each pilot time slot is obtained as $P_{\rm p} = \frac{P_{\rm p_{max}}}{M}$, where $P_{\rm p_{max}}$ denotes the maximum (total) pilot power of the tagged user.

D. Data Signal Model

At data time slot i of frame m, the received signal at the BS can be stated as follows:

$$\mathbf{y}(i) = \underbrace{\alpha\mathbf{h}(i)\sqrt{P_{\rm d}}x(i)}_{\text{tagged user}} + \underbrace{\sum_{k=2}^{K}\alpha_k\mathbf{h}_k(i)\sqrt{P_{{\rm d},k}}x_k(i)}_{\text{co-scheduled MU-MIMO users}} + \mathbf{n}_d(i) \in \mathbb{C}^{N_r\times 1}, \tag{17}$$

where $\mathbf{y}(i) \in \mathbb{C}^{N_r\times 1}$; and $x_k(i) \in \mathbb{C}^{1\times 1}$ denotes the transmitted data symbol of user k at time slot i of m-th frame with transmit data power $P_{{\rm d},k}$. Here, the data power of each data time slot is obtained as $P_{{\rm d},k} = \frac{P_{{\rm d},k_{\max}}}{\delta_M - 1 - M}$, where $P_{{\rm d},k_{\max}}$ denotes the maximum data power of user k. Furthermore, $\mathbf{n}_d(i) \sim \mathcal{CN}\left(\mathbf{0}, \sigma_{\rm d}^2\mathbf{I}_{N_r}\right)$ is the AWGN at the receiver. The sum of the maximum pilot and data powers should not exceed the total power budget, which is denoted by $P_{\rm tot}$.

III. Proposed Scheme

In this section, based on the temporal correlation information obtained in Proposition 2, we first characterize the linear MMSE (LMMSE) estimate for the time-varying Rician aging channel as a function of frame size, number of frames and pilot and data powers in Section III-A. Next, we calculate the covariance matrix of the channel estimate as a function of the aforementioned parameters, which provides the required part for the subsequent sections. In Section III-B, based on the channel estimates in III-A, we provide an optimal MMSE receiver combiner in order to estimate the data message of the tagged user exploiting the AR channel information in previous times. Next, in Section III-C, based on the channel estimate and the data estimate, we calculate the instantaneous slot-by-slot SINR of the tagged user over fast fading Rician non-stationary channels in Theorem 1. The instantaneous SINR depends on the random channel estimates obtained in Section III-A, and is a random variable. Next, we state Theorem 2 in this section that provides a deterministic expression for the instantaneous SINR, named deterministic equivalent SINR. This deterministic equivalent SINR relies solely on temporal correlation details of the Rician channel, independent of channel estimates or measurements. It forms the basis for our analysis in Section IV.

A. Linear MMSE Channel Estimation

We assume that in each data slot i of frame m, the BS utilizes the pilot signals of the current, previous and next frames to estimate the channel in data time slots of the current frame. It's important to highlight that, in the initial frame, we solely take into account one pilot before and two pilots after. Conversely, in the last frame, our consideration includes two pilots before. According to (15), the observed measurements at the pilot time slots corresponding to frames $m-1$, m and $m+1$ are given by:

$$\begin{aligned}
\mathbf{Y}\left(\delta_{m-2}\right) &= \alpha_1\sqrt{P_p}\mathbf{h}\left(\delta_{m-2}\right)\mathbf{s}^{\rm T} + \mathbf{N}\left(\delta_{m-2}\right),\\
\mathbf{Y}\left(\delta_{m-1}\right) &= \alpha_1\sqrt{P_p}\mathbf{h}\left(\delta_{m-1}\right)\mathbf{s}^{\rm T} + \mathbf{N}\left(\delta_{m-1}\right),\\
\mathbf{Y}\left(\delta_{m}\right) &= \alpha_1\sqrt{P_p}\mathbf{h}\left(\delta_{m}\right)\mathbf{s}^{\rm T} + \mathbf{N}\left(\delta_{m}\right).
\end{aligned} \tag{18}$$

By stacking the above relations in matrix-vector form, we have that

$$\mathbf{y}_{\rm p} \triangleq \begin{bmatrix}\mathbf{vec}(\mathbf{Y}(\delta_{m-2}))\\ \mathbf{vec}(\mathbf{Y}(\delta_{m-1}))\\ \mathbf{vec}(\mathbf{Y}(\delta_m))\end{bmatrix} = \quad \widetilde{\mathbf{S}}\mathbf{h}_p + \boldsymbol{\epsilon}_{\rm p} \in \mathbb{C}^{3N_r\tau_{\rm p}\times 1}, \tag{19}$$

where $\mathbf{h}_p = \begin{bmatrix}\mathbf{h}(\delta_{m-2})\\ \mathbf{h}(\delta_{m-1})\\ \mathbf{h}(\delta_m)\end{bmatrix} \in \mathbb{C}^{3N_r\times 1}$, $\widetilde{\mathbf{S}} = \mathbf{I}_{3N_r} \otimes \mathbf{s}$ and $\boldsymbol{\epsilon}_{\rm p} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\rm p}^2\mathbf{I}_{3N_r\tau_{\rm p}})$. Note that as a convention, we assumed that $q_l = 0 \ \forall l \leqslant 0$.

Having the measurements in pilot time slots at frames $m-1$, m, $m+1$, we can estimate the channel at data time slot i in frame m as is stated in the following lemma.

Lemma 1. *The LMMSE channel estimate at data time slot i of the m-th frame, which approximates the AR fast fading channel in time slot i, based on the received measurements at pilot time slots of the previous and next frames is given by:*

$$\widehat{\mathbf{h}}(\mathbf{q}, i) = \frac{1}{\alpha_1\sqrt{P_{\rm p}}\tau_{\rm p}}\mathbf{E}_m(\mathbf{q}, i)\left(\mathbf{M}_m(\mathbf{q}) + \frac{\sigma_{\rm p}^2}{\alpha_1^2 P_{\rm p}\tau_{\rm p}}\mathbf{I}_{3N_r}\right)^{-1}\widetilde{\mathbf{S}}^{\rm H}\widetilde{\mathbf{y}}_{\rm p} + \overline{\mathbf{h}}(i), \tag{20}$$

where, for $\forall i = 1, ..., \delta_M - 1, \forall m = 1, .., M$, $\mathbf{E}_m(\mathbf{q}, i)$ and $\mathbf{M}_m(\mathbf{q})$ are

$$\mathbf{E}_m(\mathbf{q}, i) \triangleq \Big[\mathbf{C}_{\mathbf{h}}(\delta_{m-2}, i) \quad \mathbf{C}_{\mathbf{h}}(\delta_{m-1}, i) \quad \mathbf{C}_{\mathbf{h}}(\delta_m, i)\Big]; \tag{21}$$

$$\mathbf{M}_m(\mathbf{q}) \triangleq \tag{22}$$

$$\begin{bmatrix} \mathbf{C}_{\mathbf{h}}(\delta_{m-2}) & \mathbf{C}_{\mathbf{h}}(\delta_{m-2},\delta_{m-1}) & \mathbf{C}_{\mathbf{h}}(\delta_{m-2},\delta_m) \\ \mathbf{C}_{\mathbf{h}}(\delta_{m-1},\delta_{m-2}) & \mathbf{C}_{\mathbf{h}}(\delta_{m-1}) & \mathbf{C}_{\mathbf{h}}(\delta_{m-1},\delta_m) \\ \mathbf{C}_{\mathbf{h}}(\delta_m,\delta_{m-2}) & \mathbf{C}_{\mathbf{h}}(\delta_m,\delta_{m-1}) & \mathbf{C}_{\mathbf{h}}(\delta_m) \end{bmatrix},$$

and $\widetilde{\mathbf{y}}_{\mathrm{p}} \triangleq \mathbf{y}_{\mathrm{p}} - \alpha_1\sqrt{P}_p\widetilde{\mathbf{S}}(\overline{\mathbf{h}}(i)\otimes\mathbf{1}_3)$.

Proof. See Appendix C.

Remark 5. (Special cases) When all frames have the same size, while lacking a LoS component in the channels from the users to the BS, the estimated channel in (20) aligns with the findings obtained in [23, Eq. 10].

Remark 6. (Optimality) If the channel follows Gaussian distribution, the LMMSE estimate (20) coincides with the MMSE estimate and is optimal in terms of minimizing mean squared error (MSE). Alternatively, when the channel exhibits sparse representations in the angular domain, compressed sensing methods can be employed. These methods exploit the temporal prior information to achieve optimal channel estimation (see e.g., [43]–[45] for more details).

According to Lemma 1, the LMMSE estimate of the channel can also be stated as:

$$\widehat{\mathbf{h}}(\mathbf{q},i) = \mathbf{E}_m(\mathbf{q},i) \left(\mathbf{M}_m(\mathbf{q}) + \tfrac{\sigma_{\mathrm{p}}^2}{\alpha^2 P_{\mathrm{p}}\tau_{\mathrm{p}}}\mathbf{I}_{3N_r}\right)^{-1}\left(\mathbf{h}_p - \overline{\mathbf{h}}(i)\otimes\mathbf{1}_3 + \boldsymbol{\epsilon}\right) + \overline{\mathbf{h}}(i), \tag{23}$$

where $\boldsymbol{\epsilon} \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma_{\mathrm{p}}^2}{\alpha^2\tau_{\mathrm{p}}P_{\mathrm{p}}}\mathbf{I}_{3N_r})$

Corollary 1. The covariance matrix of the estimated channel $\widehat{\mathbf{h}}(\mathbf{q},i)$ in Lemma 1 is given by:

$$\mathbf{C}_{\widehat{\mathbf{h}}}(i) = \mathbf{E}_m(\mathbf{q},i)\left(\mathbf{M}_m(\mathbf{q}) + \tfrac{\sigma_{\mathrm{p}}^2}{\tau_{\mathrm{p}}\alpha^2 P_{\mathrm{p}}}\mathbf{I}_{3N_r}\right)^{-1}\mathbf{E}_m^{\mathrm{H}}(\mathbf{q},i). \tag{24}$$

In what follows, we use the latter channel covariance matrix provided in (24) to calculate the optimum MMSE receiver at the BS.

B. Optimal MMSE Receiver for AR Aging Channels

In this section, we aim to calculate the optimum receiver $\mathbf{g} \in \mathbb{C}^{1\times N_r}$ that estimates the transmitted data symbol of the tagged user, which is denoted by x. Without loss of generality, we assume that the transmitted symbol of the tagged user has zero mean with unit variance. Specifically, the BS estimates the transmitted symbol of the tagged user in slot i based on two prior information with different accuracy levels provided at i and i_{p}. Note that $i_{\mathrm{p}} < i$ can be any arbitrary time slot. Both of these time slots can provide beneficial and different information about the real channel at time i. In our MMSE receiver, we will take into account both of them. The following vector collects the prior information at time instances i and i_{p}:

$$\boldsymbol{\zeta}(i) \triangleq \begin{bmatrix} \widehat{\mathbf{h}}(\mathbf{q},i) \\ \widehat{\mathbf{h}}(\mathbf{q},i_{\mathrm{p}}) \end{bmatrix} \in \mathbb{C}^{2N_r\times 1}. \tag{25}$$

Then, the optimum receiver is expressed as:

$$\mathbf{g}^{\star}(\mathbf{q},i) \triangleq \underset{\mathbf{g}\in\mathbb{C}^{1\times N_r}}{\arg\min}\, \mathbb{E}_{x|\boldsymbol{\zeta}(i)}[|\mathbf{g}\mathbf{y} - x|]. \tag{26}$$

The solution of the latter optimization problem is given by:

$$\mathbf{g}^{\star}(\mathbf{q},i) = \alpha_1\sqrt{P_{\mathrm{d}}}\mathbf{z}_1^{\mathrm{H}}\mathbf{F}(i)^{-1}, \tag{27}$$

where

$$\mathbf{F}(i) \triangleq \sum_{k=1}^{K}\alpha_k^2 P_{\mathrm{d},k}\mathbf{D}_k(i) + \sigma_{\mathrm{d}}^2\mathbf{I}_{N_r}, \mathbf{D}_k(i) \triangleq \mathbf{Q}_k(i) + \mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}, \tag{28}$$

$$\mathbf{z}_k(i) \triangleq \overline{\mathbf{h}}_k(i) + \boldsymbol{\Psi}_k\widetilde{\boldsymbol{\zeta}}_k(i), \boldsymbol{\zeta}_k(i) \triangleq \begin{bmatrix} \widehat{\mathbf{h}}_k(i) \\ \widehat{\mathbf{h}}_k(i_{\mathrm{p}}) \end{bmatrix} \in \mathbb{C}^{2N_r\times 1}, \tag{29}$$

$$\widetilde{\boldsymbol{\zeta}}_k(i) \triangleq \boldsymbol{\zeta}_k(i) - \overline{\mathbf{h}}(i)\otimes\mathbf{1}_2, \tag{30}$$

$$\boldsymbol{\Psi}_k \triangleq \begin{bmatrix} \mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i)} & \mathbf{A}(i_{\mathrm{p}})\mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i_{\mathrm{p}})} \end{bmatrix}\begin{bmatrix} \mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i)} & \mathbf{A}(i_{\mathrm{p}})\mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i_{\mathrm{p}})} \\ \mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i_{\mathrm{p}})}\mathbf{A}(i_{\mathrm{p}})^H & \mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i_{\mathrm{p}})} \end{bmatrix}, \tag{31}$$

$$\mathbf{Q}_k(i) \triangleq \mathbf{C}_{\mathbf{h}_k(i)} - \boldsymbol{\Psi}_k\begin{bmatrix} \mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i)} \\ \mathbf{C}_{\widehat{\mathbf{h}}_k(\mathbf{q},i_{\mathrm{p}})}\mathbf{A}(i_{\mathrm{p}})^{\mathrm{H}} \end{bmatrix}. \tag{32}$$

This finding extends the result presented in [15, Eq. 24] to the Rician channel, incorporating a significant LoS component alongside the NLoS components. The proof follows a similar approach.

C. SINR and SE Calculations

In this section, we first calculate the instantaneous SINR based on channel and data estimates provided in Sections III-A and III-B. Then, based on the instantaneous SINR, we obtain the random SE, where its randomness comes from the stochastic nature of the channel estimates. The following theorem whose proof is in Appendix D, provides a closed-form expression for the instantaneous SINR that is calculated at the BS side.

Theorem 1. *Assume that there is some prior information about the channel at time i collected in $\boldsymbol{\zeta}(i)$. Also, the receiver employs the MMSE combiner to estimate the data symbols of the tagged user at time slot i. Then, the instantaneous SINR of the data symbol of the tagged user at time slot i is obtained as:*

$$\begin{aligned}\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i)) &= \alpha^2 P_{\mathrm{d}}\mathbf{z}_1(i)^{\mathrm{H}}\left(\mathbf{F}(i) - \alpha_1^2 P_{\mathrm{d}}\mathbf{z}_1(i)\mathbf{z}_1(i)^{\mathrm{H}}\right)^{-1}\mathbf{z}_1 \\ &= \alpha_1^2 P_{\mathrm{d}}\mathbf{z}_1(i)^{\mathrm{H}}\left(\mathbf{F}_1^{-1}(i)\right)\mathbf{z}_1(i),\end{aligned} \tag{33}$$

where

$$\mathbf{F}_1 \triangleq \mathbf{F}_1(i) \triangleq \mathbf{F}(i) - \alpha_1^2 P_{\mathrm{d}}\mathbf{z}_1(i)\mathbf{z}_1(i)^{\mathrm{H}}. \tag{34}$$

The latter result immediately leads to finding the random SE as follows:

$$\mathrm{SE}\Big(\mathbf{q},i,\boldsymbol{\zeta}(i)\Big) \triangleq \log\Big(1+\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i))\Big), \tag{35}$$

which is a random variable. In the next theorem, leveraging concentration inequality results from random matrix theory tools as detailed in [46], we present a deterministic equivalent expression for the SE. This expression serves as a reliable approximation for the average SE, when the number of BS antennas becomes large.

Theorem 2. *Define $\mathbf{S}(i) \triangleq \sum_{k=1}^{K}\alpha_k^2 P_{\mathrm{d},k}\mathbf{Q}_k(i)$ and $\rho_{\mathrm{d}} \triangleq \sigma_{\mathrm{d}}^2$. When N_r is sufficiently large ($N_r \to \infty$), the instantaneous*

SE provided in (35) *is concentrated around a deterministic expression given by:*

$$\mathrm{SE}^{\circ}(\mathbf{q},i)\triangleq \log\left(1+\left\langle \mathbf{R}_{\mathbf{z}_1}(i),\left(\sum_{k=2}^{K}\frac{\alpha_k^2 P_{\mathrm{d},k}\mathbf{R}_{\mathbf{z}_k}(i)}{1+m_{\mathbf{B}_k}(\rho_{\mathrm{d}})}+\mathbf{S}(i)+\rho_{\mathrm{d}}\mathbf{I}_{N_r}\right)^{-1}\right\rangle\right), \tag{36}$$

where $m_{\mathbf{B}_2}(\rho_{\mathrm{d}}),\ldots,m_{\mathbf{B}_k}(\rho_{\mathrm{d}})$ *are the solution of the following system of equations for* $k=2,\ldots,K$*:*

$$m_{\mathbf{B}_k}(\rho_{\mathrm{d}})=\left\langle \mathbf{R}_{\mathbf{z}_k}(i),\left(\sum_{l=2}^{K}\frac{\mathbf{R}_{\mathbf{z}_l}(i)}{1+m_{\mathbf{B}_l}(\rho_{\mathrm{d}})}+\mathbf{S}(i)+\rho_{\mathrm{d}}\mathbf{I}_{N_r}\right)^{-1}\right\rangle \tag{37}$$

Proof. See Appendix E.

We call the deterministic expression (36) the deterministic equivalent SE. In the next section, we exploit this useful expression in order to propose optimal frame design and power control.

IV. Proposed Strategy for Frame and Power Design

In this section, we formulate an optimization problem to design the optimal values for the number of frames (denoted by $M^{\star}$), the frame size (denoted by $\mathbf{q}^{\star}$), maximum pilot power ($P_{\mathrm{p}\max}^{\star}$) and maximum data power $P_{\mathrm{d}\max}^{\star}$ under the condition that we have a limited total power budget (denoted by P_{tot}).

The objective function that we intend to maximize is averaged spectral efficiency (ASE), which is obtained by taking the average of $\mathrm{SE}^{\circ}(\mathbf{q},i)$ over all data time slots of all frames defined as

$$\mathrm{ASE}\triangleq\frac{\sum_{l=1}^{\delta_M-1}\mathrm{SE}^{\circ}(\mathbf{q},l)}{\delta_M-1},\quad \delta_M>1. \tag{38}$$

The proposed optimization problem that finds the optimal values of M, q_m, $P_{\mathrm{p}\max}$ and $P_{\mathrm{d}\max}$ is provided below:

$$\{\mathbf{q}^{\star},M_{\max}^{\star},P_{\mathrm{p}\max}^{\star},P_{\mathrm{d}\max}^{\star}\}=\arg\max_{\mathbf{q},M,P_{\mathrm{p}\max},P_{\mathrm{d}\max}}\mathrm{ASE}, \tag{39}$$

$$\text{s.t.}\quad P_{\mathrm{p}\max}+P_{\mathrm{d}\max}\leqslant P_{\mathrm{tot}},$$

where $P_{\mathrm{d}\max}=P_{\mathrm{d},1_{\max}}$ and $P_{\mathrm{p}\max}=P_{\mathrm{p},1_{\max}}$ denote the maximum data and pilot power of the tagged user, respectively.

Remark 7. (Critical factors in the proposed optimization problem (39)) In our numerical experiments (see Figure 8), we observe that while the interference components– such as interference pilot and data power of other users and interference channels including Doppler frequency and path loss of other users– affect the resulting ASE of the tagged user, they do not influence the determination the optimal frame design (i.e., $\mathbf{q}^{\star}$ and $M^{\star}$). This characteristic allows the application of our method on the transmitter side (user side) rather than on the receiver side (BS side). Another notable observation is that the optimal frame size and pilot spacing are governed by the channel correlation matrix, rather than the channel covariance matrices at individual time instances. This indicates that users, knowing their velocities, can determine the optimal pilot spacing and frame design without explicit dependence on the channel covariance matrices. However, the optimal data and pilot powers are indeed influenced by the channel covariance matrices. As explained in Remark 4, by predicting these matrices, users can subsequently determine the optimal pilot and data power levels using our proposed strategy.

A. A Heuristic Algorithm for Optimal Frame and Power Design

The optimization problem (39) is a mixed-integer nonlinear problem that is in general non-polynomial time (NP)-hard. However, in this section, we provide a heuristic algorithm named OptResource that calculates the optimal values of frame size, number of frames and pilot and data powers. The pseudo code of this algorithm is provided in Algorithm 1. For each number of frames and for each frame size, a projected gradient ascent is performed to find the optimal values of pilot and data powers (Lines 10 to 16). The projected gradient ascent consists of two steps: update step and projection. In the update step in Line 14, a regular gradient ascent is performed. The resulting variables after this update step might not be inside the feasible region (the sum of pilot and power must be less that the total power budget). The updated variables of power is then projected in Line 16 to the closest point in the feasible region. The algorithm scans through all possible values of M and q_m that lead to the maximum spectral efficiency. The final outputs of OptResource is $\mathbf{q}^{\star},M^{\star},P_{\mathrm{p}\max}^{\star},P_{\mathrm{d}\max}^{\star}$.

Remark 8. (Computational complexity) Algorithm 1 involves several computationally intensive steps. The algorithm performs a projected gradient ascent to determine the optimal pilot and data powers, while simultaneously searching through all time slots to identify the optimal number of frames and slots. The computational complexity of the projected gradient ascent depends on the dimensionality of the optimization vector, which, in our case, is two. Additionally, computing the ASE (Equation 38) at each time slot requires implementing a fixed-point algorithm with a computational complexity of $\mathcal{O}(K)$. Furthermore, obtaining the autocorrelation matrices $\mathbf{R}_{\mathbf{z}_l}(i)$ at time slot i and for each $l=1,...,K$ involves the solution of Equation 24, which requires $\mathcal{O}(N_r^3)$ computations. In summary, the overall computational complexity is $\mathcal{O}(q_{\max}KN_r^3)$.

Remark 9. (Convergence analysis) The convergence behavior of Algorithm 1 is directly tied to the convergence properties of computing the deterministic spectral efficiency in Equation (36). A detailed convergence analysis is presented in the following proposition, which is adapted from [47, Appendix I.B] and Vitali's convergence theorem [48]:

Proposition 3. *Let* $m_{\mathbf{B}_l}^{(p)}(\rho_d), l=2,...,K$ *be the p-th iteration of the following fixed-point equation*

$$m_{\mathbf{B}_k}^{(p)}(\rho_{\mathrm{d}})=\left\langle \mathbf{R}_{\mathbf{z}_k}(i),\left(\sum_{l=2}^{K}\frac{\mathbf{R}_{\mathbf{z}_l}(i)}{1+m_{\mathbf{B}_l}^{(p-1)}(\rho_{\mathrm{d}})}+\mathbf{S}(i)+\rho_{\mathrm{d}}\mathbf{I}_{N_r}\right)^{-1}\right\rangle \tag{40}$$

with the initial point $m^{(0)}_{\mathbf{B}_k}(\rho_{\rm d}) = \frac{1}{\rho_{\rm d}}$. *Then, it holds that*

$$\left|m^{(p+1)}_{\mathbf{B}_l}(\rho_d) - m^{(p)}_{\mathbf{B}_l}(\rho_d)\right| \leqslant \frac{1}{\rho_d^2} \sup_{2\leqslant l\leqslant K} \left|m^{(p)}_{\mathbf{B}_l}(\rho_d) - m^{(p-1)}_{\mathbf{B}_l}(\rho_d)\right| \tag{41}$$

and for $l = 2, ..., K$ *the iteration converges to* $\lim_{p\to\infty} m^{(p)}_{\mathbf{B}_l}(\rho_d) = m_{\mathbf{B}_l}(\rho_d)$, *defined in* (37).

Algorithm 1 Proposed algorithm for optimal frame and power design

procedure OPTRESOURCE($\mathbf{P_h}(t_1, t_2)$, Tol, $q_{\max}$, $M_{\max}$, maxiter, $P_{\rm tot}$)
  Define $\mathbf{w} \triangleq [P_{\rm p_{max}}, P_{\rm d_{max}}]^{\rm T}$,
  Choose a small constant for the steepest ascent step size $\mu_{\mathbf{w}}$.
  Objective function : $\mathrm{SE}(q_1, ..., q_M, M, \mathbf{w})$ in (38)
  Pick an initial point $\mathbf{w}^0$ for maximum pilot and data power that satisfies $\mathbf{1}^{\rm T}\mathbf{w}^0 \leqslant P_{\rm tot}$
  $M = 1, \mathrm{SE}^\star = 0$
  while $M \leqslant M_{\max}$ **do**
    for $i_1 = 1$ to$\lceil \frac{q_{\max}}{M} \rceil$ **do**
      ⋮
      for $i_M = 1$ to$\lceil \frac{q_{\max}}{M} \rceil$ **do**
        Objective function for power
        $f(\mathbf{w}) = \mathrm{SE}(i_1, ..., i_M, M, \mathbf{w})$
        $k \leftarrow 1$
        while $\|\mathbf{w}^k - \mathbf{w}^{k-1}\|_2 >$ Tol and $|f(\mathbf{w}^k) - f(\mathbf{w}^{k-1})| >$ Tol and $k <$ maxiter **do**
          Compute ascent direction by calculating $\frac{\partial \mathrm{SE}}{\partial \mathbf{w}}$ at the point $\mathbf{w}^k$
          Update according to $\widetilde{\mathbf{w}}^{k+1} \leftarrow \mathbf{w}^k + \mu_{\mathbf{w}} \frac{\partial \mathrm{SE}}{\partial \mathbf{w}}$
          Project to the closest point inside feasible region
          $\mathbf{w}^{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{x} - \widetilde{\mathbf{w}}^{k+1}\|_2$ s.t. $\mathbf{1}^{\rm T}\mathbf{x} \leqslant P_{\rm tot}$
          $k \leftarrow k + 1$
        end while
        $\mathbf{w}^\star \leftarrow \mathbf{w}^k$
        $\mathrm{SE} = \mathrm{SE}(q_1, ..., q_M, M, \mathbf{w}^\star)$
        if $\mathrm{SE} > \mathrm{SE}^\star$ **then**
          $\mathrm{SE}^\star \leftarrow SE$
          $\mathbf{q}^\star \leftarrow [i_1, , ..., i_M]^{\rm T}$
          $M^\star \leftarrow M$
        end if
      end for
      ⋮
    end for
    $M \leftarrow M + 1$
  end while
  $[P^\star_{\rm p_{max}}, P^\star_{\rm d_{max}}]^{\rm T} \leftarrow \mathbf{w}^\star$
end procedure
Outputs: $\mathbf{q}^\star, M^\star, P^\star_{\rm p_{max}}, P^\star_{\rm d_{max}}$

V. NUMERICAL RESULTS

In this section, we study the performance of our proposed method through some numerical experiments. We consider a multi-user uplink system with $K = 2$ users transmitting multiple data frames via Rician communication channels towards a BS with N_r antennas (as shown in Figure 3). These channels are composed of one Line of Sight (LOS) component and $L(t)$ Non-Line of Sight (NLOS) components as it is in Figure 2. Consequently,

$$\mathbf{h}(t) = \overline{\mathbf{h}}(t) + \widetilde{\mathbf{h}}(t) = K_f L(t) \mathrm{e}^{j(2\pi f_d^0(t)t + \beta_0)} \mathbf{a}(\theta^0_{\rm AoA}(t)) + \sum_{i=1}^{L(t)} \mathrm{e}^{j(2\pi f_d^i(t)t + \beta_i)} \mathbf{a}(\theta^i_{\rm AoA}(t)) \in \mathbb{C}^{N_r \times 1}, \tag{42}$$

where

$$\mathbf{a}(\theta) = [1, \mathrm{e}^{j\frac{2\pi d \cos(\eta-\theta)}{\lambda}}, ..., \mathrm{e}^{j\frac{2\pi d \cos(\eta-\theta)(N_r-1)}{\lambda}}]^{\rm T} \tag{43}$$

is the array response vector and β_i is the phase of the i-th NLOS path as shown in Figure 2. Here, K_f denotes the K factor of Rician distribution defined as the ratio of the LOS power to the NLOS power.

In the first experiment, we consider $q_{\max} = 12$ and $M_{\max} = 4$ and compare four cases in Table I: 1. use one frame with one pilot time slot and 11 data time slots. 2. use $M = 2$ frames with two pilot time slots at the first of each frame. 3. use $M = 3$ time slots with three pilots. 4. use $M = 4$ frames with four pilot time slots. Each of the four rows in Table I corresponds to one experiment. The first three experiments are related to the stationary settings while the rest examine the performance of our method in non-stationary scenarios. We considered $K = 2$ users and the first user is assumed to be the tagged user. The Doppler frequencies, pilot power, data power, Rician K factor, path loss and channel variances are shown respectively by $f_{di}, P_{{\rm p},i_{\max}}, P_{{\rm d},i_{\max}}, K_{f\,i}, \mathrm{PL}_i, \sigma^2_{\mathbf{h}_i}$ where $i = 1$ corresponds to the tagged user and $i = 2$ shows the interference component. The optimal values for the number M of frames, frame size $\mathbf{q}$ and their corresponding deterministic ASEs are shown by bold numbers.

The pilot and data power in this experiment are assumed to be fixed. Note that SNR is defined as $\mathrm{SNR} \triangleq 10\log(\frac{P_{\rm d}}{\sigma^2_{\rm d}}) - \mathrm{PL}_1$ where $\mathrm{PL}_1 \triangleq -20\log(\alpha_1)$ is the path loss of the channel corresponding to the tagged user. Note that the values of both the SNR and the path loss are shown in dB units. As it turns out from Table I, in the first case (first four rows), the Doppler frequency of the tagged user is small and our proposed method suggests to use the maximum data possible slots with one frame while in the third case where the tagged user is fast, it suggests to use 4 frames with optimal frame sizes $\mathbf{q}^\star = [3, 3, 3, 2]$. Also, by comparing the first and second cases, we observe that high Rician K factor leads to higher SEs. We also observe from the last case that $\mathbf{q}^\star = [5, 2]$ is optimal and it suggests that using all time slots does not necessarily lead to higher SE values in this non-stationary case. We also observe that the main key factors in determining optimal frame design is Doppler, K factor of the tagged user and other parameters do not affect the optimal number of frames and frame size.

In the second experiment, we have evaluated the accuracy of the approximation used in (38) by comparing it to both empirical spectral efficiency and the state-of-the-art method in [23]. First, we have used the empirical Monte-Carlo approximation to replace the expectation operator and obtained an empirical form of the ASE function provided in (38) as follows:

$$\frac{\sum_{l=1}^{\delta_M - 1} \mathbb{E}[\mathrm{SE}(\mathbf{q}, l, \boldsymbol{\zeta}(l))]}{\delta_M - 1} \approx \frac{\sum_{l=1}^{\delta_M - 1} \sum_{r=1}^{r_{\max}} [\mathrm{SE}(\mathbf{q}, l, \boldsymbol{\zeta}^{(r)}(l))]}{r_{\max}(\delta_M - 1)}, \quad \delta_M > 1 \tag{44}$$

where the expectation operator $\mathbb{E}[\mathrm{SE}(\mathbf{q}, l, \boldsymbol{\zeta}(l))]$ is approximated by the empirical mean $\frac{\sum_{r=1}^{r_{\max}} [\mathrm{SE}(\mathbf{q}, l, \boldsymbol{\zeta}^{(r)}(l))]}{r_{\max}}$ with $r_{\max}$ number of Monte-Carlo iterations. We also obtained the upper-bound on the ASE function in (38) based on the approach used in [23, Eq. 30]. Under different values of Doppler frequencies in Rayleigh stationary environments and with $r_{\max} = 100$ Monte-Carlo iterations, we compare the aforementioned bounds in Figure 4. This figure shows that the approximation provided in (38) is tight in different values of the Doppler frequencies.

In the third experiment we compared our optimization method provided in (39) with the state-of-the-art method [23],

which tries to maximize an upper-bound for the spectral efficiency provided in [23, Eq. 44]. By comparing the optimal frame size in Figures 5a and 5b, we observe that the two methods suggest different optimal frame sizes. For example in case of $f_{d_1} = 100$, our proposed method suggests to use $q^\star = 2$ while [23, Eq. 44] provides $q^\star = 5$. This together with Figure 4 indicates that maximizing the spectral efficiency is not equivalent with maximizing its upper-bound.

In the fourth experiment, shown in Figure 6, we examined the effects of Doppler and Rician K factor in the optimal frame design when we have one frame with $q_{\max} = 24$. In the zero K factor setting in Figure 6a, the optimal frame size is shifting to the right as the tagged user's speed decreases. It suggests to use all the power budget in initial time slots in high speed scenarios as the channel is highly dynamic. In Figure 6b, the proposed method suggests to use more time slots in good LoS conditions. The more LoS power the channel has, the more SE is achievable.

In the fifth experiment, we examined the joint effect of SNR and Doppler on the optimal frame size. It is also interesting to observe in Figure 7a that even though the Doppler is high, when the tagged channel has strong LoS condition, the proposed method suggests to use more time slots in different SNR conditions. However, Figure 7b indicates that the ASE is less compared to the lower Doppler frequency case. Figure 7c plots the ASE as a function of both the Doppler and frame size in different SNR conditions.

In the sixth experiment we investigate the effect of the parameters on the optimal frame design and power allocation. Figures 8a, 8b, 8c and 8d suggests that the path loss, the variance of the channel, the power allocation of other users and the Doppler frequency do not affect the optimal frame size for the tagged user typically. That is, while different Doppler, path loss and powers of interference and channel variance can change the resulting deterministic ASE, optimal frame size is not sensitive to these interference components. Here, $r_{p_i} = \frac{P_{\mathrm{p},i_{\max}}}{P_{\mathrm{d},i_{\max}}}$ specifies the distribution of pilot to data power for the tagged user ($i = 1$) and for the interference user ($i = 2$).

In the seventh experiment, we have examined the effect of power allocation on the spectral efficiency. In Figure 9a, we observe that the power distribution between pilot and data (captured by r_p) can affect the optimal frame size. Figure 9b shows how pilot and data power are adapted to the changes in the channel path loss of the tagged and the interference users. Figure 9c shows the results of the joint frame size and power optimization, and illustrates the effect of the interference power distribution (captured by r_{p_2}) on the optimal power allocation. We observe in Figure 9c that while r_{p_2} affects $P^\star_{\mathrm{p}_{\max}}$ and $P^\star_{\mathrm{d}_{\max}}$, it does not play significant role in determining $\mathbf{q}^\star$ and $M^\star$. Additionally, the maximum deterministic ASE occurs in the case of $M^\star = 2$, and $\mathbf{q}^\star = [3, 2]$ in the settings provided in the caption of Figure 9c.

In the last experiment, we explored the impact of inaccurate channel covariance matrices on optimal pilot spacing and power control. To simplify the analysis, we focused solely on the stationary scenario with variance $\sigma^2_{\mathbf{h}}$ for the tagged user, and assuming a fixed level of error introduced to the channel variance. Figure 10a depicts the influence of variance errors on the optimal frame size (or pilot spacing for $M = 1$).

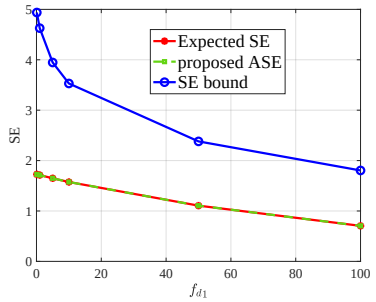
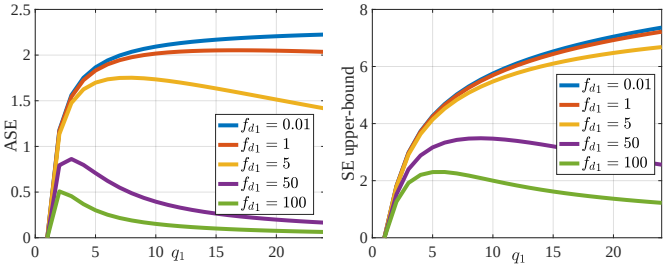


Figure 4. Comparison of the spectral efficiency approximations in (38), the upper-bound derived in [23] and the expected SE obtained from Monte-Carlo iterations. The used parameters are as follows: $N_r = 10$, $PL_1 = 20$, $\mathrm{PL}_2 = 50$, $q_{\max} = 4$, $M = 1$, SNR $= 0$dB, $f_{d_2} = 1$, $K_{f_1} = K_{f_2} = 0$.



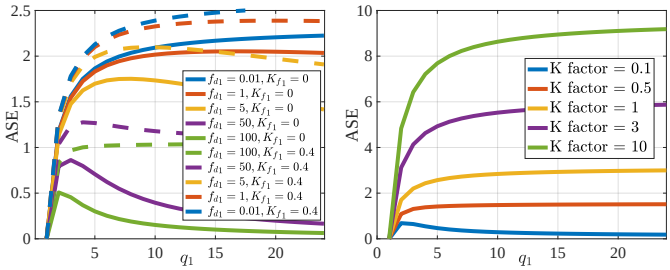
(a) (b)

Figure 5. (a) The obtained ASE in (38) is depicted versus frame size under different Doppler frequencies. (b) The upper bound of SE obtained in [23, Equation 44] is shown versus frame size. The used parameters related to these two figures are as follows: $N_r = 20, \mathbf{C}_{\mathbf{h}_1} = \mathbf{I}_{N_r}, P_{\mathrm{p}_{\max}} = P_{\mathrm{d}_{\max}} = 1, \mathrm{SNR} = 0\ dB, f_c = 1000\mathrm{Hz}, P_{\mathrm{p}_{\max}} = P_{\mathrm{d}_{\max}} = 1, \tau_p = 1, \mathrm{PL}_1 = \mathrm{PL}_2 = 0, q_{\max} = 24, M = 1, K_{f_2} = 0, K_{f_1} = 0.$

We observe that these errors do not alter the optimal frame size. However, as shown in Figure 10b, the optimal power ratio (pilot to data power), is indeed sensitive to variance error levels.

VI. Conclusions

This study investigated uplink communications systems operating over fast-fading Rician non-stationary channels that experience aging between subsequent pilot time slots. An analytical framework was proposed for determining the optimal parameters such as frame size, number of frames, and power distribution between data and pilot slots. We proposed an



(a) (b)

Figure 6. (a) ASE versus Doppler frequency and Rician K factor. (b) ASE versus Rician K factor in a fixed Doppler frequency. The parameters related to these two experiments are set as follows: $N_r = 20, \mathbf{C}_{\mathbf{h}_1} = \mathbf{I}_{N_r}, P_{\mathrm{p}_{\max}} = P_{\mathrm{d}_{\max}} = 1, \mathrm{SNR} = 0\ dB, f_c = 1000, P_{\mathrm{p}_{\max}} = P_{\mathrm{d}_{\max}} = 1, \tau_p = 1, \mathrm{PL}_1 = \mathrm{PL}_2 = 0, q_{\max} = 24, M = 1, K_{f_2} = 0$. In (b), the Doppler frequency is fixed to $f_d = 100$.

Table I. Performance of our proposed method in diverse non-stationary and stationary scenarios

$\mathbf{q}^\star$	M	SE	SNR	$P_{\mathrm{p}\max}$	$P_{\mathrm{d}\max}$	$P_{\mathrm{p}2\max}$	$P_{\mathrm{d}2\max}$	N_r	f_{d1}	f_{d2}	$K_{1\mathrm{f}}$	$K_{2\mathrm{f}}$	PL_1	PL_2	$\sigma^2_{\mathbf{h}_1}$	$\sigma^2_{\mathbf{h}_2}$
12	**1**	**13.3067**	30	1	1	1	1	20	0.1	100	1	0	1	90	1	1
[6, 6]	2	12.9756	30	1	1	1	1	20	0.1	100	1	0	1	90	1	1
[4, 4, 4]	3	11.9469	30	1	1	1	1	20	0.1	100	1	0	1	90	1	1
[3, 3, 3, 3]	4	10.7511	30	1	1	1	1	20	0.1	100	1	0	1	90	1	1
12	1	3.8722	0	.1	.1	.1	.1	20	1	10	1	0	1	90	1	1
[6, 6]	**2**	**4.3383**	0	.1	.1	.1	.1	20	1	10	0.1	0	0	90	1	1
[4, 4, 4]	3	4.0842	0	.1	.1	.1	.1	20	1	10	0.1	0	1	90	1	1
[3, 3, 3, 3]	4	3.7122	0	.1	.1	.1	.1	20	1	10	0.1	0	1	90	1	1
2	1	1.5806	30	.1	.1	.1	.1	20	100	0.01	0	1	1	90	1	1
[3, 2]	2	2.0004	30	.1	.1	.1	.1	20	100	0.01	0	1	1	90	1	1
[3, 3, 2]	3	2.1054	30	.1	.1	.1	.1	20	100	0.01	0	1	1	90	1	1
[3, 3, 3, 2]	**4**	**2.1531**	30	.1	.1	.1	.1	20	100	0.01	0	1	1	90	1	1
6	**1**	**9.1511**	50	1	1	1	1	20	$0.1t$	$0.1t$	$0.1t$	$0.1t$	50	50	$\frac{10}{t}$	$\frac{10}{t}$
[6, 4]	2	8.773	50	1	1	1	1	20	$0.1t$	$0.1t$	$0.1t$	$0.1t$	50	50	$\frac{10}{t}$	$\frac{10}{t}$
[4, 4, 4]	3	8.2585	50	1	1	1	1	20	$0.1t$	$0.1t$	$0.1t$	$0.1t$	50	50	$\frac{10}{t}$	$\frac{10}{t}$
[3, 3, 3, 3]	4	7.5917	50	1	1	1	1	20	$0.1t$	$0.1t$	$0.1t$	$0.1t$	50	50	$\frac{10}{t}$	$\frac{10}{t}$
10	**1**	**3.671**	10	10	1	10	1	20	$10(12-t)$	$10t$	$10(12-t)$	$10t$	1	1	$\frac{1}{10(12-t)}$	$\frac{1}{10t}$
[6, 6]	2	3.3593	10	10	1	10	1	20	$10(12-t)$	$10t$	$10(12-t)$	$10t$	1	1	$\frac{1}{10(12-t)}$	$\frac{1}{10t}$
[4, 4, 2]	3	3.0668	10	10	1	10	1	20	$10(12-t)$	$10t$	$10(12-t)$	$10t$	1	1	$\frac{1}{10(12-t)}$	$\frac{1}{10t}$
[3, 3, 3, 2]	4	2.7636	10	10	1	10	1	20	$10(12-t)$	$10t$	$10(12-t)$	$10t$	1	1	$\frac{1}{10(12-t)}$	$\frac{1}{10t}$
6	1	0.8886	0	1	1	1	1	20	$1(12-t)$	$10t$	0	0	50	100	0	0
[6, 6]	2	1.7832	0	1	1	1	1	20	$1(12-t)$	$10t$	0	0	50	100	0	0
[4, 4, 4]	**3**	**1.9684**	0	1	1	1	1	20	$1(12-t)$	$10t$	0	0	50	100	0	0
[3, 3, 3, 3]	4	1.9231	0	1	1	1	1	20	$1(12-t)$	$10t$	0	0	50	100	0	0
5	1	2.9595	0	0.1	0.1	0.1	0.1	20	$5t$	10	$\frac{t}{50}$	$\frac{12-t}{12}$	0	90	1	1
[5, 2]	**2**	**3.0616**	0	0.1	0.1	0.1	0.1	20	$5t$	10	$\frac{t}{50}$	$\frac{12-t}{12}$	0	90	1	1
[4, 4, 2]	3	2.9445	0	0.1	0.1	0.1	0.1	20	$5t$	10	$\frac{t}{50}$	$\frac{12-t}{12}$	0	90	1	1
[3, 3, 3, 2]	4	2.7627	0	0.1	0.1	0.1	0.1	20	$5t$	10	$\frac{t}{50}$	$\frac{12-t}{12}$	0	90	1	1

analytical framework to identify optimal parameters, including frame size, the number of frames, and power allocation between data and pilot slots. The proposed methodology revolves around optimizing a deterministic equivalent expression for spectral efficiency. Importantly, our optimization approach relied solely on the knowledge of the temporal dynamics of the channels, excluding the need for measurements or channel estimates. The core objective was to ascertain the optimal rate at which the receiver should update its CSIR within dynamic, i.e., non-stationary environments. We also demonstrated that optimal frame design can be performed at the transmitter side, while optimal pilot and power configurations can be accomplished at the receiver side. Moreover, we proposed an efficient algorithm to find the optimal parameters.

Simulation results validated the efficacy of our methodology, highlighting its impact on optimizing pilot power, data power, frame duration, and the number of frames across diverse practical scenarios. This work contributes to the advancement of uplink communication systems by providing a novel approach that leverages predictive insights into channel behavior for improved performance in dynamic and aging channel environments.

Appendix A
Proof of Proposition 1

The channel at time $t+T$ is described based on mean and covariance. According to (1), the mean of the channel at time $t+T$ is $\mathbb{E}[\mathbf{h}(t+T)] = \overline{\mathbf{h}}(t+T)$, which is a result of $\mathbb{E}[\widetilde{\mathbf{h}}(t)] = \mathbb{E}[\boldsymbol{\xi}(t+T)] = \mathbf{0}$. Also, based on the AR process in (1), the covariance matrix of the channel at time $t+T$ is obtained as follows:

$$\mathbf{C}_{\mathbf{h}}(t+T) = \mathbf{A}(t)\mathbf{C}_{\mathbf{h}}(t)\mathbf{A}^{\mathrm{H}}(t) + \boldsymbol{\Theta}(t+T) \tag{45}$$

Furthermore, based on the definition of the cross-covariance matrix, we have:

$$\mathbb{E}[\widetilde{\mathbf{h}}(t+T)\widetilde{\mathbf{h}}^{\mathrm{H}}(t)] = \mathbf{A}(t)\mathbf{C}_{\mathbf{h}}(t) = \mathbf{C}_{\mathbf{h}}(t+T,t) \tag{46}$$

which results in

$$\mathbf{A}(t) = \mathbf{C}_{\mathbf{h}}(t+T,t)\mathbf{C}_{\mathbf{h}}^{-1}(t) \tag{47}$$

By combining (45) and (47), the covariance of AR noise is characterized as below:

$$\boldsymbol{\Theta}(t+T) = \mathbf{C}_{\mathbf{h}}(t+T) - \mathbf{C}_{\mathbf{h}}(t+1,t)\mathbf{C}_{\mathbf{h}}^{-1}(t)\mathbf{C}_{\mathbf{h}}(t,t+T). \tag{48}$$

The last property that our model should satisfy is what we call the correlation decaying property. In fact, the correlation between $t+nT$ and t must be decaying as the integer distance n increases. To prove this, we first write the AR process based on our model for the normalized centered channel $\mathbf{h}'(t)$ which is obtained as follows:

$$\mathbf{h}'(t+T) = \mathbf{P}_{\mathbf{h}}(t+T,t)\mathbf{h}'(t) + \mathbf{C}_{\mathbf{h}}^{-\frac{1}{2}}(t+T)\boldsymbol{\xi}(t+T), \tag{49}$$

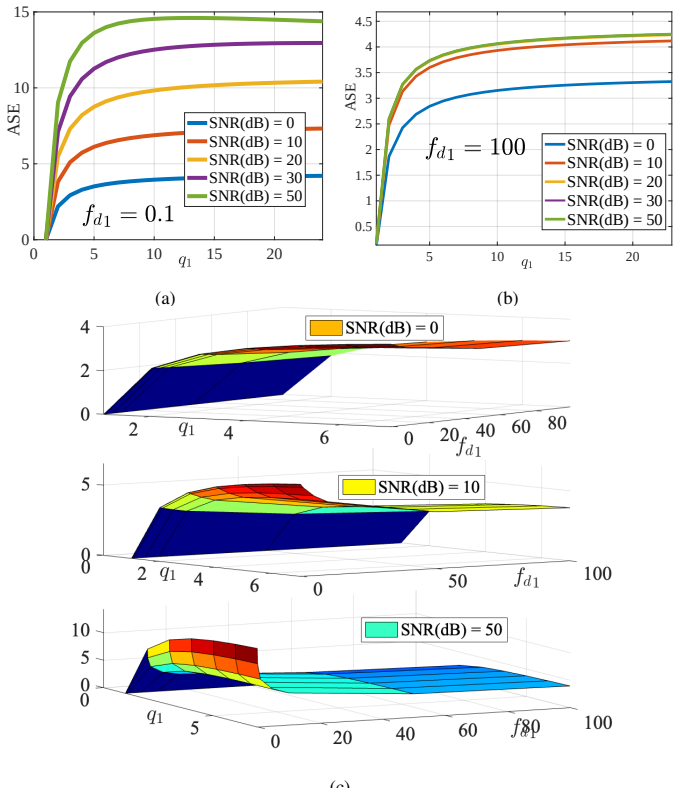


Figure 7. Figures (a) and (b) show ASE versus SNR in the case of $f_{d1} = 0.1$ and $f_{d1} = 100$, respectively. Figure (c) shows ASE versus both Doppler and SNR in different time slots. The parameters related to these experiments are set as follows: $f_{d2} = 100, f_c = 1000, \mathrm{PL}_1 = 0, \mathrm{PL}_2 = 90, q_{\max} = 24, M = 1, K_{f_1} = 1, K_{f_2} = 0, P_{\mathrm{p}\max} = P_{\mathrm{d}\max} = 1, N_r = 20, \tau_p = 1$.

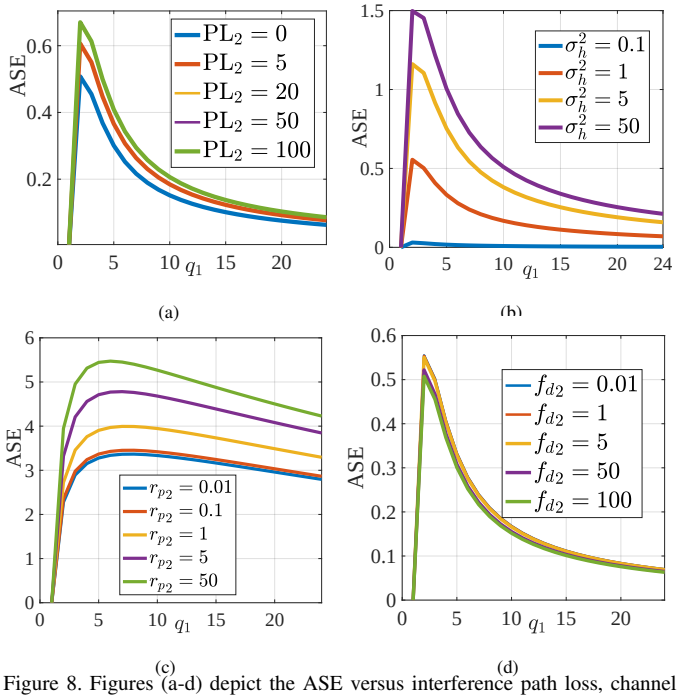


Figure 8. Figures (a-d) depict the ASE versus interference path loss, channel variance, interference power ratio and interference Doppler under the following configurations: (a): $f_{d1} = f_{d2} = 100, \mathrm{PL}_1 = 0, P_{\mathrm{p}\max} = P_{\mathrm{d}\max} = 1$, (b):$f_{d1} = 0.1, f_{d2} = 100$, (c): $f_{d1} = 5, f_{d2} = 100, \mathrm{PL}_1 = \mathrm{PL}_2 = 0$. (d):$f_{d1} = 100, \mathrm{PL}_1 = \mathrm{PL}_2 = 0, P_{\mathrm{p}\max} = P_{\mathrm{d}\max} = 1$. The remaining parameters for these experiments are consistent and set as follows: $q_{\max} = 24, M = 1, K_{f_1} = 0, K_{f_2} = 0, N_r = 20, \tau_p = 1, \sigma^2_{\mathrm{p}1} = \sigma^2_{\mathrm{p}2} = \sigma^2_{\mathrm{d}1} = 10^{-4}$.

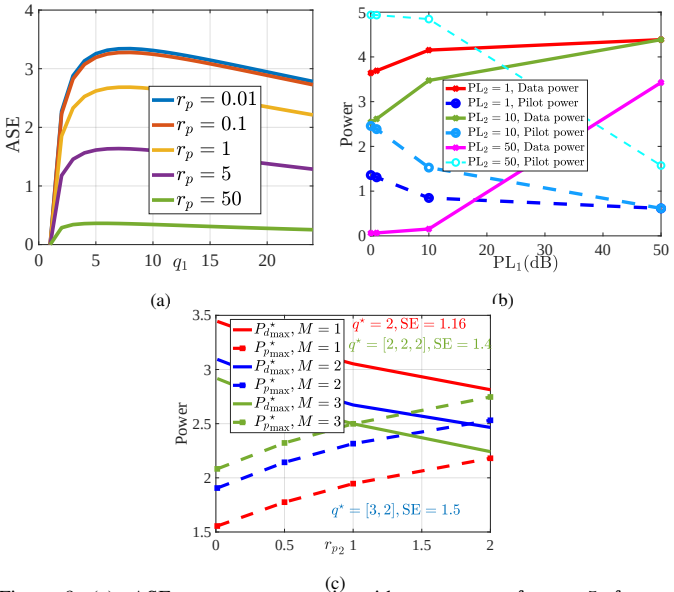


Figure 9. (a): ASE versus power ratio with parameters $f_{d1} = 5, f_{d2} = 100, \mathrm{PL}_1 = \mathrm{PL}_2 = 0, q_{\max} = 24, M = 1$.(b): This figure shows how the optimal data and pilot powers are adapted to the varying path loss of the tagged and interference channels. The parameters are as follows: $f_{d1} = 1, f_{d2} = 0.1, K_{f_1} = 0, K_{f_2} = 0, N_r = 10, \tau_p = 1$. (c): Pilot and data power distribution versus interference power with parameters $f_{d1} = 100, f_{d2} = 0.1, \mathrm{PL}_1 = 10, \mathrm{PL}_2 = 50, K_{f_1} = 0, K_{f_2} = 0, N_r = 10, \tau_p = 1, \sigma^2_{\mathrm{p}1} = 10^{-5}, \sigma^2_{\mathrm{p}_2} = 10^{-6}, \sigma^2_{\mathrm{d}} = 10^{-5}$. Also, the parameters that are similar in both experiments are as follows: $f_c = 1000, K_{f_1} = 0, K_{f_2} = 0, N_r = 20, \tau_p = 1$.

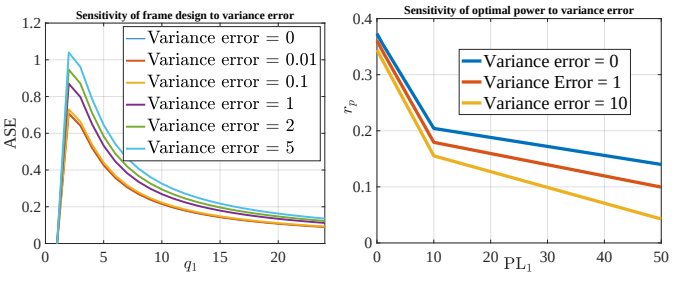


Figure 10. (a): In this figure, the ASE is shown versus the frame size under different levels of channel variance error. The used parameters are as follows: $N_r = 10, f_{d1} = 100, f_{d2} = 100, \mathrm{PL}_1 = 50, \mathrm{PL}_2 = 90, q_{\max} = 24, M = 1$ (b) This figure shows the effect of erroneous channel variance on the optimal power in different levels of path loss. Specifically, the power ratio is depicted versus path loss in different channel variance errors. The parameters are as follow: $f_{d1} = 1, f_{d2} = 0.1, K_{f_1} = 0, K_{f_2} = 0, N_r = 10, \tau_p = 1, \mathrm{PL}_2 = 1$.

where $\mathbf{P_h}(t+T,t) = \mathbf{C_h}^{-\frac{1}{2}}(t+T)\mathbf{C_h}(t+T,t)\mathbf{C_h}^{-\frac{1}{2}}(t)$ is the correlation matrix.

Then, the correlation between channels at times $t+nT$ and t will be obtained as follows:

$$\mathbf{P_h}(t+nT,t) = \mathbb{E}[\mathbf{h}'(t+nT)\mathbf{h}'^{\mathrm{H}}(t)] = \prod_{i=t}^{t+(n-1)T} \mathbf{P_h}(i+T,i). \tag{50}$$

The spectral radius of the above correlation matrix can also written as

$$\lambda_{\max}(\mathbf{P_h}(t+nT,t)) \leqslant \prod_{i=t}^{t+(n-1)T} \lambda_{\max}(\mathbf{P_h}(i+T,i)). \tag{51}$$

The decaying behaviour is ensured when the spectral radius of

the one step covariance matrices are less than one:

$$|\lambda_{\max}(\mathbf{P}_{\mathbf{h}}(t+T,t))|<1 \quad \forall t=nT, n=0,1,.. \tag{52}$$

because in this case we have

$$\lim_{n\to\infty}\lambda_{\max}(\mathbf{P}_{\mathbf{h}}(t+nT,t))=0$$

according to (51).

Appendix B
Proof of Proposition 2

At time t, the centered channel can be described as follows:

$$\widetilde{\mathbf{h}}(t)=\sum_{i=1}^{L(t)}\mathrm{e}^{j(2\pi f_d^i(t)t+\beta_i)}\mathbf{a}(\theta^i_{\mathrm{AoA}}(t)), \tag{53}$$

where $\mathbf{a}(\theta)$ is the steering vector defined in (43) and β_i is the phase of the i-th NLOS path. Also, the mean of the channel is deterministic and contains the LOS component defined as follows:

$$\overline{\mathbf{h}}(t)=K_f L(t)\mathrm{e}^{j(2\pi f_d^0(t)t+\beta_0)}\mathbf{a}(\theta^0_{\mathrm{AoA}}(t)). \tag{54}$$

Define

$$\mathbf{P}_{\mathrm{spatial}}(t_1,t_2)\triangleq \mathbb{E}_{\theta_{\mathrm{AoA}}(t_1),\theta_{\mathrm{AoA}}(t_2)}[\mathbf{a}(\theta_{\mathrm{AoA}}(t_1))\mathbf{a}^{\mathrm{H}}(\theta_{\mathrm{AoA}}(t_2))]. \tag{55}$$

Before obtaining the correlation matrix, we obtain the covariance matrix at time t:

$$\mathbf{C}_{\mathbf{h}}(t)=\mathbb{E}\left[\sum_{i=1}^{L(t)}\mathbf{a}(\theta^i_{\mathrm{AoA}}(t))\mathbf{a}^{\mathrm{H}}(\theta^i_{\mathrm{AoA}}(t))\right]\approx L(t)\mathbf{P}_{\mathrm{spatial}}(t,t), \tag{56}$$

where in the last step, we used central limit theorem for sufficiently large $L(t)$. Now, we can proceed to obtain the coefficient matrix between t_1 and t_2 as follows:

$$\begin{aligned}
&\mathbf{P}_{\mathbf{h}}(t_1,t_2)=\mathbb{E}[\mathbf{h}'(t_1)\mathbf{h}'(t_2)^{\mathrm{H}}]=\tfrac{1}{\sqrt{L(t_1)L(t_2)}}\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_1,t_1)\\
&\mathbb{E}\Big[\sum_{i=1}^{L(t_1)}\sum_{l=1}^{L(t_2)}\mathrm{e}^{j\left(2\pi\left(f_d^i(t_1)t_1-f_d^l(t_2)t_2\right)+\beta_i-\beta_l\right)}\\
&\mathbf{a}(\theta^i_{\mathrm{AoA}}(t_1))\mathbf{a}^{\mathrm{H}}(\theta^l_{\mathrm{AoA}}(t_2))\Big]\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_2,t_2)=\tfrac{1}{\sqrt{L(t_1)L(t_2)}}\\
&\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_1,t_1)\mathbb{E}\Big[\sum_{i=1}^{\min(L(t_1),L(t_2))}\mathrm{e}^{j\left(2\pi\left(f_d^i(t_1)t_1-f_d^i(t_2)t_2\right)\right)}\\
&\mathbf{a}(\theta^i_{\mathrm{AoA}}(t_1))\mathbf{a}^{\mathrm{H}}(\theta^i_{\mathrm{AoA}}(t_2))\Big]\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_2,t_2),
\end{aligned} \tag{57}$$

where in the second equality, we used the fact that the phases β_is are uniformly distributed between $[-\pi,\pi]$ and are independent from each other and also from AoAs and AoDs. This leads to the fact that $f_d^i(t_1)t_1-f_d^l(t_2)t_2+\beta_i-\beta_l$ is also uniformly distributed between $[-\pi,\pi]$ for $i\neq l$ and consequently $\mathbb{E}[\mathrm{e}^{j\left(2\pi\left(f_d^i(t_1)t_1-f_d^l(t_2)t_2\right)+\beta_i-\beta_l\right)}]=0$. When the number of scatterers is sufficiently large, we can approximate $\mathbf{P}_{\mathbf{h}}(t_1,t_2)$ in (57) with the following:

$$\mathbf{P}_{\mathbf{h}}(t_1,t_2)=\rho_{\mathrm{temporal}}(t_1,t_2)\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_1,t_1)\mathbf{P}_{\mathrm{spatial}}(t_1,t_2)\mathbf{P}^{-\frac{1}{2}}_{\mathrm{spatial}}(t_2,t_2) \tag{58}$$

where

$$\rho_{\mathrm{temporal}}(t_1,t_2)\triangleq\mathbb{E}\Big[\mathrm{e}^{j\frac{2\pi}{\lambda}\left[t_1\nu(t_1)\cos(\gamma(t_1)-\theta_{\mathrm{AoD}}(t_1))-t_2\nu(t_2)\cos(\gamma(t_2)-\theta_{\mathrm{AoD}}(t_2))\right]}\Big] \tag{59}$$

In the special cases where the user and environment do not change with time, then the parameters $\nu(t)=\nu,\gamma(t)=\gamma,\theta_{\mathrm{AoA}}(t)=\theta_{\mathrm{AoA}},\theta_{\mathrm{AoD}}(t)=\theta_{\mathrm{AoD}}$ are fixed. Define $\tau\triangleq t_2-t_1$ and $\mu\triangleq k-l$. In the specific case where AoA and AoD follows von Mises distribution given by

$$f(\theta_{\mathrm{AoD}})=\tfrac{\mathrm{e}^{\kappa_{\mathrm{AoD}}\cos(\theta_{\mathrm{AoD}}-\theta^c_{AoD})}}{2\pi J_0(j\kappa_{AoD})},f(\theta_{\mathrm{AoA}})=\tfrac{\mathrm{e}^{\kappa_{\mathrm{AoA}}\cos(\theta_{\mathrm{AoA}}-\theta^c_{AoA})}}{2\pi J_0(j\kappa_{AoA})}, \tag{60}$$

where θ^c_{AoD} and θ^c_{AoA} are the central angles showing the mean of the von Mises distribution. κ_{AoD} and κ_{AoA} are measures of concentration of the AoD and AoA around the central angles. By having this information, we can now calculate $\rho_{\mathrm{temporal}}(t_1,t_2)=\rho_{\mathrm{temporal}}(\tau)$ in closed-form as follows:

$$\begin{aligned}
&\rho_{\mathrm{temporal}}(\tau)=\mathbb{E}\Big[\mathrm{e}^{-j\frac{2\pi}{\lambda}\left[\tau\nu\cos(\gamma-\theta_{\mathrm{AoD}})\right]}\Big]=\\
&\int_{-\pi}^{\pi}\mathrm{e}^{-j\frac{2\pi}{\lambda}\left[\tau\nu\cos(\gamma-\theta)\right]}f(\theta)\mathrm{d}\theta=\\
&\int_{-\pi}^{\pi}\tfrac{\mathrm{e}^{-j\frac{2\pi}{\lambda}\tau\nu\cos(\gamma-\theta)+\kappa_{\mathrm{AoD}}\cos(\theta-\theta^c_{AoD})}}{2\pi J_0(j\kappa_{AoD})}\mathrm{d}\theta\\
&=\tfrac{J_0\left(j\sqrt{\kappa^2_{AoD}-\frac{4\pi^2}{\lambda^2}\tau^2\nu^2-j\frac{4\pi}{\lambda}\tau\nu\kappa_{AoD}\cos(\gamma-\theta^c_{AoD})}\right)}{J_0(j\kappa_{AoD})}
\end{aligned} \tag{61}$$

where we used the trigonometric property [49] and the definition of Bessel function of zero kind. With a similar approach, we can obtain spatial correlation as follows:

$$\rho_{\mathrm{spatial}}(t_1,t_2,k,l)\triangleq[\mathbf{P}_{\mathrm{spatial}}(t_1,t_2)]_{k,l}=\rho_{\mathrm{spatial}}(\mu)=\tfrac{J_0\left(j\sqrt{\kappa^2_{AoA}-\frac{4\pi^2}{\lambda^2}\mu^2d^2+j\frac{4\pi}{\lambda}\mu d\kappa_{AoA}\cos(\eta-\theta^c_{AoA})}\right)}{J_0(j\kappa_{AoA})}, \tag{62}$$

when $\kappa_{AoD}=\kappa_{AoA}=0$, we have a uniform distribution for AoD and AoA. In this simple case, the correlation becomes

$$\mathbf{P}_{\mathbf{h}}(\tau)=J_0\left(-\tfrac{2\pi}{\lambda}\tau\nu\right)\mathbf{I}_{N_r}, \tag{63}$$

which aligns with the result of [40].

Appendix C
Proof of Lemma 1

Proof. To find the linear LMMSE channel estimate, we need to solve the following optimization problem:

$$\min_{\mathbf{W},\mathbf{b}}\|\mathbf{h}(i)-\widehat{\mathbf{h}}(i)\|_2 \ \text{ s.t. } \widehat{\mathbf{h}}(i)=\mathbf{W}\mathbf{y}_p+\mathbf{b}, \tag{64}$$

which by [50] leads to

$$\widehat{\mathbf{h}}(i)=\mathbf{C}_{\mathbf{h}(i),\mathbf{y}_p}\mathbf{C}^{-1}_{\mathbf{y}_p}(\mathbf{y}_{\mathrm{p}}-\mathbb{E}[\mathbf{y}_{\mathrm{p}}])+\mathbb{E}(\mathbf{h}(i)). \tag{65}$$

With the measurements $\mathbf{y}_{\mathrm{p}}$ given in (19), the covariance matrix $\mathbf{C}_{\mathbf{h}(i),\mathbf{y}_{\mathrm{p}}}$ can be given by:

$$\begin{aligned}
&\mathbf{C}_{\mathbf{h}(i),\mathbf{y}_{\mathrm{p}}}=\alpha\sqrt{P_{\mathrm{p}}}\mathbb{E}[(\mathbf{h}(i)-\overline{\mathbf{h}}(i))(\mathbf{h}_p-\overline{\mathbf{h}}(i)\otimes\mathbf{1}_3)^{\mathrm{H}}]\widetilde{\mathbf{S}}^{\mathrm{H}}\\
&\alpha_1\sqrt{P_{\mathrm{p}}}\mathbf{C}_{\mathbf{h}(i),\mathbf{h}_p}\widetilde{\mathbf{S}}^{\mathrm{H}}=\alpha_1\sqrt{P_{\mathrm{p}}}\mathbf{E}_m(\mathbf{q},i)\widetilde{\mathbf{S}}^{\mathrm{H}},
\end{aligned} \tag{66}$$

where the last step comes from the definition (21). Assuming that the channel vector is independent of zero-mean noise

vector and that $\mathbb{E}[\mathbf{h}_p] = \overline{\mathbf{h}}(i) \otimes \mathbf{1}_3$, the autocovariance of $\mathbf{y}_\mathrm{p}$ is given by:

$$\begin{aligned}&\mathbf{C}_{\mathbf{y}_\mathrm{p}} = \mathbb{E}[\widetilde{\mathbf{y}}_\mathrm{p}\widetilde{\mathbf{y}}_\mathrm{p}^\mathrm{H}] = \\ &\alpha_1^2 P_\mathrm{p}\widetilde{\mathbf{S}}\mathbb{E}[(\mathbf{h}_p - \overline{\mathbf{h}}(i)\otimes\mathbf{1}_3)(\mathbf{h}_p - \overline{\mathbf{h}}(i)\otimes\mathbf{1}_3)^\mathrm{H}]\widetilde{\mathbf{S}}^\mathrm{H} + \\ &\sigma_\mathrm{p}^2\mathbf{I}_{3N_r\tau_\mathrm{p}} = \alpha_1^2 P_\mathrm{p}\widetilde{\mathbf{S}}\mathbf{M}_m(i)\widetilde{\mathbf{S}}^\mathrm{H} + \sigma_\mathrm{p}^2\mathbf{I}_{3N_r\tau_\mathrm{p}},\end{aligned} \tag{67}$$

where the last step uses the definition (22). By replacing (66) and (67) into (65), it follows that:

$$\begin{aligned}&\widehat{\mathbf{h}}_m(i) = \tfrac{1}{\alpha_1\sqrt{P_\mathrm{p}}}\mathbf{E}_m(\mathbf{q},i)\widetilde{\mathbf{S}}^\mathrm{H}\left(\widetilde{\mathbf{S}}\mathbf{M}_m(i)\widetilde{\mathbf{S}}^\mathrm{H} + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}}\mathbf{I}_{3N_r\tau_\mathrm{p}}\right)^{-1} \\ &\widetilde{\mathbf{y}}_\mathrm{p} + \overline{\mathbf{h}}(i)\end{aligned} \tag{68}$$

By using Woodbury matrix lemma [51], we can write:

$$\begin{aligned}&\left(\widetilde{\mathbf{S}}\mathbf{M}_m(i)\widetilde{\mathbf{S}}^\mathrm{H} + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}}\mathbf{I}_{3N_r\tau_\mathrm{p}}\right)^{-1} = \tfrac{\alpha_1^2 P_\mathrm{p}}{\sigma_\mathrm{p}^2}\mathbf{I}_{3N_r\tau_\mathrm{p}} - \\ &\left(\tfrac{\alpha_1^2 P_\mathrm{p}}{\sigma_\mathrm{p}^2}\right)^2\mathbf{I}_{3N_r\tau_\mathrm{p}}\widetilde{\mathbf{S}}\left(\mathbf{M}_m^{-1}(i) + \widetilde{\mathbf{S}}^\mathrm{H}\widetilde{\mathbf{S}}\tfrac{\alpha_1^2 P_\mathrm{p}}{\sigma_\mathrm{p}^2}\right)^{-1}\widetilde{\mathbf{S}}^\mathrm{H}\end{aligned} \tag{69}$$

Moreover, by having $\widetilde{\mathbf{S}}^\mathrm{H}\widetilde{\mathbf{S}} = \tau_\mathrm{p}\mathbf{I}_{3N_r}$ and using again Woodbury matrix lemma, it follows that:

$$\begin{aligned}&\left(\mathbf{M}_m^{-1}(i) + \tfrac{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}{\sigma_\mathrm{p}^2}\mathbf{I}_{3N_r}\right)^{-1} = \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\mathbf{I}_{3N_r} - \\ &\left(\tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\right)^2\left(\mathbf{M}_m(i) + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\mathbf{I}_{3N_r}\right)^{-1}.\end{aligned} \tag{70}$$

By replacing (70) into (69), it holds that

$$\begin{aligned}&\left(\widetilde{\mathbf{S}}\mathbf{M}_m(i)\widetilde{\mathbf{S}}^\mathrm{H} + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}}\mathbf{I}_{3N_r\tau_\mathrm{p}}\right)^{-1} = \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\mathbf{I} - \\ &\tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^\mathrm{H} + \tfrac{1}{\tau_\mathrm{p}^2}\widetilde{\mathbf{S}}\left(\mathbf{M}_m(i) + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\mathbf{I}\right)^{-1}\widetilde{\mathbf{S}}^\mathrm{H}.\end{aligned} \tag{71}$$

Multiplying the latter relation by $\widetilde{\mathbf{S}}^\mathrm{H}$ leads to

$$\begin{aligned}&\widetilde{\mathbf{S}}^\mathrm{H}\left(\widetilde{\mathbf{S}}\mathbf{M}_m(i)\widetilde{\mathbf{S}}^\mathrm{H} + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}}\mathbf{I}_{3N_r\tau_\mathrm{p}}\right)^{-1} = \\ &\tfrac{1}{\tau_\mathrm{p}}\left(\mathbf{M}_m(i) + \tfrac{\sigma_\mathrm{p}^2}{\alpha_1^2 P_\mathrm{p}\tau_\mathrm{p}}\mathbf{I}_{3N_r}\right)^{-1}\widetilde{\mathbf{S}}^\mathrm{H}.\end{aligned} \tag{72}$$

This together with (68) gives the final result. ■

Appendix D
Proof of Theorem 1

Based on the received measurements at data time slots, the BS employs the optimal MMSE receiver to estimate the transmitted data vector of the tagged user in time slot i of the m-th frame, i.e., $\widehat{x}(i) = \mathbf{g}^\star(\mathbf{q},i)\mathbf{y}$. The expected power of $\widehat{x}(i)$ conditioned on the prior information provided in $\boldsymbol{\zeta}(i)$, is given by:

$$\mathbb{E}_{\widehat{x},\mathbf{n}_\mathrm{d},\mathbf{h}(i)|\boldsymbol{\zeta}(i)}\left[|\widehat{x}|^2\right] = \left\langle \mathbf{g}^\mathrm{H}\mathbf{g}, \mathbb{E}\left[\mathbf{y}\mathbf{y}^\mathrm{H}|\boldsymbol{\zeta}(i)\right]\right\rangle. \tag{73}$$

The expression $\mathbb{E}\mathbf{y}\mathbf{y}^\mathrm{H}|\boldsymbol{\zeta}(i)$ in the above formula can be stated as

$$\mathbb{E}[\mathbf{y}\mathbf{y}^\mathrm{H}|\boldsymbol{\zeta}_k(i)] = \sum_{k=1}^{K}\alpha_k^2 P_{\mathrm{d},k}\mathbf{D}_k(i) + \sigma_\mathrm{d}^2\mathbf{I}_{N_r}, \tag{74}$$

where

$$\begin{aligned}&\mathbf{D}_k(i) \triangleq \mathbb{E}\left[\mathbf{h}_k(i)\mathbf{h}_k(i)^\mathrm{H}|\boldsymbol{\zeta}_k(i)\right] = \mathbf{C}_{\mathbf{h}_k(i)|\boldsymbol{\zeta}_k(i)} + \\ &\mathbb{E}[\mathbf{h}_k(i)|\boldsymbol{\zeta}_k(i)]\mathbb{E}[\mathbf{h}_k(i)|\boldsymbol{\zeta}_k(i)]^\mathrm{H} = \mathbf{Q}_k(i) + \mathbf{z}_k\mathbf{z}_k^\mathrm{H},\end{aligned} \tag{75}$$

and $\mathbf{z}_k = \overline{\mathbf{h}}_k(i) + \boldsymbol{\Psi}_k(\boldsymbol{\zeta}_k(i) - \overline{\mathbf{h}}(i)\otimes\mathbf{1}_2)$. The last equality in (75) comes from [15, Proposition 1]. By replacing (74) into (73), it follows that:

$$\mathbb{E}_{\widehat{x},\mathbf{n}_\mathrm{d},\mathbf{h}(i)|\boldsymbol{\zeta}(i)}\left[|\widehat{x}|^2\right] = \sum_{k=1}^{K}\alpha_k^2 P_{\mathrm{d},k}\langle\mathbf{g}^\mathrm{H}\mathbf{g}, \mathbf{D}_k(i)\rangle + \sigma_\mathrm{d}^2\|\mathbf{g}\|_2^2. \tag{76}$$

We can also rewrite (76) as follows:

$$\begin{aligned}&\mathbb{E}_{\widehat{x},\mathbf{n}_\mathrm{d},\mathbf{h}(i)|\boldsymbol{\zeta}(i)}\left[|\widehat{x}|^2\right] = \alpha^2 P_{d1}\left\langle\mathbf{z}_1\mathbf{z}_1^\mathrm{H}, \mathbf{g}^\mathrm{H}\mathbf{g}\right\rangle + \sigma_\mathrm{d}^2\|\mathbf{g}\|_2^2 + \\ &\sum_{k=2}^{K}\alpha_k^2 P_{\mathrm{d},k}\left\langle\mathbf{z}_k\mathbf{z}_k^\mathrm{H}, \mathbf{g}^\mathrm{H}\mathbf{g}\right\rangle + \sum_{k=1}^{K}\alpha_k^2 P_{\mathrm{d},k}\left\langle\mathbf{Q}_k(i), \mathbf{g}^\mathrm{H}\mathbf{g}\right\rangle.\end{aligned} \tag{77}$$

By having the expected power of estimated data symbols, we can now form the instantaneous SINR of the data estimate in slot i of frame m corresponding to the tagged user as in (78).

$$\gamma(\mathbf{q},i,\zeta(i)) \triangleq \frac{\alpha^2 P\left\langle\mathbf{z}_1\mathbf{z}_1^\mathrm{H}, \mathbf{g}^\mathrm{H}\mathbf{g}\right\rangle}{\sum_{k=2}^{K}\alpha_k^2 P_{\mathrm{d},k}\left\langle\mathbf{z}_k\mathbf{z}_k^\mathrm{H}, \mathbf{g}^\mathrm{H}\mathbf{g}\right\rangle + \sum_{k=1}^{K}\alpha_k^2 P_{\mathrm{d},k}\langle\mathbf{Q}_k(i), \mathbf{g}^\mathrm{H}\mathbf{g}\rangle + \sigma_\mathrm{d}^2\|\mathbf{g}\|_2^2} \tag{78}$$

By having (34) and (76) in mind, the instantaneous SINR can be rewritten as

$$\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i)) = \frac{\left\langle\mathbf{g}^\mathrm{H}\mathbf{g}, \mathbf{F}-\mathbf{F}_1\right\rangle}{\left\langle\mathbf{g}^\mathrm{H}\mathbf{g}, \mathbf{F}_1\right\rangle}. \tag{79}$$

It then follows that based on (27), $\mathbf{g}^\mathrm{H}\mathbf{g}$ can be stated as:

$$\mathbf{g}^\mathrm{H}\mathbf{g} = \alpha_1^2 P_\mathrm{d}\mathbf{F}^{-1}\mathbf{z}_1\mathbf{z}_1^\mathrm{H}\mathbf{F}^{-1}. \tag{80}$$

By incorporating (80) into (79), we reach to the following equation.

$$\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i)) = \frac{\alpha_1^2 P_1\left\langle\mathbf{F}^{-1}\mathbf{z}_1\mathbf{z}_1^\mathrm{H}\mathbf{F}^{-1}, \mathbf{z}_1\mathbf{z}_1^\mathrm{H}\right\rangle}{\left\langle\mathbf{F}^{-1}\mathbf{z}_1\mathbf{z}_1^\mathrm{H}\mathbf{F}^{-1}, \mathbf{F}\right\rangle - \alpha_1^2 P_1\left\langle\mathbf{F}^{-1}\mathbf{z}_1\mathbf{z}_1^\mathrm{H}\mathbf{F}^{-1}, \mathbf{z}_1\mathbf{z}_1^\mathrm{H}\right\rangle}. \tag{81}$$

By some algebraic manipulations, the SINR expression in (81) can be written as

$$\gamma(\mathbf{q},i,\zeta(i)) = \frac{\alpha_1^2 P_\mathrm{d}\mathbf{z}_1^\mathrm{H}\mathbf{F}^{-1}\mathbf{z}_1}{1-\alpha_1^2 P_\mathrm{d}\mathbf{z}_1^\mathrm{H}\mathbf{F}^{-1}\mathbf{z}_1} \tag{82}$$

Ultimately, by leveraging the insights from [52, Lemma 1], the aforementioned expression can be more succinctly streamlined to (33).

Appendix E
Proof of Theorem 2

Define the expectation of SE and SINR, by

$$\overline{\mathrm{SE}}(\mathbf{q},i) \triangleq \mathbb{E}\left[\mathrm{SE}(\mathbf{q},i,\boldsymbol{\zeta}(i))\right], \overline{\gamma}(\mathbf{q},i) \triangleq \mathbb{E}\left[\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i))\right]. \tag{83}$$

The aim is to prove that the following limit holds with high probability.

$$\tfrac{1}{N_r}\mathrm{SE}(\mathbf{q},i,\boldsymbol{\zeta}(i)) \xrightarrow{N_r\to\infty} \tfrac{1}{N_r}\mathrm{SE}^\circ(\mathbf{q},i), \tag{84}$$

where $\mathrm{SE}^\circ(\mathbf{q},i) \triangleq \log(1+\gamma^\circ(\mathbf{q},i))$ and $\gamma^\circ(\mathbf{q},i)$ is an expression to be defined later. To show (84), first, by taking the expectation with respect to the random channel estimates and using Jensen's inequality bounds for concave functions borrowed from [53], we have

$$\tfrac{1}{N_r}\left|\overline{\mathrm{SE}}(\mathbf{q},i)-\log(1+\gamma^{\circ}(\mathbf{q},i))\right|\leqslant\tfrac{c'}{N_r}\left[\left|\overline{\gamma}(\mathbf{q},i)-\gamma^{\circ}(\mathbf{q},i)\right|\right], \tag{85}$$

where c' is some constant term independent of the frame parameters. Now, it remains to prove that $\frac{1}{N_r}\overline{\gamma}(\mathbf{q},i)$ is well approximated by $\frac{1}{N_r}\gamma^{\circ}(\mathbf{q},i)$ when N_r tends to infinity. We start by this fact that the instantaneous SINR given in Theorem 1 concentrates around its mean. Since the interference channel and tagged user channel are assumed to be independent, we can separate the expectations over the channels corresponding to tagged user and interference. If we fix $\mathbf{z}_2,\ldots,\mathbf{z}_K$ and only take expectation over $\mathbf{z}_1$, due to concentration inequality provided in [46, Lemma B.26], we have

$$\begin{aligned}&\tfrac{1}{N_r}\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i))\xrightarrow{N_r\to\infty}\\&\quad\tfrac{1}{N_r}\overline{\gamma}(\mathbf{q},i)=\tfrac{1}{N_r}\mathbb{E}\left[\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i))\Big|\mathbf{z}_2(i),\ldots,\mathbf{z}_K(i)\right]\\&\quad=\tfrac{1}{N_r}\left\langle\mathbf{R}_{\mathbf{z}_1}(i),\mathbb{E}_{\mathbf{z}_2,\ldots,\mathbf{z}_K}[\mathbf{F}_1^{-1}]\right\rangle,\end{aligned} \tag{86}$$

where $\mathbf{R}_{\mathbf{z}_1}(i)=\mathbb{E}[\mathbf{z}_1(i)\mathbf{z}_1(i)^{\mathrm{H}}]$. The above result is an extended version of [46, Lemma B.26] to the Rician case. Define $\mathbf{B}=\sum_{k=2}^{K}\alpha_k^2P_{\mathrm{d},k}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}+\mathbf{S}$. By (34), it holds that $\mathbf{F}_1=\mathbf{B}+\mathbf{S}+\rho_{\mathrm{d}}\mathbf{I}_{N_r}$. Now, we prove that the expression on the right-hand side of (86) indeed tends to a stationary point when the number of BS antennas tends to infinity. It then follows that

$$\begin{aligned}&\tfrac{1}{N_r}\left\langle\mathbf{R}_{\mathbf{z}_1}(i),\mathbf{F}_1^{-1}\right\rangle\xrightarrow{N_r\to\infty}\tfrac{1}{N_r}\left\langle\mathbf{R}_{\mathbf{z}_1}(i),\mathbb{E}_{\mathbf{z}_k,k=2,\ldots,K}\mathbf{F}_1^{-1}\right\rangle\\&\xrightarrow{N_r\to\infty}\tfrac{1}{N_r}\left\langle\mathbf{R}_{\mathbf{z}_1}(i),\mathbf{D}^{-1}\right\rangle\triangleq\tfrac{1}{N_r}\gamma^{\circ}(\mathbf{q},i),\end{aligned} \tag{87}$$

where $\mathbf{D}\triangleq\boldsymbol{\Gamma}+\mathbf{S}+\rho_{\mathrm{d}}I_{N_r}$ and $\boldsymbol{\Gamma}$ is a matrix depending on the Stieltjes transform corresponding to the empirical distribution of $\mathbf{B}$ and is given by

$$\boldsymbol{\Gamma}=\sum_{k=2}^{K}\tfrac{\alpha_k^2P_{\mathrm{d},k}\mathbf{R}_{\mathbf{z}_k}(i)}{1+m_{\mathbf{B}_k}(\rho_{\mathrm{d}})}, \tag{88}$$

where

$$m_{\mathbf{B}_k}(\rho_{\mathrm{d}})\triangleq\langle\mathbf{R}_{\mathbf{z}_k}(i),(\mathbf{B}_k+\rho_{\mathrm{d}}\mathbf{I}_{N_r})^{-1}\rangle, \tag{89}$$

and $\mathbf{B}_k\triangleq\mathbf{B}-\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}$. The approximation error in (87) is given by:

$$w_N=\tfrac{1}{N_r}\left\langle\mathbf{R}_{\mathbf{z}_1}(i),(\mathbf{B}+\rho_{\mathrm{d}}\mathbf{I}_{N_r})^{-1}\right\rangle-\tfrac{1}{N_r}\left\langle\mathbf{R}_{\mathbf{z}_1}(i),\mathbf{D}^{-1}\right\rangle. \tag{90}$$

By extending the findings of [10, Theorem 1] to encompass the Rician channel with a LoS component, and borrowing results of concentration inequalities in random matrix theory [46], it can be shown that the magnitude of w_N approaches zero as N_r becomes sufficiently large. Thus, the relations (87) and (86) show that $\frac{1}{N_r}\gamma(\mathbf{q},i,\boldsymbol{\zeta}(i))$ tends to $\frac{1}{N_r}\overline{\gamma}(\mathbf{q},i)$ with high probability and $\frac{1}{N_r}\overline{\gamma}(\mathbf{q},i)$ is asymptotically approximated by $\frac{1}{N_r}\gamma^{\circ}(\mathbf{q},i)$. This together with (85) shows the relation (84).

REFERENCES

[1] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[2] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links ?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[3] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2572–2583, September 2004.

[4] T. Marzetta, "How much training is needed for multiuser MIMO ?" *IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, pp. 359–363, June 2006.

[5] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot Contamination Precoding for Interference Reduction in Large Scale Antenna Systems," 51^{th} *Annual Allerton Conference*, vol. Allerton House, Illinois, USA, pp. 226–232, October 2013.

[6] T. Kim and J. G. Andrews, "Optimal pilot-to-data power ratio for MIMO-OFDM," *IEEE Globecom*, pp. 1481–1485, December 2005.

[7] T. Kim and J. Andrews, "Cth03-2: Balancing pilot and data power for adaptive mimo-ofdm systems," in *IEEE Globecom 2006*, 2006, pp. 1–5.

[8] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhear in multipath fading channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, October 2010.

[9] G. Fodor, P. D. Marco, and M. Telek, "On the impact of antenna correlation and CSI errors on the pilot-to-data power ratio," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2622–2633, 2016.

[10] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need ?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[11] R. K. Mallik, M. R. Bhatnagar, and S. P. Dash, "Fractional pilot duration optimization for SIMO in Rayleigh fading with MPSK and imperfect CSI," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1732–1744, 2018.

[12] L. Hanlen and A. Grant, "Capacity analysis of correlated MIMO channels," *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6773–6787, 2012.

[13] R. Couillet, J. Hoydis, and M. Debbah, "Random beamforming over quasi-static and fading channels: A deterministic equivalent approach," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6392–6425, 2012.

[14] C. Wen, G. Pan, K. Wong, M. Guo, and J. Chen, "A deterministic equivalent for the analysis of non-Gaussian correlated MIMO multiple access channels," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 329–352, 2013.

[15] G. Fodor, S. Fodor, and M. Telek, "Performance analysis of a linear MMSE receiver in time-variant rayleigh fading channels," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 4098–4112, 2021.

[16] G. Fodor, S.Fodor, and M.Telek, "MU-MIMO receiver design and performance analysis in time-varying Rayleigh fading," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1214–1228, 2022.

[17] H. Abeida, "Data-aided SNR estimation in time-variant Rayleigh fading channels," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5496–5507, Nov. 2010.

[18] H. Hijazi and L. Ros, "Joint data QR-detection and Kalman estimation for OFDM time-varying Rayleigh channel complex gains," *IEEE Transactions on Communications*, vol. 58, no. 1, pp. 170–177, Jan. 2010.

[19] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, 2013.

[20] C. Kong, C. Zhong, A. K. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4879–4893, 2015.

[21] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2960–2973, 2020.

[22] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Transactions on Communications*, pp. 1–1, 2020, early access.

[23] S. Fodor, G. Fodor, D. Gürgünoğlu, and M. Telek, "Optimizing pilot spacing in MU-MIMO systems operating over aging channels," *IEEE Transactions on Communications*, vol. 71, pp. 3708–3720, 2023.

[24] S. Savazzi and U. Spagnolini, "On the pilot spacing constraints for continuous time-varying fading channels," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3209–3213, 2009.

[25] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1098–1112, 2009.

[26] Z. Lian, L. Jiang, C. He, and D. He, "A non-stationary 3-D wideband GBSM for HAP-MIMO communication systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1128–1139, 2018.

[27] S. Banerjee, R. Bhattacharjee, and A. Sinha, "Fundamental limits of age-of-information in stationary and non-stationary environments," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1741–1746.

[28] H. Iimori, G. T. F. deAbreu, D. González G., and O. Gonsa, "Mitigating channel aging and phase noise in millimeter wave MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 7237–7242, 2021.

[29] D. Löschenbrand, M. Hofer, L. Bernado, S. Zelenbaba, and T. Zemen, "Towards cell-free massive MIMO: A measurement-based analysis," *IEEE Access*, vol. 10, pp. 89 232–89 247, 2022.

[30] Z. Song, T. Yang, X. Wu, H. Feng, and B. Hu, "Regret of age-of-information bandits in nonstationary wireless networks," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2415–2419, 2022.

[31] X. Cheng, Z. Huang, and L. Bai, "Channel nonstationarity and consistency for beyond 5g and 6g: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1634–1669, 2022.

[32] Z. Zou, M. Careem, A. Dutta, and N. Thawdar, "Joint spatio-temporal precoding for practical non-stationary wireless channels," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2396–2409, 2023.

[33] J. Bian, C.-X. Wang, X. Gao, X. You, and M. Zhang, "A general 3D non-stationary wireless channel model for 5G and beyond," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 3211–3224, 2021.

[34] M. A. A. Careem and A. Dutta, "Real-time prediction of non-stationary wireless channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 7836–7850, 2020.

[35] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[36] K. E. Baddour and N. C. Beaulieu, "Accurate simulation of multiple cross-correlated rician fading channels," *IEEE transactions on communications*, vol. 52, no. 11, pp. 1980–1987, 2004.

[37] W. C. Jakes, "Mobile microwave communication," 1974.

[38] G. Matz, "On non-WSSUS wireless fading channels," *IEEE transactions on wireless communications*, vol. 4, no. 5, pp. 2465–2478, 2005.

[39] A. Paier, T. Zemen, L. Bernadó, G. Matz, J. Karedal, N. Czink, C. Dumard, F. Tufvesson, A. F. Molisch, and C. F. Mecklenbrauker, "Non-WSSUS vehicular channel characterization in highway and urban scenarios at 5.2 GHz using the local scattering function," in *2008 international itg workshop on smart antennas*. IEEE, 2008, pp. 9–15.

[40] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. Wiley-IEEE press, 1994.

[41] C. Park and B. Lee, "Online compressive covariance sensing," *Signal Processing*, vol. 162, pp. 1–9, 2019.

[42] Y. Liu, C. Li, J. Wang, and M. Long, "Koopa: Learning non-stationary time series dynamics with koopman predictors," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[43] S. Daei, F. Haddadi, and A. Amini, "Exploiting prior information in block-sparse signals," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 5093–5102, 2019.

[44] S.Daei, F.Haddadi, and A.Amini, "Living near the edge: A lower-bound on the phase transition of total variation minimization," *IEEE Transactions on Information Theory*, vol. 66, no. 5, pp. 3261–3267, 2019.

[45] S. Daei, F. Haddadi, A. Amini, and M. Lotz, "On the error in phase transition computations for compressed sensing," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6620–6632, 2019.

[46] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer, 2010, vol. 20.

[47] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[48] E. C. Titchmarsh, "The theory of functions," 1939.

[49] E. W. Weisstein, "Harmonic addition theorem," *Mathworld-A Wolfram Web Resource*, 2017.

[50] S. M. Kay, *Modern Spectral Estimation Theory and Application*. Prentice Hall, 1988.

[51] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2013.

[52] A. Abrardo, G. Fodor, M. Moretti, and M. Telek, "MMSE receiver design and SINR calculation in MU-MIMO systems with imperfect CSI," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 269–272, Feb. 2019.

[53] X. Gao, M. Sitharam, and A. E. Roitberg, "Bounds on the jensen gap, and implications for mean-concentrated distributions," *arXiv preprint arXiv:1712.05267*, 2017.

Sajad Daei Sajad Daei (Member, IEEE) received the B.Sc. degree in electronic engineering from Guilan University, Guilan, Iran, in 2011, the M.Sc. degree in communication engineering from the Sharif University of Technology, Tehran, Iran, in 2013, and the Ph.D. degree in communication engineering from the Iran University of Science and Technology, Tehran, in 2020. He was a Researcher of Electronics Research Institute (ERI) from 2019 to 2021. He was a Postdoc Researcher with the Eurecom working on semantic and goal-oriented signal processing. He is currently a Postdoc Researcher with the KTH Royal Institute of Technology, Stockholm, Sweden, working on integrated sensing and communications. He was the recipient of the 2020 Outstanding Ph.D. Thesis Award of IEEE (Iran section). His main research interests include various topics in signal processing and wireless communications particularly inverse problems, compressed sensing, blind deconvolution, super-resolution, optimization, massive MIMO, random access, goal-oriented inverse problems.

Gabor Fodor Gabor Fodor (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Budapest University of Technology and Economics in 1998 and the D.Sc. degree from the Hungarian Academy of Sciences (Doctor of MTA) in 2019. He is currently a Master Researcher with the Ericsson Research and a Docent and an Adjunct Professor with the KTH Royal Institute of Technology, Stockholm, Sweden. He has authored or coauthored more than 100 refereed journal articles and conference papers and seven book chapters and holds more than 100 European and U.S. granted patents. From 2017 to 2020, he was a member of the board of the IEEE Sweden joint Communications, Information Theory, and Vehicle Technology Chapter. He was a co-recipient of the IEEE Communications Society Stephen O. Rice Prize in 2018 and the Best Student Conference Paper Award from the IEEE Sweden VT/COM/IT Chapter in 2018. He is currently serving as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE WIRELESS COMMUNICATIONS

Mikael Skoglund Mikael Skoglund (Fellow, IEEE) received the Ph.D. degree from Chalmers University of Technology, Sweden, in 1997. In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Division of Information Science and Engineering and the Department of Intelligent Systems. He has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory, information-theoretic security, information theory for statistics and learning, information and control, and signal processing. He has authored and coauthored around 200 journals and more than 420 conference papers. Dr. Skoglund is an elected member of the IEEE Information Theory Society Board of Governors. He was the General Co-Chair of IEEE ITW 2019 and the TPC Co-Chair of IEEE ISIT 2022. He has served on numerous technical program committees for IEEE-sponsored conferences. From 2003 to 2008, he was an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS. From 2008 to 2012, he was on the editorial board of IEEE TRANSACTIONS ON INFORMATION THEORY, where he joined in Fall 2021.

Miklos Telek Miklos Telek received the Ph.D. degree in electrical engineering from the Budapest University of Technology and Economics, Hungary, in 1994 and the MTA Doctor degree from the Hungarian Academy of Sciences in 2004. He is currently a Professor with the Department of Networked Systems and Services, Budapest University of Technology and Economics. Since 1997, he has been the Head of the Stochastic Modeling Laboratory, which participates in international research projects on traffic modeling of mobile telecommunication networks. Since 2012, he has also been the Head of the ELKH-BME Information Systems Research Group, Eotvos Lorand Research Network. He is the coauthor of the textbook, Introduction to Queueing Systems with Telecommunication Applications, whose second edition got published in 2019. His research interests include various aspects of stochastic performance modeling and analysis of computer and communication systems.