

Zeroing Unknown Terms: A Novel Clustering Architecture for Low Network Overhead in Distributed Massive MIMO

SUPUNI GUNASEKARA¹ (Graduate Student Member, IEEE), RAJITHA SENANAYAKE¹ (Member, IEEE), PETER SMITH² (Fellow, IEEE), AND MARGRETA KUIJPER¹ (Senior Member, IEEE)

¹Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia

²School of Mathematics and Statistics, Victoria University of Wellington, Wellington 6140, New Zealand

CORRESPONDING AUTHOR: S. GUNASEKARA (e-mail: supuni.gunasekara@student.unimelb.edu.au)

This work was supported in part by the Australian Research Council (ARC) through the ARC Discovery Early Career Researcher Award under Grant DE180100501.

ABSTRACT We propose a novel user-centric clustering architecture for distributed massive multiple-input-multiple-output networks. We examine the uplink of a general multi-cell scenario in which a cluster of base stations (BSs) with large antenna arrays detect the signals of multiple users simultaneously. As opposed to traditional clustering schemes, where the channel coefficients of all the users within the cooperating cluster are learned at the BSs, in the proposed approach, BSs in the network learn the channels of only the strongest users based on the received signal-to-noise ratio. We consider two linear receivers, namely, zero forcing and linear minimum mean squared error receivers at the central processing unit. The two receivers operate with only partial knowledge of the user channels and the unknown terms are set to zero. We conduct a thorough analytical investigation and derive novel approximations for the instantaneous received signal-to-interference-and-noise ratio of an arbitrary user processed by a cluster of BSs. Furthermore, we develop simple closed-form expressions for the achievable rate and symbol error probability of an arbitrary user processed by a cluster of BSs. Numerical examples are used to illustrate the accuracy of the proposed approach and also to compare it to existing approaches. We show that our proposed architecture outperforms traditional clustering methods, particularly for cluster edge users.

INDEX TERMS Cooperative communications, distributed antenna systems, LMMSE receiver, massive MIMO, ZF receiver.

I. INTRODUCTION

THE EVER-INCREASING demand for wireless communication has necessitated the exploration of advanced technologies to accommodate the growing capacity requirements. Among these technologies, massive multiple-input multiple-output (MIMO) has emerged as a promising solution due to its potential to improve spectral efficiency and enhance overall system performance [1], [2]. By utilizing the large number of antennas at its disposal, this technology exhibits a remarkable increase in network capacity, often exceeding tenfold or more [3].

However, traditional massive MIMO systems are unable to ensure uniform data rates within the network. This limitation

has prompted researchers to explore alternative approaches, leading to the emergence of distributed antenna systems. In such a distributed system, the antennas are geographically distributed over a wide area and they cooperate via connection to a central processing unit (CPU) by fibre or coaxial cable [4], [5], [6]. Compared to its co-located counterpart, the distributed system brings forth numerous advantages. These include increased channel de-correlation for users even in close proximity, enhanced coverage area [7], and higher capacity [5]. Extensive studies on the cooperative massive MIMO approach have consistently shown its superior performance over non-cooperative small-cell systems, particularly in terms of throughput [8], [9].

Within the realm of distributed massive MIMO systems, two distinct paradigms exist namely the base station (BS) centric approach and user-centric approach. However, the significant variation in signal-to-noise ratio (SNR) within a cell or cluster of cells remains unresolved in BS-centric approaches due to two primary factors: the rapid decay of received signal power with propagation distance and interference from neighboring non-cooperating BSs. These factors pose significant challenges in meeting the needs of users who require consistent and high-speed connectivity throughout their coverage area [10]. In light of these limitations, this paper focuses on the user-centric paradigm, aiming to further enhance the performance of distributed massive MIMO systems.

A cell-free system emerges when the associations between users and cells are released, and the rigid boundaries of individual cells are dissolved [11], [12]. In a cell-free system, users are no longer constrained to specific cells, resulting in enhanced flexibility and collaborative potential within the network [10], [13]. User-centric cell-free massive MIMO, alternatively referred to as user-centric distributed massive MIMO, advocates for a clustering approach based on the users' channels [10], [13]. In such a user-centric approach, each user is served by a cluster of BSs. Hence, any particular BS may cooperate with different sets of BSs to serve different users [14], [15]. Several existing works have explored user-centric distributed massive MIMO systems, contributing valuable insights to this field. For instance, a user-centric virtual cell massive MIMO approach was introduced in [13] leading to reduced backhaul overhead when employing maximal ratio combining. However, this approach uses a lower level of cooperation, where local estimates at the APs are combined at the CPU. The study undertaken in [16] demonstrates that a centralized implementation maximizes spectral efficiency and reduces fronthaul signaling in cell-free massive MIMO networks. The user-centric framework proposed in [15] reformulates the dynamic cooperation clustering approach [14] for cell-free networks and is able to utilize more advanced interference-aware receiver combining techniques. In [17], two user-centric clustering schemes were investigated. One, fixing the number of BSs serving a user, and the other, fixing the radius of the user's serving cluster. The evaluation was based on the data rates of the cell-edge users. An approach where users are served by multiple BS-centric clusters which in turn cooperate to create user-centric clusters was studied in [18].

Nevertheless, a system where user data is detected centrally, as discussed above, imposes a significant strain on the fronthaul, because the CPU requires all the channels and the received signals at the BSs. Consequently, in systems where multiple BSs cooperate, the number of user channels that need to be learned at a BS increases rapidly. To acquire these channels, BSs would be tapping into the resources which would have otherwise been spent on data transmission. It should also be noted that combining these signals at the CPU will require additional processing power. Efforts

have been made to introduce low-complexity approaches within the user-centric distributed massive MIMO context. Some examples are [11], [15], [19] which emphasize the joint processing of observations from selected users and BSs to enhance the system efficiency.

Despite these enhancements, certain limitations persist, which motivates the present paper. These limitations are discussed herein. The user-centric approach outlined in [15], [19], describes a scalable combining scheme based on the MMSE receiver where a BS learns the channels of the users who have chosen that BS as a serving BS. However, the scheme requires the BSs to learn the channels from a user even if that BS is not a BS serving the said user. For example, assuming the n -th user, U_n , is served by BS₁ and BS₂ and U_{n+1} is served by only BS₁, the approach proposed in [15] needs the channel from U_{n+1} to both BS₁ and BS₂. With this setup, if all the BSs serving U_n share U_n as the common user while serving entirely distinct user sets otherwise, then all the BSs might have to learn the channels of most of the users in the network. In [11], the authors explore scalability in the cell-free architecture by introducing both BS-centric and user-centric minimum mean squared error (MMSE) receivers. The system under consideration consists of single antenna BSs serving a subset of users in an ultra-dense deployment scenario. The proposed model is said to approach network-wide MMSE in which all the BSs cooperate to serve every user in the system. However, we note that to achieve the network-wide MMSE performance, the user is served by a large number of BSs. According to the outlined user subset selection policy this results in a BS serving (and in turn learning the channels of) nearly every user in the system. As such, both these approaches suffer from having to learn a large number of user channels in the limit. As a consequence, the BSs are required to exchange more information with the CPU(s) increasing the network overhead significantly. Moreover, the large number of user channels used to aid in detection contributes to a notable increase in receiver complexity, yet fails to yield a substantial enhancement in performance.

Building on these works, this paper aims to fill those gaps and presents a novel approach that mitigates network overhead while maintaining a low-complexity design. We propose a user-centric approach based on the notion that a BS learns the channels of the strongest interfering users in addition to the users for whom it serves as the strongest BS. The detection of user signals takes place at the CPU, where the received signals from cooperating BSs are jointly processed. Unknown terms in the channel matrix at the CPU will be set to zero.

We focus on the uplink of a distributed massive MIMO network where several BSs equipped with a large number of antennas serve single antenna users dispersed at random across the network. These receiving BSs cooperate via a backhaul network that incorporates multiple CPUs, providing computational support for their joint operations. From a practical point of view, this cooperation is restricted to a set

of geographically close BSs that form a cooperating cluster. In the traditional cooperation scenarios, this leads to two classes of network users: in-cluster users, that are jointly detected by the cooperating BSs, and out-of-cluster users, that cause interference. However, due to the nature of our approach, we segregate the users in the system using slightly different terms. User segregation is based on additional constraints proposed in Section III which become instrumental in reducing the overhead introduced to the network through cooperative communications. The corresponding terms to traditional in-cluster and out-of-cluster users are associated and non-associated users, respectively. Also, the set of users who are served by a cluster of BSs are defined as serving users. We explain these terms with a simple example in Section III. As such, this paper extends the results of our earlier conference version [20] and makes the following contributions:

- We propose a modified user-centric clustering architecture with reduced computational complexity and reduced overhead added to the network which also improves the rate uniformity across the network. In contrast to the conventional clustering schemes, where the channel coefficients between every user and BS in the cooperative cluster are required at the CPU for detection, our proposed approach only requires the channels of the strongest users to each BS in the cooperating cluster. The identification of the strongest users is determined by a pre-determined threshold.
- We derive simple closed-form expressions to accurately approximate the instantaneous received SINR of an arbitrary serving user in a distributed massive MIMO network. Our approximations are derived for two linear receivers namely ZF and linear MMSE (LMMSE) receivers. While ZF and LMMSE receivers are commonly applied in cooperative systems [21], [22], [23], in this work, they are utilized differently. This is due to their operation in a modified scenario where the CPU lacks the full channel matrix, working instead with partial user channel information provided by the BSs. Our analysis considers imperfect CSI in which the channels of the strongest users are estimated through uplink training symbols. Non-associated user interference is taken into account in both approaches.
- Based on the derived approximations for SINR, we conduct a performance analysis and obtain remarkably accurate approximations for the achievable rate and symbol error probability (SEP) of an arbitrary serving user, under our proposed architecture.
- We present a comprehensive complexity analysis, demonstrating a reduction in both complexity and network overhead in the proposed architecture compared to the reference architectures.

We highlight the advantages of the proposed architecture by comparing the achievable rate of a user processed by our method against two traditional clustering architectures

namely non-cooperation and static clustering. It is important to note that in the special case when perfect CSI is accessible at the CPU the results derived for ZF in this study converge to that of [20] in which only the ZF was considered.

The remainder of the paper is organized as follows. Section II describes the system model used throughout this analysis. In Section III, we introduce the modified user-centric approach. Then, Section IV derives approximations for the received SINR for both receiver types. Section V formulates the approximations for the achievable rate and SEP. A comprehensive complexity analysis is provided in Section VI. Numerical examples are provided in Sections VII and VIII concludes the paper.

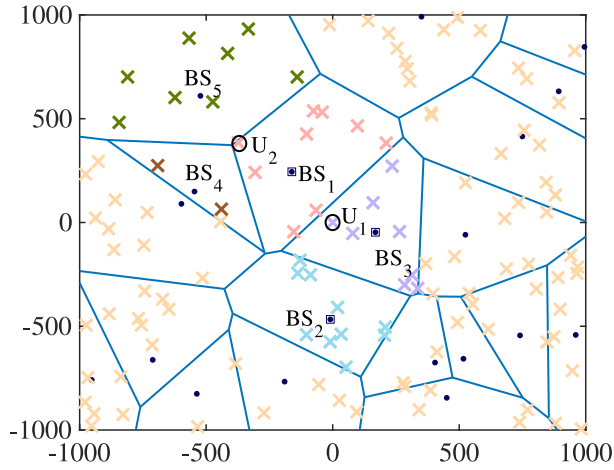
II. SYSTEM MODEL

In this section, we introduce the system model used throughout the paper, followed by a simple example that will be used to describe the different clustering methods. We then proceed to present three baseline clustering methods, namely non-cooperation, static cooperation and dynamic cooperation prior to introducing the new modified user-centric clustering method in the subsequent section.

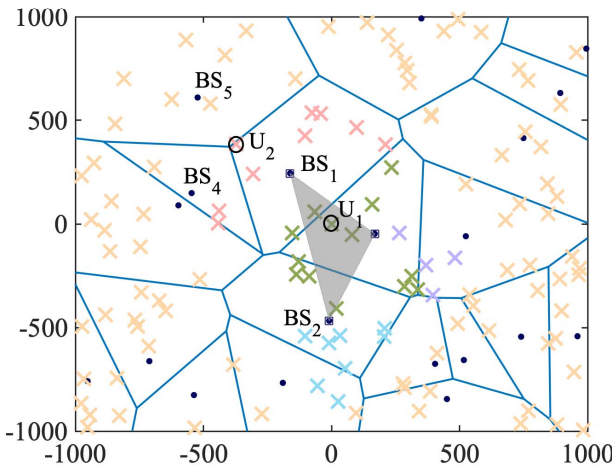
We consider the uplink of a communication network where BSs are deployed randomly based on stochastic geometry. More specifically, we model the BSs in \mathbb{R}^2 , with a homogeneous Poisson Point Process (PPP), Φ_b , of density λ_b . The i^{th} BS (BS_{*i*}) is equipped with a large antenna array with M_i antennas. We further consider a population of single antenna users distributed across the network. Locations of these users are generated according to another homogeneous PPP, Φ_u , which has a density λ_u and is independent of Φ_b . By tessellating the plane into Voronoi regions around each BS, we can define cells in a traditional manner. Let L be the number of BSs in the system where the i^{th} cell encompasses BS_{*i*} along with the collection of users in set \mathcal{N}_i . The total number of users in the system is $N_{\text{sys}} = \sum_{i=1}^L |\mathcal{N}_i|$. An example cellular network is illustrated in Fig. 1(a) where $\lambda_b = 20$ and $\lambda_u = 120$. The channels between the users and the receive antennas are modelled as independent Rayleigh channels with path loss. Large-scale fading coefficients of the user channels are assumed to be known a priori at each BS. This assumption is reasonable due to the fact that large-scale fading coefficients vary slowly, usually over the range of several coherence intervals [24]. In [25], it has been shown that in the presence of log-normal-based shadowing, a network in which the BSs are arranged at fixed positions in a hexagonal grid appear as though they were realized from a Poisson-distributed process. Leveraging this observation, the channels in the network also approximate the effects of shadow fading.

A. NON-COOPERATION

In the traditional non-cooperative approach, users assigned to a specific cell, i , would be detected at BS_{*i*}. In the example network in Fig. 1(a), BS_{*i*} serves the users in the set \mathcal{N}_i and all other users in the system are interferers. For example, pink,



(a) BS clustering with traditional approaches.



(b) PJP cluster made by BS₁ – BS₂ – BS₃ which cooperate to serve the users inside the gray triangle (serving users).

FIGURE 1. Example of BS positions from a homogeneous PPP and users uniformly distributed in the corresponding Poisson-Voronoi cells. Small dots represent BSs and crosses represent users.

blue, purple and brown users are served by BS₁, BS₂, BS₃ and BS₄, respectively, with each user's channels exclusively learned at their respective serving BS. For such a network, the $\mathcal{C}^{M_i \times 1}$ received signal vector at BS_{*i*} can be written as

$$\mathbf{y}_i^{\text{NC}} = \mathbf{H}_i^{\text{NC}} \mathbf{s}^{\text{NC}} + \mathbf{z}_i^{\text{NC}} \quad \text{for } i = 1, 2, \dots, L, \quad (1)$$

where the superscript NC is used to denote the non-cooperative approach and the $\mathcal{C}^{|\mathcal{N}_i| \times 1}$ vector \mathbf{s}^{NC} contains the data symbols from users in the set \mathcal{N}_i . The data symbols are independent and normalized such that $\mathbb{E}[\mathbf{s}^{\text{NC}}(\mathbf{s}^{\text{NC}})^H] = \mathbf{I}$. The notation $(\mathbf{X})^H$ denotes the Hermitian transpose of matrix \mathbf{X} . The vector \mathbf{z}_i^{NC} denotes the inter-cell interference plus noise at BS_{*i*}. We assume that the noise is additive white Gaussian, where each noise component is characterized by a complex Gaussian distribution with a mean of zero and a variance of σ^2 . The channels connecting the users with the receive antennas are modelled as independent Rayleigh

channels with path loss. Therefore, the entries of the $\mathcal{C}^{M_i \times |\mathcal{N}_i|}$ channel matrix, \mathbf{H}_i^{NC} , are independent yet non-identically distributed across the columns. The element in the (m, n) -th position of the channel matrix \mathbf{H}_i^{NC} , denoted as h_{imn} , follows a complex Gaussian distribution with a mean of zero and a variance of $P_{i,n}$. Here, $P_{i,n}$ represents the large-scale fading coefficient between user n and BS_{*i*} which is due to the distance dependant path gain. Note that, even though we do not explicitly model shadowing, shadowing is already incorporated into the location information (through the use of Poisson distributed BS setup [25]), and consequently, in $P_{i,n}$.

This approach is severely hindered by the interference from neighboring cells and a significant drop in data rates can be experienced by users located near the edge of a cell [10].

B. STATIC COOPERATION

The data rate of cell edge users can be increased through multi-cell cooperation which exploits inter-cell interference [26]. When we consider static clustering (one such cooperation technique), the BSs in a pre-defined cluster would be cooperating together to detect users in the vicinity. For example, Fig. 1(a), considers a cluster of three BSs, namely BS₁, BS₂ and BS₃, who cooperate to detect all $N^{\text{SC}} = |\mathcal{N}_1| + |\mathcal{N}_2| + |\mathcal{N}_3|$ users located in the combined three cell zone using $M^{\text{SC}} = M_1 + M_2 + M_3$ cluster antennas. The three BSs in the cluster learn the channels of the users in blue, pink and purple, whereas all the other users are interfering users. The N^{SC} users are defined as *in-cluster* users of the said cluster and the cluster BSs send the received signals and the channel coefficients from the N^{SC} users to a CPU which detects the signals. These in-cluster users are subject to interference from $K^{\text{SC}} = \sum_{i=4}^L |\mathcal{N}_i|$ *out-of-cluster* users and the instantaneous channel gains of these users are unknown to the CPU. For the static clustering approach, the vertically stacked $\mathcal{C}^{M^{\text{SC}} \times 1}$ received vector at the CPU is

$$\mathbf{y}^{\text{SC}} = \mathbf{H}^{\text{SC}} \mathbf{s}^{\text{SC}} + \mathbf{z}^{\text{SC}}, \quad (2)$$

where the superscript SC is used to denote the static clustering approach and the aggregated channel matrix, \mathbf{H}^{SC} , contains channels from all the in-cluster users to M^{SC} cluster antennas. Also, $\mathbf{z}^{\text{SC}} = ((\mathbf{z}_1^{\text{SC}})^T, (\mathbf{z}_2^{\text{SC}})^T, (\mathbf{z}_3^{\text{SC}})^T)^T$, is the aggregated out-of-cluster interference and noise. Note that \mathbf{z}_i^{SC} contains the out-of-cluster interference from the K^{SC} users and the noise perceived at BS_{*i*}. The data symbols from the users in the set N^{SC} is carried in \mathbf{s}^{SC} . From Fig. 1(a), we see that user U₁, who is located in the cluster-center, is well served by this clustering. However, user U₂, who is at the cluster edge, suffers from strong interference by the out-of-cluster users. A better choice for U₂ would have been a cluster arrangement of BS₁, BS₄ and BS₅. Clearly, this cooperative cell concept with disjointly clustered BSs cannot solve the problem of cluster edge users having low data rates due to out-of-cluster interference.

C. DYNAMIC COOPERATION

To encourage uniform data rates throughout the network, research is shifting from static clustering to user-centric clustering which is dynamic as a user's cluster is chosen based on the user location. User-centric clustering increases fairness and provides a better service to all users in the system [10]. In a system where user-centric clustering is enforced, each BS needs to learn the channels of the users in neighboring cells in addition to learning the channels of the users in its own cell. For example, BS₁, which learned the channels of the users in the three cells in a *three-cell static clustering* architecture, now needs to learn users in seven cells to operate in a dynamic architecture. As such, user-centric clustering suffers from a significantly higher network overhead due to the overlapping BS clusters as the channel information required at the CPU is increased [10]. Furthermore, limiting the fronthaul overhead is imperative to achieve scalability [27].

In the next section, we propose a modified user-centric approach which reduces the computational complexity and overhead added to the network while providing good performance throughout the system. In contrast to the traditional clustering schemes discussed above, our proposed approach is grounded on the concept that a BS only learns the channels of the strongest interfering users in addition to the users for whom it acts as the strongest BS.

III. THE MODIFIED USER-CENTRIC APPROACH

The first contribution of our paper, which is the design of a new user-centric clustering approach, is presented in this section. Our modified user-centric approach is based on a partial joint processing (PJP) technique to cluster BSs [20]. Let us explain the proposed method using a toy example. We define a PJP cluster as a group of three^{1,2} neighboring BSs cooperating together. The rationale behind choosing a three-cell cluster is twofold. Firstly, it naturally arises in a Voronoi tessellation.³ Secondly, for a typical user, the optimal cluster cardinality is small, as this minimizes overhead by limiting the number of cooperating cells [28], [29]. The small cluster size recommended in this approach and the localization of the BS clusters eliminates the necessity for complex large-scale BS scheduling. As such, the overall network size will not hinder the scalability and efficiency of the method. The three BSs in the PJP cluster are named BS_i, BS_j and BS_k, and they are equipped with M_i , M_j and M_k antennas, respectively.

It is assumed that the BSs are connected to one or more CPUs via high-capacity, delay-free links, through which the received signal vector and the available channel

¹Although we have chosen three as the cluster cardinality, the approach can be readily extended to larger cluster sizes as needed.

²In [20], only two BSs collaborated in a PJP cluster due to the 1D simulation setup in which the BSs were arranged in a row.

³While we employ a Voronoi tessellated cell structure to explain the modified user-centric approach, we emphasize that this method relies solely on utilizing long-term power information. Therefore, the proposed approach can essentially be applied in any general massive MIMO network.

coefficients at each BS are sent to the CPU(s). The CPU processes the received signals of each user by consolidating received signals from BSs in the BS cluster serving that specific user. It should be noted that the BSs themselves can function as CPUs by collecting received signals and performing combining. In this configuration, a given PJP cluster of BSs designates one BS to execute the CPU functionality, while another BS in the same cluster may provide the CPU functionality for a separate PJP cluster to which it belongs. As such, a BS might have to transmit the obtained CSI to multiple CPUs (BSs). However, as we demonstrate in Section VII, the number of user channels estimated in PJP is lower than that of 3-cell static clustering. Consequently, the network overhead associated with PJP will also be comparatively lower. It should be noted that, while there are some deployment challenges compared to non-cooperative systems, the proposed approach introduces fewer challenges not only compared to fully cooperative systems but also to systems with fixed or dynamic clustering involving a similar number of BSs as PJP.⁴ Some deployment challenges linked to the proposed method include ensuring that every BS is connected to each neighboring BS and configuring the master node or CPU for each cluster of BSs.

Contrasting with traditional architectures, a BS operating in the PJP architecture learns the channels of all the users for whom it functions as the dominant BS, alongside some additional users with strong interference. That is, from the perspective of BS_i, it learns the channels of two kinds of users. The first group of users are the users located in the i^{th} cell and the next group of users are the users whose channels are learned by BS_i but who are not located in the cell i . These second group of users of BS_i are picked out of the set $\{1 \dots N_{\text{sys}}\} \setminus \mathcal{N}_i$ if they maintain a strong average received SNR to BS_i. More specifically, the channel of $U_n \in \{1 \dots N_{\text{sys}}\} \setminus \mathcal{N}_i$ is learned at BS_i if the following constraints are satisfied.

- 1) The received SNR of U_n at BS_i, $\bar{\gamma}_{i,n} = P_{i,n}/\sigma^2$, is such that

$$\bar{\gamma}_{i,n} \geq \lambda, \quad (3)$$

where λ is a pre-defined performance threshold.

- 2) BS_i does not already serve another user on the same pilot as U_n . If BS_i already serves a user on the same pilot as U_n , the user surpassing the threshold in (3) by the largest margin will be selected.
- 3) The set of users whose user channels are known at BS_i, denoted as $\mathcal{N}_i^{\text{PJP}} \subset \{1, \dots, N_{\text{sys}}\}$, is such that

$$|\mathcal{N}_i^{\text{PJP}}| + 1 \leq N_{\text{max}}, \quad (4)$$

where the superscript PJP indicates the partial joint processing approach and N_{max} represents the maximum number of users that could be learnt at a BS.

Note that the second constraint is introduced in order to avoid pilot contamination within the PJP cluster.⁴ Overall, constraints 1 to 3 ensure the selection of a limited number of users with strong channels, effectively preventing pilot contamination. The ability to select λ and N_{\max} that best enhance the system performance, is a valuable design attribute. This feature enables the system to be adaptable to varying BS and user densities. For example, in a network characterized by high user and BS density, decreasing λ and increasing N_{\max} would contribute to enhancing system performance.

Fig. 1(b) depicts an example scenario where a PJP cluster is formed by BS₁, BS₂ and BS₃. Considering only the BSs within this PJP cluster, the channels of the users represented by the colors pink, blue, and purple are only learned at BS₁, BS₂ and BS₃, respectively, whereas the channels of the users in green are learned at more than one BS. Note that the channels of some of these users will be learned at other BSs outside the considered cluster of BSs. However, the BS cluster made by BS₁, BS₂ and BS₃ cooperating, is not a user-centric cluster for all the users in the sets $\mathcal{N}_1^{\text{PJP}}$, $\mathcal{N}_2^{\text{PJP}}$ and $\mathcal{N}_3^{\text{PJP}}$. For example, U₂, located at the edge of this cluster is positioned close to the center of the BS₁ – BS₄ – BS₅ cluster. The channels of the users which strongly interfere with U₂, are known by BSs 1, 4 and 5 contributing to better performance for U₂ if served by this BS cluster (BS₁ – BS₄ – BS₅). As such, in a PJP cluster, we only serve users who are inside the triangle which is constructed by connecting the three cooperating BSs. More specifically, if a user is located inside a Poisson-Delaunay triangulation, the three BSs at the vertices of the triangle are chosen as its serving BSs. Note that, the Poisson-Delaunay triangulation is dual to the Poisson-Voronoi tessellation and is uniquely determined once the BS positions are known [30]. As BSs in PJP learn channels of additional users who are not served by the cluster, it might appear that they learn more channels than in traditional static clustering. However, in Section VII we show that PJP BSs learn fewer channels while delivering higher data rates. Thus, the three BSs (BS₁ – BS₂ – BS₃) cooperate to serve the collection of users in set $\mathcal{N}_{[1,2,3]}^{\text{PJP}}$ who are located in the shaded region in Fig. 1(b). These users are defined as the *serving users* of the said cluster. The serving users of a particular PJP cluster are the collection of users whose data symbols are cooperatively detected by the BSs in that PJP cluster.

For a generic PJP cluster, where BSs i , j and k are cooperating, the three BSs learn a total of $N = |\mathcal{N}_i^{\text{PJP}}| + |\mathcal{N}_j^{\text{PJP}}| + |\mathcal{N}_k^{\text{PJP}}|$ user channels and serve the collection of users in set $\mathcal{N}_{[i,j,k]}^{\text{PJP}}$ using $M = M_i + M_j + M_k$ cluster antennas.

⁴By introducing intelligent user scheduling approaches, PJP can efficiently serve a large number of users without encountering pilot contamination within the PJP cluster. Nevertheless, PJP can be modified to accommodate and function in the presence of pilot contamination if required. Considering that the receivers in a PJP system operate with partial information by zeroing weak user channels, the directed interference caused by pilot contamination could affect PJP less than in traditional approaches.

The defined N users are the *associated users* for the specific cluster made by BS _{i} – BS _{j} – BS _{k} . Note that all the users learned by the set of BSs in a specific PJP cluster are termed as associated users for the said cluster. These users are subject to interference from the users in the set $\mathcal{K} = \{1, \dots, N_{\text{sys}}\} \setminus \{\mathcal{N}_i^{\text{PJP}} \cup \mathcal{N}_j^{\text{PJP}} \cup \mathcal{N}_k^{\text{PJP}}\}$, who are defined as *non-associated users* of the cluster. For the example PJP cluster in Fig. 1(b), non-associated users are in yellow and associated users are shown in other colors.

The user channels learned at each BS are then transmitted to the CPU, where the channel matrix is assembled by setting the unknown channels of the associated users to zero. Hence the term *Zeroing Unknown Terms* (ZUT). It is important to note that at the BSs, complete information about the channels of the users in $\mathcal{N}_i^{\text{PJP}}$ (for BS _{i}) is available. In contrast, at the CPU, some user channels to certain BSs are missing due to the nature of this approach. For such a network, the $\mathcal{C}^{M \times 1}$ received signal vector at the CPU is given by,

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{H}'\mathbf{s}' + \mathbf{n}, \quad (5)$$

where the $\mathcal{C}^{M \times N}$ channel matrix, \mathbf{H} , contains the channels of N associated users and \mathbf{H}' denotes the non-associated user channel matrix. Also, $\mathbf{s} = (s_1, \dots, s_N)^T$ and $\mathbf{s}' = (s'_1, \dots, s'_K)^T$ are vectors which contain data symbols from the N associated users and $K = |\mathcal{K}|$ non-associated users, respectively. The noise, distributed as $\mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$, is represented by \mathbf{n} . It should be noted that \mathbf{H} is not completely available at the CPU as not all BSs in the PJP cluster learn the same set of user channels. For example, in Fig. 1(b) only the channels of the green colored users are learned at more than one BS. By separating out the unknown terms, we can re-express (5) as

$$\mathbf{y} = \mathbf{H}_0\mathbf{s} + \underbrace{\mathbf{E}\mathbf{s} + \mathbf{H}'\mathbf{s}'}_{\mathbf{z}} + \mathbf{n}, \quad (6)$$

where \mathbf{H}_0 , contains only the known channel coefficients at the CPU. The unknown channels represented by zeros in \mathbf{H}_0 according to the ZUT procedure are contained in $\mathbf{E} = \mathbf{H} - \mathbf{H}_0$. The r^{th} column of \mathbf{H}_0 and \mathbf{E} correspond to the channels of the r^{th} associated user and are defined as $(\mathbf{h}_0)_r$ and \mathbf{e}_r .

IV. RECEIVED SINR ANALYSIS

In our previous work [20] we have assumed that perfect CSI is available a priori wherever required. In this paper, the system is analyzed under imperfect CSI conditions where the channels are estimated utilizing the LMMSE estimator. We derive approximations for the received SINR for two predominant linear receivers, namely ZF and LMMSE. For the massive MIMO scenario, the approximations for both ZF and LMMSE receivers are highly precise.

A. CHANNEL ESTIMATION

We assume that the small-scale fading varies in time according to a block-fading model [31], where the channel is constant over each frame of length τ_c channel uses, and

changes independently from frame to frame. As mentioned in Section II, we assume that the large-scale fading coefficients of the users are known a-priori at each BS. Since the small-scale fading gains are not known to the receiver, each BS estimates the channel vectors based on a sequence of known training symbols. We have taken the channel estimation approach outlined in [10]. However, for completeness, we have included the important steps of the derivation and introduced some additional terminology to account for channel estimation. Let τ_p be the number of symbols reserved for pilot transmissions in each coherence interval with the assumption that each user transmits a pilot sequence that spans these τ_p symbols. The channel of each user can be estimated at each BS using the received signals. Because the pilots are τ_p dimensional vectors, there are at most τ_p mutually orthogonal sequences. In this context, we consider that the network uses τ_p mutually orthogonal pilot sequences $\phi_1, \dots, \phi_{\tau_p} \in \mathbb{C}^{\tau_p \times 1}$ which satisfy the following constraint,

$$\phi_t^H \phi_s = \begin{cases} \tau_p & t = s \\ 0 & t \neq s. \end{cases} \quad (7)$$

Elements of ϕ_t will be transmitted over τ_p consecutive channel uses and the index of the pilot assigned to user n is defined as $p_n \in \{1, \dots, \tau_p\}$. Given the limited number of τ_p orthogonal pilot sequences, in cases where $\tau_p < N_{\text{sys}}$, multiple users will share the same pilot. We define the set of users that use the same pilot as user n as $\chi_n = \{r: p_r = p_n, r \in \{1, \dots, N_{\text{sys}}\}\} \subset \{1, \dots, N_{\text{sys}}\}$. Considering BS i , the $\mathbb{C}^{1 \times \tau_p}$ received sequence at antenna m is,

$$\mathbf{y}_m = \sum_{r=1}^{N_{\text{sys}}} \sqrt{P_{i,r}} g_{m,r} \phi_{p_r}^T + \mathbf{n}_m, \quad (8)$$

where $g_{m,r}$ is the Rayleigh faded channel from user r to antenna m . $\mathbf{n}_m \in \mathbb{C}^{1 \times \tau_p}$ is the i.i.d distributed receiver noise with variance σ^2 . Note that the pilot sequence $\phi_{p_r} = \phi_{p_q}$ when $\chi_r = \chi_q$.

To estimate the Rayleigh faded channel of user n , we first remove the interference from users using orthogonal pilots by multiplying the received signal y_m with the normalized conjugate of the pilot, $\phi_{p_n}^*$, that is $\phi_{p_n}^* / \sqrt{\tau_p}$, resulting in,

$$\begin{aligned} y_m^{\text{pilot}} &= y_m \frac{\phi_{p_n}^*}{\sqrt{\tau_p}} \\ &= \sqrt{P_{i,n} \tau_p} g_{m,n} + \sum_{r \in \chi_n \setminus \{n\}} \sqrt{P_{i,r} \tau_p} g_{m,r} + n_{p_n m}. \end{aligned} \quad (9)$$

Estimating the true channel, $g_{m,n} \sim \mathcal{CN}(0, 1)$, will be based on y_m^{pilot} . In (9), the first term is the desired channel $g_{m,n}$ scaled by $\sqrt{P_{i,n} \tau_p}$ and the second term contains the interference generated by the pilot sharing users. The third term is the noise with variance unchanged because $\frac{\phi_{p_n}^*}{\sqrt{\tau_p}}$ is a vector with unit norm. By concatenating the unwanted terms we re-express (9) as,

$$y_m^{\text{pilot}} = \sqrt{P_{i,n} \tau_p} g_{m,n} + w_m, \quad (10)$$

where w_m denotes the sum of the interference of the pilot sharing users and noise with variance $\sigma_{w_m}^2 = \sum_{r \in \chi_n \setminus \{n\}} P_{i,r} \tau_p + \sigma^2$. The variance of y_m^{pilot} can be calculated as,

$$\mathbb{E}[|y_m^{\text{pilot}}|^2] = P_{i,n} \tau_p + \sigma_{w_m}^2. \quad (11)$$

Therefore, the LMMSE estimate of $g_{m,n}$ at antenna m is [10],

$$\hat{g}_{m,n} = \left(\frac{\sqrt{P_{i,n} \tau_p}}{P_{i,n} \tau_p + \sigma_{w_m}^2} \right) y_m^{\text{pilot}}. \quad (12)$$

The channel estimate $\hat{g}_{m,n}$ is complex Gaussian with zero mean and variance

$$\begin{aligned} \mathbb{E}[|\hat{g}_{m,n}|^2] &= \frac{P_{i,n} \tau_p}{(P_{i,n} \tau_p + \sigma_{w_m}^2)^2} \mathbb{E}[|y_m^{\text{pilot}}|^2] \\ &= \frac{P_{i,n} \tau_p}{P_{i,n} \tau_p + \sigma_{w_m}^2} = \frac{P_{i,n} \tau_p}{\sum_{r \in \chi_n} P_{i,r} \tau_p + \sigma^2}. \end{aligned} \quad (13)$$

The exact channel, $g_{m,n}$, can be written as

$$g_{m,n} = \hat{g}_{m,n} + \tilde{g}_{m,n}, \quad (14)$$

where $\tilde{g}_{m,n}$ is the channel estimation error, which is independent of $\hat{g}_{m,n}$ and is complex Gaussian with mean zero and variance $\sigma_{\tilde{g}_{m,n}}^2 = \frac{\sigma_{w_m}^2}{P_{i,n} \tau_p + \sigma_{w_m}^2}$ [10].

Based on (14), (6) can be expressed as,

$$\mathbf{y} = \hat{\mathbf{H}}_0 \mathbf{s} + \underbrace{\tilde{\mathbf{H}}_0 \mathbf{s} + \mathbf{E} \mathbf{s} + \mathbf{H}' \mathbf{s}' + \mathbf{n}}_{\mathbf{z}}. \quad (15)$$

The estimated $\mathbb{C}^{M \times N}$ matrix, $\hat{\mathbf{H}}_0$, and the $\mathbb{C}^{M \times N}$ estimation error matrix, $\tilde{\mathbf{H}}_0$, contains independent horizontally stacked columns, $(\hat{\mathbf{h}}_r)_0 = \mathbf{P}_r^{1/2} \hat{\mathbf{g}}_r$, and $(\tilde{\mathbf{h}}_r)_0 = \mathbf{P}_r^{1/2} \tilde{\mathbf{g}}_r$, respectively. Here, \mathbf{P}_r defined in block-diagonal form as $\mathbf{P}_r = \text{diag}(P_{i,r} \mathbf{I}_{M_i}, P_{j,r} \mathbf{I}_{M_j}, P_{k,r} \mathbf{I}_{M_k})$ contains the long term powers of user r . Note that $\hat{\mathbf{g}}_r = [\hat{g}_{1,r} \dots \hat{g}_{M,r}]^T \sim \mathcal{CN}(0, \mathbf{I}_M)$ and $\tilde{\mathbf{g}}_r = [\tilde{g}_{1,r} \dots \tilde{g}_{M,r}]^T \sim \mathcal{CN}(0, \mathbf{I}_M)$. The covariance of the estimated channel, $\hat{\mathbf{g}}_r$, can be expressed as $\Sigma_{\hat{\mathbf{g}}_r} = \tau_p \mathbf{P}_r (\Sigma_w^2 + \tau_p \mathbf{P}_r)^{-1}$, where $\Sigma_w^2 = \text{diag}(\sigma_{w_1}^2, \sigma_{w_2}^2, \dots, \sigma_{w_M}^2)$. The full expression for $\Sigma_{\hat{\mathbf{g}}_r}$ is shown in (16) shown at the bottom of the next page.

It is important to note that in (15), we have defined \mathbf{z} to contain the impact of estimation error via the term $\tilde{\mathbf{H}}_0 \mathbf{s}$. Due to the ZUT process, some blocks of $\hat{\mathbf{H}}_0$ will be zero depending on whether the BS learns the corresponding channels. In the remainder of the paper, we evaluate the performance of the proposed architecture by considering an arbitrary user n in the set $\mathcal{N}_{[i,k]}^{\text{PJP}}$.

B. ZF RECEIVER

We first examine the received SINR when the ZF receiver is used at the CPU for joint detection of the transmitted signals. The channel matrix containing the known channels at the CPU, $\hat{\mathbf{H}}_0$, is utilized to perform ZF. However, as \mathbf{z} is not white, we employ a whitening filter $\Theta^{-1/2}$ as outlined in [20] where

$$\Theta = \mathbb{E}[\mathbf{z} \mathbf{z}^H] = \text{diag}\{\theta_1, \dots, \theta_M\}. \quad (17)$$

In (17), the value of θ_m is associated with antenna m and is expressed as

$$\theta_m = \sum_{r \in \mathcal{N}_i^{\text{PJP}}} P_{i,r} \sigma_{g_{m,r}}^2 + \sum_{r \in \{1, \dots, N_{\text{sys}}\} \setminus \{\mathcal{N}_i^{\text{PJP}}\}} P_{i,r} + \sigma^2. \quad (18)$$

Note that the expectation in (17) is taken over the small-scale fading of the channel estimation error, unknown channel coefficients, out-of-cluster interference and noise. The ZF receiver determines the output of the combiner as:

$$\hat{\mathbf{s}} = \mathbf{W}_0^H \boldsymbol{\Theta}^{-1/2} \mathbf{y}, \quad (19)$$

where the weight matrix $\mathbf{W}_0 = \boldsymbol{\Theta}^{-1/2} \hat{\mathbf{H}}_0 (\hat{\mathbf{H}}_0^H \boldsymbol{\Theta}^{-1} \hat{\mathbf{H}}_0)^{-1}$ is obtained based on $\hat{\mathbf{H}}_0$ and $\boldsymbol{\Theta}$ which are known to the central processor. An expression for the instantaneous received SINR of an arbitrary serving user n can be formulated using (19) as:

$$\gamma_n^{\text{ZF}} = \frac{1}{\left[(\hat{\mathbf{H}}_0^H \boldsymbol{\Theta}^{-1} \hat{\mathbf{H}}_0)^{-1} \hat{\mathbf{H}}_0^H \boldsymbol{\Theta}^{-1} \boldsymbol{\Omega}^{\text{ZF}} \boldsymbol{\Theta}^{-1} \hat{\mathbf{H}}_0 (\hat{\mathbf{H}}_0^H \boldsymbol{\Theta}^{-1} \hat{\mathbf{H}}_0)^{-1} \right]_{nn}}, \quad (20)$$

where $\boldsymbol{\Omega}^{\text{ZF}} = \tilde{\mathbf{H}}_0 \tilde{\mathbf{H}}_0^H + \mathbf{E} \mathbf{E}^H + \mathbf{H}' \mathbf{H}'^H + \sigma^2 \mathbf{I}$. The denominator in (20) contains the instantaneous channel gains of both associated and non-associated users of a given BS. In the subsequent analysis, we adopt a similar methodology as in [20] to obtain an approximate of the received SINR for an arbitrary user in a PJP system. However, for the sake of completeness, we include the critical steps of the process.

We note that the instantaneous channel gains of the associated users tend to be larger than those of the interfering non-associated users. Hence, we approximate the instantaneous channel gains of the non-associated users and unknown channels of the associated users with their expected values [32]. Furthermore, by carefully choosing an appropriate number of pilots, we can ensure that the estimation error remains negligible in comparison to the channel estimates [32]. As such, we can replace $\boldsymbol{\Omega}^{\text{ZF}}$ in (20) with its expected value, $\mathbb{E}[\boldsymbol{\Omega}^{\text{ZF}}] = \boldsymbol{\Theta}$, to obtain an approximation as

$$\gamma_n^{\text{ZF}} \approx \frac{1}{\left[(\hat{\mathbf{H}}_0^H \boldsymbol{\Theta}^{-1} \hat{\mathbf{H}}_0)^{-1} \right]_{nn}}. \quad (22)$$

According to [33], the approximation given in (22) can be expanded as follows:

$$\gamma_n^{\text{ZF}} \approx \left(\hat{\mathbf{h}}_0 \right)_n^H \boldsymbol{\Theta}^{-\frac{1}{2}} (\mathbf{I} - \mathbf{X}) \boldsymbol{\Theta}^{-\frac{1}{2}} \left(\hat{\mathbf{h}}_0 \right)_n, \quad (23)$$

where $(\hat{\mathbf{h}}_0)_n$ is the n^{th} column of $\hat{\mathbf{H}}_0$ and

$$\mathbf{X} = \boldsymbol{\Theta}^{-\frac{1}{2}} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \left(\left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H \boldsymbol{\Theta}^{-1} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \right)^{-1} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H \boldsymbol{\Theta}^{-\frac{1}{2}}, \quad (24)$$

where $(\hat{\mathbf{H}}_0)_{\bar{n}}$ represents $\hat{\mathbf{H}}_0$ with column vector $(\hat{\mathbf{h}}_0)_n$ removed. Given that the influence of small-scale fading reduces as the number of antennas grows [34], we proceed to analyze $\mathbb{E}[\mathbf{X}]$ by using Lemma 1 given below.

Lemma 1: The expected value of a matrix in the form of $\mathbf{R}(\mathbf{R}^H \mathbf{R} + \beta \mathbf{I}_{N-1})^{-1} \mathbf{R}^H$ where \mathbf{R} is a $\mathbb{C}^{M \times N-1}$ matrix and β is a real scalar, can be given by

$$\mathbf{D} = \mathbb{E} \left[\mathbf{R}(\mathbf{R}^H \mathbf{R} + \beta \mathbf{I}_{N-1})^{-1} \mathbf{R}^H \right] = \text{diag}(d_1, d_2, \dots, d_{N-1}), \quad (25)$$

where d_1, d_2, \dots, d_{N-1} are real scalars.

Proof of Lemma 1 is given in the Appendix A.

Therefore, when $\mathbf{R} = \boldsymbol{\Theta}^{-\frac{1}{2}} (\hat{\mathbf{H}}_0)_{\bar{n}}$ and $\beta = 0$, $\mathbb{E}[\mathbf{X}]$ can be expressed as a diagonal matrix. We also note that \mathbf{X} is an idempotent matrix (Appendix B) and therefore the trace of \mathbf{X} is equal to its rank [32]. Accordingly, following a similar approach to [21], we approximate \mathbf{X} by a simple diagonal matrix as

$$\mathbf{X} = \hat{\mathbf{X}} + \boldsymbol{\Xi}_X, \quad (26)$$

where we have defined $\hat{\mathbf{X}}$ by spreading the trace of the idempotent matrix \mathbf{X} equally along the diagonal as, $\hat{\mathbf{X}} = \frac{N-1}{M} \mathbf{I}_M$. Now, by substituting (26) in (20) where $\boldsymbol{\Xi}_X$ is an error matrix we can rephrase (20) as

$$\gamma_n^{\text{ZF}} \approx \left(\hat{\mathbf{h}}_0 \right)_n^H \boldsymbol{\Theta}^{-1/2} (\mathbf{I} - \hat{\mathbf{X}}) \boldsymbol{\Theta}^{-1/2} \left(\hat{\mathbf{h}}_0 \right)_n (\mathbf{I} - \Psi), \quad (27)$$

where

$$\Psi = \frac{\left(\hat{\mathbf{h}}_0 \right)_n^H \boldsymbol{\Theta}^{-1/2} \boldsymbol{\Xi}_X \boldsymbol{\Theta}^{-1/2} \left(\hat{\mathbf{h}}_0 \right)_n}{\left(\hat{\mathbf{h}}_0 \right)_n^H \boldsymbol{\Theta}^{-1/2} (\mathbf{I} - \hat{\mathbf{X}}) \boldsymbol{\Theta}^{-1/2} \left(\hat{\mathbf{h}}_0 \right)_n}. \quad (28)$$

It is worth noting that both the numerator and denominator in equation (28) consist of quadratic forms that encompass the sum of M terms. By dividing both the numerator and denominator of Ψ by M , we can analyze them. According to the law of large numbers for non-identical variables [35], as M approaches infinity, the denominator of Ψ converges to its mean. This convergence is guaranteed as all channel gains are finite. As such, we can write,

$$\frac{\left(\hat{\mathbf{h}}_0 \right)_n^H \boldsymbol{\Theta}^{-1/2} (\mathbf{I} - \hat{\mathbf{X}}) \boldsymbol{\Theta}^{-1/2} \left(\hat{\mathbf{h}}_0 \right)_n}{M} \rightarrow \frac{\text{tr}[\boldsymbol{\Theta}^{-1} \mathbf{P}_n \boldsymbol{\Sigma}_{\hat{\mathbf{g}}_n}]}{M} \triangleq \mu^{\text{ZF}} > 0. \quad (29)$$

$$\boldsymbol{\Sigma}_{\hat{\mathbf{g}}_r} = \text{diag} \left(\left(\frac{P_{i,r} \tau_p}{\sum_{q \in \mathcal{X}_r} P_{i,q} \tau_p + \sigma^2} \right) \mathbf{I}_{M_i}, \left(\frac{P_{j,r} \tau_p}{\sum_{r \in \mathcal{X}_r} P_{j,q} \tau_p + \sigma^2} \right) \mathbf{I}_{M_j}, \left(\frac{P_{k,r} \tau_p}{\sum_{r \in \mathcal{X}_r} P_{k,q} \tau_p + \sigma^2} \right) \mathbf{I}_{M_k} \right). \quad (16)$$

Some terms of \mathbf{P}_n in (29) could be zero depending on whether the channel of the user n is learned at BS_i , BS_j or BS_k . However, it is guaranteed that at least $\min(M_i, M_j, M_k)$ of the diagonal terms in \mathbf{P}_n are non-zero as user n is known by at least one of the BSs in the PJP cluster. Hence, for large M , the denominator of (29) divided by M converges to its mean value, μ_{ZF} .

Analyzing the numerator of (28) is more complex than the denominator. Despite the Hermitian nature of Ξ_X , it is random with interdependencies among its elements. Consequently, to facilitate our analysis, we shift our focus to mean-square convergence criteria. We then proceed to examine the convergence of the numerator in a mean-squared sense, using Lemma 2 as defined below.

Lemma 2: Let ξ denote a scalar in the form of $\xi = \frac{(\mathbf{u}^H \Delta \mathbf{u})}{M}$ where $\mathbf{u} \sim \mathcal{CN}(0, \Sigma_{\mathbf{u}})$ is an $M \times 1$ vector with diagonal covariance matrix $\Sigma_{\mathbf{u}}$ and Δ is an $M \times M$ matrix. The expected value of the mean square of ξ (i.e., $\mathbb{E}[\xi^2]$) converges to zero as M approaches infinity.

We provide the proof of Lemma 2 in Appendix C. Due to Lemma 2, when $M \gg N$,

$$\mathbb{E} \left(\frac{\left| \left(\hat{\mathbf{h}}_0 \right)_n^H \Theta^{-\frac{1}{2}} \Xi_X \Theta^{-\frac{1}{2}} \left(\hat{\mathbf{h}}_0 \right)_n \right|^2}{M^2} \right) \rightarrow 0, \quad (30)$$

Therefore we can conclude that for large M ,

$$0 < |\Psi| \ll 1. \quad (31)$$

Using (31), we can approximate γ_n^{ZF} in (27) for $M \gg N$ as

$$\gamma_n^{\text{ZF}} \approx C^{\text{ZF}} \left(\hat{\mathbf{h}}_0 \right)_n^H \Theta^{-1} \left(\hat{\mathbf{h}}_0 \right)_n, \quad (32)$$

where $C^{\text{ZF}} = \frac{M-N+1}{M}$.

C. LMMSE RECEIVER

As discussed in our previous work [20], the possible noise amplification in a ZF receiver can be mitigated by adopting a linear alternative like the LMMSE receiver. Additionally, investigating the LMMSE receiver holds significance as it stands as the optimal linear receiver which maximizes the SINR [11]. Accordingly, we analyze the LMMSE receiver in this subsection. To do so, we first separate out the desired user, n , from the rest of the signals and concatenate signals from the unknown channels of associated users, the non-associated user signals, and noise into a single term. As such, (15) can be rephrased as,

$$\mathbf{y} = \left(\hat{\mathbf{h}}_0 \right)_n s_n + \underbrace{\left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \mathbf{s}_{\bar{n}} + \mathbf{z}}_{\mathbf{z}'}. \quad (33)$$

In (33), $\mathbf{s}_{\bar{n}}$ contains signals from the $N-1$ associated users except user n , who is the user under consideration. With LMMSE receiver combining, the estimated symbol for user n is, $\hat{s}_n = \mathbf{V}_n^H \mathbf{y}$, where

$$\mathbf{V}_n = \left(\hat{\mathbf{H}}_0 \hat{\mathbf{H}}_0^H + \Theta \right)^{-1} \left(\hat{\mathbf{h}}_0 \right)_n, \quad (34)$$

is the weight vector for user n which gives the minimum mean squared error [36]. In (34), Θ contains the correlation of the noise plus non-associated user interference, unknown channel coefficients as well as the channel estimation error of the estimated channels. Due to the independence of the terms in \mathbf{z} , Θ simplifies to a diagonal matrix which can be expressed as,

$$\begin{aligned} \Theta &= \mathbb{E} \left[\mathbf{z} \mathbf{z}^H \right] \\ &= \mathbb{E} \left[\left(\tilde{\mathbf{H}}_0 \right)_0 \left(\tilde{\mathbf{H}}_0 \right)_0^H \right] + \mathbb{E} \left[\mathbf{E} \mathbf{E}^H \right] + \mathbb{E} \left[\mathbf{H}' \mathbf{H}'^H \right] + \sigma^2 \mathbf{I} \\ &= \text{diag}(\theta_1, \theta_2, \dots, \theta_M), \end{aligned} \quad (35)$$

where the value of θ_m is associated with antenna m and is the same as in (18). The instantaneous received SINR for user n when the LMMSE receiver is utilized is,

$$\gamma_n^{\text{LMMSE}} = \frac{\mathbf{V}_n^H \left(\hat{\mathbf{h}}_0 \right)_n \left(\hat{\mathbf{h}}_0 \right)_n^H \mathbf{V}_n}{\mathbf{V}_n^H \left[\left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H + \left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}} \left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}}^H + \mathbf{E} \mathbf{E}^H \right] \mathbf{V}_n + \Omega^{\text{LMMSE}} + \mathbf{H}' \mathbf{H}'^H + \sigma^2 \mathbf{I}}. \quad (36)$$

In (36), Ω^{LMMSE} is defined in (37) shown at the bottom of the page, whereas $\left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}}$ and $\mathbf{E}_{\bar{n}}$ represent $\tilde{\mathbf{H}}_0$ and \mathbf{E} with column vectors $\left(\tilde{\mathbf{h}}_0 \right)_n$, \mathbf{e}_n removed, respectively. Note that the r^{th} column of \mathbf{E} corresponds to the unknown channel of the r^{th} associated user and is indicated by \mathbf{e}_r .

Similar to Section IV-B, we proceed to obtain an approximation for the instantaneous SINR in (36). As before, we approximate the instantaneous channel gains of the non-associated users and unknown channels of the associated users with their expected values. As a result of the fact that $\left(\hat{\mathbf{h}}_0 \right)_r$, $\left(\tilde{\mathbf{h}}_0 \right)_r$, \mathbf{e}_r are all independent and zero mean, $\mathbb{E}[\Omega^{\text{LMMSE}}]$ reduces to zero in (37). Therefore, we can derive an approximation for (36) as follows:

$$\begin{aligned} \gamma_n^{\text{LMMSE}} &\approx \frac{\mathbf{V}_n^H \left(\hat{\mathbf{h}}_0 \right)_n \left(\hat{\mathbf{h}}_0 \right)_n^H \mathbf{V}_n}{\mathbf{V}_n^H \left[\left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H + \Theta \right] \mathbf{V}_n} \\ &= \left(\hat{\mathbf{h}}_0 \right)_n^H \Sigma^{-1} \left(\hat{\mathbf{h}}_0 \right)_n, \end{aligned} \quad (38)$$

where the received interference-plus-noise correlation matrix Σ is given by

$$\Sigma = \mathbb{E} \left[\mathbf{z}' \mathbf{z}'^H \right] = \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H + \Theta. \quad (39)$$

$$\Omega^{\text{LMMSE}} = \left(\tilde{\mathbf{h}}_0 \right)_n^H \mathbf{e}_n + \mathbf{e}_n^H \left(\tilde{\mathbf{h}}_0 \right)_n + \mathbf{E}_{\bar{n}} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H + \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \mathbf{E}_{\bar{n}}^H + \mathbf{E}_{\bar{n}} \left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}}^H + \left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}} \mathbf{E}_{\bar{n}}^H + \left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}} \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}}^H + \left(\hat{\mathbf{H}}_0 \right)_{\bar{n}} \left(\tilde{\mathbf{H}}_0 \right)_{\bar{n}}^H. \quad (37)$$

The expectation in (39) is taken over the small-scale fading of the unknown channel coefficients, channel estimation error and non-associated interference and noise. Using the matrix inversion lemma, we rephrase (39) as shown in (40) shown at the bottom of the page. We now introduce the variables

$$\iota = \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}^H \Theta^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}, \quad (41)$$

and

$$\kappa = \Theta^{-1/2} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}} \left(\mathbf{I}_{N-1} + \iota\right)^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}^H \Theta^{-1/2}, \quad (42)$$

for ease of notation. By substituting (41) and (42) into (40) we get,

$$\Sigma^{-1} = \Theta^{-1/2} (\mathbf{I} - \kappa) \Theta^{-1/2}. \quad (43)$$

From Lemma 1, when $\mathbf{R} = \Theta^{-1/2} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}$ and $\beta = 1$, $\mathbb{E}[\kappa]$ can also be expressed as a diagonal matrix. As such, following a similar approach to IV-B, we approximate κ by a simple diagonal matrix that has the same trace as κ :

$$\hat{\kappa} = \frac{\text{tr}(\kappa)}{M} \mathbf{I}_M. \quad (44)$$

By defining an error matrix Ξ_κ in terms of the trace of κ , we express κ as

$$\kappa = \hat{\kappa} + \Xi_\kappa. \quad (45)$$

Based on (43) and (45), the approximation in (38) can be further simplified as

$$\gamma_n^{\text{LMMSE}} \approx \left(\mathbf{h}_0\right)_n^H \Theta^{-1/2} (\mathbf{I} - \hat{\kappa}) \Theta^{-1/2} \left(\mathbf{h}_0\right)_n (\mathbf{I} - \Psi), \quad (46)$$

where

$$\Psi = \frac{\left(\mathbf{h}_0\right)_n^H \Theta^{-1/2} \Xi_\kappa \Theta^{-1/2} \left(\mathbf{h}_0\right)_n}{\left(\mathbf{h}_0\right)_n^H \Theta^{-1/2} (\mathbf{I} - \hat{\kappa}) \Theta^{-1/2} \left(\mathbf{h}_0\right)_n}. \quad (47)$$

Notice that Ψ in (47) is in the same form as (28). Therefore we can use the same approach as in Section IV-B to show that $\Psi \approx 0$. However, since κ is different from \mathbf{X} , we include the subsequent analysis for completeness.

First, we set out to show that the denominator of Ψ in (47) divided by M approaches a constant as $M \rightarrow \infty$. Similar to Section IV-B, following the law of large numbers, the

denominator of Ψ in (47) divided by M , approaches its mean as M tends towards infinity:

$$\begin{aligned} & \frac{\left(\mathbf{h}_0\right)_n^H \Theta^{-1/2} (\mathbf{I} - \hat{\kappa}) \Theta^{-1/2} \left(\mathbf{h}_0\right)_n}{M} \\ & \approx \frac{\text{tr}\left[(\mathbf{I} - \hat{\kappa}) \Theta^{-1} \mathbf{P}_n \Sigma_{\hat{\mathbf{g}}_n}\right]}{M}. \end{aligned} \quad (48)$$

However, (48) contains $\hat{\kappa}$, the nature of which remains unknown at this point. As such, to further simplify (48), we examine $\frac{\text{tr}(\kappa)}{M}$, as $\hat{\kappa}$ is defined in terms of $\text{tr}(\kappa)$.

$$\begin{aligned} \frac{\text{tr}(\kappa)}{M} &= \frac{1}{M} \text{tr}\left[\Theta^{-1/2} \left(\mathbf{H}_0\right)_{\bar{n}} \left(\mathbf{I}_{N-1} + \iota\right)^{-1} \left(\mathbf{H}_0\right)_{\bar{n}}^H \Theta^{-1/2}\right] \\ &= \frac{1}{M} \text{tr}\left[\left(\mathbf{I}_{N-1} + \iota\right)^{-1} \left(\mathbf{H}_0\right)_{\bar{n}}^H \Theta^{-1} \left(\mathbf{H}_0\right)_{\bar{n}}\right] \\ &= \frac{1}{M} \text{tr}\left[\left(\mathbf{I}_{N-1} + \iota^{-1}\right)^{-1}\right] \end{aligned} \quad (49)$$

To further simplify (49), we use the fact that,

$$\frac{\iota}{M} \approx \frac{\mathbb{E}[\iota]}{M} \quad \text{when } M \rightarrow \infty, \quad (50)$$

the proof of which is available in Appendix D. On account of channels being independent and Rayleigh distributed, $\mathbb{E}[\iota]$ reduces to a diagonal matrix as shown in (51) shown at the bottom of the page, in which the expectation is taken over the small-scale fading components in $\left(\hat{\mathbf{H}}_0\right)_{\bar{n}}$. Utilizing (50), (49) can be approximated as,

$$\begin{aligned} \frac{\text{tr}(\kappa)}{M} &\approx \text{tr}\left[\left(M \mathbf{I}_{N-1} + \left(\frac{\mathbb{E}[\iota]}{M}\right)^{-1}\right)^{-1}\right] \\ &= \frac{1}{M} \left(N - 1 + \sum_{\substack{r=1 \\ r \neq n}}^N \frac{1}{\text{tr}\left(\Sigma_{\hat{\mathbf{g}}_r} \mathbf{P}_r \Theta^{-1}\right)}\right). \end{aligned} \quad (52)$$

Substituting for $\hat{\kappa}$ from (44) and (52), we arrive at the following simplification for (48),

$$\begin{aligned} & \frac{\left(\mathbf{h}_0\right)_n^H \Theta^{-1/2} (\mathbf{I} - \hat{\kappa}) \Theta^{-1/2} \left(\mathbf{h}_0\right)_n}{M} \\ & \approx \frac{M - N + 1 - \sum_{\substack{r=1 \\ r \neq n}}^N \frac{1}{\text{tr}\left(\Sigma_{\hat{\mathbf{g}}_r} \mathbf{P}_r \Theta^{-1}\right)}}{M^2} \text{tr}\left(\Sigma_{\hat{\mathbf{g}}_n} \mathbf{P}_n \Theta^{-1}\right) \\ & \rightarrow \text{tr}\left(\Sigma_{\hat{\mathbf{g}}_n} \mathbf{P}_n \Theta^{-1}\right) \triangleq \mu^{\text{LMMSE}} > 0. \end{aligned} \quad (53)$$

Therefore, as M becomes large, the denominator of (47) divided by M approaches its mean value, μ^{LMMSE} .

$$\begin{aligned} \Sigma^{-1} &= \Theta^{-1} - \Theta^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}} \left(\mathbf{I}_{N-1} + \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}^H \Theta^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}\right)^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}^H \Theta^{-1} \\ &= \Theta^{-1/2} \left(\mathbf{I} - \Theta^{-1/2} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}} \left(\mathbf{I}_{N-1} + \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}^H \Theta^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}\right)^{-1} \left(\hat{\mathbf{H}}_0\right)_{\bar{n}}^H \Theta^{-1/2}\right) \Theta^{-1/2}. \end{aligned} \quad (40)$$

$$\mathbb{E}(\iota) = \text{diag}\left(\text{tr}\left(\Sigma_{\hat{\mathbf{g}}_1} \mathbf{P}_1 \Theta^{-1}\right), \dots, \text{tr}\left(\Sigma_{\hat{\mathbf{g}}_{n-1}} \mathbf{P}_{n-1} \Theta^{-1}\right), \text{tr}\left(\Sigma_{\hat{\mathbf{g}}_{n+1}} \mathbf{P}_{n+1} \Theta^{-1}\right), \dots, \text{tr}\left(\Sigma_{\hat{\mathbf{g}}_N} \mathbf{P}_N \Theta^{-1}\right)\right). \quad (51)$$

Now, we shift our focus to the numerator of Ψ in (47). As per Lemma 2, when $M \gg N$:

$$\mathbb{E} \left(\frac{\left| (\mathbf{h}_0)_n^H \mathbf{\Theta}^{-\frac{1}{2}} \mathbf{\Xi}_\kappa \mathbf{\Theta}^{-\frac{1}{2}} (\mathbf{h}_0)_n \right|^2}{M^2} \right) \rightarrow 0. \quad (54)$$

From (53) and (54) we can conclude that for large M ,

$$0 < |\Psi| \ll 1. \quad (55)$$

Using (44), (52) and (55), we finally approximate γ_n^{LMMSE} in (46) for $M \gg N$ as

$$\gamma_n^{\text{LMMSE}} \approx C^{\text{LMMSE}} (\hat{\mathbf{h}}_0)_n^H \mathbf{\Theta}^{-1} (\hat{\mathbf{h}}_0)_n, \quad (56)$$

where

$$C^{\text{LMMSE}} = \left(\frac{M - N + 1 - \sum_{r \neq n}^N \frac{1}{\text{tr}(\mathbf{\Sigma}_{\hat{\mathbf{g}}_r} \mathbf{P}_r \mathbf{\Theta}^{-1})}}{M} \right). \quad (57)$$

V. PERFORMANCE ANALYSIS

This section is focused on analyzing the performance of the proposed architecture by utilizing the derived approximations for the SINR in Section IV. We derive novel approximations for achievable rate and SEP for the considered two receivers. In this section, γ_n and C correspond to received SINR for the considered receiver (i.e., γ_n^{ZF} or γ_n^{LMMSE}) and the constant defined for that respective receiver (i.e., C^{ZF} or C^{LMMSE}).

A. ACHIEVABLE RATE ANALYSIS

The achievable rate of user n is given by

$$R_n = (1 - \alpha) \mathbb{E}[\log_2(1 + \gamma_n)], \quad (58)$$

where $\alpha = \tau_p / \tau_c$, and τ_c is the coherence interval in symbols. In scenarios with perfect CSI, the entire coherence block is used for data transmission, and (58) reduces to $R_n = \mathbb{E}[\log_2(1 + \gamma_n)]$.

We note that the Jensen's inequality provides a simple upper bound for R_n [21], given by

$$R_n \leq (1 - \alpha) \log_2(1 + \mathbb{E}\{\gamma_n\}). \quad (59)$$

The expected value of γ_n in the form of (32) and (56) can be computed as,

$$\begin{aligned} \mathbb{E}[\gamma_n] &= C \mathbb{E} \left[(\hat{\mathbf{h}}_0)_n^H \mathbf{\Sigma}^{-1} (\hat{\mathbf{h}}_0)_n \right] \\ &= C \sum_{l \in \{i,j,k\}} \sum_{m=1}^{M_l} \frac{P_{l,n}^2 \tau_p}{\theta_m (P_{l,n} \tau_p + \sigma_{w_m}^2)}. \end{aligned} \quad (60)$$

When deriving (60), we have used (11). Substituting (60) into (59) results in the closed form expression given by

$$R_n \approx (1 - \alpha) \log_2 \left(1 + C \sum_{l \in \{i,j,k\}} \sum_{m=1}^{M_l} \frac{P_{l,n}^2 \tau_p}{\theta_m (P_{l,n} \tau_p + \sigma_{w_m}^2)} \right). \quad (61)$$

Increasing τ_p facilitates the learning of more user channels but reduces the time allocated for data transmission, impacting the achievable rate (61). In practice, the selection of τ_p should aim for an optimal point, striking a balance between channel estimation and achievable rate [37].

B. SYMBOL ERROR PROBABILITY ANALYSIS

We use a characteristic function (CF) based approach to derive the SEP of user n . Noting that the channel estimates from user n to each receiving antenna are independent and Gaussian, we first write the joint PDF of the channel gains in $(\hat{\mathbf{h}}_0)_n$ as

$$f \left[(\hat{\mathbf{h}}_0)_n \right] = \frac{1}{\pi^M \det(\mathbf{P}_n \mathbf{\Sigma}_{\hat{\mathbf{g}}_n})} e^{-\left(\hat{\mathbf{h}}_0 \right)_n^H (\mathbf{P}_n \mathbf{\Sigma}_{\hat{\mathbf{g}}_n})^{-1} (\hat{\mathbf{h}}_0)_n}, \quad (62)$$

where $(\hat{\mathbf{h}}_0)_n = \mathbf{P}_n^{1/2} \hat{\mathbf{g}}_n$. By substituting (62) into the definition of the CF of γ_n in [38] we obtain the characteristic function for γ_n which is given in (63) shown at the bottom of the page. In (63), $(\hat{h}_0)_{m,n}$ is the estimated channel from user n to the m^{th} antenna in the cluster. The multifold integral in (63) can be solved by using [39, Lemma 2] on (63),

$$\phi_{\gamma_n}(t) \approx \frac{1}{|1 - jtC\mathbf{\Theta}^{-1} \mathbf{P}_n \mathbf{\Sigma}_{\hat{\mathbf{g}}_n}|}. \quad (64)$$

Based upon (64), we use the moment generating function (MGF) approach in [38] to derive an expression for the MGF of γ_n , i.e., $\mathcal{M}_{\gamma_n}(s)$ as

$$\mathcal{M}_{\gamma_n}(s) \approx \frac{1}{\prod_{m=1}^M (1 - sC(\vartheta_n)_m)}. \quad (65)$$

In (65), $(\vartheta_n)_m$, which is the (m, m) -th element of $\mathbf{\Theta}^{-1} \mathbf{P}_n \mathbf{\Sigma}_{\hat{\mathbf{g}}_n}$, is associated with antenna m . Therefore, when m is an antenna of BS $_i$,

$$(\vartheta_n)_m = \frac{\tau_p P_{i,n}^2}{(\sigma_{w_m}^2 + \tau_p P_{i,n}) \theta_m}. \quad (66)$$

Accordingly, the SEP with coherent M_{PSK} phase shift keying modulation for user n , can be written as [38],

$$P_e^{(n)} = \frac{1}{\pi} \int_0^T \mathcal{M}_{\gamma_n}(s) \left(\frac{-Z}{\sin^2 \theta} \right) d\theta, \quad (67)$$

$$\phi_{\gamma_n}(t) = \mathbb{E} \left[e^{jt\gamma_n} \right] = \frac{1}{\pi^M \det(\mathbf{P}_n \mathbf{\Sigma}_{\hat{\mathbf{g}}_n})} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\left(\hat{\mathbf{h}}_0 \right)_n^H \left[(\mathbf{P}_n \mathbf{\Sigma}_{\hat{\mathbf{g}}_n})^{-1} - jtC\mathbf{\Theta}^{-1} \right] (\hat{\mathbf{h}}_0)_n} d(\hat{\mathbf{h}}_0)_{1,n} \dots d(\hat{\mathbf{h}}_0)_{M,n}. \quad (63)$$

where $Z = \sin^2(\pi/M_{\text{PSK}})$ and $T = \frac{(M_{\text{PSK}}-1)\pi}{M_{\text{PSK}}}$. The expression in (67) can be solved by using the tight approximation in [40] to produce

$$P_e^{(n)} \approx \left(\frac{T}{2\pi} - \frac{1}{6}\right) \mathcal{M}_{\gamma_n}(-Z) + \frac{1}{4} \mathcal{M}_{\gamma_n}\left(-\frac{4Z}{3}\right) + \left(\frac{T}{2\pi} - \frac{1}{4}\right) \mathcal{M}_{\gamma_n}\left(-\frac{Z}{\sin^2 T}\right). \quad (68)$$

In Section VII, we present simulations showing that (61) and (68) give accurate approximations for the achievable rate and SEP, respectively, for an arbitrary serving user n .

VI. COMPLEXITY ANALYSIS AND NETWORK OVERHEAD

In this section, we analyze the computational complexity and network overhead per coherence block for the proposed approach, comparing it with the two clustering schemes outlined in Section II. The complexity is analyzed in two aspects: combining vector computation and channel estimation complexity, whereas the network overhead is measured by the number of user channels transmitted from the BSs to the CPU.

The computational complexity associated with combining vector computation is generally similar for identical receiver designs, contingent upon the size of the channel matrix, which, in turn, relies on the number of users processed and BSs in the cooperating cluster. Quantifying this is challenging due to the situation-specific nature of the parameters associated. However, our proposed approach introduces a block structure to the channel matrix, resulting from the zeroing of the channels of the weak users. The introduced additional structure contributes to lowering the complexity associated with computing combining vectors in the proposed approach compared to traditional static clustering with the same number of cooperating BSs but a full channel matrix. This efficiency arises from the more streamlined calculation of the inverse of a block matrix through the utilization of the Schur complement [41].

The channel estimation complexity and the network overhead depend on the total number of user channels estimated (\mathcal{T}_C) in a coherence block. Note that the number of user channels estimated at a BS in the proposed architecture depends on the predefined threshold λ and N_{max} and can be adjusted to best suit the system. For the three approaches considered, the metric \mathcal{T}_C is given below.

$$\mathcal{T}_C^{\text{SC}} = (|\mathcal{N}_i| + |\mathcal{N}_j| + |\mathcal{N}_k|)(M_i + M_j + M_k), \quad (69a)$$

$$\mathcal{T}_C^{\text{PJP}} = |\mathcal{N}_i^{\text{PJP}}|M_i + |\mathcal{N}_j^{\text{PJP}}|M_j + |\mathcal{N}_k^{\text{PJP}}|M_k, \quad (69b)$$

$$\mathcal{T}_C^{\text{NC}} = |\mathcal{N}_i|M_i. \quad (69c)$$

While a specific channel estimation scheme has been employed in Section III, it is essential to note that the proposed system is independent of the chosen channel estimation approach. Thus, assuming that the estimation of a user channel requires executing \mathcal{C} complex operations, the

channel estimation complexity would be $\mathcal{C}\mathcal{T}_C$. Additionally, the network overhead amounts to \mathcal{T}_C complex symbols per coherence block.

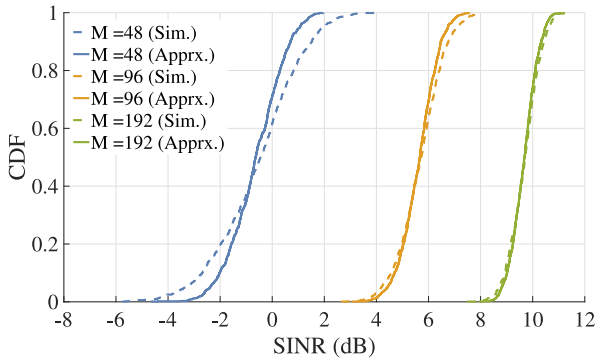
In Section VII, we use the metric, \mathcal{T}_C , in (69) to gain insights into the complexity and network overheads by providing numerical results for the example scenario shown in Fig. 1. Despite the superior performance of PJP compared to the traditional static clustering architecture (Fig. 5), the number of user channels learned in PJP is lower (Table 1) supporting our claims of low network overhead and complexity. Furthermore, we discuss how adjusting the predefined threshold λ will increase (or decrease) the number of user channels learned. This adjustment, in turn, has a direct influence on the overall system performance.

VII. SIMULATIONS AND NUMERICAL RESULTS

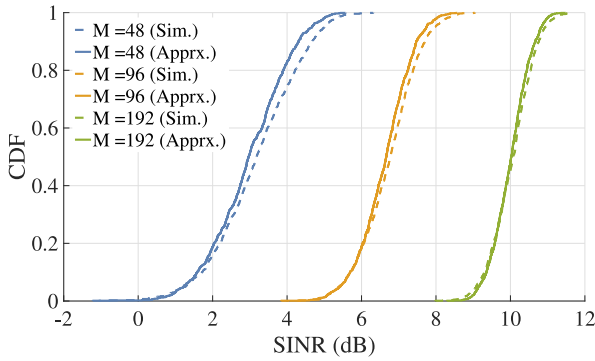
In this section, we verify the analysis presented in previous sections through Monte Carlo simulations. While our results apply to any general setup as detailed in Section II, for verification purposes we use the example scenario in Fig. 1, where the total number of users in the system, $N_{\text{sys}} = 119$. We illustrate the accuracy of our approximations for SINR, achievable rate and SEP for both considered receivers. We use a distance-dependent path loss model as detailed in [11] and set the path loss exponent to 4. We assume that all the BSs in the PJP cluster have an equal number of antennas (i.e., $M_i = M_j = M_k$).

In this setting, within each coherence interval, a share of resources is set aside for the pilot transmissions from the users. As such, we set $\tau_c = 200$ and $\tau_p = 40$. In the subsequent simulations, we divide the τ_p pilots into disjoint groups and assign a set of disjoint pilot sets to each cell. The classical four-color theorem [42] is utilized to ensure no two neighboring cells share the same pilot group and therefore, each cell is assigned 10 pilots each. We use the same pilot assignment scheme for all three clustering methods.⁵ As the four-color theorem restricts adjacent cells to have different sets of pilots, BSs in both the 3-cell static clustering approach and PJP can learn additional users outside of the assigned cell without having two users sharing the same pilot. However, when it comes to learning the channels of users who are located outside the designated cell area, the BSs in these two approaches choose different users. In 3-cell static clustering, a BS learns the channels of users located within the cell areas of the other two cooperative BSs. On the other hand, in PJP, a BS chooses to learn the channel of a user surpassing

⁵Although we consider the aforementioned pilot assignment scheme in numerical results, our approach is not constrained to this specific scheme. This pilot assignment was deliberately chosen to enable a fair comparison between the proposed scheme and the 3-cell static clustering approach. By adopting this specific pilot assignment strategy, both approaches can operate without being adversely affected by pilot contamination. This ensures that any performance differences observed between the two methods are solely attributed to their inherent characteristics rather than the coherent interference caused by pilot contamination.



(a) ZF receiver.



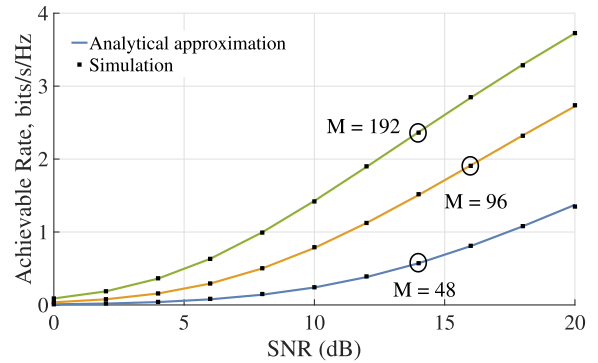
(b) LMMSE receiver.

FIGURE 2. The CDF plot of the received SINR U_1 , when processed by the proposed approach.

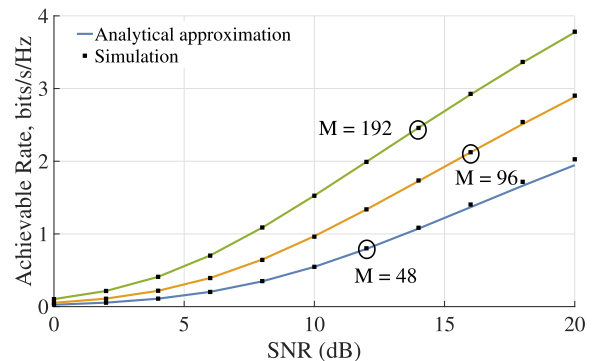
the threshold in (3), and these users are not required to be located in a specific cell, as long as the BS is not currently serving a user on that particular pilot. In the following plots, we consider U_1 as the user of interest when obtaining results unless otherwise stated.

First, we demonstrate the validity of our approximations for the instantaneous received SINR by plotting its cumulative distribution function (CDF) for the considered user. In Fig. 2, we vary the number of antennas at each BS as $M_i = 16, 32, 64$ and plot the CDFs of the approximate received SINR in (32) and (56) alongside those of the exact SINR in (20) and (36), respectively. The plots clearly illustrate that our analytical expressions in (32) and (56) very closely approximate the exact received SINR and the accuracy of the approximation increases significantly as we increase the number of antennas at each BS. Interestingly, even when we have a highly loaded system with the ratio between the number of antennas and users being 48 : 37 (i.e., $M_i = 16$), the approximation still holds well. In a heavily loaded system, we wouldn't typically expect such accurate approximations to apply effectively to massive MIMO scenarios. Therefore, these results are very promising.

Next, Fig. 3 and Fig. 4 consider PJP and plot the achievable rate and SEP, respectively, of the user under consideration, versus the received SNR. When obtaining



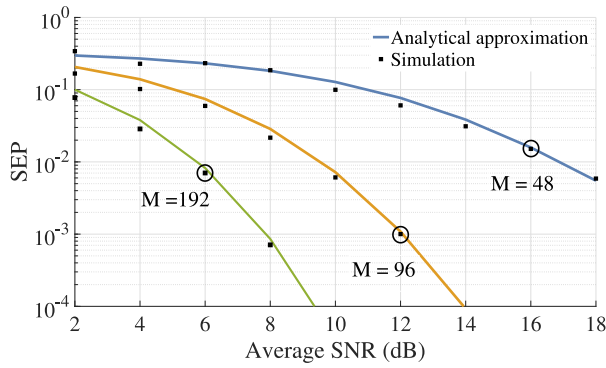
(a) ZF receiver.



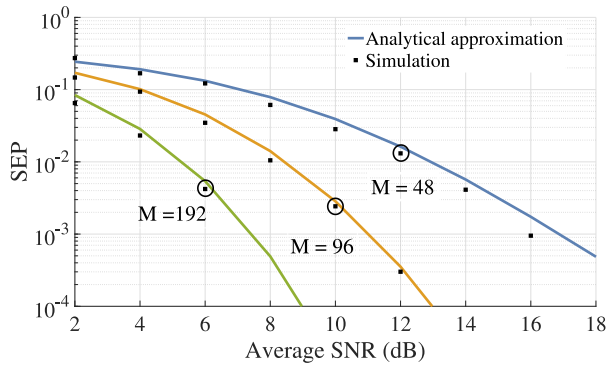
(b) LMMSE receiver.

FIGURE 3. Achievable rate of U_1 , when processed by the proposed approach.

SEP results, we assume that the transmitted symbols are modulated using binary phase shift keying (BPSK). We consider three different scenarios by changing the number of antennas located at each BS as $M_i = 16, 32, 64$. Fig. 3 shows that when the number of antennas is large, our approximate achievable rate results, generated using (61), very accurately predict the simulation points, produced via Monte Carlo simulation, throughout the full range of SNRs. It should be noted that reducing the number of cooperative antennas below $M = 39$ is not viable as it would result in a scenario where the number of users served in the PJP system exceeds the number of antennas available to serve them. In this overloaded scenario, the ZF receiver cannot be used as the matrix requiring inversion is singular. Interestingly, the approximation still holds remarkably well even as the number of antennas approaches this limit (i.e., $M = 48$). We also observe that the achievable rate increases with the increasing number of antennas. Furthermore, Fig. 4 clearly shows that the approximate error probability results, generated using (68), closely align with the values obtained through Monte Carlo simulation. Moreover, it is evident that the error probability performance improves as the number of antennas increases. The gaps between the analytical approximations and the simulation points in both plots are very small even when M_i is 16.



(a) ZF receiver.



(b) LMMSE receiver.

FIGURE 4. SEP of U_1 when processed by the proposed approach.

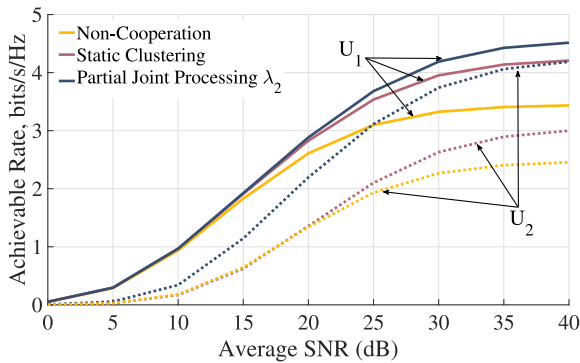


FIGURE 5. Achievable rate of U_1 and U_2 when processed by different architectures.

To emphasize the advantages of the suggested method, we contrast its performance with that of two clustering techniques described in Section II.6 This comparison, depicted in Fig. 5, is based on the achievable rate when the LMMSE receiver is used. We evaluate the achievable rates of U_1 who is located at the cluster-center and U_2 who is located at the cluster edge in the static clustering architecture. Results for

⁶Due to the requirement of a large number of orthogonal pilots to operate without pilot contamination, the associated excessive network overhead, and the computational complexity linked with dynamic cooperation, we refrain from comparing the proposed approach with dynamic clustering. However, it is worth noting that as the BSs possess CSI about more users, dynamic clustering is expected to offer superior performance.

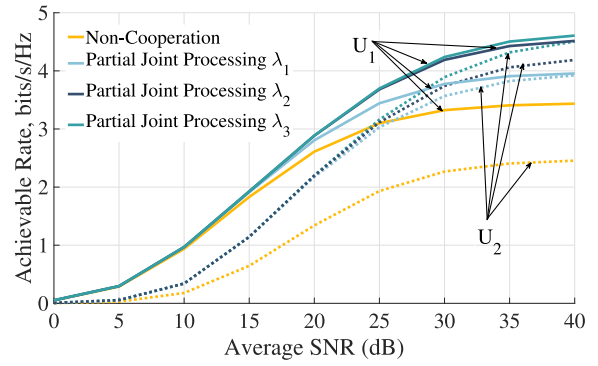


FIGURE 6. Achievable rate of U_1 and U_2 at different λ values.

the PJP approach are shown for threshold λ_2 , for which the selected users are depicted in Fig. 1(b). Both U_1 and U_2 are processed by the static cluster made by BS₁ – BS₂ – BS₃ cooperating. Note that U_1 would be served by the BS₁ – BS₂ – BS₃ cluster in the PJP approach and by BS₃ in the non-cooperating approach. Similarly, BS₁ – BS₄ – BS₅ cluster processes U_2 in the PJP approach and BS₁ processes U_2 in the non-cooperating approach. Results for static clustering and non-cooperating approaches are based on [23] and [36], respectively. The plot clearly depicts a higher performance for both users when processed by the proposed architecture. In particular, it can be observed that U_2 , who is a cluster-edge user, has a significantly better achievable rate when processed by the proposed approach. This performance gain is expected as our modified user-centric architecture perceives the traditionally defined cluster-edge users as cluster-center users of a different PJP cluster. Note that the slight performance gap evident for the two users in the PJP architecture is due to the fact that in our simulation setup, U_2 is interfered with a higher number of users than U_1 . Moreover, the effect of non-associated user interference brings forth a saturation regime at high SINRs where a further increase in SINR does not notably improve the achievable rate. It should be noted that this phenomenon is not unique to the proposed approach but appears in traditional clustering approaches as well [32]. Consequently, traditional out-of-cluster interference mitigation techniques could be applied to the proposed approach to alleviate such interference, particularly in high-density networks.

In Fig. 6, we have examined the PJP approach under three different threshold values where $\lambda_1 > \lambda_2 > \lambda_3$. We have also tabulated the number of user channels learned by each architecture and at different λ values in Table 1. From Fig. 6 and Table 1, it is evident that there exists a trade-off between the chosen threshold and the number of user channels learned by a BS, influencing the overall performance of the system. Opting for a lower threshold leads to learning a higher number of user channels by the BSs, thereby enhancing the system's overall performance. Conversely, setting λ to a large value prevents BSs from

TABLE 1. Comparison between the number of channels learned by a BS in different architectures.

BS	NC	SC	PJP- λ_1	PJP- λ_2	PJP- λ_3
1	9	27	10	17	22
2	10	27	10	15	19
3	8	27	10	15	24
Avg.	9	27	10	15.6	21.6
\mathcal{T}_C	320	2592	960	1504	2080

learning channels of users beyond the designated cell area. In such a scenario, cooperation becomes redundant, as each BS only learns channels of the users within its assigned cell area. Consequently, the system reverts to a non-cooperative setup and the performance of the PJP system would collapse to that of Non-cooperation. However, using (56) with the PJP channel matrix (\mathbf{H}_0) to derive approximations would result in errors. This discrepancy arises because, even though the channel of each user is learned only at a single BS, (57) contains terms for $N = |\mathcal{N}^{\text{PJP}}_i| + |\mathcal{N}^{\text{PJP}}_j| + |\mathcal{N}^{\text{PJP}}_k|$ users. Hence, when the system reverts to a non-cooperative setup, \mathbf{H}_i^{NC} should be used in the analysis. With this adjustment, the achievable rate of the PJP approach would correspond to the yellow curve in Fig. 6.

As is evident from Table 1, to provide similar performance for U_1 , a BS operating in the proposed approach needs to learn around 16 user channels on average whereas the same BS operating in the static clustering approach learns 27 user channels. The total number of user channels estimated in a coherence block, \mathcal{T}_C , for each architecture is presented in Table 1. These are obtained by substituting the corresponding values to (69a), (69b) and (69c) when $M_i = M_j = M_k = 32$. For non-cooperation, \mathcal{T}_C is given for BS₂, which learns the highest number of user channels. As such, a BS in the proposed approach sends a lower number of user channels to the CPU. Consequently, the proposed approach exhibits a 40% reduction in the network overhead and channel estimation complexity for this example. Accordingly, it can be seen that the proposed architecture improves the achievable rate of traditional cluster-edge users while also reducing the overhead to the network.

VIII. CONCLUSION

In this paper, we proposed a modified user-centric approach to BS cooperation. We consider the uplink of a multi-cell scenario where a cluster of large antenna arrays jointly detect the signals from multiple serving users. We obtained novel closed-form expressions to approximate the instantaneous received SINR, the achievable rate and the SEP of an arbitrary serving user. These expressions are based on ZF and LMMSE receivers. Our approximations were derived for Rayleigh fading channels and took into consideration the impact of channel estimation errors as well as interference stemming from non-associated users. The accuracy of the

derived approximations was illustrated through simulations. The performance of the proposed scheme was compared to more traditional clustering schemes using numerical examples showing the advantages of the proposed scheme. Exploring the impact of pilot contamination on the proposed scheme presents an intriguing avenue for future extensions of this research.

APPENDIX A PROOF OF LEMMA 1

In the following we show that $\mathbf{D} = \mathbb{E}[\mathbf{R}(\mathbf{R}^H\mathbf{R} + \beta\mathbf{I}_{N-1})^{-1}\mathbf{R}^H]$ is a diagonal matrix by proving that the off-diagonal elements are zero. First, let us write an off-diagonal element of \mathbf{D} as

$$\mathbf{D}_{i,j} = \mathbb{E}\left[\mathbf{r}_i(\mathbf{r}_i^H\mathbf{r}_i + \mathbf{Q})^{-1}\mathbf{r}_j^H\right], \quad (\text{A.1})$$

where $i \neq j$, \mathbf{r}_i is the i -th row of \mathbf{R} and $\mathbf{Q} = \mathbf{R}_i^H\mathbf{R}_i + \beta\mathbf{I}_{N-1}$ with \mathbf{R}_i representing \mathbf{R} with row vector \mathbf{r}_i removed. Conditioned on \mathbf{Q} and \mathbf{r}_j the expectation in (A.1) can be re-expressed as

$$\mathbf{D}_{i,j} = \mathbb{E}\left[\underbrace{\mathbb{E}\left[\mathbf{r}_i(\mathbf{r}_i^H\mathbf{r}_i + \mathbf{Q})^{-1}\mathbf{r}_j^H \mid \mathbf{Q}, \mathbf{r}_j\right]}_{T_{i,j}}\right], \quad (\text{A.2})$$

where the inside expectation, $T_{i,j}$, is over the variables in the row vector \mathbf{r}_i . We rearrange the inverse in $T_{i,j}$ as (A.3) by applying the inverse of a small rank adjustment [43].

$$\begin{aligned} T_{i,j} &= \mathbb{E}\left[\mathbf{r}_i\left[\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{r}_i^H\mathbf{r}_i\mathbf{Q}^{-1}}{1 + \mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H}\right]\mathbf{r}_j^H \mid \mathbf{Q}, \mathbf{r}_j\right] \\ &= \mathbb{E}\left[\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H \mid \mathbf{Q}, \mathbf{r}_j\right] \\ &\quad - \mathbb{E}\left[\frac{\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H}{1 + \mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H} \mid \mathbf{Q}, \mathbf{r}_j\right]. \end{aligned} \quad (\text{A.3})$$

As $\mathbb{E}[\mathbf{r}_i] = 0$, the first expectation goes to zero and (A.3) can be further simplified to produce

$$\begin{aligned} T_{i,j} &= -\mathbb{E}\left[\frac{\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H}{1 + \mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H} \mid \mathbf{Q}, \mathbf{r}_j\right] \\ &= -\mathbb{E}\left[\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H - \frac{\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H}{1 + \mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H} \mid \mathbf{Q}, \mathbf{r}_j\right] \\ &= \mathbb{E}\left[\frac{\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H}{1 + \mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H} \mid \mathbf{Q}, \mathbf{r}_j\right]. \end{aligned} \quad (\text{A.4})$$

Let $\Upsilon = \frac{\mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_j^H}{1 + \mathbf{r}_i\mathbf{Q}^{-1}\mathbf{r}_i^H}$, then $T_{i,j} = \mathbb{E}[\Upsilon \mid \mathbf{Q}, \mathbf{r}_j]$. Next, we substitute $\mathbf{r}_i = \mathbf{v}_i\mathbf{G}_i$, where the row vector $\mathbf{v}_i \sim \mathcal{CN}(0, \mathbf{I})$ and \mathbf{G}_i is a diagonal matrix to re-express Υ as

$$\Upsilon = \frac{\mathbf{v}_i\mathbf{G}_i\mathbf{Q}^{-1}\mathbf{r}_j^H}{1 + \mathbf{v}_i\mathbf{G}_i\mathbf{Q}^{-1}\mathbf{G}_i^H\mathbf{v}_i^H}. \quad (\text{A.5})$$

As \mathbf{Q} is Hermitian matrix, it can be written in terms of a unitary matrix Φ and a diagonal matrix Ω such that $\mathbf{Q} = \Phi\Omega\Phi^H$ which results in

$$\Upsilon = \frac{\mathbf{v}_i \mathbf{G}_i \Phi \Omega^{-1} \Phi^H \mathbf{r}_j^H}{1 + \mathbf{v}_i \mathbf{G}_i \Phi \Omega^{-1} \Phi^H \mathbf{G}_i^H \mathbf{v}_i^H}. \quad (\text{A.6})$$

Let $\mathbf{W} = (\mathbf{G}_i \Phi \Omega^{-1} \Phi^H \mathbf{G}_i^H)$, then \mathbf{W} is Hermitian. Therefore, $\mathbf{W} = \Psi \mathbf{B} \Psi^H$, where Ψ is a unitary matrix and \mathbf{B} is a diagonal matrix. Hence,

$$\Upsilon = \frac{\mathbf{v}_i \Psi \mathbf{B} \Psi^H (\mathbf{G}_i^H)^{-1} \mathbf{r}_j^H}{1 + \mathbf{v}_i \Psi \mathbf{B} \Psi^H \mathbf{v}_i^H}. \quad (\text{A.7})$$

Now (A.7) can be expressed in a simpler notation as

$$\Upsilon = \frac{\mathbf{x} (\mathbf{B} \Psi^H (\mathbf{G}_i^H)^{-1} \mathbf{r}_j^H)}{1 + \mathbf{x} \mathbf{B} \mathbf{x}^H}, \quad (\text{A.8})$$

where the row vector $\mathbf{x} = \mathbf{v}_i \Psi$ and each element in \mathbf{x} has a complex normal distribution. In (A.8), the denominator is an even function of \mathbf{x} whereas the numerator is odd. As such, the distribution of Υ is symmetric around zero. Therefore, the expectation of Υ is zero, i.e., $T_{i,j} = 0$. Finally, it follows that matrix \mathbf{D} is diagonal as $\mathbf{D}_{i,j} = 0$ for all $i \neq j$.

APPENDIX B PROOF FOR THE IDEMPOTENT NATURE OF \mathbf{X}

The idempotent nature [43] of matrix \mathbf{X} can be shown from the below simplifications.

$$\begin{aligned} \mathbf{X}\mathbf{X} &= \Theta^{-\frac{1}{2}} (\hat{\mathbf{H}}_0)_{\bar{n}} \left(\left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}} \right)^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-\frac{1}{2}} \right. \\ &\quad \left. \Theta^{-\frac{1}{2}} (\hat{\mathbf{H}}_0)_{\bar{n}} \left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}} \right)^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-\frac{1}{2}} \right) \\ &= \Theta^{-\frac{1}{2}} (\hat{\mathbf{H}}_0)_{\bar{n}} \left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}} \right)^{-1} \\ &\quad \underbrace{\left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}} \left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}} \right)^{-1} \right)}_{\mathbf{I}} \\ &\quad \left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-\frac{1}{2}} \right) \\ &= \Theta^{-\frac{1}{2}} (\hat{\mathbf{H}}_0)_{\bar{n}} \left((\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}} \right)^{-1} (\hat{\mathbf{H}}_0)_{\bar{n}}^H \Theta^{-\frac{1}{2}} \\ &= \mathbf{X}. \end{aligned} \quad (\text{B.1})$$

APPENDIX C PROOF OF LEMMA 2

Consider scalar ξ defined as $\xi = \frac{(\mathbf{u}^H \Delta \mathbf{u})}{M}$ where $\mathbf{u} \sim \mathcal{CN}(0, \Sigma_{\mathbf{u}})$ is a $1 \times M$ vector with diagonal covariance matrix $\Sigma_{\mathbf{u}}$ and Δ is an $M \times M$ diagonal matrix. The expectation of the mean square of ξ is,

$$\mathbb{E}[\xi^2] = \left(\frac{|\mathbf{u}^H \Delta \mathbf{u}|^2}{M^2} \right). \quad (\text{C.1})$$

Expanding (C.1),

$$\mathbb{E}[\xi^2] = \mathbb{E} \left(\frac{\sum_j \sum_r u_r^* \Delta_{r,j} u_j \times \sum_k \sum_s u_s \Delta_{s,k}^* u_k^*}{M^2} \right) \quad (\text{C.2a})$$

$$\begin{aligned} &= \frac{1}{M^2} \mathbb{E} \left(\sum_j \Delta_{j,j} |u_j|^2 \sum_k \Delta_{k,k}^* |u_k|^2 \right. \\ &\quad \left. + \sum_j \sum_{r \neq j} u_r^* \Delta_{r,j} u_j u_r \Delta_{r,j}^* u_j^* \right), \end{aligned} \quad (\text{C.2b})$$

where $(\cdot)^*$ denotes the complex conjugate, $\Delta_{r,j}$ denotes the (r, j) -th element of Δ and u_r denotes the r th element of \mathbf{u} . The expression in (C.2b) is obtained by noticing that when $s \neq k$, $\mathbb{E}[u_s u_k] = 0$ and rearranging the terms in (C.2a). Since $|u_j|^2$ has an exponential distribution, $\mathbb{E}[|u_j|^4] = 2\sigma_j^2$, where σ_j^2 is the entry at (j, j) -th position of $\Sigma_{\mathbf{u}}$. Now we can further simplify (C.2b) as (C.3) shown at the bottom of the page.

To evaluate the first term in (C.3), consider a diagonal element of $\Delta \Delta^H$ given by,

$$\begin{aligned} \frac{[\Delta \Delta^H]_{r,r}}{M^2} &= \frac{\left[\mathbf{P}_n^{\frac{1}{2}} \Sigma^{-\frac{1}{2}} \Xi_{\kappa} \Sigma^{-\frac{1}{2}} \mathbf{P}_n \Sigma^{-\frac{1}{2}} \Xi_{\kappa}^H \Sigma^{-\frac{1}{2}} \mathbf{P}_n^{\frac{1}{2}} \right]_{r,r}}{M^2} \\ &= \frac{[\mathbf{P}_n]_{r,r} (\Sigma^{-1})_{r,r} (\Xi_{\kappa})_r (\Sigma^{-1} \mathbf{P}_n) (\Xi_{\kappa})_r^H}{M^2}. \end{aligned} \quad (\text{C.4})$$

In (C.4), $(\Xi_{\kappa})_r$ is the r th row of the matrix Ξ_{κ} and $(\Sigma)_{r,r}^{-\frac{1}{2}}$ and $(\mathbf{P}_n)_{r,r}^{\frac{1}{2}}$ denotes the r th diagonal element of the corresponding matrix. We proceed to upper bound the expected value of (C.4) by replacing the diagonal matrix \mathbf{P}_n by $(\mathbf{P}_n)_{\max} \mathbf{I}$ with $(\mathbf{P}_n)_{\max}$ representing the maximum element in \mathbf{P}_n , and $(\Sigma^{-1})_{r,r} \leq \frac{1}{\sigma^2}$ to write,

$$\begin{aligned} \frac{(\Delta \Delta^H)_{r,r}}{M^2} &\leq \frac{(\mathbf{P}_n)_{\max}}{\sigma^2} \frac{(\Xi_{\kappa})_r \Sigma^{-1} \mathbf{P}_n (\Xi_{\kappa})_r^H}{M^2} \\ &\leq \left(\frac{(\mathbf{P}_n)_{\max}}{\sigma^2} \right)^2 (\Xi_{\kappa})_r (\Xi_{\kappa})_r^H. \end{aligned} \quad (\text{C.5})$$

$$\begin{aligned} \mathbb{E}[\xi^2] &= \frac{1}{M^2} \mathbb{E} \left(2 \sum_j |\Delta_{j,j}|^2 + \sum_j \sum_{k \neq j} \Delta_{j,j} \Delta_{k,k}^* + \sum_j \sum_{r \neq j} \Delta_{r,j} \Delta_{r,j}^* \right) \\ &= \frac{1}{M^2} \mathbb{E} \left(\sum_j \sum_r |\Delta_{r,j}|^2 + \sum_j \sum_k \Delta_{j,j} \Delta_{k,k}^* \right) = \frac{1}{M^2} \mathbb{E} \left(\text{tr}(\Delta \Delta^H) + \text{tr}(\Delta) \text{tr}(\Delta^H) \right). \end{aligned} \quad (\text{C.3})$$

Also, by substituting $(\mathbf{\Xi}_k)_r = \kappa_r - \hat{\kappa}_r$ and $\hat{\kappa}_r = \mathbf{C}\mathbf{I}$,

$$\hat{\kappa}_r \hat{\kappa}_r^H = \left(\kappa \kappa^H \right)_{r,r} - C \kappa_{r,r}^* - C \kappa_{r,r} + C^2, \quad (\text{C.6})$$

where $\kappa_{r,r} \approx N - 1 - \sum_{\substack{r=1 \\ r \neq n}}^N \frac{1}{1 + \text{tr}(\mathbf{P}_r \mathbf{\Sigma}^{-1})}$ is the (r, r) -th element of κ and $\hat{\kappa}_r \hat{\kappa}_r^H$ is a quadratic function in $N - 1$ variables. Therefore, we can write, when $M \rightarrow \infty$,

$$\mathbb{E} \left[\frac{(\mathbf{\Delta} \mathbf{\Delta}^H)_{r,r}}{M^2} \right] \leq \left(\frac{(\mathbf{P}_n)_{\max}}{\sigma^2} \right) \hat{\kappa}_r \hat{\kappa}_r^H \rightarrow 0. \quad (\text{C.7})$$

Now, to evaluate the second term of (C.3), we first define $\mathbf{\Theta} = \mathbf{P}_n^{\frac{1}{2}} \mathbf{\Sigma}^{-\frac{1}{2}}$. Therefore, $\mathbf{\Delta} = \mathbf{P}_n^{\frac{1}{2}} \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{\Xi} \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{P}_n^{\frac{1}{2}} = \mathbf{\Delta}^H = \mathbf{\Theta} \mathbf{\Xi} \mathbf{\Theta}^H$, and $\text{tr}(\mathbf{\Theta} \mathbf{\Xi} \mathbf{\Theta}^H) = \text{tr}(\mathbf{\Theta} \mathbf{X} \mathbf{\Theta}^H) - \text{tr}(\mathbf{\Theta} \tilde{\mathbf{X}} \mathbf{\Theta}^H)$. As $\text{tr}(\mathbf{\Theta} \tilde{\mathbf{X}} \mathbf{\Theta}^H)$ is quadratic, $\text{tr}(\mathbf{\Theta} \mathbf{\Xi} \mathbf{\Theta}^H) \leq \text{tr}(\mathbf{\Theta} \mathbf{X} \mathbf{\Theta}^H) = \sum_{r=1}^M \Theta_{r,r}^2 X_{r,r}$. As before we take the upper-bounds, $(\mathbf{P}_n)_{r,r} \leq (\mathbf{P}_n)_{\max}$ and $\mathbf{\Sigma}^{-1} \leq \frac{1}{\sigma^2} \mathbf{I}$. As such, $\Theta_{r,r}^2 = \frac{(\mathbf{P}_n)_{\max}}{\sigma^2}$ and,

$$\begin{aligned} \text{tr}(\mathbf{\Theta} \mathbf{\Xi} \mathbf{\Theta}^H) &\leq \frac{(\mathbf{P}_n)_{\max}}{\sigma^2} \sum_{r=1}^M X_{r,r} \\ &\leq \frac{(\mathbf{P}_n)_{\max}}{\sigma^2} \left(N - 1 - \sum_{\substack{r=1 \\ r \neq n}}^N \frac{1}{1 + \text{tr}(\mathbf{P}_r \mathbf{\Sigma}^{-1})} \right). \end{aligned} \quad (\text{C.8})$$

Accordingly, as $M \rightarrow \infty$,

$$\begin{aligned} &\frac{\mathbb{E}[\text{tr}(\mathbf{\Delta}) \text{tr}(\mathbf{\Delta}^H)]}{M^2} \\ &\leq \left(\frac{(\mathbf{P}_n)_{\max}}{\sigma^2} \left(N - 1 - \sum_{\substack{r=1 \\ r \neq n}}^N \frac{1}{1 + \text{tr}(\mathbf{P}_r \mathbf{\Sigma}^{-1})} \right) \right)^2 \frac{1}{M^2} \rightarrow 0. \end{aligned} \quad (\text{C.9})$$

As (C.1) is the sum of (C.7) and (C.9) we can write

$$\mathbb{E}[\xi^2] \rightarrow 0, \quad \text{when } M \rightarrow \infty, \quad (\text{C.10})$$

which concludes the proof of mean-square convergence in (55).

APPENDIX D APPROXIMATING ι AS M TENDS TO INFINITY

Let us first consider the (r, s) -th element of ι ,⁷

$$\iota_{r,s} = (\mathbf{h}_0)_r^H \mathbf{\Theta}^{-1} (\mathbf{h}_0)_s. \quad (\text{D.1})$$

As $\iota_{r,s}$ is a sum of M terms, when M is large, the law of large numbers for non-identical variables [35] gives:

$$\frac{\iota_{r,s}}{M} \approx \frac{\mathbb{E}\{(\mathbf{h}_0)_r^H \mathbf{\Theta}^{-1} (\mathbf{h}_0)_s\}}{M}. \quad (\text{D.2})$$

⁷In (D.1) $(\mathbf{h}_0)_r, (\mathbf{h}_0)_s \neq (\mathbf{h}_0)_n$, as ι is defined in terms of $(\hat{\mathbf{H}}_0)_{\bar{n}}$.

Therefore, when $r = s$, $\hat{\mathbf{g}}_r \sim \mathcal{CN}(0, \mathbf{\Sigma}_{\hat{\mathbf{g}}_r})$,

$$\frac{\iota_{r,r}}{M} \approx \frac{\mathbb{E}\left\{(\hat{\mathbf{h}}_0)_r^H \mathbf{\Theta}^{-1} (\hat{\mathbf{h}}_0)_r\right\}}{M} = \frac{\text{tr}(\mathbf{\Sigma}_{\hat{\mathbf{g}}_r} \mathbf{P}_r \mathbf{\Theta}^{-1})}{M}, \quad (\text{D.3})$$

and when $r \neq s$, $\mathbb{E}[\mathbf{g}_r^H \mathbf{g}_s] = 0$,

$$\frac{\iota_{r,s}}{M} \approx \frac{\mathbb{E}\left\{(\hat{\mathbf{h}}_0)_r^H \mathbf{\Theta}^{-1} (\hat{\mathbf{h}}_0)_s\right\}}{M} = 0. \quad (\text{D.4})$$

Let us define $\mathbf{\Xi}_\iota$, in terms of the expected value of ι an error term $\mathbf{\Xi}_\iota$ as,

$$\iota = \mathbb{E}[\iota] + \mathbf{\Xi}_\iota, \quad (\text{D.5})$$

We next examine the (r, s) -th element of the error matrix of $\mathbf{\Xi}_\iota$, $(\mathbf{\Xi}_\iota)_{r,s}$. From (51) and (D.5), elements of $\mathbf{\Xi}_\iota$ are defined by

$$(\mathbf{\Xi}_\iota)_{r,s} = \begin{cases} (\mathbf{h}_0)_r^H \mathbf{\Theta}^{-1} (\mathbf{h}_0)_r - \text{tr}(\mathbf{\Sigma}_{\hat{\mathbf{g}}_r} \mathbf{P}_r \mathbf{\Theta}^{-1}) & r = s \\ (\mathbf{h}_0)_r^H \mathbf{\Theta}^{-1} (\mathbf{h}_0)_s & r \neq s. \end{cases} \quad (\text{D.6})$$

Now, as $M \rightarrow \infty$, from (D.6) and (D.3),

$$\frac{(\mathbf{\Xi}_\iota)_{r,r}}{M} \rightarrow 0, \quad (\text{D.7})$$

and from (D.6) and (D.4),

$$\frac{(\mathbf{\Xi}_\iota)_{r,s}}{M} \rightarrow 0. \quad (\text{D.8})$$

Finally, from (D.7) and (D.8) as $M \rightarrow \infty$,

$$\frac{\mathbf{\Xi}_\iota}{M} \rightarrow 0. \quad (\text{D.9})$$

Therefore, from (D.5) and (D.9) as $M \rightarrow \infty$, we can approximate ι as, $\frac{\iota}{M} \approx \frac{\mathbb{E}[\iota]}{M}$.

REFERENCES

- [1] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [2] T. L. Marzetta, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
- [3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [4] K. T. Truong and R. W. Heath, "The viability of distributed antennas for massive MIMO systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2013, pp. 1318–1323.
- [5] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69–73, Jan. 2007.
- [6] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 161–167, Apr. 2013.
- [7] C.-M. Chen et al., "Distributed massive MIMO: A diversity combining method for TDD reciprocity calibration," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2017, pp. 1–7.
- [8] H. Q. Ngo, A. E. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: A uniformly great service for everyone," in *Proc. 16th IEEE Int. Workshop Signal Process. Adv. Wireless Commun. SPAWC*, 2015, pp. 201–205.
- [9] K. Hosseini, W. Yu, and R. S. Adve, "Large-scale MIMO versus network MIMO for multicell interference mitigation," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 930–941, Oct. 2014.

- [10] Ö. T. Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*. Boston, MA, USA: Now Publ., 2021.
- [11] M. Attarifar, A. Abbasfar, and A. Lozano, "Subset MMSE receivers for cell-free networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4183–4194, Jun. 2020.
- [12] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.
- [13] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.
- [14] E. Björnson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec. 2011.
- [15] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.
- [16] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.
- [17] W. Bao and B. Liang, "Optimizing cluster size through handoff analysis in user-centric cooperative wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 766–778, Feb. 2018.
- [18] G. Interdonato, P. K. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–6.
- [19] E. Björnson and L. Sanguinetti, "A new look at cell-free massive MIMO: Making it practical with dynamic cooperation," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2019, pp. 1–6.
- [20] S. Gunasekara, R. Senanayake, P. Smith, and M. Kuijper, "A novel partial joint processing architecture for distributed massive MIMO," in *Proc. IEEE 95th Veh. Technol. Conf.*, 2022, pp. 1–6.
- [21] R. Senanayake, P. J. Smith, P. L. Yeoh, and J. S. Evans, "An SNR approximation for distributed massive MIMO with zero forcing," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1885–1888, Nov. 2015.
- [22] R. Senanayake, P. L. Yeoh, and J. Evans, "Distributed LMMSE estimation in cooperative cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 4083–4088.
- [23] R. Senanayake, P. L. Yeoh, and J. S. Evans, "Performance analysis of centralized and partially decentralized co-operative networks," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 863–875, Feb. 2016.
- [24] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4758–4774, Jul. 2020.
- [25] B. Blaszczyszyn, M. K. Karray, and H. P. Keeler, "Wireless networks appear Poissonian due to strong shadowing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4379–4390, Aug. 2015.
- [26] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [27] S. Chen, J. Zhang, J. Zhang, E. Björnson, and B. Ai, "A survey on user-centric cell-free massive MIMO systems," *Digit. Commun. Netw.*, vol. 8, no. 5, pp. 695–719, 2022.
- [28] N. Lee, D. Morales-Jiménez, A. E. Lozano, and R. W. H. Jr., "Spectral efficiency of dynamic coordinated beamforming: A stochastic geometry approach," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 230–241, Jan. 2015.
- [29] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.
- [30] Y. Li, M. Xia, and S. Aïssa, "Coordinated multi-point transmission: A Poisson-Delaunay triangulation based approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2946–2959, May 2020.
- [31] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [32] R. Senanayake, A. Lozano, P. J. Smith, and J. S. Evans, "Analytical handle for ZF reception in distributed massive MIMO," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, ACSSC, 2016, pp. 16–20.
- [33] M. Kiessling and J. Speidel, "Analytical performance of MIMO zero-forcing receivers in correlated Rayleigh fading environments," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Rome, Italy, 2003, pp. 383–387.
- [34] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [35] P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics*, 1st ed. Boca Raton, FL, USA: Chapman Hall Inc., 1993.
- [36] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, pp. 666–672, May 1998.
- [37] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.
- [38] M. K. Simon and M. S. Alouini, *Digital Communication Over Fading Channels*, 2nd ed. Newark, NJ, USA: Wiley, 2005.
- [39] D. A. Basnayaka, P. J. Smith, and P. A. Martin, "Performance analysis of dual-user macrodiversity MIMO systems with linear receivers in flat Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4394–4404, Dec. 2012.
- [40] M. R. McKay, A. Zanella, I. B. Collings, and M. Chiani, "Error probability and SINR analysis of optimum combining in Rician fading," *IEEE Trans. Commun.*, vol. 57, pp. 676–687, Mar. 2009.
- [41] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [42] G. Gonthier, "Formal proof—the four-color theorem," *Not. AMS*, vol. 55, no. 11, pp. 1382–1393, 2008.
- [43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.



SUPUNI GUNASEKARA (Graduate Student Member, IEEE) received the B.Sc. degree in electronic and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2018. She is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Melbourne. She was with London Stock Exchange Group Technologies from 2018 to 2021. Her research interests are in cooperative communications, distributed antenna systems, and machine learning.



RAJITHA SENANAYAKE (Member, IEEE) received the B.E. degree in electrical and electronics engineering from the University of Peradeniya, Sri Lanka, and the bachelor's degree in information technology from the University of Colombo, Sri Lanka, in 2009 and 2010, respectively, and the Ph.D. degree in electrical and electronics engineering from The University of Melbourne, Australia, in 2015. She was with the Research and Development Team with Excel Technology, Sri Lanka, from 2009 to 2011, and the Department

of Electrical and Computer Systems Engineering, Monash University, Australia, from 2015 to 2016. She is currently a Senior Lecturer with the Department of Electrical and Electronics Engineering, University of Melbourne. Her research interests are in cooperative communications, distributed antenna systems and joint communications, and sensing. She is a recipient of the Australian Research Council Discovery Early Career Researcher Award.



PETER SMITH (Fellow, IEEE) received the B.Sc. degree in mathematics and the Ph.D. degree in statistics from the University of London, London, U.K., in 1983 and 1988, respectively. From 1983 to 1986, he was with the Telecommunications Laboratories, GEC Hirst Research Centre. From 1988 to 2001, he was a Lecturer of Statistics with the Victoria University of Wellington, New Zealand. From 2001 to 2015, he worked in electrical and computer engineering with the University of Canterbury. In 2015, he joined the Victoria

University of Wellington as a Professor of Statistics. He is also an Adjunct Professor of Electrical and Computer Engineering with the University of Canterbury, New Zealand, and an Honorary Professor with the School of Electronics, Electrical Engineering and Computer Science, Queens University Belfast. His research interests include the statistical aspects of design, modeling and analysis for communication systems, especially antenna arrays, MIMO, cognitive radio, massive MIMO, mmWave systems, reconfigurable intelligent surfaces, and the fusion of radar sensing and communications. From 2018 to 2019, he was awarded Visiting Fellowships at the University of Bologna, the University of Bristol, and the University of Melbourne. He was awarded a Distinguished Visiting Fellowship by the U.K.-based Royal Academy of Engineering at Queens University Belfast in 2017.



MARGRETA KUIJPER (Senior Member, IEEE) received the master's degree (cum laude) in mathematics from Vrije Universiteit (Free University), Amsterdam, and the Ph.D. degree in the area of behavioral systems and control from the Centrum voor Wiskunde & Informatica (Centre for Mathematics and Computer Science), Amsterdam, in 1985 and 1992, respectively.

She was with National Aerospace Laboratory, Amsterdam, from 1985 to 1988. From 1988 to 1992, she worked with the Centrum voor Wiskunde & Informatica, Amsterdam. From 1992 to 1995, she was a Postdoctoral Fellow in the area of behavioral systems and control with the Mathematics Department, University of Groningen, The Netherlands. She is currently a Professor with the Department of Electrical and Electronic Engineering, The University of Melbourne, Australia. In 1995, she joined The University of Melbourne. Her main research interests evolve around the interplay between systems theory and coding theory. Her research interests include behavioral system theory, polynomial matrix theory, convolutional coding, Reed-Solomon coding, systems and codes over finite fields and rings, erasure coding for streaming, distributed storage, MIMO wireless communications, and system cyber security.