Multi-User Pilot-Domain NOMA Under Coherence Disparity and Channel State Feedback

Mehdi Karbalayghareh, *Member, IEEE*, and Aria Nosratinia, *Fellow, IEEE*

***Abstract*— Non-orthogonal transmission of data and pilots using product superposition is known to be highly efficient under unequal coherence conditions in a downlink channel, improving achievable rates and degrees of freedom (DoF). However, these techniques have not been used when transmit beamforming is present, as product superposition measures and utilizes composite (virtual) link gains, while beamforming requires knowledge of true (physical) link gains at the transmitter. This paper presents new techniques that enable the gains of pilot-domain non-orthogonal multiple access (NOMA) product superposition to be combined with transmit beamforming gain. The technical novelty of this paper lies in reconciling the requirements of transmit beamforming and product superposition under perfect or imperfect channel state feedback, and demonstrating its effectiveness under multi-user scenarios. The paper begins with a multi-user generalization of product superposition rate analysis in the absence of feedback. Then, a novel non-orthogonal scheme is proposed that harmoniously combines with either perfect or imperfect feedback under disparity in coherence time or coherence bandwidth among users. The proposed scheme includes efficient pilot placement strategies under multi-user scenarios with arbitrary coherence time and coherence bandwidth for different users. Numerical results illustrate the effectiveness of the proposed techniques.**

***Index Terms*— Non-orthogonal transmission, product superposition, coherence disparity, multi-user MIMO, channel state feedback, zero-forcing.**

I. Introduction

FOR downlink channel estimation, all receivers are served by the same pilots, thus pilot time slots and pilot power are identical for all users [1], [2]. This remains true even though the coherence time and coherence bandwidth of different users may be non-identical. Since the channel state for some links varies more rapidly (in time or frequency or both) than for some other links, the pilot sequence that is geared toward some links may be either inadequate or excessive for other links. Efficiency can be restored if the users employ different pilot duty cycles, but then the temporal orthogonality of pilots and data must be relinquished. In the literature, this is known as *pilot-domain NOMA* (non-orthogonal multiple access), and requires coexistence of some data and some pilots on some times or frequencies.

Recent work has shown that non-orthogonal pilot/data transmission via *product superposition* [3] can achieve gains in a two-user downlink channel in which one user's fading is static, and the other is dynamic. This signaling structure is designed to manage the interference between data and pilots under non-orthogonal transmission, and has been shown to yield not only rate gains, but also gains in degrees of freedom[1] (DoF), compared with when pilots are orthogonal to both users' data. Thus far, these gains have been demonstrated only in the absence of channel state information at the transmitter (CSIT). Product superposition under channel state feedback was first introduced in the conference version of this paper [4] where we considered a multiple-input single-output (MISO) broadcast channel whose links are frequency-flat and have integer coherence time ratios.

The present work proposes pilot-domain NOMA techniques for a frequency-selective multi-user downlink multiple-input multiple-output (MIMO) having arbitrary coherence conditions in time, bandwidth or both in the presence of either perfect or imperfect channel state feedback. Our non-orthogonal transmission scheme combines product superposition and zero-forcing beamforming. The key to making this possible is a reconciliation of the distinct requirements of product superposition and beamforming, and facilitating their joint operation. Product superposition operates by presenting to one user a virtual channel that is a product of its link gain with another user's data [3]. The user with the virtual channel is unable to measure the true (physical) link gain, a quantity that is needed for beamforming. The resolution of this issue results in significant performance improvements, including in degrees of freedom, which has not been demonstrated for pilot superposition techniques other than product superposition.

We briefly review the relevant literature to set the stage and to highlight the contribution of the present work. Pilot-domain NOMA has been explored in the literature via superimposed pilots, where data and pilots are transmitted via an additive superposition. The idea of such a scheme goes back at least to [5] and [6] where a known pilot sequence is added to

Manuscript received 30 May 2023; revised 25 September 2023 and 3 January 2024; accepted 3 March 2024. Date of publication 26 March 2024; date of current version 12 September 2024. This work was supported in part by the National Science Foundation under Grant 1718551 and Grant 2148211. An earlier version of this paper was presented in part at the IEEE Information Theory Workshop (ITW), Mumbai, India, November 2022 [DOI: 10.1109/ITW54588.2022.9965758]. The associate editor coordinating the review of this article and approving it for publication was Y. Liu. *(Corresponding author: Aria Nosratinia.)*

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: mehdi.karbalayghareh@utdallas.edu; aria@utdallas.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2024.3378996.

Digital Object Identifier 10.1109/TWC.2024.3378996

[1]Pre-log factor of capacity.

an unknown data sequence for the purpose of phase synchronization. Later, this technique was applied for orthogonal frequency division multiplexing (OFDM) frame synchronization [7] and joint time and frequency synchronization [8]. This idea was used for channel synchronization by Hoeher and Tufvesson [9]. Tugnait and Meng [10] optimized the power allocation for superimposed signals and derived the data aided channel estimation variance. Coldrey and Bohlin [11] compared the rate achieved by superimposed pilot scheme against conventional pilot schemes. Zhang et al. [12] calculated the sum rate of a multi-cell MIMO uplink system under channel estimation with superimposed pilots. Ma et al. [13] investigated rate gains under high mobility conditions using pilot sequence optimization. For massive MIMO sparse uplink channels, Mansoor et al. [14] proposed channel estimation techniques for superimposed pilots and demonstrated its gains. Upadhya et al. [15] investigated a hybrid technique involving superimposed pilots as well as time-multiplexed pilots. Jiao et al. [16] considered both pilot and code domain NOMA in satellite-based internet of things, based on a ridge regression algorithm. Jing et al. [17] optimized superimposed pilots and achieved spectral efficiency gain in multi-user massive MIMO systems. Upadhya et al. [18] explored downlink throughput in the time division duplexing (TDD) multi-cell massive MIMO systems when superimposed pilots are used for channel estimation in the uplink. Zhang et al. [19] investigated spectral efficiency gains in cell-free massive MIMO systems with superimposed pilots. The performance of wireless-energy-transfer enabled massive MIMO systems with superimposed pilots was explored in [20]. For recent contributions, superimposed pilots have been proposed to achieve spectral efficiency gains by improving the accuracy of channel estimation in orthogonal time frequency space (OTFS) systems [21], [22], [23], [24], [25], [26].

In the absence of CSI feedback, a highly efficient pilot-domain NOMA was proposed [3], [27], [28] for two users under coherence disparity. The corresponding multi-user DoF region was investigated in [29] and [30] again assuming no CSI feedback. The corresponding multi-user *achievable rates* have been unavailable and are addressed in this paper. When CSI feedback is available and the transmitter employs beamforming, the applicability of product superposition has thus far been unclear;[2] this is the main subject of the present paper.

To summarize, the main contributions of this paper include: (a) A generalization of the results of [3], in the absence of CSI feedback, to a multi-user scenario with arbitrary coherence times, (b) Synthesizing a mutually consistent method of product superposition and transmit beamforming, and demonstrating its efficacy, and (c) Extending the results to a general multi-user system under disparity in coherence time, frequency, or both, considering perfect or imperfect feedback, and proposing efficient pilot placement strategies.

[2]A narrowly-defined two-user case, where one of the two users has long coherence intervals, enjoys free CSIR, and has no need for feedback, was studied in [31].

The remainder of the paper is organized as follows: Section II describes the system model. Section III investigates product superposition in a downlink channel with multiple receivers and derive the achievable rate expressions assuming no channel state feedback. Section IV introduces a non-orthogonal transmission scheme for a two-user downlink channel with coherence disparity and derive the achievable rate expressions under channel state feedback. Section V analyzes the multi-user downlink channel under coherence disparity and channel state feedback. Section VI contains numerical results, and Section VII offers concluding remarks.

II. Notation and System Model

Matrices and vectors are denoted by bold capital letters and bold small letters, respectively. Their elements are denoted by small letters. The superscripts $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ respectively stand for the transpose, Hermitian and conjugate operations. $\mathbf{I}_k$ denotes the $k \times k$ identity matrix. $\mathbb{C}^{p\times q}$ denotes the set of $p \times q$ complex matrices. $\mathcal{CN}(m, n)$ denotes the circularly symmetric complex Gaussian distribution with mean m and variance n. Furthermore, diag $(\mathbf{a})$ denotes a diagonal matrix whose entries are the elements of the vector $\mathbf{a}$, tr$(\cdot)$ denotes the trace, and $\mathbb{E}(\cdot)$ denotes the expectation. The least common multiple of integers is denoted with $\mathrm{lcm}(\cdot, \cdot)$.

We consider an OFDM downlink channel whose M-antenna transmitter serves L receivers, each equipped with N antennas (see Fig. 1). Throughout the paper, the receiving terminals are denoted 'receiver' or 'user.' The system operates under block-fading, where the link gain for each user remains constant within one coherence block (see Fig. 2). User ℓ has coherence time T_ℓ and coherence bandwidth B_ℓ. The link gains for each user are statistically independent in different blocks. Channels for different users are also assumed statistically independent. We denote with $\mathbf{H}_{\ell,k} \in \mathbb{C}^{N\times M}$ the MIMO link gains for User ℓ at subcarrier k. The entries of $\mathbf{H}_{\ell,k}$ are independent identically distributed (i.i.d.) obeying $\mathcal{CN}(0, 1)$. The received signal by User ℓ at subcarrier k is given as

$$\mathbf{Y}_{\ell,k} = \mathbf{H}_{\ell,k}\mathbf{X}_k + \mathbf{W}_{\ell,k},\ k = 1, \cdots, K, \tag{1}$$

where $\mathbf{X}_k$ is the transmitted signal at subcarrier k, $\mathbf{W}_{\ell,k}$ is the additive Gaussian noise matrix whose elements are i.i.d. with zero mean and variance N_0, and K denotes the number of subcarriers. With a slight abuse of notation, the same variables in Eq. (1) continue to be utilized under different block lengths even though matrix dimensions will change. Often, block lengths of T_ℓ are used, in which case $\mathbf{X}_k \in \mathbb{C}^{M\times T_\ell}$, and $\mathbf{W}_{\ell,k} \in \mathbb{C}^{N\times T_\ell}$. Sometimes, shorter sub-blocks of length M are discussed, in which case signals and noise have corresponding dimensions $M \times M$, and $N \times M$ that are clear from context, and are suppressed in the notation.

The transmitter is assumed to have an average power constraint ρ at each subcarrier:

$$\mathbb{E}\Big[\sum_{i=1}^{M} \mathrm{tr}\big(\mathbf{x}_i\mathbf{x}_i^H\big)\Big] \le \rho T_\ell, \tag{2}$$

where $\mathbf{x}_i \in \mathbb{C}^{T_\ell\times 1}$ is the signal vector sent by the antenna i.

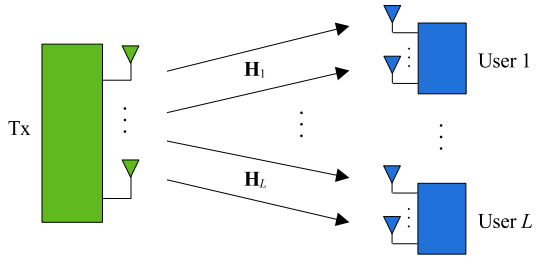

Fig. 1. Multi-user downlink MIMO.

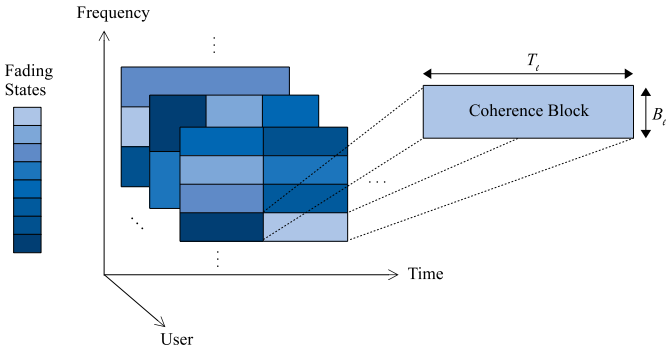

Fig. 2. Coherence blocks for different users.

III. New Results on Coherence Disparity Without CSI Feedback

To build a foundation for the results of the paper to follow, this section develops crucial generalizations of [3] and [29] *without CSI feedback*. Analytical methods of [3] were critically dependent on one coherence interval being infinity, which are generalized to arbitrary coherence times in this section. The results of [29] were limited to DoF results, while the present chapter analyses the rate region. This section also extends the feedback-free results to a general case of a multi-user downlink channel with L receivers having arbitrary coherence intervals.

The achievable rates in this section have been calculated for one OFDM subcarrier, where the dependence on the subcarrier index has been suppressed for brevity of notation. The total rate for a user can be calculated by a summation over the subcarriers, a straight forward fact whose mention is omitted in the remainder of this section. The notation required for representing different coherence bandwidths does not have a direct bearing on the development of ideas in this section. The subcarrier index will appear in the analysis of the general channel scenarios in Sections IV, and V under feedback.

A. Transmission Scheme

Consider a downlink channel with two receivers where, without loss of generality, we assume the coherence times $T_1 > T_2$. We also assume an integer coherence ratio $\frac{T_1}{T_2}$ that allows us to limit the discussions to a time period of length T_1, after which everything will repeat (see Fig. 3). We divide this time period into blocks of length T_2. Over the first length-T_2 block, neither user has knowledge of its channel state, therefore the transmitter emits a pilot intended for both users. During this time, both users perform conventional channel estimation and coherent data detection, with pilot time slots followed by data time slots that are time-shared across users. For brevity, we omit a description of this well-known orthogonal signaling structure, and concentrate on the subsequent blocks where non-orthogonal pilots make an appearance.

In the subsequent length-T_2 blocks, the channel of User 1 remains unchanged, so it does not need to estimate the channel again, but User 2 needs to update its channel estimate. Therefore, over each of the remaining $(\frac{T_1}{T_2}-1)$ blocks, a pilot will be transmitted for User 2, but these pilot time slots will also carry data for User 1. To implement this strategy, the transmitted signal is

$$\mathbf{X} = \mathbf{X}_1\tilde{\mathbf{X}}_2, \tag{3}$$

where $\mathbf{X}_1 \in \mathbb{C}^{M\times N}$ denotes the information carrying signal for User 1 and has i.i.d. entries $\mathcal{CN}(0,1)$, and $\tilde{\mathbf{X}}_2 \in \mathbb{C}^{N\times T_2}$ is the total signaling component intended for User 2, and is given by

$$\tilde{\mathbf{X}}_2 = \big[\mathbf{\Pi}\,, \mathbf{X}_2\big], \tag{4}$$

where $\mathbf{\Pi} \in \mathbb{C}^{N\times N}$ denotes the pilot matrix, which is *unitary*, and $\mathbf{X}_2 \in \mathbb{C}^{N\times(T_2-N)}$ denotes the information carrying signal for User 2 with i.i.d. entries $\mathcal{CN}(0,1)$ and has power ρ_2 at each time slot. Let ρ_p denote the overall power at each pilot time slot. Given the transmit signal in (3), the constants ρ_p and ρ_2 satisfy the power constraint (2):

$$\frac{N}{T_2}\rho_p + \frac{T_2-N}{T_2}(MN\rho_2) \le \rho. \tag{5}$$

Remark 1: The power of User 2, ρ_2, is multiplied with MN. This is due to the fact that $\mathbf{X}_2$ multiplies $\mathbf{X}_1$, whose $\mathcal{CN}(0,1)$ entries give rise to overall power MN. A similar phenomenon appears later in multi-user scenarios in the sequel.

The received signal at User 2 is

$$\mathbf{Y}_2 = \mathbf{H}_2\mathbf{X}_1\big[\mathbf{\Pi}\,, \mathbf{X}_2\big] + \mathbf{W}_2 = \big[\mathbf{G}_2\mathbf{\Pi}\,, \mathbf{G}_2\mathbf{X}_2\big] + \mathbf{W}_2, \tag{6}$$

where $\mathbf{G}_2 \triangleq \mathbf{H}_2\mathbf{X}_1$ denotes the virtual channel for User 2 that is the product of its link gain with the signal of User 1. User 2 estimates the virtual channel $\mathbf{G}_2$ over the first N time slots and then decodes $\mathbf{X}_2$ coherently.

The received signal at User 1 is

$$\mathbf{Y}_1 = \mathbf{H}_1\mathbf{X} + \mathbf{W}_1 = \big[\mathbf{H}_1\mathbf{X}_1\mathbf{\Pi}\,, \mathbf{H}_1\mathbf{X}_1\mathbf{X}_2\big] + \mathbf{W}_1. \tag{7}$$

The receiver knows the pilot matrix $\mathbf{\Pi}$, and also the estimate of the channel $\mathbf{H}_1$ (denoted by $\overline{\mathbf{H}}_1$) is known at this receiver because the channel has remained unchanged since the last time it was estimated. Therefore, the receiver will perform coherent decoding of $\mathbf{X}_1$ from the first N columns of $\mathbf{Y}_1$.

B. Achievable Rates

In the proposed scheme, in the first length-T_2 block, both users estimate the channel during the first N time slots. In this block, no superposition is applied and single-user transmission is employed. For completeness, we allow the data duration in this block to be used for the two users with a time-sharing ratio $\gamma \in [0,1]$. In the subsequent blocks, User 1 does not need to do any channel estimation, and receives data during

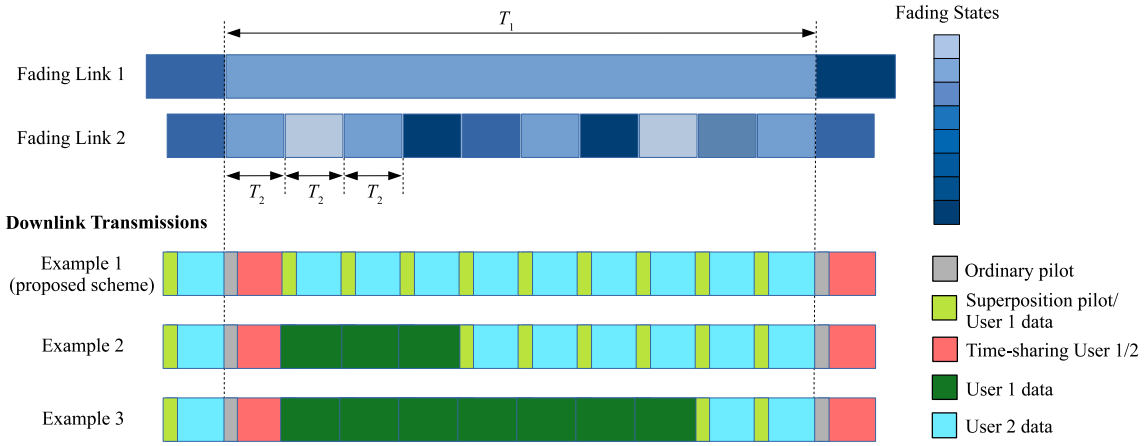

Fig. 3. Example 1 represents the rate 'corner point' calculated in Eqs. (8) and (9). Examples 2 and 3 represent other operating points on the boundary of the rate region, as explained in Remark 2.

the N pilot time slots in each block, for $(\frac{T_1}{T_2}-1)$ blocks. Thus, the rate per channel use for User 1 is

$$R_1 = \gamma\Big(\frac{T_2}{T_1}\Big)\Big(1-\frac{N}{T_2}\Big)\mathbb{E}\Big[\text{logdet}\big(\mathbf{I}+\frac{\rho}{\lambda_1}\overline{\mathbf{H}}_1\overline{\mathbf{H}}_1^H\big)\Big] + \Big(1-\frac{T_2}{T_1}\Big)\Big(\frac{N}{T_2}\Big)\Big]\mathbb{E}\Big[\text{logdet}\big(\mathbf{I}+\frac{\rho_p}{\lambda_1}\overline{\mathbf{H}}_1\overline{\mathbf{H}}_1^H\big)\Big], \quad (8)$$

where λ_1 denotes the equivalent receiver noise power, which is the sum of receiver noise and channel estimation error, and $\overline{\mathbf{H}}_1$ is the estimate of true channel $\mathbf{H}_1$ for User 1. The first term in (8) is the achieved rate at User 1 over the first block, i.e., $\gamma(T_2-N)$ time slots. The second term in (8) is the achieved rate at User 1 during the pilot slots of the subsequent blocks, i.e., $(\frac{T_1}{T_2}-1)N$ time slots.

For User 2, data is transmitted over $(1-\gamma)(T_2-N)$ time slots in the first block where the receiver estimates the true channel $\mathbf{H}_2$, and $(\frac{T_1}{T_2}-1)(T_2-N)$ time slots in subsequent blocks where the receiver estimates the virtual channel $\mathbf{G}_2$. These appear as constants in the first term and second term in achievable rate for User 2 in Eq. (9):

$$R_2 = (1-\gamma)\Big(\frac{T_2}{T_1}\Big)\Big(1-\frac{N}{T_2}\Big)\mathbb{E}\Big[\text{logdet}\big(\mathbf{I}+\frac{\rho}{\lambda_2}\overline{\mathbf{H}}_2\overline{\mathbf{H}}_2^H\big)\Big] + \Big(1-\frac{T_2}{T_1}\Big)\Big(1-\frac{N}{T_2}\Big)\mathbb{E}\Big[\text{logdet}\big(\mathbf{I}+\frac{\rho_2}{\tilde{\lambda}_2}\overline{\mathbf{G}}_2\overline{\mathbf{G}}_2^H\big)\Big], \quad (9)$$

where $\overline{\mathbf{H}}_2$ and $\overline{\mathbf{G}}_2$ respectively denote the estimate of the true channel and the virtual channel for User 2. λ_2 and $\tilde{\lambda}_2$ denote the total noise power at the receiver over the first block and subsequent blocks, respectively.

Remark 2: The rate pair (8), (9) essentially describes a corner point in the rate region, representing the operating regime when both users are active. Obviously, in any given coherence interval, the system can also operate in a single-user mode, and in that case pilots are only necessary if the active user's channel state needs updating. The single-user rates are straight forward and therefore are not mentioned in the interest of brevity. The overall rate region consists of the convex hull of the rate pair under superposition (8), (9), and the single-user rates. The signaling and operation according to these variations are shown in Fig. 3.

Remark 3: The two-user model in [3] had a fundamentally asymmetric framework due to one user not requiring channel training. The model in the present paper, while allowing for differences in both the link gains and link dynamics, has a fundamental symmetry in terms of requiring channel training for all users. The results of [3] can be recovered from the rate pair (8), (9) by taking the limit $T_1 \to \infty$. Another facet of our model putting the users on equal footing is manifested by the time-sharing variable γ, whose effect appears once every T_1 time slots.

We now extend this result to a multi-user downlink channel whose links experience arbitrary coherence times. We also demonstrate the application of product superposition under non-integer coherence time ratios, using the notion of a super interval. To manage the complexity of explanations, we develop the details of multi-user analysis in the context of a three-user system. We then present the expressions for the general multi-user case as a direct extension, without further elaboration.

Consider a three-user downlink system with arbitrary coherence times $T_1 > T_2 > T_3$. Since coherence times are no longer multiples of each other, we construct a periodic channel structure by considering a super interval of length $T_1T_2T_3$. The smallest coherence time is T_3, therefore User 3 experiences T_1T_2 different channel realizations in the super interval, and needs as many pilots in that duration. User 1 and User 2 respectively need T_2T_3 and T_1T_3 pilots during this time. Product superposition uses $T_1(T_2-T_3)$ and $T_3(T_1-T_2)$ pilot intervals for data transmission to User 2 and User 1, respectively. Without loss of generality, we assume the canonical pilot sequence $\mathbf{\Pi}=\mathbf{I}$. In the super interval, the T_1T_2 blocks of length T_3 fall into three signaling categories:

1) In T_2T_3 blocks, all three users need to update their channel estimate. The signaling consists of orthogonal pilot and data. Each user estimates its channel from the pilot. The signaling is conventional, i.e., no superposition is used. The data in these blocks can serve any of the three users in any proportion, i.e., time-sharing.
2) In $T_3(T_1-T_2)$ blocks, the slowest channel (User 1) does not require a CSI update, but both User 2 and

User 3 need an update. In that case, the signaling is:

$$\mathbf{X} = [\mathbf{X}_1\ , \mathbf{X}_1\mathbf{X}_3], \tag{10}$$

where $\mathbf{X}_1 \in \mathbb{C}^{M\times N}$ and $\mathbf{X}_3 \in \mathbb{C}^{N\times(T_3-N)}$ denote the intended messages for User 1 and User 3, respectively. The received signals at the three users are:

$$\mathbf{Y}_\ell = \mathbf{H}_\ell\mathbf{X} + \mathbf{W}_\ell = \big[\mathbf{H}_\ell\mathbf{X}_1, \mathbf{H}_\ell\mathbf{X}_1\mathbf{X}_3\big] + \mathbf{W}_\ell,\ \ell = 1, 2, 3. \tag{11}$$

User 1, whose channel state is unchanged from the last block, decodes its message from the component $\mathbf{H}_1\mathbf{X}_1$. User 2 refreshes its CSI via estimating the virtual channel $\mathbf{G}_2 \triangleq \mathbf{H}_2\mathbf{X}_1$ during the first component of the block. This virtual channel estimate is kept in store for future blocks, because User 2 has no data transmission in this block. User 3, similarly, estimates its virtual channel $\mathbf{G}_3 \triangleq \mathbf{H}_3\mathbf{X}_1$ during the first component of the block, and decodes its message from the second component of the block, $(\mathbf{H}_3\mathbf{X}_1)\mathbf{X}_3$, which is composed of $(T_3 - N)$ time slots. The transmit signal in (10) satisfies the power constraint at the transmitter:

$$\frac{N}{T_3}\rho_p + \frac{T_3 - N}{T_3}(MN\rho_3) \le \rho, \tag{12}$$

where ρ_3 is the power of User 3 at each time slot.

3) In $T_1(T_2 - T_3)$ blocks, only User 3 needs a CSI update. In that case, the signaling is:

$$\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2\ , \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3], \tag{13}$$

where $\mathbf{X}_2 \in \mathbb{C}^{N\times N}$ denotes the intended message for User 2. The received signals at the three users are:

$$\mathbf{Y}_\ell = \big[\mathbf{H}_\ell\mathbf{X}_1\mathbf{X}_2, \mathbf{H}_\ell\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3\big] + \mathbf{W}_\ell, \ell = 1, 2, 3. \tag{14}$$

User 2 knows its virtual channel $\mathbf{G}_2 \triangleq \mathbf{H}_2\mathbf{X}_1$ and thus it decodes its intended signal from the first component $\mathbf{H}_2\mathbf{X}_1\mathbf{X}_2$ in the block. User 3, on the other hand, refreshes its CSI via estimating the virtual channel $\mathbf{F}_3 \triangleq \mathbf{H}_3\mathbf{X}_1\mathbf{X}_2$ during the first component of *its* block, and then decodes $\mathbf{X}_3$ during the second component of the block, namely $(\mathbf{H}_3\mathbf{X}_1\mathbf{X}_2)\mathbf{X}_3$. The transmit signal in (13) satisfies the power constraint at the transmitter:

$$\frac{N}{T_3}\rho_p + \frac{T_3 - N}{T_3}(MN^3\hat{\rho}_3) \le \rho, \tag{15}$$

where $\hat{\rho}_3$ is the power of User 3 at each time slot. For the underlying reasons for the normalization of $\hat{\rho}_3$ resulting in the factor MN^3, please see Remark 1.

Using this signaling strategy, normalizing over the entire super interval of $T_1T_2T_3$ time slots, the following rates per channel use for User 1, User 2 and User 3 are respectively achieved:

$$R_1 = \gamma_1\Big(\frac{T_3}{T_1}\Big)\Big(1 - \frac{N}{T_3}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\rho}{\lambda_1}\overline{\mathbf{H}}_1\overline{\mathbf{H}}_1^H\big)\Big] + \Big(\frac{N}{T_2} - \frac{N}{T_1}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\rho_p}{\lambda_1}\overline{\mathbf{H}}_1\overline{\mathbf{H}}_1^H\big)\Big], \tag{16}$$

$$R_2 = \gamma_2\Big(\frac{T_3}{T_1}\Big)\Big(1 - \frac{N}{T_3}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\rho}{\lambda_2}\overline{\mathbf{H}}_2\overline{\mathbf{H}}_2^H\big)\Big] + \Big(\frac{N}{T_3} - \frac{N}{T_2}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\rho_p}{\hat{\lambda}_2}\overline{\mathbf{G}}_2\overline{\mathbf{G}}_2^H\big)\Big], \tag{17}$$

$$R_3 = \Big(1 - \frac{N}{T_3}\Big)\Big(\gamma_3\Big(\frac{T_3}{T_1}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\rho}{\lambda_3}\overline{\mathbf{H}}_3\overline{\mathbf{H}}_3^H\big)\Big] + \Big(\frac{T_3}{T_2} - \frac{T_3}{T_1}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\rho_3}{\hat{\lambda}_3}\overline{\mathbf{G}}_3\overline{\mathbf{G}}_3^H\big)\Big] + \Big(1 - \frac{T_3}{T_2}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I} + \frac{\hat{\rho}_3}{\tilde{\lambda}_3}\overline{\mathbf{F}}_3\overline{\mathbf{F}}_3^H\big)\Big]\Big), \tag{18}$$

where $\overline{\mathbf{H}}_3$, $\overline{\mathbf{G}}_3$ and $\overline{\mathbf{F}}_3$ respectively denote the estimate of true channel $\mathbf{H}_3$, the estimate of virtual channel $\mathbf{G}_3$ and the estimate of virtual channel $\mathbf{F}_3$ for User 3. Also, λ_3, $\hat{\lambda}_3$ and $\tilde{\lambda}_3$ respectively denote the total noise power at receiver 3 over the blocks in which all receivers need pilots, both User 2 and User 3 need pilots, and only User 3 needs pilots. The first term in Equations (16), (17) and (18) are the achieved rates by each user over T_2T_3 blocks in which no pilot interval is reused for carrying data, and the data portion is time-shared among three users with time-sharing factors $\gamma_1, \gamma_2, \gamma_3 \in [0, 1]$ satisfying $\gamma_1 + \gamma_2 + \gamma_3 = 1$. The second term in (16) is the achieved rate by User 1 over $T_3(T_1 - T_2)$ pilot intervals with length N. The second term in (17) is the achieved rate by User 2 over $T_1(T_2 - T_3)$ pilot intervals with length N. The second term in (18) is the achieved rate by User 3 over $T_3(T_1 - T_2)$ blocks in which the pilot intervals are reused for carrying data of User 1. The third term in (18) is the achieved rate by User 3 over the remaining $T_1(T_2 - T_3)$ blocks.

Remark 4: The rates (16), (17) and (18) represent a corner point of the overall rate region, at which all three users are active via superposition. Other operating regimes are also possible, including single-user transmission to each of the three users, and two-user transmission via product superposition, reflected via (8), (9), when one user is turned off. The overall rate region is the convex hull of the single-user rates, the two-user rates, and the three-user rate highlighted via Equations (16), (17) and (18).

In the following, we present an augmented notation and rate expressions for multiple users without elaboration. This economy of expression is made possible because the ideas and tools required for the multi-user case are adequately manifested in the three-user case we developed above. Let $\overline{\mathbf{H}}_\ell$ and $\overline{\mathbf{G}}_\ell$ respectively denote the estimate of the true channel and virtual channel for User ℓ. λ_ℓ and $\hat{\lambda}_\ell$ respectively denote the total noise power at receiver ℓ over the blocks in which all receivers need pilot, and the blocks with virtual channels. $\overline{\mathbf{G}}_{L,i}$ denotes the estimate of virtual channel for User L whose distribution is the product of i Gaussian distributions. $\hat{\rho}_{L,i}$ and $\hat{\lambda}_{L,i}$ respectively denote the transmit power and the total noise power at receiver L in the blocks with virtual channel $\mathbf{G}_{L,i}$. γ_ℓ is the time-sharing factor for User ℓ satisfying $\sum_{\ell=1}^{L}\gamma_\ell = 1$. By employing the strategies and analyses mentioned above in a general case of multi-user downlink system with L receivers, the rate expressions in Eqs. (19), (20) and (21), shown at the bottom of the next page, have been derived.

IV. Achievable Rates Under Coherence Disparity & CSI Feedback: Foundations

Under perfect or imperfect channel state feedback, we propose a novel pilot-domain NOMA by reconciling the distinct requirements of product superposition and beamforming.

Consider a two-user downlink channel whose links have disparity in coherence time. Without loss of generality, we assume $T_1 > T_2$. The two links are wideband with K subcarriers, and both have the same coherence bandwidth. We construct a super interval of length T_1T_2, with all operations repeating every T_1T_2 time slots. User 2 experiences T_1 different channel realizations in the super interval, therefore, User 2 needs T_1 pilots in that duration. User 1, on the other hand, needs only T_2 pilots during this time. The two users' distinct channel estimation needs are met efficiently as follows: Consider the length-T_2 fading blocks of User 2 inside the super interval. In $T_2 < T_1$ of these blocks, both users need to estimate the channel, and the transmitter emits a pilot intended for both users. In these blocks, the transmitted signal at each subcarrier is

$$\mathbf{X}_k = \left[\mathbf{\Pi}, \sum_{\ell=1}^{2} \mathbf{V}_{\ell,k}\sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}\right], \quad k = 1, \ldots, K, \tag{22}$$

where $\mathbf{\Pi}$ denotes the pilot matrix, $\mathbf{S}_{\ell,k} \in \mathbb{C}^{N\times(T_2-M)}$ contains the normalized i.i.d. Gaussian symbols for User ℓ, $\mathbf{V}_{\ell,k} \in \mathbb{C}^{M\times N}$ is the zero-forcing beamforming matrix consisting of unit-norm beamforming vectors for User ℓ [32], [33], and $\rho_{\ell,k}$ is the allocated power to User ℓ satisfying the power constraint ρ per time slot per subcarrier as $N(\rho_{1,k} + \rho_{2,k}) \le \rho$. Let $\mathbf{\Pi} \triangleq \sqrt{\rho}\,\mathbf{I}$. During the first M slots of the block, each user estimates its K subcarrier channels from observations given by Eq. (1). The minimum mean square error (MMSE) estimate for channel gains $\mathbf{H}_{\ell,k}$ is obtained as [34]

$$\overline{\mathbf{H}}_{\ell,k} = \mathbb{E}\big[\mathbf{H}_{\ell,k}\mathbf{Y}_{\ell,k}^H\big]\,\mathbb{E}\big[\mathbf{Y}_{\ell,k}\mathbf{Y}_{\ell,k}^H\big]^{-1}\mathbf{Y}_{\ell,k} = \frac{\sqrt{\rho}}{\rho+N_0}\mathbf{Y}_{\ell,k}. \tag{23}$$

The estimation error is denoted $\mathbf{E}_{\ell,k} = \mathbf{H}_{\ell,k} - \overline{\mathbf{H}}_{\ell,k}$ which is Gaussian with covariance $\sigma^2_{\ell,k}\mathbf{I}$:

$$\sigma^2_{\ell,k} = \frac{N_0}{\rho+N_0}. \tag{24}$$

Each user shares its estimated channel state with the transmitter through a feedback link, which may be perfect or imperfect.

A. Perfect Feedback

We begin by considering a feedback link that is free of noise or errors, even though the channel estimation at each receiver has an error which is modeled as Gaussian noise. Using the channel estimate $\overline{\mathbf{H}}_{\ell,k}$, the transmitter designs the precoder and transmits data over $T_2 - M$ time slots. The received signal at User ℓ is given by

$$\begin{aligned}\mathbf{Y}_{\ell,k} &= \mathbf{H}_{\ell,k}\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{\ell,k}\\ &= \big[\overline{\mathbf{H}}_{\ell,k} + \mathbf{E}_{\ell,k}\big]\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{\ell,k}\\ &\triangleq \sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k} + \mathbf{\Omega}_{\ell,k}, \quad \ell = 1,2,\end{aligned} \tag{25}$$

where $\mathbf{\Omega}_{\ell,k} \triangleq \mathbf{E}_{\ell,k}\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{\ell,k}$ denotes the sum of the additive noise $\mathbf{W}_{\ell,k}$ and the residual channel estimation error. At each time slot, the covariance of $\mathbf{\Omega}_k$ is

$$\begin{aligned}\mathbb{E}\Big[\mathbf{\Omega}_{\ell,k}\,\mathbf{\Omega}_{\ell,k}^H\Big] &= \mathbb{E}\Bigg[\Big(\mathbf{E}_{\ell,k}\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{\ell,k}\Big)\\ &\qquad\times\Big(\mathbf{E}_{\ell,k}\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{\ell,k}\Big)^H\Bigg]\\ &= \mathbb{E}\Bigg[\sum_{i=1}^{2}\mathrm{tr}\Big(\rho_{i,k}\mathbf{V}_{i,k}\mathbf{S}_{i,k}\mathbf{S}_{i,k}^H\mathbf{V}_{i,k}^H\Big)\Bigg]\sigma_k^2\mathbf{I} + N_0\mathbf{I}\\ &= \left(\frac{2\rho N_0 + N_0^2}{\rho+N_0}\right)\mathbf{I},\end{aligned} \tag{26}$$

where we employ the identity $\mathbb{E}[\mathbf{\Psi}\mathbf{A}\mathbf{\Psi}^H] = \mathrm{tr}(\mathbf{A})\mathbf{I}_N$ for any $\mathbf{A} \in \mathbb{C}^{N\times N}$ when the entries of $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ are i.i.d. distributed $\mathcal{CN}(0,1)$ [35], [36]. We further use $\mathbb{E}\big[\mathbf{W}_{\ell,k}\mathbf{W}_{\ell,k}^H\big] = N_0\mathbf{I}$, and $\mathbb{E}\big[\sum_{i=1}^{2}\mathrm{tr}\big(\rho_{i,k}\mathbf{V}_{i,k}\mathbf{S}_{i,k}\mathbf{S}_{i,k}^H\mathbf{V}_{i,k}^H\big)\big] = \rho$ due to the power constraint at the transmitter.

In the remaining $T_1 - T_2$ blocks, each with length T_2, the channel remains unchanged for User l, while User 2 needs to refresh its channel estimate. Therefore, the pilot slots of User 2 can be reused to transmit data for User 1. The transmitted signal is

$$\mathbf{X}_k = \left[\mathbf{X}_{1,k}, \sum_{\ell=1}^{2}\hat{\mathbf{V}}_{\ell,k}\sqrt{\hat{\rho}_{\ell,k}}\,\mathbf{S}_{\ell,k}\right], \tag{27}$$

where $\mathbf{X}_{1,k} \in \mathbb{C}^{M\times M}$ is the signal matrix for User 1,[3] $\mathbf{S}_{\ell,k} \in \mathbb{C}^{N\times(T_2-M)}$ contains the normalized i.i.d. Gaussian symbols for User ℓ, $\hat{\mathbf{V}}_{\ell,k} \in \mathbb{C}^{M\times N}$ is the zero-forcing beamforming

[3] We note that $\mathbf{X}_{1,k}$ is a full-rank square matrix with i.i.d. $\mathcal{CN}(0,1)$ entries.

$$R_1 = \gamma_1\Big(\frac{T_L}{T_1}\Big)\Big(1-\frac{N}{T_L}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I}+\frac{\rho}{\lambda_1}\overline{\mathbf{H}}_1\overline{\mathbf{H}}_1^H\big)\Big] + \Big(\frac{N}{T_2}-\frac{N}{T_1}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I}+\frac{\rho_p}{\lambda_1}\overline{\mathbf{H}}_1\overline{\mathbf{H}}_1^H\big)\Big] \tag{19}$$

$$R_\ell = \gamma_\ell\Big(\frac{T_L}{T_1}\Big)\Big(1-\frac{N}{T_L}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I}+\frac{\rho}{\lambda_\ell}\overline{\mathbf{H}}_\ell\overline{\mathbf{H}}_\ell^H\big)\Big] + \Big(\frac{N}{T_{\ell+1}}-\frac{N}{T_\ell}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I}+\frac{\rho_p}{\hat{\lambda}_\ell}\overline{\mathbf{G}}_\ell\overline{\mathbf{G}}_\ell^H\big)\Big],\ \ \ell = 2,\cdots,L-1 \tag{20}$$

$$R_L = \gamma_L\Big(\frac{T_L}{T_1}\Big)\Big(1-\frac{N}{T_L}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I}+\frac{\rho}{\lambda_L}\overline{\mathbf{H}}_L\overline{\mathbf{H}}_L^H\big)\Big] + \sum_{i=2}^{L}\Big(\frac{T_L}{T_i}-\frac{T_L}{T_{i-1}}\Big)\Big(1-\frac{N}{T_L}\Big)\mathbb{E}\Big[\mathrm{logdet}\big(\mathbf{I}+\frac{\hat{\rho}_{L,i}}{\hat{\lambda}_{L,i}}\overline{\mathbf{G}}_{L,i}\overline{\mathbf{G}}_{L,i}^H\big)\Big] \tag{21}$$

matrix for User ℓ, and $\hat{\rho}_{\ell,k}$ is the allocated power to User ℓ satisfying the power constraint ρ per time slot per subcarrier as $N(\hat{\rho}_{1,k}+\hat{\rho}_{2,k}) \le \rho$. During the pilot interval, User 1 receives a noisy version of $\mathbf{H}_{1,k}\mathbf{X}_{1,k}$ and decodes $\mathbf{X}_{1,k}$, because it already knows $\mathbf{H}_{1,k}$. However, User 2 receives a noisy version of its virtual channel $\mathbf{H}_{2,k}\mathbf{X}_{1,k}$ during this time and attempts to estimate it. Let $\mathbf{G}_{2,k} \triangleq \mathbf{H}_{2,k}\mathbf{X}_{1,k}$. Then, the MMSE estimate of $\mathbf{G}_{2,k}$ is denoted $\overline{\mathbf{G}}_{2,k}$:

$$\overline{\mathbf{G}}_{2,k} = \frac{\rho}{\rho+N_0}\big(\mathbf{G}_{2,k}+\mathbf{W}_{2,k}\big). \tag{28}$$

The estimate of the virtual channel in (28) is perfectly returned to the transmitter. The transmitter obtains the estimate $\hat{\mathbf{H}}_{2,k}$ of the true channel $\mathbf{H}_{2,k}$ multiplying $\overline{\mathbf{G}}_{2,k}$ by $\mathbf{X}_{1,k}^{-1}$ as

$$\hat{\mathbf{H}}_{2,k} = \frac{\rho}{\rho+N_0}\big(\mathbf{H}_{2,k}+\mathbf{W}_{2,k}\mathbf{X}_{1,k}^{-1}\big), \tag{29}$$

with the estimation error

$$\hat{\mathbf{E}}_{2,k} = \mathbf{H}_{2,k}-\hat{\mathbf{H}}_{2,k} = \frac{N_0}{\rho+N_0}\mathbf{H}_{2,k} - \frac{\rho}{\rho+N_0}\mathbf{W}_{2,k}\mathbf{X}_{1,k}^{-1}. \tag{30}$$

The covariance of the error is calculated as

$$\mathbb{E}\Big[\hat{\mathbf{E}}_{2,k}\,\hat{\mathbf{E}}_{2,k}^{H}\Big] = \left(\frac{N_0^2}{(\rho+N_0)^2} + \frac{\rho^2 N_0}{(\rho+N_0)^2}\mathbb{E}\Big[\mathrm{tr}\big(\mathbf{X}_{1,k}^{*}\mathbf{X}_{1,k}\big)^{-1}\Big]\right)\mathbf{I}. \tag{31}$$

This constitutes the covariance of channel estimation error in the final T_1-T_2 blocks, which is the counterpart of $\mathbf{E}_{2,k}$ in T_2 blocks. The noise power calculation for $\mathbf{\Omega}_{2,k}$ is still according to (26), wherein $\mathbf{E}_{2,k}$ is replaced with $\hat{\mathbf{E}}_{2,k}$. This results in Eq. (32), shown at the bottom of the next page. The estimation error for User 1 remains unchanged, i.e., $\hat{\mathbf{\Omega}}_{1,k} = \mathbf{\Omega}_{1,k}$. However, the estimation error for User 2 increases in the last T_1-T_2 blocks, compared with the first T_2 blocks.

B. Imperfect Feedback

We now consider analog feedback [37]: a scaled version of each user's downlink pilot observation is emitted back to the transmitter, and observed under (further) additive Gaussian noise. During T_2 blocks, each with length T_2, both users need downlink training for channel estimation. Both users transmit the scaled version of their pilot observation, given in (1), on the feedback channel. The received signal at the transmitter is given by [37]

$$\begin{aligned}\mathbf{Z}_{\ell,k} &= \frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\Big(\sqrt{\rho}\,\mathbf{H}_{\ell,k}+\mathbf{W}_{\ell,k}\Big)+\mathbf{\Gamma}_k \\ &\triangleq \frac{\rho}{\sqrt{\rho+N_0}}\mathbf{H}_{\ell,k}+\tilde{\mathbf{\Gamma}}_{\ell,k}, \quad \ell=1,2,\end{aligned} \tag{33}$$

where $\mathbf{\Gamma}_k$ denotes the feedback noise and has i.i.d. entries $\mathcal{CN}(0,N_0)$. $\mathbf{\Gamma}_k$ is independent of the pilot observation noise $\mathbf{W}_{\ell,k}$ and the channel gain. The total noise contaminating the channel state knowledge at the transmitter is $\tilde{\mathbf{\Gamma}}_{\ell,k} \triangleq \frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\mathbf{W}_{\ell,k}+\mathbf{\Gamma}_k$. The covariance of $\tilde{\mathbf{\Gamma}}_k$ is

$$\begin{aligned}\mathbb{E}\Big[\tilde{\mathbf{\Gamma}}_{\ell,k}\,\tilde{\mathbf{\Gamma}}_{\ell,k}^{H}\Big] &= \Big(\frac{\rho}{\rho+N_0}\Big)\mathbb{E}\Big[\mathbf{W}_{\ell,k}\mathbf{W}_{\ell,k}^{H}\Big]+\mathbb{E}\Big[\mathbf{\Gamma}_k\mathbf{\Gamma}_k^{H}\Big] \\ &= N_0\Big(\frac{2\rho+N_0}{\rho+N_0}\Big)\mathbf{I}.\end{aligned} \tag{34}$$

The transmitter computes the MMSE estimate of the channel $\mathbf{H}_{\ell,k}$ based on observation $\mathbf{Z}_{\ell,k}$:

$$\overline{\mathbf{H}}_{\ell,k} = \frac{\rho}{\sqrt{\rho+N_0}(\rho+N_0)}\mathbf{Z}_{\ell,k}. \tag{35}$$

The estimation has error

$$\begin{aligned}\tilde{\mathbf{E}}_{\ell,k} = \mathbf{H}_{\ell,k}-\overline{\mathbf{H}}_{\ell,k} &= \frac{2\rho N_0+N_0^2}{(\rho+N_0)^2}\mathbf{H}_{\ell,k} \\ &\quad - \frac{\rho}{\sqrt{\rho+N_0}(\rho+N_0)}\tilde{\mathbf{\Gamma}}_k,\end{aligned} \tag{36}$$

with covariance

$$\mathbb{E}\Big[\tilde{\mathbf{E}}_{\ell,k}\,\tilde{\mathbf{E}}_{\ell,k}^{H}\Big] = \frac{(2\rho N_0+N_0)^2+\rho^2 N_0(2\rho+N_0)}{(\rho+N_0)^4}\mathbf{I}. \tag{37}$$

The transmitter designs a zero-forcing precoder based on $\overline{\mathbf{H}}_{\ell,k}$ to transmit data over T_2-M time slots. Given the transmitted signal in (22), the received signal at User ℓ is given by

$$\begin{aligned}\mathbf{Y}_{\ell,k} &= \mathbf{H}_{\ell,k}\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k}+\mathbf{W}_{\ell,k} \\ &= \big[\overline{\mathbf{H}}_{\ell,k}+\tilde{\mathbf{E}}_{\ell,k}\big]\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k}+\mathbf{W}_{\ell,k} \\ &\triangleq \sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}+\tilde{\mathbf{\Omega}}_{\ell,k}, \quad \ell=1,2,\end{aligned} \tag{38}$$

where $\tilde{\mathbf{\Omega}}_{\ell,k} \triangleq \tilde{\mathbf{E}}_{\ell,k}\sum_{i=1}^{2}\mathbf{V}_{i,k}\sqrt{\rho_{i,k}}\,\mathbf{S}_{i,k}+\mathbf{W}_{\ell,k}$ denotes the sum of the additive noise $\mathbf{W}_{\ell,k}$ and residual channel estimation error, whose covariance is calculated similarly to (26):

$$\mathbb{E}\Big[\tilde{\mathbf{\Omega}}_{\ell,k}\,\tilde{\mathbf{\Omega}}_{\ell,k}^{H}\Big] = \left(\frac{\rho(2\rho N_0+N_0)^2+\rho^3 N_0(2\rho+N_0)}{(\rho+N_0)^4}+N_0\right)\mathbf{I}. \tag{39}$$

In the remaining T_1-T_2 blocks, each with length T_2, the structure of the transmitted signal is expressed by (27) where the message of User 1 is sent over the pilot slots of User 2. During the pilot interval, User 1 receives a noisy version of $\mathbf{H}_{1,k}\mathbf{X}_{1,k}$ and decodes the message therein. User 2 estimates the virtual channel $\mathbf{G}_{2,k}$ during this time and returns the estimate to the transmitter via an analog feedback. The received signal at the transmitter is given by

$$\begin{aligned}\mathbf{Z}_{2,k} &= \frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}(\mathbf{G}_{2,k}+\mathbf{W}_{2,k})+\mathbf{\Gamma}_k \\ &\triangleq \frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\mathbf{G}_{2,k}+\tilde{\mathbf{\Gamma}}_{2,k},\end{aligned} \tag{40}$$

where $\tilde{\mathbf{\Gamma}}_{2,k} \triangleq \frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\mathbf{W}_{2,k}+\mathbf{\Gamma}_k$ denotes the total noise at the transmitter and its covariance is calculated in (34). The transmitter first computes MMSE estimate of the virtual channel

$$\check{\mathbf{G}}_{2,k} = \frac{\rho\sqrt{\rho(\rho+N_0)}}{\rho^2+N_0(2\rho+N_0)}\Big(\frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\mathbf{G}_{2,k}+\tilde{\mathbf{\Gamma}}_{2,k}\Big). \tag{41}$$

Then, it obtains the estimate of the true channel $\mathbf{H}_{2,k}$ multiplying $\check{\mathbf{G}}_{2,k}$ by $\mathbf{X}_{1,k}^{-1}$

$$\check{\mathbf{H}}_{2,k} = \frac{\rho\sqrt{\rho(\rho+N_0)}}{\rho^2+N_0(2\rho+N_0)}\Big(\frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\mathbf{H}_{2,k}+\tilde{\mathbf{\Gamma}}_{2,k}\mathbf{X}_{1,k}^{-1}\Big), \tag{42}$$

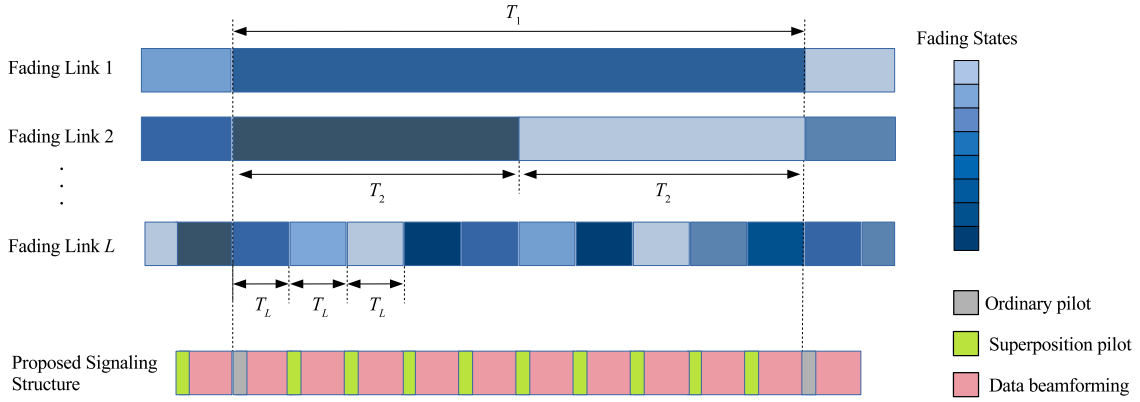

Fig. 4. Coherence time and signaling structure in the presence of CSI feedback. All transmitted data, including during superimposed pilots, employ beamforming. Achievable rates in this structure are represented by Eqs. (59) and (60), and its full scope is highlighted in Remark 5.

with the estimation error

$$\check{\mathbf{E}}_{2,k} = \mathbf{H}_{2,k} - \check{\mathbf{H}}_{2,k} = \frac{N_0(2\rho+N_0)}{\rho^2+N_0(2\rho+N_0)}\mathbf{H}_{2,k} - \frac{\rho\sqrt{\rho(\rho+N_0)}}{\rho^2+N_0(2\rho+N_0)}\tilde{\mathbf{\Gamma}}_{2,k}\mathbf{X}_{1,k}^{-1}. \quad (43)$$

The covariance of the error is calculated in Eq. (44), shown at the bottom of the next page. This represents the covariance of channel estimation error in the final $T_1 - T_2$ blocks. The noise power calculation for User 2 is still based on (32), but with $\hat{\mathbf{E}}_{2,k}$ replaced by $\check{\mathbf{E}}_{2,k}$. This results in Eq. (45), shown at the bottom of the next page. The estimation error for User 1 remains unchanged, i.e., $\check{\mathbf{\Omega}}_{1,k} = \tilde{\mathbf{\Omega}}_{1,k}$.

We now outline an accounting of time slots required for rate calculations. The first M time slots in every length-T_2 block are referred to as *pilot phase,* while the remaining $T_2 - M$ time slots are referred to as *data phase*. User 1, within one super interval with length T_1T_2, has $T_1(T_2-M)$ time slots in data phase, but also can receive data during $(T_1-T_2)M$ time slots under pilot phase. User 2 only has $T_1(T_2-M)$ time slots in data phase within the super interval. Combining all this, the achievable rates for User 1 and User 2 are

$$R_1 = \sum_{k=1}^{K} \frac{T_2}{T_1}\Big(1-\frac{M}{T_2}\Big)\log\Big(1+\frac{\rho_{1,k}}{\lambda_{1,k}}\Big) + \Big(1-\frac{T_2}{T_1}\Big)\Big(1-\frac{M}{T_2}\Big)\log\Big(1+\frac{\hat{\rho}_{1,k}}{\hat{\lambda}_{1,k}}\Big) + \Big(1-\frac{T_2}{T_1}\Big)\Big(\frac{M}{T_2}\Big)\log\Big(1+\frac{\rho}{\sigma_{1,k}^2+N_0}\Big), \quad (46)$$

$$R_2 = \sum_{k=1}^{K} \frac{T_2}{T_1}(1-\frac{M}{T_2})\log\Big(1+\frac{\rho_{2,k}}{\lambda_{2,k}}\Big) + \Big(1-\frac{T_2}{T_1}\Big)\Big(1-\frac{M}{T_2}\Big)\log\Big(1+\frac{\hat{\rho}_{2,k}}{\hat{\lambda}_{2,k}}\Big), \quad (47)$$

where $\lambda_{\ell,k}$ and $\hat{\lambda}_{\ell,k}$ denote the total noise power (i.e., the sum of receiver noise and channel estimation error) at receiver ℓ, respectively during T_2 blocks and the remaining $T_1 - T_2$ blocks. The rate in (46) consists of three terms: the first term refers to the achieved rate over the data phase of T_2 blocks in which both users update their channel state. The second and third terms refer to the rates over the data and pilot phases of $T_1 - T_2$ blocks in which only User 2 updates its channel state. The rate in (47) has two terms that represent the rate over the data phase of T_2 blocks and remaining $T_1 - T_2$ blocks, respectively.

V. Coherence Disparity and CSI Feedback: Multi-User Results

This section presents achievable rate results under coherence disparity and CSI feedback, with two important new features: first, the number of users is generalized from two users to L users, which exposes certain signaling combinations that were not available in Section IV. Second, we generalize the operation of the system to the case where the multiple users have unequal coherence time *as well as* unequal coherence bandwidth.

We should caution that the signaling schemes shown in Fig. 3 were developed in the absence of CSI for two users, and are no longer valid for this section. The scope and complexity of multi-user signaling does not lend itself to producing a similar, simple, diagram. As is common in other multi-user results in the literature, we rely on rate expressions that are further elaborated and clarified via follow-up remarks.

A. Mismatched Coherence Times

Consider a multi-user downlink system with L receivers, where the links have the same coherence bandwidths but unequal coherence times. The users are ordered based on their coherence times, in descending order (see Fig. 4). To focus on

$$\mathbb{E}\Big[\hat{\mathbf{\Omega}}_{2,k}\,\hat{\mathbf{\Omega}}_{2,k}^H\Big] = \mathbb{E}\Bigg[\Big(\hat{\mathbf{E}}_{2,k}\sum_{i=1}^{2}\hat{\mathbf{V}}_{i,k}\sqrt{\hat{\rho}_{i,k}}\,\mathbf{S}_{i,k}+\mathbf{W}_{2,k}\Big)\Big(\hat{\mathbf{E}}_{2,k}\sum_{i=1}^{2}\hat{\mathbf{V}}_{i,k}\sqrt{\hat{\rho}_{i,k}}\,\mathbf{S}_{i,k}+\mathbf{W}_{2,k}\Big)^H\Bigg] = \frac{N_0}{(\rho+N_0)^2}\Big((\rho+N_0)^2+\rho N_0+\rho^3\mathbb{E}\Big[\mathrm{tr}\big(\mathbf{X}_{1,k}^*\mathbf{X}_{1,k}\big)^{-1}\Big]\Big)\mathbf{I} \quad (32)$$

the essential ideas, we initially assume integer coherence time ratios; this assumption is later removed. The integer coherence time ratios allow us to concentrate on a single time period of length T_1 at each subcarrier, with all operations repeating every T_1 time slots. User L with the smallest coherence time T_L experiences $\frac{T_1}{T_L}$ different channel realizations during this duration. User 1, on the other hand, only needs one pilot during this time since its channel realization remains unchanged. Other users fall somewhere in between. After each pilot interval, some users will refresh their channel estimates, and all new channel estimates are fed back to the transmitter. The transmitter then updates its beamforming matrix based on these estimated channels.

Whenever a user reuses a pilot interval for data transmission, it will be unable to estimate the channel during that interval. Therefore, we need to keep track of the pilot intervals required for each user, as well as the number of data transmission opportunities available for that user. At each candidate pilot interval, any User ℓ may have a channel that either transitioned to a new value since the last channel estimation or remained the same, depending on T_ℓ. If the link for User ℓ experiences a new realization since the last refresh, the user needs to estimate the new channel and therefore requires the pilot, and cannot reuse the pilot interval for data transmission. However, if the channel remains unchanged since the last refresh, User ℓ does not require another pilot at that time. The same effect holds for all users. With this, when all users need a pilot (i.e., during first T_L time slots), the transmitted signal is

$$\mathbf{X}_k = \left[\mathbf{\Pi}, \sum_{\ell=1}^{L} \mathbf{\Psi}_{\ell,k}\sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}\right], \tag{48}$$

where $\mathbf{\Pi} \in \mathbb{C}^{M\times M}$ is a pilot matrix, $\mathbf{S}_{\ell,k} \in \mathbb{C}^{N\times(T_L-M)}$ is the information carrying signal for User ℓ and contains normalized i.i.d. Gaussian symbols, $\mathbf{\Psi}_{\ell,k} \in \mathbb{C}^{M\times N}$ is the zero-forcing beamforming matrix for User ℓ, and $\rho_{\ell,k}$ is the allocated power to User ℓ satisfying the power constraint ρ per time slot per subcarrier as $N\sum_{\ell=1}^{L}\rho_{\ell,k} \le \rho$. The received signal at User ℓ is

$$\mathbf{Y}_{\ell,k} = \left[\mathbf{H}_{\ell,k}\mathbf{\Pi}, \mathbf{H}_{\ell,k}\sum_{\ell=1}^{L} \mathbf{\Psi}_{\ell,k}\sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}\right] + \mathbf{W}_{\ell,k}. \tag{49}$$

During the first M time slots of the block, User ℓ estimates its link gain and sends it back to the transmitter. The MMSE estimate of channel $\mathbf{H}_{\ell,k}$ is given in Eq. (23). The transmitter then designs the beamforming matrices $\mathbf{\Psi}_{\ell,k}$ to transmit data to all users during the data phase of the block with length $T_L - M$.

The remaining $T_1 - T_L$ time slots are divided into $\frac{T_1}{T_L} - 1$ blocks, each with length T_L (see Fig. 4). In each of these blocks, the transmitted signal has the following form:

$$\mathbf{X}_k = \left[\mathbf{U}_k, \sum_{\ell=1}^{L} \mathbf{\Psi}_{\ell,k}\sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}\right]. \tag{50}$$

The main difference between this equation and Eq. (48) is the presence of $\mathbf{U}_k$, a signal component designed to serve two purposes: carrying data for some users with unchanged channels and enabling channel estimation for other users. To achieve this, we construct the full-rank $M \times M$ matrix $\mathbf{U}_k$ using zero-forcing beamforming over pilot slots:

$$\mathbf{U}_k = \sum_{\ell=1}^{L} \hat{\mathbf{\Psi}}_{\ell,k}\sqrt{\hat{\rho}_{\ell,k}}\,\hat{\mathbf{S}}_{\ell,k}, \tag{51}$$

where $\hat{\mathbf{\Psi}}_k$ denotes the zero-forcing beamforming matrix in the pilot phase. For the users that are *not* participating in data transmission in this pilot slot, we assume $\hat{\mathbf{S}}_{\ell,k} = \mathbf{0}$. The corresponding received values at User ℓ is

$$\mathbf{Y}_{\ell,k} = \left[\mathbf{H}_{\ell,k}\mathbf{U}_k, \mathbf{H}_{\ell,k}\sum_{\ell=1}^{L} \mathbf{\Psi}_{\ell,k}\sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}\right] + \mathbf{W}_{\ell,k}. \tag{52}$$

When the transmitter is emitting $\mathbf{U}_k$, User ℓ receives a noisy version of $\mathbf{H}_{\ell,k}\mathbf{U}_k$. If User ℓ already knows $\mathbf{H}_{\ell,k}$ and does not require channel estimation at this time, it can attempt to decode $\mathbf{U}_k$. If User ℓ does *not* know $\mathbf{H}_{\ell,k}$ at this time, it will attempt to estimate $\mathbf{\Theta}_{\ell,k} = \mathbf{H}_{\ell,k}\mathbf{U}_k$ and feed it back to the transmitter. The transmitter has full knowledge of transmitted value $\mathbf{U}_k$, therefore can estimate the true channel $\mathbf{H}_{\ell,k}$ and use it for beamforming. Here, the estimation of the virtual and true channels follows similar ideas as those discussed in Section IV under both perfect and imperfect feedback links, and is omitted for brevity. Utilizing an approach similar to Section IV, during blocks with reused pilots, the MMSE channel estimate $\hat{\mathbf{H}}_{\ell,k}$ under perfect feedback and the channel estimate $\check{\mathbf{H}}_{\ell,k}$ under imperfect feedback are given by:

$$\hat{\mathbf{H}}_{\ell,k} = \frac{\rho}{\rho+N_0}\left(\mathbf{H}_{\ell,k} + \mathbf{W}_{\ell,k}\mathbf{U}_k^{-1}\right), \tag{53}$$

$$\check{\mathbf{H}}_{\ell,k} = \frac{\rho\sqrt{\rho(\rho+N_0)}}{\rho^2+N_0(2\rho+N_0)} \times \left(\frac{\sqrt{\rho}}{\sqrt{\rho+N_0}}\mathbf{H}_{\ell,k} + \tilde{\mathbf{\Gamma}}_k\mathbf{U}_k^{-1}\right). \tag{54}$$

$$\mathbb{E}\left[\check{\mathbf{E}}_{2,k}\,\check{\mathbf{E}}_{2,k}^H\right] = \left(\left(\frac{N_0(2\rho+N_0)}{\rho^2+N_0(2\rho+N_0)}\right)^2 + \left(\frac{\rho\sqrt{\rho(\rho+N_0)}}{\rho^2+N_0(2\rho+N_0)}\right)^2\left(\frac{2\rho N_0+N_0^2}{\rho+N_0}\right)\mathbb{E}\left[\mathrm{tr}\left(\mathbf{X}_{1,k}^*\mathbf{X}_{1,k}^T\right)^{-1}\right]\right)\mathbf{I} \tag{44}$$

$$\begin{aligned}\mathbb{E}\left[\check{\mathbf{\Omega}}_{2,k}\,\check{\mathbf{\Omega}}_{2,k}^H\right] &= \mathbb{E}\left[\left(\check{\mathbf{E}}_{2,k}\sum_{i=1}^{2}\hat{\mathbf{V}}_{i,k}\sqrt{\hat{\rho}_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{2,k}\right)\left(\check{\mathbf{E}}_{2,k}\sum_{i=1}^{2}\hat{\mathbf{V}}_{i,k}\sqrt{\hat{\rho}_{i,k}}\,\mathbf{S}_{i,k} + \mathbf{W}_{2,k}\right)^H\right]\\ &= \left(\left(\frac{\sqrt{\rho}N_0(2\rho+N_0)}{\rho^2+N_0(2\rho+N_0)}\right)^2 + \left(\frac{\rho^2\sqrt{2\rho N_0+N_0^2}}{\rho^2+N_0(2\rho+N_0)}\right)^2\mathbb{E}\left[\mathrm{tr}\left(\mathbf{X}_{1,k}^*\mathbf{X}_{1,k}^T\right)^{-1}\right] + N_0\right)\mathbf{I}\end{aligned} \tag{45}$$

Let $\hat{\mathbf{E}}_{\ell,k} \triangleq \mathbf{H}_{\ell,k} - \hat{\mathbf{H}}_{\ell,k}$ and $\check{\mathbf{E}}_{\ell,k} \triangleq \mathbf{H}_{\ell,k} - \check{\mathbf{H}}_{\ell,k}$. Then, we have

$$\mathbb{E}\Big[\hat{\mathbf{E}}_{\ell,k}\hat{\mathbf{E}}_{\ell,k}^H\Big] = \frac{N_0}{(\rho+N_0)^2}\Big(N_0 + \rho^2\,\mathbb{E}\Big[\mathrm{tr}\big(\mathbf{U}_k^*\mathbf{U}_k^T\big)^{-1}\Big]\Big)\mathbf{I}, \tag{55}$$

$$\mathbb{E}\Big[\check{\mathbf{E}}_{\ell,k}\check{\mathbf{E}}_{\ell,k}^H\Big] = \frac{N_0(2\rho+N_0)}{(\rho+N_0)^4} \times \Big(N_0(2\rho+N_0) + \rho^3\mathbb{E}\Big[\mathrm{tr}\Big(\mathbf{U}_k^*\mathbf{U}_k^T\Big)^{-1}\Big]\Big)\mathbf{I}, \tag{56}$$

which represent the covariance of the channel estimation errors in the blocks where pilots and data are superimposed. Similar to (32) and (45), the noise powers are calculated as follows:

$$\mathbb{E}\Big[\hat{\mathbf{\Omega}}_{\ell,k}\hat{\mathbf{\Omega}}_{\ell,k}^H\Big] = \frac{N_0}{(\rho+N_0)^2} \times \Big(\rho^2+N_0^2+3\rho N_0+\rho^3\mathbb{E}\Big[\mathrm{tr}\big(\mathbf{U}_k^*\mathbf{U}_k^T\big)^{-1}\Big]\Big)\mathbf{I}, \tag{57}$$

$$\mathbb{E}\Big[\check{\mathbf{\Omega}}_{\ell,k}\check{\mathbf{\Omega}}_{\ell,k}^H\Big] = \Big(N_0 + \Big(\frac{\rho N_0(2\rho+N_0)}{(\rho+N_0)^4}\Big) \times \Big(N_0^2 + 2\rho N_0 + \rho^3\,\mathbb{E}\Big[\mathrm{tr}\big(\mathbf{U}_k^*\mathbf{U}_k^T\big)^{-1}\Big]\Big)\Big)\mathbf{I}. \tag{58}$$

For calculating the achievable rates per channel use, we must account for the number of time slots in each signaling category. User ℓ has $\frac{T_\ell}{T_L}$ blocks, each with length T_L, within its coherence time T_ℓ. It receives data over $\frac{T_\ell}{T_L}(T_L - M)$ data slots. The proposed scheme further provides the opportunity for User ℓ to receive data during $(\frac{T_\ell}{T_L} - 1)M$ pilot slots. With this, the rate per channel use for User ℓ is achieved as

$$R_\ell = \frac{M}{T_\ell}\sum_{k=1}^{K}\sum_{i=1}^{\frac{T_\ell}{T_L}-1}\log\big(1+\frac{\hat{\rho}_{i,k}}{\hat{\lambda}_{i,k}}\big) + \Big(\frac{T_L-M}{T_\ell}\Big)\sum_{k=1}^{K}\sum_{j=1}^{\frac{T_\ell}{T_L}}\log\big(1+\frac{\rho_{j,k}}{\lambda_{j,k}}\big), \tag{59}$$

where $\lambda_{j,k}$ denotes the total noise power at User ℓ during the data phase of block j, $\hat{\lambda}_{i,k}$ denotes the total noise power at User ℓ during pilot i. Both $\lambda_{j,k}$ and $\hat{\lambda}_{i,k}$ need to be calculated in each block based on the receiver noise and channel estimation error. $\hat{\rho}_{i,k}$ and $\rho_{j,k}$ denote the allocated powers to User ℓ over pilot i and the data phase of block j, respectively. In Eq. (59), the first term corresponds to the achieved rate over $(\frac{T_\ell}{T_L} - 1)$ pilot phases with length M. The second term represents the achieved rate over the data phases, which can be obtained by averaging over $\frac{T_\ell}{T_L}$ data phases each with length $T_L - M$. User L, which has the shortest coherence time T_L, does not receive data over the pilot phase, as its channel estimate needs to be refreshed every T_L time slots. Therefore, the achievable rate per channel use for User L is

$$R_L = \Big(1 - \frac{M}{T_L}\Big)\sum_{k=1}^{K}\log\big(1+\frac{\rho_{L,k}}{\lambda_{L,k}}\big), \tag{60}$$

where $\lambda_{L,k}$ denotes the total noise power at the receiver.

Remark 5: Equations (59) and (60) together describe individual rates for all users. By varying the allocated powers subject to the overall power constraint ρ, the Equations (59) and (60) yield the overall rate region.

Remark 6: The rates in (59) and (60) are calculated under the assumption that the number of transmit antennas is larger than the number of users. However, our proposed method remains applicable when the number of users exceeds the number of transmit antennas, via the following generalization: When $L > M$, the scheduling algorithm will choose M users that, together, will employ beamforming and therefore will not interfere on each other. For scheduling purposes, these M users will constitute one "virtual" user for the purposes of time slot occupation. Other users will interfere on each other and on this "virtual" user, and their operation will be according to time-sharing, or more generally, according to the principles outlined in Section III. For example, either the "virtual" user or one of the other users can at any one time utilize the pilot slot. The combining of rate expressions in Section III and IV is straightforward but tedious and is omitted in the interest of brevity.

B. Mismatched Coherence Bandwidths

Consider a multi-user downlink channel with L receivers, where the links exhibit disparity only in coherence bandwidths. We separate and denote the set of users with frequency-flat channels:

$$\mathbb{J} \triangleq \{j \mid \mathbf{H}_{j,k} = \mathbf{H}_{j,k'}\ \forall k,k'\}.$$

If $j \in \mathbb{J}$, then $\mathbf{H}_{j,k}$ is constant across subcarriers and can be measured on any subcarrier $k = d$. However, if $j \notin \mathbb{J}$, then $\mathbf{H}_{j,k}$ is not constant across subcarriers, and in principle needs to be measured across all subcarriers.[4] At any subcarrier $k \neq d$, the pilot slots of the users with frequency-selective channels can be reused for data transmission to the users in the set $\mathbb{J}$. We assume that all links have the same coherence time T. Therefore, the transmitted signal is

$$\mathbf{X}_k = \Big[\mathbf{U}_k, \sum_{\ell=1}^{L}\mathbf{\Psi}_{\ell,k}\sqrt{\rho_{\ell,k}}\,\mathbf{S}_{\ell,k}\Big], \quad k \neq d, \tag{61}$$

where $\mathbf{U}_k$ is the combined pilot/data transmission in the pilot slots, constructed according to Eq. (51).

During the first M time slots, frequency-flat users decode their messages, while frequency-selective Users $j \notin \mathbb{J}$ estimate their virtual channel $\mathbf{H}_{j,k}\mathbf{U}_k$ and return it to the transmitter. The transmitter then calculates the beamforming matrix $\mathbf{\Psi}_k$ and transmits to all users during the remaining $T - M$ time

[4] In the model utilized in this paper, frequency-selective channels have subcarrier link gains that are statistically independent. This paper eschews consideration of correlated subcarriers for economy of expression, and concentrating on essential ideas. For an example of analyzing correlated subcarriers in the context of product superposition, see [28].

slots of the block. Similar to Section V-A, the rates per channel use are derived as Eq. (62), shown at the bottom of the page.

C. Arbitrary Coherence Times and Coherence Bandwidths

Consider a multi-user downlink channel with L receivers experiencing arbitrary coherence times and coherence bandwidths. Every User ℓ has a constant channel within a block with duration T_ℓ and bandwidth B_ℓ (see Fig. 2). For signaling, we consider a super block whose duration and bandwidth are the least common multiple (LCM) of coherence times and coherence bandwidths, respectively. Let $T_t \triangleq \text{lcm}(T_1, \cdots, T_L)$ and $B_t \triangleq \text{lcm}(B_1, \cdots, B_L)$ denote the duration and the bandwidth of the super block, respectively. Without loss of generality, we assume the users are indexed such that $T_L B_L \leq T_{L-1} B_{L-1} \leq \cdots \leq T_1 B_1$. By inspection, it is obvious that if two users do not have identical coherence time-bandwidth blocks, their boundaries will not be identical either, and transition boundaries of some users will potentially fall within the coherence time-bandwidth blocks of others. It is in these occasions, where one user experiences a coherence transition while another does not, that product superposition can be gainfully applied for pilot-domain NOMA.

We begin by studying a three-user scenario, the simplest multi-user system where the main features and interactions of the users manifest themselves. We present the transmission scheme and resulting achievable rates for this special case to highlight the ideas and intuition behind our approach. We then extend these ideas to describe the achievable rates in the general L-receiver scenario.

In this three-user downlink channel, whose links have unequal coherence time-bandwidth blocks, User 3 has the smallest block with coherence time T_3 and coherence bandwidth B_3. The proposed transmission scheme begins by building a super block with duration $T_t = \text{lcm}(T_1, T_2, T_3)$ and bandwidth $B_t = \text{lcm}(B_1, B_2, B_3)$. As User 3 has the smallest block, we require $\Lambda_3 \triangleq \frac{T_t B_t}{T_3 B_3}$ pilot transmissions within the super block. We design the transmission scheme over Λ_3 different sub-blocks, each containing $T_3 B_3$ channel uses. In the first sub-block, no prior CSI information exists, and all receivers utilize the M transmitted pilots to estimate their respective channels. In our model, the channel gains are fed back to the transmitter instantaneously. The transmitter then calculates a precoder for data transmission. In each of the remaining $\Lambda_3 - 1$ sub-blocks, the channel state changes for User 3, but at each instance the channel gain may or may not change for User 1 and User 2. In each instance when a channel gain need not be updated, we utilize product superposition to reuse the pilot slot for data transmission. The rate of User 1 and User 2, per channel use, is given by:

$$R_\ell = \frac{1}{\Lambda_\ell}\Big(\frac{M}{T_\ell}\Big)\sum_{i=1}^{\Lambda_3-\Lambda_\ell}\log\Big(1+\frac{\hat{\rho}_{\ell,i}}{\hat{\lambda}_{\ell,i}}\Big) + \frac{1}{\Lambda_\ell}\sum_{j=1}^{\Lambda_\ell}\Big(\frac{T_\ell-\alpha_{\ell,j}M}{T_\ell}\Big)\log\Big(1+\frac{\rho_{\ell,j}}{\lambda_{\ell,j}}\Big),\ \ell=1,2, \tag{63}$$

where $\Lambda_\ell \triangleq \frac{T_t B_t}{T_\ell B_\ell}$ is the number of coherence block for User ℓ within the super block. For each User ℓ, $\lambda_{\ell,j}$ is the sum of the noise power and channel estimation error in the data phase of the block j, and $\hat{\lambda}_{\ell,i}$ is the corresponding value for the pilot phase i. $\alpha_{\ell,j}$ is the number of pilot phases, and $\alpha_{\ell,j}M$ is the total number of pilot symbols, used for User ℓ in block j. $\rho_{\ell,j}$ denotes the allocated power to User ℓ over the data phase of block j and $\hat{\rho}_{\ell,i}$ denotes the allocated power to User ℓ during pilot i. User 3 needs all channel updates and does not receive data over any pilot phase. Thus, its rate per channel use is

$$R_3 = \frac{1}{\Lambda_3}\Big(1-\frac{M}{T_3}\Big)\sum_{i=1}^{\Lambda_3}\log\Big(1+\frac{\rho_{3,i}}{\lambda_{3,i}}\Big), \tag{64}$$

where $\lambda_{3,i}$ denotes the total noise power at User 3 and $\rho_{3,i}$ denotes the allocated power to that user. Equations (64) and (63) together give the achievable rate region.

To reveal the essential features of the achievable rates calculated above, and highlight the required pilot placements, including superimposed pilots, we provide an example. A three-user downlink channel is considered, whose coherence blocks are depicted in Fig. 5. User 3 has the smallest block consisting of $2T_0B_0$ channel uses, while User 2 and User 1 respectively have blocks of $3B_0T_0$ and $4B_0T_0$ channel uses. To create the super block, we first calculate its duration T_t and bandwidth B_t:

$$T_t = \text{lcm}(T_1, T_2, T_3) = 6T_0$$
$$B_t = \text{lcm}(B_1, B_2, B_3) = 2B_0.$$

The super block (see Fig. 5) contains six coherence blocks for User 3, three coherence blocks for User 1, and four coherence blocks for User 2. Fig. 5 shows the placement of pilots within the super block.

User 3 needs all six pilot phases to refresh its channel estimates, while User 1 and User 2 require three and four pilot phases, respectively. Therefore, for User 1 and User 2, any unneeded pilot slots can be reused for data transmission, through product superposition. The rates achieved for User 1 and User 2 in this example are as follows:

$$R_1 = \frac{1}{3}\Big(\frac{M}{T_1}\Big)\sum_{i=1}^{3}\log\Big(1+\frac{\hat{\rho}_{1,i}}{\hat{\lambda}_{1,i}}\Big)$$

$$R_j = \begin{cases} \sum\limits_{k\neq d}\Big(1-\frac{M}{T}\Big)\log\Big(1+\frac{\rho_{j,k}}{\lambda_{j,k}}\Big)+\Big(1-\frac{M}{T}\Big)\log\Big(1+\frac{\rho_{j,d}}{\lambda_{j,d}}\Big), & j\notin\mathbb{J} \\ \sum\limits_{k\neq d}\frac{M}{T}\log\Big(1+\frac{\hat{\rho}_{j,k}}{\sigma^2_{j,k}+N_0}\Big)+\sum\limits_{k\neq d}\Big(1-\frac{M}{T}\Big)\log\Big(1+\frac{\rho_{j,k}}{\lambda_{j,k}}\Big)+\Big(1-\frac{M}{T}\Big)\log\Big(1+\frac{\rho_{j,d}}{\lambda_{j,d}}\Big), & j\in\mathbb{J} \end{cases} \tag{62}$$

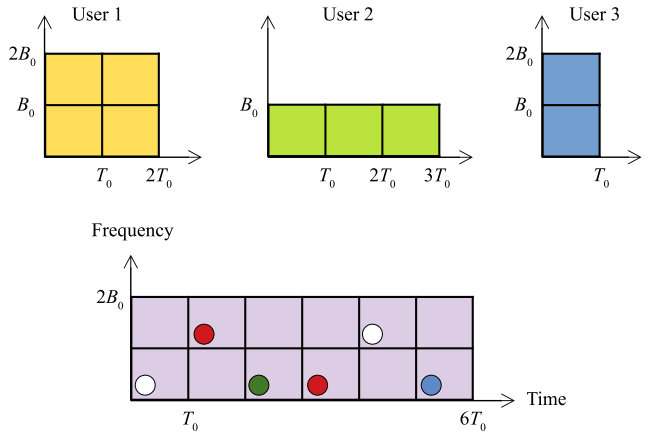

Fig. 5. Individual coherence blocks and pilot placement inside the super block. The pilots are color-coded as: (a) White indicates the pilot slots that do not admit reuse, (b) Red indicates the pilot slots reused for data transmission to User 1, (c) Green represents the pilot slots reused for data transmission to User 2, and (d) Blue represents the pilot slots reused for data transmission to Users 1 and 2.

$$+\frac{1}{3}\Big(\frac{T_1-2M}{T_1}\Big)\sum_{j=1}^{3}\log\Big(1+\frac{\rho_{1,j}}{\lambda_{1,j}}\Big), \tag{65}$$

$$R_2=\frac{1}{4}\Big(\frac{M}{T_2}\Big)\sum_{i=1}^{2}\log\Big(1+\frac{\hat{\rho}_{2,i}}{\hat{\lambda}_{2,i}}\Big)$$
$$+\frac{1}{4}\Big(\frac{T_2-2M}{T_2}\Big)\sum_{j=1}^{4}\log\Big(1+\frac{\rho_{2,j}}{\lambda_{2,j}}\Big). \tag{66}$$

User 3 has the following rate:

$$R_3=\frac{1}{6}\Big(1-\frac{M}{T_3}\Big)\sum_{i=1}^{6}\log\Big(1+\frac{\rho_{3,i}}{\lambda_{3,i}}\Big). \tag{67}$$

Following the same approach, the rate expressions for L receivers having arbitrary coherence blocks can be derived. Without loss of generality, $T_LB_L \le \cdots \le T_1B_1$. A super block with duration $T_t = \mathrm{lcm}(T_1,\cdots,T_L)$ and bandwidth $B_t=\mathrm{lcm}(B_1,\cdots,B_L)$ is considered, divided into $\frac{T_tB_t}{T_LB_L}$ sub-blocks, each with T_LB_L channel uses. User ℓ has $\Lambda_\ell \triangleq \frac{T_tB_t}{T_\ell B_\ell}$ coherence blocks within the super block. User L cannot reuse pilot slots and has rate:

$$R_L=\frac{1}{\Lambda_L}\Big(1-\frac{M}{T_L}\Big)\sum_{i=1}^{\Lambda_L}\log\Big(1+\frac{\rho_{L,i}}{\lambda_{L,i}}\Big). \tag{68}$$

The rate for User $\ell \neq L$ is

$$R_\ell=\frac{1}{\Lambda_\ell}\Big(\frac{M}{T_\ell}\Big)\sum_{i=1}^{\Lambda_L-\Lambda_\ell}\log\Big(1+\frac{\hat{\rho}_{\ell,i}}{\hat{\lambda}_{\ell,i}}\Big)$$
$$+\frac{1}{\Lambda_\ell}\sum_{j=1}^{\Lambda_\ell}\Big(\frac{T_\ell-\alpha_{\ell,j}M}{T_\ell}\Big)\log\Big(1+\frac{\rho_{\ell,j}}{\lambda_{\ell,j}}\Big). \tag{69}$$

The details of the derivation are omitted, since they closely follow the three-user case.

Remark 7: The complexity of our technique is $\mathcal{O}(LMN+M^2+\ell' M^2+2\ell' MN^3)$. The first term corresponds to legacy beamforming over the data phase. The second term represents encoding the $M\times M$ matrix $\mathbf{U}_k$ over pilot slots, the third term is for computing the channel estimates at the transmitter for ℓ' users (Eq. (53) for perfect feedback, Eq. (54) for imperfect feedback). The fourth term is for MMSE estimation of the virtual channel at the transmitter and receiver. The additional complexity in our technique is represented in the final three terms, which is directly proportional to additional rate provided. Therefore, on a per-rate basis, our technique has comparable complexity with legacy techniques.

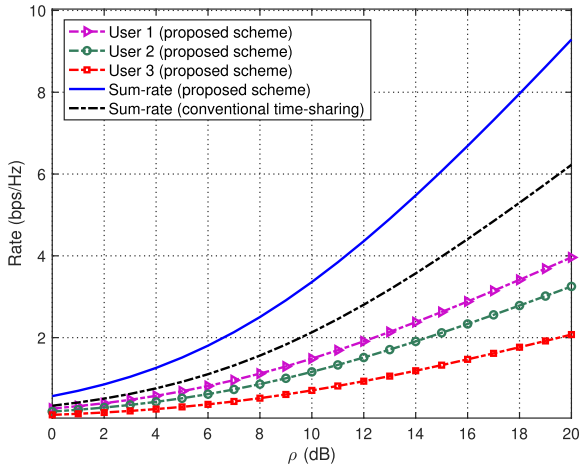

Fig. 6. Achievable rates via the proposed transmission scheme for three users in the absence of CSI feedback.

VI. Numerical Results

Throughout this section, unless stated otherwise, $N_0=1$ W, $\rho=10$ dB, $K=10$, and time-sharing factors are equally proportioned among users.

Fig. 6 shows the rates achieved through the proposed product superposition for three users in the absence of channel state feedback. Here, the channel is frequency-flat, $M=6$, $N=2$, $T_1=48$, $T_2=24$, and $T_3=12$. Product superposition provides a significant gain over the conventional transmission scheme, where pilot time slots are not reused, and data time slots are shared across users. For example, at the target sum-rate of 6 bps/Hz, the proposed method provides 4.3 dB gain over conventional transmission scheme.

Fig. 7 compares rate regions achieved through the proposed transmission scheme under perfect and imperfect feedback channels. Here, $T_1=20$, $T_2=10$, $M=4$ and $N=2$. Moreover, both users have frequency-flat channels and thus we employ product superposition over time. The achievable rate region is degraded when the feedback channel is imperfect, which is due to the fact that an imperfect feedback channel increases the noise on available channel estimates at the transmitter.

Fig. 8 compares the achievable rate regions through the proposed transmission schemes and conventional zero-forcing. Here, the feedback channel is perfect, $T_1=20$ and $T_2=10$, $B_1=1$, $B_2=2$, $M=4$ and $N=2$. Since User 1 has longer coherence time and frequency-flat channel, it can receive data through the product superposition over time or frequency. While both proposed schemes almost achieve the same performance, they provide a significant gain over the conventional zero-forcing.

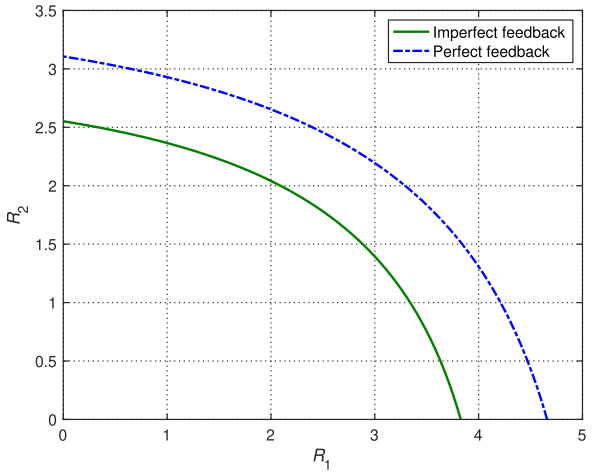

Fig. 7. Comparison of perfect feedback channel with imperfect feedback channel under coherence time disparity.

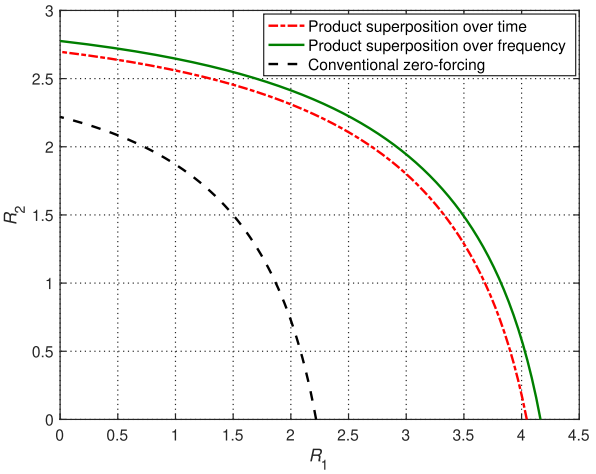

Fig. 8. Rate regions achieved via the product superposition over time, frequency, and the conventional zero-forcing scheme.

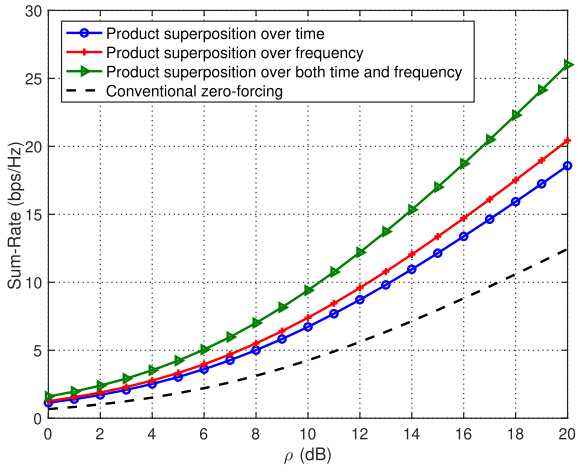

Fig. 9. Sum-rates achieved via the proposed transmission scheme and the conventional zero-forcing under different coherence disparities.

Fig. 9 shows the sum-rate achieved through the product superposition over time, product superposition over frequency, and product superposition over both time and frequency for three users. Here, the feedback channel is imperfect, $T_1 = 24$, $T_2 = 12$, $T_3 = 6$, $B_1 = 4$, $B_2 = 4$, $B_3 = 2$, $M = 6$ and $N = 2$. The proposed schemes provide significant gains over the conventional zero-forcing. For example, at the target sum-rate of 10 bps/Hz, product superposition over both time and frequency provides a gain of 6.2 dB. Furthermore, product superposition in both time and frequency outperforms the other two proposed schemes.

VII. Conclusion

This paper proposes a new and efficient pilot-domain non-orthogonal multiple access (NOMA) signaling scheme for multiple-input multiple-output (MIMO) multi-user channels that have different coherence times and/or bandwidths, a common condition in practical systems. The channel state feedback is available, either perfectly or imperfectly. The contributions of this paper enable the combination of the gains from product superposition with the gains arising from transmit beamforming, which was not possible with earlier approaches. The technical novelty of the paper lies in reconciling the requirements of product superposition and beamforming through novel methods, allowing for simultaneous harvesting of both classes of gains. We demonstrate the advantages of the proposed approach in terms of achievable rates. Overall, this paper presents a significant contribution to the field of MIMO-NOMA signaling for multi-user channels with varying coherence times and/or bandwidths.

References

[1] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[2] A. Lozano and N. Jindal, "Optimum pilot overhead in wireless communication: A unified treatment of continuous and block-fading channels," in *Proc. IEEE Eur. Wireless Conf.*, Apr. 2010, pp. 725–732.

[3] Y. Li and A. Nosratinia, "Coherent product superposition for downlink multiuser MIMO," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1746–1754, Mar. 2015.

[4] M. Karbalayghareh and A. Nosratinia, "Interaction of pilot reuse and channel state feedback under coherence disparity," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2022, pp. 190–195.

[5] D. Makrakis and K. Feher, "A novel pilot insertion-extraction method based on spread spectrum techniques," in *Proc. MIAMI Technicon*, 1987, pp. 129–132.

[6] T. P. Holden and K. Feher, "A spread spectrum based system technique for synchronization of digital mobile communication systems," in *Proc. IEEE 39th Veh. Technol. Conf.*, May 1989, pp. 780–787.

[7] A. Steingass, A. J. van Wijngaarden, and W. G. Teich, "Frame synchronization using superimposed sequences," in *Proc. IEEE Int. Symp. Inf. Theory*, Jan. 1997, p. 489.

[8] F. Tufvesson, M. Faulkner, P. Hoeher, and O. Edfors, "OFDM time and frequency synchronization by spread spectrum pilot technique," in *Proc. IEEE Commun. Theory Mini-Conf.*, Jun. 1999, pp. 115–119.

[9] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, vol. 4, Dec. 1999, pp. 2162–2166.

[10] J. K. Tugnait and X. Meng, "On superimposed training for channel estimation: Performance analysis, training power allocation, and frame synchronization," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 752–765, Feb. 2006.

[11] M. Coldrey and P. Bohlin, "Training-based MIMO systems—Part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[12] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multicell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[13] J. Ma, C. Liang, C. Xu, and L. Ping, "On orthogonal and superimposed pilot schemes in massive MIMO NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2696–2707, Dec. 2017.

[14] B. Mansoor, S. Nawaz, and S. Gulfam, "Massive-MIMO sparse uplink channel estimation using implicit training and compressed sensing," *Appl. Sci.*, vol. 7, no. 1, p. 63, Jan. 2017.

[15] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[16] J. Jiao, J. Zhou, S. Wu, and Q. Zhang, "Superimposed pilot code-domain NOMA scheme for satellite-based Internet of Things," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2732–2743, Jun. 2021.

[17] X. Jing, M. Li, H. Liu, S. Li, and G. Pan, "Superimposed pilot optimization design and channel estimation for multiuser massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11818–11832, Dec. 2018.

[18] K. Upadhya, S. A. Vorobyov, and M. Vehkaperä, "Downlink performance of superimposed pilots in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6630–6644, Oct. 2018.

[19] Y. Zhang, X. Qiao, L. Yang, and H. Zhu, "Superimposed pilots are beneficial for mitigating pilot contamination in cell-free massive MIMO," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 279–283, Jan. 2021.

[20] J. Li, H. Zhang, D. Li, and H. Chen, "On the performance of wireless-energy-transfer-enabled massive MIMO systems with superimposed pilot-aided channel estimation," *IEEE Access*, vol. 3, pp. 2014–2027, 2015.

[21] W. Yuan, S. Li, Z. Wei, J. Yuan, and D. W. K. Ng, "Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1954–1958, Sep. 2021.

[22] H. B. Mishra, P. Singh, A. K. Prasad, and R. Budhiraja, "OTFS channel estimation and data detection designs with superimposed pilots," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2258–2274, Apr. 2022.

[23] C. Yang, J. Wang, Z. Pan, and S. Shimamoto, "Delay-Doppler frequency domain-aided superimposing pilot OTFS channel estimation based on deep learning," in *Proc. IEEE 96th Veh. Technol. Conf.*, Sep. 2022, pp. 1–6.

[24] Y. Liu, Y. L. Guan, and D. González G., "BEM OTFS receiver with superimposed pilots over channels with Doppler and delay spread," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 2411–2416.

[25] H. B. Mishra, P. Singh, A. K. Prasad, and R. Budhiraja, "Iterative channel estimation and data detection in OTFS using superimposed pilots," in *Proc. IEEE Int. Conf. Commun. Workshops*, Jun. 2021, pp. 1–6.

[26] W. Liu, L. Zou, B. Bai, and T. Sun, "Low PAPR channel estimation for OTFS with scattered superimposed pilots," *China Commun.*, vol. 20, no. 1, pp. 79–87, Jan. 2023.

[27] M. Fadel and A. Nosratinia, "Coherent, non-coherent, and mixed-CSIR broadcast channels: Multiuser degrees of freedom," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2574–2578.

[28] M. Fadel and A. Nosratinia, "Frequency-selective multiuser downlink channels under mismatched coherence conditions," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2393–2404, Mar. 2019.

[29] M. Fadel and A. Nosratinia, "Coherence disparity in broadcast and multiple access channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7383–7401, Dec. 2016.

[30] M. Fadel and A. Nosratinia, "Broadcast channel under unequal coherence intervals," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 275–279.

[31] M. Fadel Shady and A. Nosratinia, "MISO broadcast channel under unequal link coherence times and channel state information," *Entropy*, vol. 22, no. 9, p. 976, Sep. 2020.

[32] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[33] A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[34] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[35] C. Wang, E. K. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 805–810, Mar. 2007.

[36] Z. Wang and W. Chen, "Regularized zero-forcing for multiantenna broadcast channels with user selection," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 129–132, Apr. 2012.

[37] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

Mehdi Karbalayghareh (Member, IEEE) received the B.Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2015, and the M.Sc. degree in electrical engineering from Ozyegin University, İstanbul, Turkey, in 2019. He is currently pursuing the Ph.D. degree in electrical engineering with The University of Texas at Dallas, Richardson, TX, USA. His research interests include wireless communications, information theory, and machine learning. He was a recipient of the 2023 Excellence in Education Doctoral Fellowship and the 2024 Research Excellence Award at The University of Texas at Dallas.

Aria Nosratinia (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign in 1996. He had visiting appointments at Princeton University, Rice University, and UCLA. His research interests include information theory and signal processing, with applications in wireless communications and data security and privacy. He is a fellow of IEEE for contributions to multimedia and wireless communications. He has received the National Science Foundation Career Award and the Outstanding Service Award from the IEEE Signal Processing Society, Dallas Chapter. He is a Registered Professional Engineer in the State of Texas and a Clarivate Analytics Highly Cited Researcher. He has served as an Editor and an Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Editor for IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SIGNAL PROCESSING LETTERS, *IEEE Wireless Communications Magazine*, and *Journal of Circuits, Systems, and Computers*.