Secure Transmission in Cell-Free Massive MIMO under Active Eavesdropping

Yasseen Sadoon Atiya, Zahra Mobini, *Member, IEEE,* Hien Quoc Ngo, *Senior Member, IEEE,* and Michail Matthaiou, *Fellow, IEEE*

***Abstract*—We study secure communications in cell-free massive multiple-input multiple-output (CF-mMIMO) systems with multi-antenna access points (APs) and protective partial zero-forcing (PPZF) precoding. In particular, we consider an active eavesdropping attack, where an eavesdropper contaminates the uplink channel estimation phase by sending an identical pilot sequence with a legitimate user of interest. We formulate an optimization problem for maximizing the received signal-to-noise ratio (SINR) at the legitimate user, subject to a maximum allowable SINR at the eavesdropper and maximum transmit power at each AP, while guaranteeing specific SINR requirements on other legitimate users. The optimization problem is solved using a path-following algorithm. We also propose a large-scale-based greedy AP selection scheme to improve the secrecy spectral efficiency (SSE). Finally, we propose a simple method for identifying the presence of an eavesdropper within the system. Our findings show that PPZF can substantially outperform the conventional maximum-ratio transmission (MRT) scheme by providing around 2-fold improvement in the SSE compared to the MRT scheme. More importantly, for PPZF precoding scheme, our proposed AP selection can achieve a remarkable SSE gain of up to 220%, while our power optimization approach can provide an additional gain of up to 55% compared with a CF-mMIMO system with equal power allocation.**

***Index Terms*—Access point selection, active eavesdropping, cell-free massive multiple-input multiple-output, physical layer security, power control, secrecy.**

I. Introduction

Cell-free massive multiple-input multiple-output (CF-mMIMO) has been envisaged as one of the promising technologies for the next generation wireless networks. By breaking the concept of cell boundaries, deploying a large number of geographically distributed access points (APs), and coherently serving users in the same time-frequency resources, it avails of all the benefits granted by the state of the art techniques including massive MIMO, distributed antenna systems, and coordinated multipoint with joint transmission. CF-mMIMO systems bring the APs geographically closer to the users and, hence, can provide seamless and handover free services for all users by exploiting macro-diversity gains and low path losses [2]–[4].

While this underpinning, distributed network architecture results in ubiquitous coverage with high spectral efficiency (SE), it also escalates the vulnerability of CF-mMIMO to malicious eavesdroppers, especially when the number of APs and users grows [5]. More precisely, since the APs are densely distributed over the area of coverage, the distances between the APs and users or the potential eavesdroppers are shortened, which can increase the risk of confidential information leakage. Therefore, the security of CF-mMIMO against eavesdropping and cyber-physical attacks is of great practical significance. Eavesdropping is typically classified into two main paradigms [6]: 1) passive eavesdropping and 2) active eavesdropping. In passive eavesdropping, eavesdroppers silently overhear the information delivery between APs and the targeted legitimate users without sending any pilot or interference signals [6], [7], while in the active eavesdropping, the active eavesdroppers intervene in the communication by either sending jamming signals and/or sending spoofing pilot sequences [8]. In a pilot spoofing attack, the uplink pilot training phase of the users of interest will be attacked by active eavesdroppers. More specifically, because the pilot sequences are publicly available and follow standardization, malicious eavesdroppers have the ability to actively transmit spoofing pilot sequences which results in a pilot contamination attack and, hence, information leakage. It has been shown that the adverse effects of active eavesdropping attacks are much more detrimental than those of passive attacks [6].

It is now worth noting the substantial amount of research interest that has been sparked in recent years in the implementation of physical-layer security techniques in massive MIMO systems. In particular, various methods for active pilot spoofing attack detection have been proposed in [9]–[12]. Additionally, the authors in [13]–[22] sought to enhance the security of massive MIMO systems either by using cooperative jamming [13] and artificial noise [14]–[17] for degrading the eavesdropping rate or by using resource allocation techniques [18], [19] and beamforming designs [20]–[22] for strengthening the legitimate links. However, in the context of secure CF-mMIMO systems, there have been only a few recent works [23]–[26]. In particular, Timilsina *et al.* [23] derived the secrecy spectral efficiency (SSE) expressions for CF-mMIMO under active pilot attacks and compared them

The authors are with the Centre for Wireless Innovation (CWI), Queen's University Belfast, BT3 9DT Belfast, U.K. email:{yhimiari01, zahra.mobini, hien.ngo, m.matthaiou}@qub.ac.uk. Yasseen Sadoon Atiya is also a lecturer at Imam Alkadhim University College.

This work is a contribution by Project REASON, a UK Government funded project under the Future Open Networks Research Challenge (FONRC) sponsored by the Department of Science Innovation and Technology (DSIT). It was also supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) (grant No. EP/X04047X/1). The work of Z. Mobini and H. Q. Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/X010635/1, and a research grant from the Department for the Economy Northern Ireland under the US-Ireland R&D Partnership Programme. The work of M. Matthaiou has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101001331). Parts of this paper were presented at IEEE VTC2023-Spring [1].

with those of co-located massive MIMO systems. For the same system setup of [23], the authors in [24] discussed power allocation problems either to maximize the achievable rate of the attacked legitimate user or to maximize the achievable SSE. Later, the authors in [25] investigated the effect of hardware impairments on the SSE of a CF-mMIMO network under pilot spoofing attacks. In [26], the problem of joint power and data transfer in a CF-mMIMO system with active eavesdropping was investigated. Moreover, the secrecy performance of CF-mMIMO with non-orthogonal pilot sequences for the uplink channel estimation was studied in [27]. However, current studies tend to investigate the secrecy performance of CF-mMIMO systems with single-antenna APs, while CF-mMIMO can better reap the channel hardening effect of cellular massive MIMO, when deploying multiple antennas at the APs [3]. Therefore, the recent works of [28], [29] studied the secrecy performance of multi-antenna CF-mMIMO networks under active eavesdropping. However, they only focused on the simple maximum-ratio transmission (MRT) precoding scheme for downlink transmissions which is incapable of mitigating inter-user interference. Providing an analytical framework to characterize the secrecy performance of CF-mMIMO over active eavesdropping by applying more advanced distributed precoding techniques is, therefore, of paramount importance and is one main goal of this paper.

Given that not all the serving APs in CF-mMIMO will contribute equally to the SE per user, due to the path loss which sharply changes with the geographical distance, several AP selection schemes for the downlink transmissions have been proposed in [30]–[33]. It has been shown that AP selection can effectively improve both the SE and energy efficiency while preserving the system scalability. However, a common assumption in the current papers on secure CF-mMIMO systems is that all APs transmit to all the users in the coverage area. In practice, there are some APs which are located far away from the legitimate users and will not contribute much to the SE. Therefore, how to efficiently select APs to serve the users under an active spoofing attack is of practical importance, but is still an open problem.

We would like to highlight that this work is an extension of [1]. Specifically, [1] pursued a performance analysis of CF-mMIMO systems with multi-antenna APs experiencing an active eavesdropping attack, while the quality-of-service (QoS) SE requirements for users and maximum allowable SINR at the eavesdropper were ignored, and the impact of AP selection was not investigated either. More importantly, we also provide an efficient scheme to detect the presence of eavesdroppers and determine which user is attacked. The main contributions of this paper are as follows:

- We develop a framework for a CF-mMIMO system with multiple-antenna APs employing protective partial ZF-based (PPZF) precoding [34] under an active spoofing attack during the uplink training phase, while we also consider channel estimation errors.
- We derive closed-form expressions for the SSE which sheds useful insights into the system's performance. We then propose a greedy large-scale-based AP selection scheme to improve the SSE.

TABLE I: List of Notations

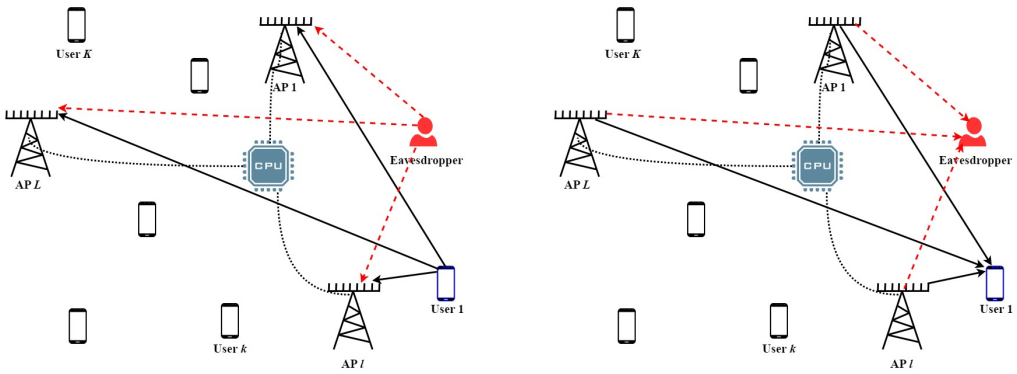
Notation	Description
L	Number of APs
M	Number of AP antennas
$\beta_{l,k}$	Large-scale fading coefficient for AP l and UE k
$\mathbf{g}_{l,k}$	Small-scale fading coefficient for AP l and UE k
$\gamma_{l,k}$	Variance of channel estimation for AP l and UE k
$\boldsymbol{\phi}_k$	Pilot sequence of the k-th user
τ_p	Uplink training duration
$\boldsymbol{\phi}_E$	Pilot sequence sent by the eavesdropper
$\mathbf{h}_{l,k}$	Channel vector between AP l and UE k
$\mathbf{h}_{l,E}$	Channel vector between AP l and the eavesdropper
$\mathbf{w}_{l,k}^{\text{PZF}}$	PZF precoding vector
$\mathbf{w}_{l,k}^{\text{PMRT}}$	PMRT precoding vector
$\rho_{l,k}$	Power control coefficient for AP l and UE k
R_{sec}	Secrecy spectral efficiency

- We formulate an optimization problem for maximizing the received SINR at the user under attack, subject to a maximum allowable SINR at the eavesdropper and maximum transmit power at each AP while guaranteeing the specific QoS requirements on other users. We solve the optimization problem using the path-following algorithm. We propose a novel and simple scheme to detect the presence of eavesdroppers in our system and determine which user is attacked. The method is based on the sample average power of the received pilot signals and can be implemented distributively at each AP, providing a detection probability of nearly one. We also propose a pilot re-transmission scheme to suppress the effect of the pilot spoofing attack. We then extend our results to the multiple-antenna eavesdropper scenario. Finally, numerical results are presented to support our findings and investigate the effect of different system parameters, including the number of antennas and APs and the position of the eavesdropper on the SSE performance, of CF-mMIMO systems.

Notation: We use bold upper case letters to denote matrices, and lower case letters to denote vectors. The superscript $(\cdot)^{\text{H}}$ stands for the conjugate-transpose (Hermitian); $\mathbb{C}^{L\times N}$ denotes a $L\times N$ matrix; $\mathbf{I}_M$ represents the $M\times M$ identity matrix; $\text{tr}(\cdot)$ denotes the trace operation. A zero mean circular symmetric complex Gaussian distribution having variance σ^2 is denoted by $\mathcal{CN}(0,\sigma^2)$. Finally, $\mathbb{E}\{\cdot\}$ denotes the statistical expectation.

II. System Model

We consider a CF-mMIMO system comprising L APs and K single-antenna users. To deal with multiple-antenna users, under the assumption of independent channels, one possible approach is to consider each user's antenna as an independent user [35]. This system also contends with a single-antenna active eavesdropper, denoted as E, as illustrated in Fig. 1. All APs cooperate to send data to all K users, while the eavesdropper attempts to intercept the information intended



(a) Uplink training phase

(b) Downlink data transmission phase

Fig. 1: CF-mMIMO with L multiple-antenna APs and K legitimate users under an active spoofing attack, where an eavesdropper contaminates the uplink channel estimation phase by sending an identical pilot sequence with the legitimate user 1.

for one of K users by a pilot spoofing attack scheme.[1] More specifically, the eavesdropper sends an identical pilot sequence with a legitimate user of interest. By sending the same pilot with the legitimate user, the channel estimation at the APs is biased and contains the components of the eavesdropping channel. In the downlink data transmission phase, since the precoding vectors at the APs associated with the targeted legitimate user are designed based on this biased channel estimation, the transmitted data signals may divert away from the legitimate user under attack and towards the direction of the eavesdropper. Therefore, the pilot spoofing attack leads to performance degradation of the legitimate transmission and, more severely, to information leakage toward the eavesdropper. We assume that the system knows the presence of an eavesdropper, and knows which user is being targeted. The method to achieve the above information is discussed in Section V. In addition, we consider an intelligent eavesdropper with protocol knowledge and the ability to synchronize and target the channel estimation phase. In particular, given the recent advancements in software defined radios (SDRs), it is straightforward for smart eavesdroppers to target the pilots in a synchronous fashion for synchronous protocols [37]. Without loss of generality, we assume that the eavesdropper targets user 1.

The sets of APs and users are denoted by $\mathcal{L} = \{1, \ldots, L\}$ and $\mathcal{K} = \{1, \ldots, K\}$, respectively. Each AP is equipped with M antennas such that $LM > K$ and all APs are distributed over a certain area. The channels of our considered system are modeled as follows:

- The $M \times 1$ channel vector between the l-th AP and the k-th user is:

$$\mathbf{h}_{l,k} = \sqrt{\beta_{l,k}}\mathbf{g}_{l,k}, \tag{1}$$

where $\beta_{l,k}$ is the large-scale fading coefficient, and $\mathbf{g}_{l,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the small-scale fading vector.

- The $M \times 1$ channel vector between the l-th AP and the eavesdropper is:

$$\mathbf{h}_{l,E} = \sqrt{\beta_{l,E}}\mathbf{g}_{l,E}, \tag{2}$$

where $\beta_{l,E}$ represents the large-scale fading coefficient and $\mathbf{g}_{l,E} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is small-scale fading vector.

We assume that the considered CF-mMIMO system operates under time-division duplex (TDD) operation, where each coherence block includes three main phases: uplink training phase for channel estimation, downlink payload data transmission, and uplink payload data transmission. In this work, we focus on the downlink transmission, and hence, the uplink payload transmission phase is ignored.

A. Uplink Training

Uplink training phase is required for channel acquisition at the APs. These acquired channel estimates play a crucial role in both downlink precoding and uplink combining designs. In this phase, all users send pilot signals to the APs. Accordingly, each AP can estimate the corresponding channels to all users using the obtained pilot signals. Let us assume that the k-th user sends a pilot sequence $\boldsymbol{\phi}_k \in \mathbb{C}^{\tau_p \times 1}$. Here, τ_p represents the training duration. We consider orthonormal pilot assignment, i.e., $\boldsymbol{\phi}_k^{\mathrm{H}}\boldsymbol{\phi}_{k'} = 0$ for $k \neq k'$ and $\|\boldsymbol{\phi}_k\|^2 = 1$. This requires $\tau_p \geq K$.

As previously mentioned, the eavesdropper aims at overhearing the confidential information intended for user 1. This is accomplished by the eavesdropper transmitting a pilot sequence denoted as $\boldsymbol{\phi}_E$, deliberately matching the pilot sequence sent by user 1, i.e., $\boldsymbol{\phi}_E = \boldsymbol{\phi}_1$. For most practical applications, the pilot sequences are publicly known and typically specified in the standard. Moreover, in some scenarios,

[1] Active full-duplex (FD) eavesdroppers are capable of both jamming as well as eavesdropping and/or sending spoofing pilot sequences [36]. In this paper, we consider that an active half-duplex (HD) eavesdropper focuses on sending the spoofing pilot sequences. The reason is that by attacking only the channel estimation phase, the eavesdropper remains fairly covert and power conservative as she only needs to operate during the estimation phase, which is typically a small fraction of a payload data transmission phase.

some untrusted (malicious) nodes in the network may be regarded as eavesdroppers. In these scenarios, they are operating with known network protocols [8], [24]. Accordingly, it is reasonable to assume that the pilot sequences associated with legitimate users are known to malicious users. Therefore, the pilot sequences $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ associated with user 1 to K can be easily obtained by the eavesdropper. Then, the received pilot matrix at the l-th AP is given by

$$\mathbf{Y}_{p,l} = \sqrt{\tau_p \rho_u} \sum_{k=1}^{K} \mathbf{h}_{l,k} \boldsymbol{\phi}_k^{\mathrm{H}} + \sqrt{\tau_p \rho_E} \mathbf{h}_{l,E} \boldsymbol{\phi}_1^{\mathrm{H}} + \mathbf{N}_l, \quad (3)$$

where ρ_u and ρ_E are the transmit signal-to-noise ratios (SNRs) at each user and the eavesdropper, respectively. More precisely, $\rho_u \triangleq P_u/N_0$ and $\rho_E \triangleq P_E/N_0$, where P_u and P_E are the transmit powers, while N_0 is the average noise power. In addition, the noise matrix $\mathbf{N}_l \in \mathbb{C}^{M \times \tau_p}$ includes independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ elements. From the received pilot signal (3), AP l uses the minimum mean squared (MMSE) estimation technique to estimate the channels to all users. The MMSE channel estimate of $\mathbf{h}_{l,k}$ is given by

$$\hat{\mathbf{h}}_{l,k} = \begin{cases} \frac{\sqrt{\tau_p \rho_u} \beta_{l,k}}{\tau_p \rho_u \beta_{l,k} + 1} \mathbf{y}_{l,k}, & k \neq 1, \\ \frac{\sqrt{\tau_p \rho_u} \beta_{l,1}}{\tau_p \rho_u \beta_{l,1} + \tau_p \rho_E \beta_{l,E} + 1} \mathbf{y}_{l,1}, & k = 1, \end{cases} \quad (4)$$

where $\mathbf{y}_{l,k} = \mathbf{Y}_{p,l} \boldsymbol{\phi}_k$. From (4), we can see that $\hat{\mathbf{h}}_{l,k}$ includes M i.i.d. $\mathcal{CN}(0, \gamma_{l,k})$ elements, where

$$\gamma_{l,k} = \begin{cases} \frac{\tau_p \rho_u \beta_{l,k}^2}{\tau_p \rho_u \beta_{l,k} + 1}, & k \neq 1, \\ \frac{\tau_p \rho_u \beta_{l,1}^2}{\tau_p \rho_u \beta_{l,1} + \tau_p \rho_E \beta_{l,E} + 1}, & k = 1. \end{cases} \quad (5)$$

Furthermore, let $\tilde{\mathbf{h}}_{l,k}$ be the channel estimation error, i.e., $\tilde{\mathbf{h}}_{l,k} = \mathbf{h}_{l,k} - \hat{\mathbf{h}}_{l,k}$. Then, $\tilde{\mathbf{h}}_{l,k}$ is independent of $\hat{\mathbf{h}}_{l,k}$, and includes i.i.d. $\mathcal{CN}(0, \beta_{l,k} - \gamma_{l,k})$ elements.

B. Downlink Data Transmission

The channels, which are estimated during the uplink training phase, will serve as the basis for precoding the data symbols intended for all K users. The precoded signal transmitted by the l-th AP can be expressed as

$$\mathbf{x}_l = \sum\nolimits_{k=1}^{K} \sqrt{\rho_{l,k}} \mathbf{w}_{l,k} s_k, \quad (6)$$

where $\mathbf{w}_{l,k} \in \mathbb{C}^{M \times 1}$, where $\mathbb{E}\{\|\mathbf{w}_{l,k}\|^2\} = 1$, is the precoding vector used by AP l towards user k, and $\rho_{l,k}$ is the power control coefficient. Moreover, s_k denotes the data symbol intended for the k-th user, where $\mathbb{E}\{|s_k|^2\} = 1$. Thus, the received signals at the k-th user and the eavesdropper are respectively given by

$$z_k = \sum\nolimits_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{x}_l + n_k, \quad (7)$$

and

$$z_E = \sum\nolimits_{l=1}^{L} \mathbf{h}_{l,E}^{\mathrm{H}} \mathbf{x}_l + n_E, \quad (8)$$

where $n_k \sim \mathcal{CN}(0,1)$ and $n_E \sim \mathcal{CN}(0,1)$ are the corresponding additive noise terms.

C. Precoding Design

Classical ZF precoding for CF-mMIMO systems suffers from modest array gain due to the fact that almost all the available degrees of freedom (DoF) are used to mitigate the interference. On the other hand, while MRT effectively sustains the system scalability, it is unable to cancel inter-user interference. For these reasons, we adopt the PPZF precoding scheme [34] to design the precoding vector $\mathbf{w}_{l,k}$ in (6). The main idea of PPZF is that each AP only mitigates the interference it causes to the strongest users, namely the users with the largest channel gain, while the interference towards the weak users, namely the users with the smallest channel gain, is tolerated. Consequently, PPZF can provide an acceptable trade-off between interference cancellation and boosting the desired signal.

More specifically, in PPZF scheme, each AP l virtually divides users into two groups: *1)* strong user set, and *2)* weak user set based on their channel gains, $\beta_{l,k}$, $\forall k \in \mathcal{K}$. The user grouping can adhere to various criteria. For example, one criterion could be the mean square of the channel gain: user k is assigned to the strong group for AP l if $\beta_{l,k}$ exceeds a predetermined threshold; otherwise, user k belongs to a weak group. Now, let us denote, for AP l, the set of indices of strong users by $\mathcal{S}_l \subset \mathcal{K}$, and the set of indices of weak users by $\mathcal{W}_l \subset \mathcal{K}$, respectively, where $\mathcal{S}_l \cap \mathcal{W}_l = \varnothing$, and $|\mathcal{S}_l| + |\mathcal{W}_l| = K$. Then, AP l transmits to the users in $\mathcal{S}_l$ by using partial ZF, and to the users in $\mathcal{W}_l$ by using protective MRT (PMRT) to avoid interference to the users in $\mathcal{S}_l$. Here, we note that to implement PPZF, the number of transmit antennas at each AP, must meet the requirement $M \geq |\mathcal{S}_l|$.

Let $\hat{\mathbf{H}}_l = [\hat{\mathbf{h}}_{l,1}, \ldots, \hat{\mathbf{h}}_{l,K}] \in \mathbb{C}^{M \times K}$ be the collective channel estimation matrix from AP l to all users, $\mathbf{E}_{\mathcal{S}_l} = \left[\mathbf{e}_i : i \in \mathcal{S}_l\right] \in \mathbb{C}^{K \times |\mathcal{S}_l|}$, where $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_K$. In addition, let user k correspond to the j-th element of set $\mathcal{S}_l$, $j \in \{1, \ldots, |\mathcal{S}_l|\}$. Then, we define $\boldsymbol{\pi}_k$ as the j-th column of $\mathbf{I}_{|\mathcal{S}_l|}$. With PPZF, the precoding vector in (6) can be expressed as

$$\mathbf{w}_{l,k} = \begin{cases} \mathbf{w}_{l,k}^{\mathtt{PZF}}, & \text{if } k \in \mathcal{S}_l, \\ \mathbf{w}_{l,k}^{\mathtt{PMRT}}, & \text{if } k \in \mathcal{W}_l, \end{cases} \quad (9)$$

where $\mathbf{w}_{l,k}^{\mathtt{PZF}}$ is the PZF precoding vector and $\mathbf{w}_{l,k}^{\mathtt{PMRT}}$ is the PMRT precoding vector. More precisely,

- PZF precoding vector:

$$\mathbf{w}_{l,k}^{\mathtt{PZF}} = \frac{\hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \left(\mathbf{E}_{\mathcal{S}_l}^{\mathrm{H}} \hat{\mathbf{H}}_l^{\mathrm{H}} \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \right)^{-1} \boldsymbol{\pi}_k}{\sqrt{\mathbb{E}\left\{ \left\| \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \left(\mathbf{E}_{\mathcal{S}_l}^{\mathrm{H}} \hat{\mathbf{H}}_l^{\mathrm{H}} \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \right)^{-1} \boldsymbol{\pi}_k \right\|^2 \right\}}}, \quad (10)$$

  where the denominator of (10) is the normalization term, which is given in closed-form as

$$\mathbb{E}\left\{ \left\| \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \left(\mathbf{E}_{\mathcal{S}_l}^{\mathrm{H}} \hat{\mathbf{H}}_l^{\mathrm{H}} \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \right)^{-1} \boldsymbol{\pi}_k \right\|^2 \right\} = \frac{1}{(M - |\mathcal{S}_l|)\, \gamma_{l,k}}. \quad (11)$$

Therefore, for any pair of users $k, t \in \mathcal{S}_l$ we have

$$\alpha_{l,k,t}^{\texttt{PZF}} \triangleq \hat{\mathbf{h}}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,t}^{\texttt{PZF}} = \begin{cases} 0, & t \neq k, \\ \sqrt{(M - |\mathcal{S}_l|)\gamma_{l,k}}, & t = k. \end{cases} \tag{12}$$

- PMRT precoding vector, from AP l to user j, $j \in \mathcal{W}_l$:

$$\mathbf{w}_{l,j}^{\texttt{PMRT}} = \frac{\mathbf{B}_l \hat{\mathbf{H}}_l \mathbf{e}_j}{\sqrt{\mathbb{E}\{\|\mathbf{B}_l \hat{\mathbf{H}}_l \mathbf{e}_j\|^2\}}} = \frac{\mathbf{B}_l \hat{\mathbf{H}}_l \mathbf{e}_j}{\sqrt{(M - |\mathcal{S}_l|)\gamma_{l,j}}}, \tag{13}$$

where

$$\mathbf{B}_l = \mathbf{I}_M - \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \left(\mathbf{E}_{\mathcal{S}_l}^{\mathrm{H}} \hat{\mathbf{H}}_l^{\mathrm{H}} \hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l} \right)^{-1} \mathbf{E}_{\mathcal{S}_l}^{\mathrm{H}} \hat{\mathbf{H}}_l^{\mathrm{H}}, \tag{14}$$

which is the null space of $\hat{\mathbf{H}}_l \mathbf{E}_{\mathcal{S}_l}$. For any pair of users $k, t \in \mathcal{W}_l$ we have

$$\mathbb{E}\left\{ \hat{\mathbf{h}}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,t}^{\texttt{PMRT}} \right\} = \begin{cases} 0, & t \neq k, \\ \sqrt{(M - |\mathcal{S}_l|)\gamma_{l,k}}, & t = k. \end{cases} \tag{15}$$

Then, the transmit signals at AP l is

$$\mathbf{x}_l = \sum\nolimits_{k \in \mathcal{S}_l} \sqrt{\rho_{l,k}} \mathbf{w}_{l,k}^{\texttt{PZF}} s_k + \sum\nolimits_{j \in \mathcal{W}_l} \sqrt{\rho_{l,j}} \mathbf{w}_{l,j}^{\texttt{PMRT}} s_j. \tag{16}$$

III. Secrecy Performance Analysis

In this section, we evaluate the secrecy performance provided by a CF-mMIMO system with PPZF precoding under an active eavesdropping attack. First, let us define $\mathcal{Z}_k$ and $\mathcal{M}_k$ as the set of indices of APs that transmit to the user k by using PZF and PMRT, respectively, as

$$\mathcal{Z}_k \triangleq \{ l : k \in \mathcal{S}_l, l = 1, \ldots, L \}, \tag{17}$$

and

$$\mathcal{M}_k \triangleq \{ l : k \in \mathcal{W}_l, l = 1, \ldots, L \}, \tag{18}$$

with $\mathcal{Z}_k \cap \mathcal{M}_k = \varnothing$, and $|\mathcal{Z}_k| + |\mathcal{M}_k| = L$.

A. Spectral Efficiency of the Legitimate Links

Using (17) and (18), the received signal at the k-th user in (7) can be rewritten as

$$\begin{aligned} z_k &= \left(\sum_{l \in \mathcal{Z}_k} \sqrt{\rho_{l,k}} \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,k}^{\texttt{PZF}} + \sum_{p \in \mathcal{M}_k} \sqrt{\rho_{p,k}} \mathbf{h}_{p,k}^{\mathrm{H}} \mathbf{w}_{p,k}^{\texttt{PMRT}} \right) s_k \\ &+ \sum_{\substack{t=1 \\ t \neq k}}^{K} \left(\sum_{l \in \mathcal{Z}_t} \sqrt{\rho_{l,t}} \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,t}^{\texttt{PZF}} + \sum_{p \in \mathcal{M}_t} \sqrt{\rho_{p,t}} \mathbf{h}_{p,k}^{\mathrm{H}} \mathbf{w}_{p,t}^{\texttt{PMRT}} \right) s_t + n_k, \\ &= \mathrm{CP}_k \times s_k + \mathrm{PU}_k \times s_k + \sum\nolimits_{\substack{t=1 \\ t \neq k}}^{K} \mathrm{UI}_{k,t} \times s_t + n_k, \end{aligned} \tag{19}$$

where CP_k, PU_k, and $\mathrm{UI}_{k,t}$ show the coherent precoding gain, precoding gain uncertainty, and multi-user interference, respectively, defined as

$$\mathrm{CP}_k = \sum_{l \in \mathcal{Z}_k} \sqrt{\rho_{l,k}}\, \mathbb{E}\left\{ \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,k}^{\texttt{PZF}} \right\} + \sum_{p \in \mathcal{M}_k} \sqrt{\rho_{p,k}}\, \mathbb{E}\left\{ \mathbf{h}_{p,k}^{\mathrm{H}} \mathbf{w}_{p,k}^{\texttt{PMRT}} \right\}, \tag{20}$$

$$\mathrm{PU}_k = \sum_{l \in \mathcal{Z}_k} \left(\sqrt{\rho_{l,k}} \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,k}^{\texttt{PZF}} - \sqrt{\rho_{l,k}}\, \mathbb{E}\left\{ \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,k}^{\texttt{PZF}} \right\} \right) + \sum_{p \in \mathcal{M}_k} \left(\sqrt{\rho_{l,k}} \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{p,k}^{\texttt{PMRT}} - \sqrt{\rho_{l,k}}\, \mathbb{E}\left\{ \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{p,k}^{\texttt{PMRT}} \right\} \right), \tag{21}$$

$$\mathrm{UI}_{k,t} = \sum_{l \in \mathcal{Z}_k} \sqrt{\rho_{l,t}} \mathbf{h}_{l,k}^{\mathrm{H}} \mathbf{w}_{l,t}^{\texttt{PZF}} + \sum_{p \in \mathcal{M}_k} \sqrt{\rho_{p,t}} \mathbf{h}_{p,k}^{\mathrm{H}} \mathbf{w}_{p,t}^{\texttt{PMRT}}. \tag{22}$$

Each user k effectively sees a deterministic channel (CP_k) with some unknown noise and, hence, to detect the intended symbol from the received signal, it relies only on the statistical channel state information (CSI). More specifically, since s_k and s_t are uncorrelated for any $t \neq k$, the first term in (19) is uncorrelated with the third term. Additionally, since s_k is independent of PU_k, the first and second terms are also uncorrelated. The fourth term. i.e, noise, is independent of the first term in (19). Accordingly, the sum of the second, third, and fourth terms in (19) can be considered as an uncorrelated effective noise. Following the discussion in [35, Sec. 2.3.2], an achievable downlink SE for user k can be written as

$$R_k = \log_2 \left(1 + \mathrm{SINR}_k \right), \tag{23}$$

where

$$\mathrm{SINR}_k = \frac{|\mathrm{CP}_k|^2}{\mathbb{E}\left\{ |\mathrm{PU}_k|^2 \right\} + \sum_{t \neq k}^{K} \mathbb{E}\left\{ |\mathrm{UI}_{k,t}|^2 \right\} + 1}, \tag{24}$$

which can be re-expressed as (25) at the top of the next page. We next provide a closed-form SE expression for user k.

Proposition 1. *The closed-form expression for the SE of user k with PPZF precoding under an active eavesdropping attack can be expressed as* (23)*, where*

$$\mathrm{SINR}_k = \frac{\left(\sum_{l=1}^{L} \sqrt{(M - |\mathcal{S}_l|)\, \rho_{l,k} \gamma_{l,k}} \right)^2}{\sum_{t=1}^{K} \sum_{l=1}^{L} \rho_{l,t} \left(\beta_{l,k} - \delta_{l,k} \gamma_{l,k} \right) + 1}, \tag{26}$$

where $\delta_{l,k} \triangleq 1$ if $k \in \mathcal{S}_l$ and $\delta_{l,k} \triangleq 0$ if $k \in \mathcal{W}_l$.

Proof. See Appendix A. □

B. Spectral Efficiency of the Eavesdropper

Hereafter, we assume that the eavesdropper has perfect CSI knowledge which results in the worst case SSE. In particular, the received signal at eavesdropper (8) can be represented in the form of

$$\begin{aligned} z_E &= \sum_{l=1}^{L} \sqrt{\rho_{l,1}} \mathbf{h}_{l,E}^{\mathrm{H}} \mathbf{w}_{l,1} s_1 + \sum_{t \neq 1}^{K} \sum_{l=1}^{L} \sqrt{\rho_{l,t}} \mathbf{h}_{l,E}^{\mathrm{H}} \mathbf{w}_{l,t} s_t + n_E, \\ &= \mathrm{BU}_{E,1} \times s_1 + \underbrace{\sum_{t \neq 1}^{K} \mathrm{UI}_{E,t} \times s_t + n_E}_{\text{treated as noise}}, \end{aligned} \tag{27}$$

where $\mathrm{BU}_{E,1}$ represents the strength of the desired signal s_1, while $\mathrm{UI}_{E,t}$ denotes the interference caused by the remaining users ($t \neq k$), and they can be expressed as

$$\mathrm{BU}_{E,1} \triangleq \sum_{l \in \mathcal{Z}_1} \sqrt{\rho_{l,1}} \mathbf{h}_{l,E}^{\mathrm{H}} \mathbf{w}_{l,1}^{\texttt{PZF}} + \sum_{p \in \mathcal{M}_1} \sqrt{\rho_{p,1}} \mathbf{h}_{p,E}^{\mathrm{H}} \mathbf{w}_{p,1}^{\texttt{PMRT}}, \tag{28}$$

$$\mathrm{SINR}_k = \frac{\left| \sum\limits_{l\in\mathcal{Z}_k} \sqrt{\rho_{l,k}}\mathbb{E}\left\{\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,k}^{\mathtt{PZF}}\right\} + \sum\limits_{p\in\mathcal{M}_k} \sqrt{\rho_{p,k}}\mathbb{E}\left\{\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,k}^{\mathtt{PMRT}}\right\}\right|^2}{\sum\limits_{t=1}^{K}\mathbb{E}\left\{\left| \sum\limits_{l\in\mathcal{Z}_t} \sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}} + \sum\limits_{p\in\mathcal{M}_t} \sqrt{\rho_{p,t}}\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathtt{PMRT}}\right|^2\right\} - \left| \sum\limits_{l\in\mathcal{Z}_k} \sqrt{\rho_{l,k}}\mathbb{E}\left\{\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,k}^{\mathtt{PZF}}\right\} + \sum\limits_{p\in\mathcal{M}_k} \sqrt{\rho_{p,k}}\mathbb{E}\left\{\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,k}^{\mathtt{PMRT}}\right\}\right|^2 + 1}. \tag{25}$$

$$\mathrm{UI}_{E,t} \triangleq \sum_{l\in\mathcal{Z}_t} \sqrt{\rho_{l,t}}\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}} + \sum_{p\in\mathcal{M}_t} \sqrt{\rho_{p,t}}\mathbf{h}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathtt{PMRT}}. \tag{29}$$

Let us denote the mutual information between s_1 and z_E by $I_E(s_1; z_E)$. Then, an upper-bound for $I_E(s_k; z_E)$ is given by

$$\begin{aligned} I_{\mathrm{E}}(s_1; z_E) &\overset{(a)}{\leq} I_E\left(s_1; z_E \mid \{h_{l,k}\}_{l,k}, \left\{\hat{h}_{l,k}\right\}_{l,k}, \{h_{l,E}\}_l\right) \\ &= \mathbb{E}\left\{\log_2\left(1 + \frac{|\mathrm{BU}_{E,1}|^2}{\sum_{t\neq 1}^{K}|\mathrm{UI}_{E,t}|^2 + 1}\right)\right\}, \\ &\overset{(b)}{\approx} \log_2\left(1 + \frac{\mathbb{E}\left\{|\mathrm{BU}_{E,1}|^2\right\}}{\sum_{t\neq 1}^{K}\mathbb{E}\left\{|\mathrm{UI}_{E,t}|^2\right\} + 1}\right), \end{aligned} \tag{30}$$

where the inequality (a) comes from the fact that the eavesdropper has perfect CSI knowledge, while the approximation in (b) follows from [38, Lemma 1].

Proposition 2. *The SE of the eavesdropper, which overhears the confidential messages destined for user* 1 *with PPZF precoding, can be approximated as* $R_E \approx \log_2(1 + \mathrm{SINR}_E)$*, where*

$$\mathrm{SINR}_E = \frac{\left(\sum\limits_{l=1}^{L}\sqrt{\rho_{l,1}(M - |\mathcal{S}_l|)\gamma_{l,E}}\right)^2 + \sum\limits_{l=1}^{L}\rho_{l,1}\beta_{l,E} - \sum\limits_{l\in\mathcal{Z}_1}\rho_{l,1}\gamma_{l,E}}{\sum\limits_{t\neq 1}^{K}\sum\limits_{l=1}^{L}\rho_{l,t}(\beta_{l,E} - \delta_{l,1}\gamma_{l,E}) + 1}. \tag{31}$$

Proof. See Appendix B. □

C. Secrecy Spectral Efficiency

Using the derived SE expressions, we can now calculate the SSE associated with user 1 as

$$R_{sec} = [R_1 - R_E]^+ \approx \left[\log_2\left(\frac{1 + \mathrm{SINR}_1}{1 + \mathrm{SINR}_E}\right)\right]^+, \tag{32}$$

where $[x]^+ = \max\{0, x\}$.

IV. Access Point Selection and Power Optimization

A. Access Point Selection

In this subsection, we propose an AP selection scheme which can increase the SSE. The proposed scheme is based on the following observations: Firstly, for user 1, there are some APs which are located far away and will not add significantly to its overall SE. Secondly, there are some APs which are in close vicinity of eavesdropper. Serving user 1 by these APs may result in high values for the overheard SINR.

Algorithm 1 Greedy AP Selection

1: **Initialize**: Set $\mathcal{T} = \emptyset$ and $R_{sec}^{0,\star} = 0$.
2: Calculate the ratio $\zeta(l) = \frac{\beta_{l,1}}{\beta_{l,E}}$ for all APs. Then, order the APs based on the ratio $\zeta(l)$ in a descending order and create the set $\mathcal{A} = \{l^{(1)}, \ldots, l^{(L)}\}$.
3: **for** $i = 1$ to L **do**
4: Calculate $R_{sec}^{i,\star} = R_{sec}(\mathcal{T} \cup l^{(i)})$
5: **if** $R_{sec}^{i,\star} > R_{sec}^{i-1,\star}$ **then**
6: Update $\mathcal{T} = \mathcal{T} \cup \{l^{(1)}\}$
7: **end if**
8: **end for**

Therefore, in order to increase the SSE, user 1 should not be served by all APs. To this end, we propose a greedy large-scale-based AP selection scheme for choosing a group of APs for serving user 1 in Algorithm 1. We consider the ratio $\zeta(l) = \frac{\beta_{l,1}}{\beta_{l,E}}$ as the criterion, order the APs in a descending order based on $\zeta(l)$, and then create the set $\mathcal{A} = \{l^{(1)}, \ldots, l^{(L)}\}$. Let $\mathcal{T}$ be the set of assigned APs to user 1 and $R_{sec}(\mathcal{T})$ denote the dependence of the SSE on the different choices of $\mathcal{T}$. The key idea of Algorithm 1 is to iteratively select a subset of APs out of the ordered AP set $\mathcal{A}$ on the condition that a new AP assignment at each iterative step improves the SSE. A key characteristic of the proposed AP selection is that it only requires the large scale fading (path loss) between the APs and eavesdropper. Additionally, since it is performed only on a large-scale time scale, it avoids the need of frequently performing the AP selection.

Remark 1. *In Algorithm 1, the dominant term in calculating* R_{sec} *given in* (32) *is the double summation in the denominator of* (26) *whose complexity scales as* $\mathcal{O}(LK)$*. The operation of computing* $\zeta(l)$ *has complexity* $\mathcal{O}(L)$*, while sorting the order index of* L *APs is performed with complexity* $\mathcal{O}(L \log L)$*. Therefore, the computational complexity of the algorithm is in the order of* $\mathcal{O}(LK)$*.*

B. Power Optimization

Here, we aim at optimally selecting the power coefficients $\rho_{l,k}$ to maximize the SE (and hence, the SINR) at user 1, subject to a maximum allowable SINR at the eavesdropper and maximum transmit power at each AP while guaranteeing the specific QoS requirements on each user k, $k \in \mathcal{K}\backslash\{1\}$. Note that we consider average normalized power constraint at each AP, i.e., $\mathbb{E}\left\{\|\mathbf{x}_l\|^2\right\} \leq \rho_{\max}$ with $\rho_{\max} = P_{\max}/N_0$,

which, using (6), can be further expressed as the following per-AP power constraint

$$\sum\nolimits_{k=1}^{K} \rho_{l,k} \leq \rho_{\max}. \tag{33}$$

More precisely, the power optimization problem is formulated as

$$\max_{\{\rho_{l,k}\}} \quad \mathrm{SINR}_1, \tag{34a}$$

$$\text{s.t.} \quad \sum\nolimits_{k=1}^{K} \rho_{l,k} \leq \rho_{\max}, \qquad \forall k, l, \tag{34b}$$

$$\mathrm{SINR}_k \geq \theta_k, \qquad \forall k, k \neq 1, \tag{34c}$$

$$\mathrm{SINR}_E \leq \theta_E, \tag{34d}$$

where θ_k and θ_E are the minimum SINRs required by user k and the maximum allowable overheard SINR at eavesdropper, respectively. In order to facilitate further analysis, let us denote the power allocation coefficient matrix by $\mathbf{\Psi}$ with elements $\mathbf{\Psi}(l,k) = \sqrt{\rho_{l,k}} \quad \forall l, k$. The k-th column vector of $\mathbf{\Psi}$ is denoted as

$$\mathbf{u}_k = \mathbf{\Psi}(:,k) = \left[\sqrt{\rho_{1,k}}, \sqrt{\rho_{2,k}}, \ldots, \sqrt{\rho_{L,k}}\right]^T. \tag{35}$$

Furthermore, we define the following matrices and vectors

$$\mathbf{a}_k = \Big[\sqrt{(M-\mathcal{S}_1)\gamma_{1,k}}, \sqrt{(M-\mathcal{S}_2)\gamma_{2,k}}, \ldots, \sqrt{(M-\mathcal{S}_L)\gamma_{L,k}}\Big]^T, \tag{36}$$

$$\mathbf{A}_{k,k} = \mathrm{diag}\left(\sqrt{\beta_{1,k}-\delta_{1,k}\gamma_{1,k}}, \ldots, \sqrt{\beta_{L,k}-\delta_{L,k}\gamma_{L,k}}\right), \tag{37}$$

$$\mathbf{b}_{\mathrm{E}} = \Big[\sqrt{(M-\mathcal{S}_1)\gamma_{1,E}}, \sqrt{(M-\mathcal{S}_2)\gamma_{2,E}}, \ldots, \sqrt{(M-\mathcal{S}_L)\gamma_{L,E}}\Big]^T, \tag{38}$$

and

$$\mathbf{B}_E = \mathrm{diag}\left(\sqrt{\beta_{1,E}-\delta_{1,1}\gamma_{1,E}}, \ldots, \sqrt{\beta_{L,E}-\delta_{L,1}\gamma_{L,E}}\right). \tag{39}$$

Accordingly, using (35)-(39), the received SINRs at user 1 and eavesdropper in (26) and (31) can be rewritten as

$$\mathrm{SINR}_k = \frac{\left(\mathbf{a}_k^T\mathbf{u}_k\right)^2}{\varphi_k(\mathbf{\Psi})}, \tag{40}$$

and

$$\mathrm{SINR}_E = \frac{(\mathbf{b}_{\mathrm{E}}\mathbf{u}_1)^2 + \|\mathbf{B}_E\mathbf{u}_1\|^2}{\varphi_E(\mathbf{\Psi})}, \tag{41}$$

respectively, where

$$\varphi_k(\mathbf{\Psi}) = \sum\nolimits_{t=1}^{K} \|\mathbf{A}_{k,k}\mathbf{u}_t\|^2 + 1, \quad k \in \mathcal{K}, \tag{42}$$

$$\varphi_E(\mathbf{\Psi}) = \sum\nolimits_{t=2}^{K} \|\mathbf{B}_E\mathbf{u}_t\|^2 + 1. \tag{43}$$

Now, problem (34) can be transformed into a more tractable form as follows

$$\max_{\mathbf{\Psi}} \quad \frac{\left(\mathbf{a}_1^T\mathbf{u}_1\right)^2}{\varphi_1(\mathbf{\Psi})} \tag{44a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \mathbf{\Psi}^2(l,k) \leq \rho_{\max}, \quad l \in \mathcal{L}, \tag{44b}$$

$$\frac{\left(\mathbf{a}_k^T\mathbf{u}_k\right)^2}{\varphi_k(\mathbf{\Psi})} \geq \theta_k \qquad \forall k, k \neq 1, \tag{44c}$$

$$\frac{(\mathbf{b}_{\mathrm{E}}\mathbf{u}_1)^2 + \|\mathbf{B}_E\mathbf{u}_1\|^2}{\varphi_{\mathrm{E}}(\mathbf{\Psi})} \leq \theta_E. \tag{44d}$$

Problem (44) is a nonconvex optimization problem. Therefore, in what follows, we use a path-following iterative algorithm [39] to solve the problem efficiently. The first constraint in (44) is obviously convex, while the second constraint can be written as

$$\frac{1}{\sqrt{\theta_k}}\mathbf{a}_k^T\mathbf{u}_k \geq \sqrt{\varphi_k(\mathbf{\Psi})}, \quad k \in \mathcal{K}\backslash\{1\}, \tag{45}$$

which is a second-order cone (SOC) constraint and is convex. Let $\mathbf{\Psi}^{(\kappa)}$ denote a feasible point for (44) found from the $(\kappa-1)$-th iteration. By invoking the following upper bound

$$\frac{x^2}{y} \geq 2\frac{\bar{x}}{\bar{y}}x - \frac{\bar{x}^2}{\bar{y}^2}y, \quad \forall x>0, y>0, \bar{x}>0, \bar{y}>0, \tag{46}$$

we obtain

$$\frac{\left(\mathbf{a}_1^T\mathbf{u}_1\right)^2}{\varphi_1(\mathbf{\Psi})} \geq f_1^{(\kappa)}(\mathbf{\Psi}) \triangleq a^{(\kappa)}\mathbf{a}_1^T\mathbf{u}_1 - b^{(\kappa)}\varphi_1(\mathbf{\Psi}), \tag{47}$$

with

$$a^{(\kappa)} = 2\frac{\left(\mathbf{a}_1^T\mathbf{u}_1^{(\kappa)}\right)^2}{\varphi_1\left(\mathbf{\Psi}^{(\kappa)}\right)}, \quad b^{(\kappa)} = \left(a^{(\kappa)}/2\right)^2. \tag{48}$$

Therefore, the objective function $\left(\mathbf{a}_1^T\mathbf{u}_1\right)^2/\varphi_1(\mathbf{\Psi})$ in (44) can be replaced by $f_1^{(\kappa)}(\mathbf{\Psi})$. In addition, since the function $\varphi_{\mathrm{E}}(\mathbf{\Psi})$ is convex, we can use the first-order Taylor approximation of $\varphi_{\mathrm{E}}(\mathbf{\Psi})$ near $\mathbf{\Psi}^{(k)}$ to handle the non-convex constraint (44d). In particular, constraint (44d) can be approximated by the convex quadratic constraint

$$\frac{(\mathbf{b}_{\mathrm{E}}\mathbf{u}_1)^2 + \|\mathbf{B}_E\mathbf{u}_1\|^2}{\theta_E} \leq \varphi_E^{(\kappa)}(\mathbf{\Psi}), \tag{49}$$

with

$$\varphi_{\mathrm{E}}^{(\kappa)}(\mathbf{\Psi}) \triangleq \sum\nolimits_{k=2}^{K}\left[\mathbf{u}_k^{(\kappa)^T}\mathbf{B}_E^2\left(2\mathbf{u}_k - \mathbf{u}_k^{(\kappa)}\right)\right] + 1. \tag{50}$$

At iteration $(\kappa+1)$, for a given point $\mathbf{\Psi}^{(\kappa)}$, the optimization problem (44) can finally be approximated by the following convex problem:

$$\max_{\mathbf{\Psi}} \quad f_1^{(\kappa)}(\mathbf{\Psi}), \tag{51a}$$

Algorithm 2 Path-Following Algorithm for Solving (51)

1: **Initialize**: $\kappa = 0$, a feasible point $\mathbf{\Psi}^{(0)}$ for (51).
2: **repeat**
3: Update $\kappa := \kappa + 1$.
4: Solve (51) to obtain the optimized solution $\mathbf{\Psi}^*$.
5: Update $\mathbf{\Psi}^{(\kappa)} = \mathbf{\Psi}^*$
6: **until** Convergence.
7: Return $\mathbf{\Psi}^{(\kappa)}$.

$$\text{s.t.} \quad \sum_{k=1}^{K} \mathbf{\Psi}^2(l,k) \leq \rho_{\max}, \quad l \in \mathcal{L}, \tag{51b}$$

$$\frac{1}{\sqrt{\theta_k}} \mathbf{a}_k^T \mathbf{u}_k \geq \sqrt{\varphi_k(\mathbf{\Psi})}, \quad k \in \mathcal{K} \backslash \{1\}, \tag{51c}$$

$$\frac{(\mathbf{b}_{\mathrm{E}} \mathbf{u}_1)^2 + \|\mathbf{B}_E \mathbf{u}_1\|^2}{\theta_E} \leq \varphi_E^{(\kappa)}(\mathbf{\Psi}). \tag{51d}$$

Starting with a feasible point $\mathbf{\Psi}^{(0)}$, we solve (51) to obtain its optimized solution $\mathbf{\Psi}^*$, and use $\mathbf{\Psi}^*$ as an initial point in the next iteration. The detailed algorithm for solving (51) is provided in Algorithm 2 where the algorithm terminates when an accuracy level is obtained.

Remark 2. *Algorithm 2 requires a feasible point $\mathbf{\Psi}^{(0)}$ to initialize the procedure. To this end, we first find a feasible point $\mathbf{\Psi}^{(0)}$ for the constraints* (44b) *and* (44c) *and then iteratively solve the optimization problem*

$$\min_{\mathbf{\Psi}} \quad \frac{(\mathbf{b}_{\mathrm{E}} \mathbf{u}_1)^2 + \|\mathbf{B}_E \mathbf{u}_1\|^2}{\theta_E} - \varphi_{\mathrm{E}}^{(\kappa)}(\mathbf{\Psi}), \tag{52a}$$

$$\textit{s.t.} \quad \text{(44b)} - \text{(44c)}, \tag{52b}$$

until the following requirement has been satisfied.

$$((\mathbf{b}_{\mathrm{E}} \mathbf{u}_1)^2 + \|\mathbf{B}_E \mathbf{u}_1\|^2)/\theta_E - \varphi_{\mathrm{E}}\left(\mathbf{\Psi}^{(\kappa)}\right) \leq 0. \tag{53}$$

In this case, $\mathbf{\Psi}^{(\kappa)}$ is a feasible solution for Problem (51).

Remark 3. *Algorithm 2 solves a convex problem at each iteration which involves $A_s = LK$ real-valued scalar variables and $A_q = L+K$ quadratic constraints. Therefore, the per-iteration complexity of Algorithm 2 is $\mathcal{O}\big((A_s)^2 A_q^{2.5} + A_q^{3.5}\big)$ [40].*

V. Detection of Eavesdropping Attack

In order to effectively design the AP selection and power allocation as in Section IV, it is crucial to identify the presence of eavesdroppers in our system and determine which user is being targeted. Thus, it is of importance to have some robust mechanism for detecting eavesdropping attacks. To date, there are a few works on pilot spoofing attack detection. Among them, the authors in [41] used Gaussian mixture models to identify spoofing attacks, while an additional random training phase after the conventional uplink pilot training phase was considered in [42] to identify the attack. In addition, a two-way training method was proposed in [43] for attack detection which requires additional downlink training from AP to the legitimate user. The authors in [44] utilized the secret key arrangement protocol to detect the eavesdropping attack, while the proposed protocol requires several uplink and downlink transmissions among legitimate transmitter and user. A three-step training scheme was proposed in [45] for detecting pilot spoofing attacks in reconfigurable intelligent surface (RIS)-assisted systems, while a two-step training scheme were proposed in [46] and [47] for detection of pilot spoofing attacks in massive MIMO systems, respectively. Although the above works make important steps towards detecting the pilot spoofing attack, they investigate simple setups concerning the APs and/or users. More specifically, a popular assumption in the aforementioned literature is that there is a single AP and/or a single user in the network. However, in more complicated scenarios such as CF-mMIMO systems, there are multiple cooperating APs serving multiple users. Therefore, we cannot straightforwardly apply the proposed detection schemes in the above mentioned works to detect the eavesdropper in our system. In [24], the authors proposed a simple method to identify abnormality in pilot training. However, this scheme requires all APs to exchange their received pilot signals, leading to substantial fronthaul demands and increased information exchange overhead. In addition, it is not clear how to obtain the average power of the received pilot signals at the central processing unit (CPU). In this work, we propose a new and efficient method which is also based on the received pilot signals, but can be implemented at each AP. Moreover, the APs do not possess prior knowledge of the users' instantaneous CSI to implement our proposed method.

To identify if user k is overheard by an eavesdropper or not, we consider two hypotheses $\mathcal{H}_{k,0}$ and $\mathcal{H}_{k,1}$, where the latter represents the scenario with an active eavesdropper, while the former represents the scenario without any active eavesdropper. Let $\mathbf{y}_{l,k}$ denote the projection of the received pilot vector at AP l onto $\boldsymbol{\phi}_k$ (i.e. $\mathbf{y}_{l,k} = \mathbf{Y}_{p,l}\boldsymbol{\phi}_k$), given by

$$\mathbf{y}_{l,k} = \begin{cases} \sqrt{\tau_p \rho_u}\mathbf{h}_{l,k} + \mathbf{N}_l \boldsymbol{\phi}_k, & \mathcal{H}_{k,0}, \\ \sqrt{\tau_p \rho_u}\mathbf{h}_{l,k} + \sqrt{\tau_p \rho_E}\mathbf{h}_{l,E} + \mathbf{N}_l \boldsymbol{\phi}_k, & \mathcal{H}_{k,1}. \end{cases} \tag{54}$$

Then, the average power of the m-th element of $\mathbf{y}_{l,k}$ is given by

$$\mathbb{E}\left\{|y_{l,k,m}|^2\right\} = \begin{cases} \tau_p \rho_u \beta_{l,k} + 1, & \mathcal{H}_{k,0}, \\ \tau_p \rho_u \beta_{l,k} + \tau_p \rho_E \beta_{l,E} + 1, & \mathcal{H}_{k,1}. \end{cases} \tag{55}$$

From (55), if AP l knows $\mathbb{E}\left\{|y_{l,k,m}|^2\right\}$, then it can determine if there exists an eavesdropper in the system. More precisely, AP l conducts a comparison between $\mathbb{E}\left\{|y_{l,k,m}|^2\right\}$ and $\tau_p \rho_u \beta_{l,k} + 1$. If these values do not equate, then it indicates the existence of an eavesdropping attempt targeted at user k. Otherwise, there is no such eavesdropper.

In practice, AP l does not exactly know $\mathbb{E}\left\{|y_{l,k,m}|^2\right\}$. Thus, we propose a simple method to estimate $\mathbb{E}\left\{|y_{l,k,m}|^2\right\}$. The proposed scheme uses the sample average power of $y_{l,k,m}$ as follows:

$$\xi_{l,k,m} = \frac{\sum_{n=1}^{N_{\mathrm{cb}}} \|\mathbf{y}_{l,k}(n)\|^2}{M N_{\mathrm{cb}}}, \tag{56}$$

where N_{cb} is the number of coherence bandwidth intervals within a whole system bandwidth. Since $\mathbf{y}_{l,k}(n)$ includes i.i.d. elements, $\xi_{l,k,m} \to \mathbb{E}\left\{|y_{l,k,m}|^2\right\}$ as $MN_{\mathrm{cb}} \to \infty$.

Remark 4. *In the 5G NR structure, the system bandwidth is 100 MHz and coherence bandwidth is about 360 KHz. Thus, the number of coherence bandwidth intervals is $N_{cb} = 100 \times 10^3/360 = 277$. In addition, the number of antennas per each APs, M, typically ranges from 2 to 10. As a consequence, MN_{cb} is around 544-2770 which is sufficiently large enough to ensure that $\xi_{l,k,m}$ is very close to $\mathbb{E}\left\{|y_{l,k,m}|^2\right\}$.*

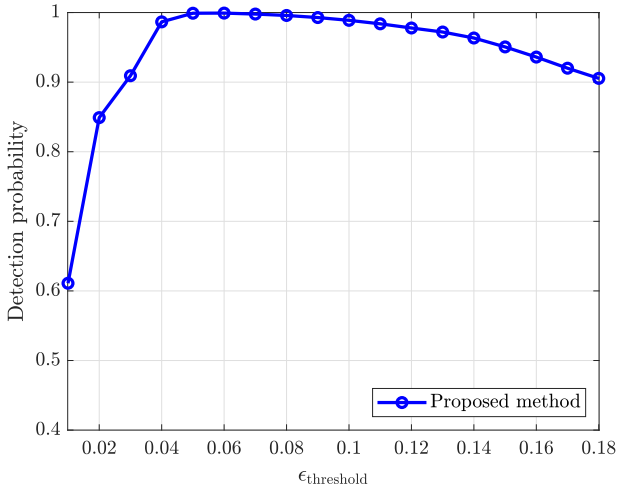


Fig. 2: Probability of active eavesdropping attack detection as a function of $\epsilon_{\text{threshold}}$.

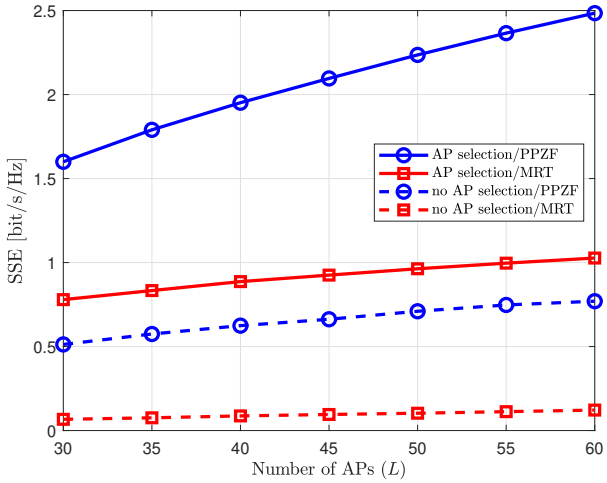


Fig. 3: SSE versus L for PZZF and MRT precoding schemes with AP selection and without it. Here, $M = 4$, $K = 10$, and $r = 100$ m.

By using (56), we can estimate $\mathbb{E}\left\{|y_{l,k,m}|^2\right\}$. Intuitively, if $\xi_{l,k,m}$ is close to $\tau_p \rho_u \beta_{l,k} + 1$, then we can conclude that there is no eavesdropper attacking user k. Building upon this insight, we propose a simple method to identify the presence of eavesdroppers in our system and determine which user is being targeted, outlined as follows:

- **Step 1**: At AP l, compute $\xi_{l,k,m}$ using (56).
- **Step 2**: Compute the ratio

$$\upsilon_{l,k,m} = \xi_{l,k,m} / \left(\tau_p \rho_u \beta_{l,k} + 1\right),$$

  and choose a threshold $\epsilon_{\text{threshold}}$. If $|\upsilon_{l,k,m} - 1| \leq \epsilon_{\text{threshold}}$, then AP l decides that there is no eavesdropper attacking user k. Otherwise, user k is overheard by an eavesdropper.
- **Step 3**: If a majority of the APs determine that user k is being overheard by an eavesdropper, then we conclude that the system is overheard by an eavesdropper, and user k is specifically targeted. Otherwise, we can conclude that there is no eavesdropper present in the system attacking user k.

Figure 2 shows the detection probability of an active eavesdropping attack relying on our proposed method for various values of $\epsilon_{\text{threshold}}$. The simulation settings resemble those explained in Section VI with $L = 10$, $M = 2$, and $K = 4$. It is seen that the proposed method works very well (i.e., the detection probability is close to 1). This implies that our assumption that the system knows the presence of an eavesdropper is reasonable. In addition, we observe that a trade-off exists between $\epsilon_{\text{threshold}}$ and the detection probability. More specifically, first, as $\epsilon_{\text{threshold}}$ increases, the detection probability increases but it then starts decreasing as $\epsilon_{\text{threshold}}$ increases beyond the optimized value. The intuitive reason is that a large $\epsilon_{\text{threshold}}$ can lead to misidentifying a user as attacked due to the fact that for the majority of the APs we would have $|\upsilon_{l,k,m} - 1| \leq \epsilon_{\text{threshold}}$, while a small $\epsilon_{\text{threshold}}$ may result in falsely dismissing the presence of an eavesdropper. Therefore, it is important to optimize $\epsilon_{\text{threshold}}$ to maximize the detection probability. For this network setup, its turns out that the optimized values for $\epsilon_{\text{threshold}}$ are between 0.05 to 0.08.

VI. Simulation Results

In this section, we provide numerical results to study the secrecy performance of CF-mMIMO with PPZF precoding in scenarios involving an active eavesdropper, as well as to verify the benefit of our AP selection and power allocation schemes. We also include MRT precoding as a benchmark for comparison. The APs and the users are randomly located within a square of size 1×1 km^2, which is wrapped around at the edges to avoid the boundary effects. Moreover, the eavesdropper is randomly located inside a circle with radius r around user 1. The large-scale fading is modelled as

$$\beta_{l,k} = 10^{\frac{\text{PL}_{l,k}^d}{10}} 10^{\frac{F_{l,k}}{10}}, \tag{57}$$

where $10^{\frac{\text{PL}_{l,k}^d}{10}}$ denotes the path loss, and $10^{\frac{F_{l,k}}{10}}$ depicts the shadowing effect with $F_{l,k} \in \mathcal{N}(0, 4^2)$ (in dB). Moreover, $\text{PL}_{l,k}^d$ (in dB) is given by

$$\text{PL}_{l,k}^d = -30.5 - 36.7 \log_{10}\left(\frac{d_{l,k}}{1\,\text{m}}\right), \tag{58}$$

and the correlation among the shadowing terms from the AP l to different users k is given by

$$\mathbb{E}\{F_{l,k} F_{j,k'}\} \triangleq \begin{cases} 4^2 2^{-\zeta_{k,k'}/9\,\text{m}}, & j = l, \\ 0, & \text{otherwise}, \end{cases} \tag{59}$$

where $\zeta_{k,k'}$ is the physical distance between users k and k' [34]. In addition, we choose the bandwidth $B = 20$ MHz, a noise power equal to -92 dBm, and the maximum transmit power for each AP and each user 200 mW and 100 mW, respectively. Also, $\rho_E = \rho_u$ and $\tau_p = K$. In all figures, we show the average SSE, while the average is taken over the large-scale fading.

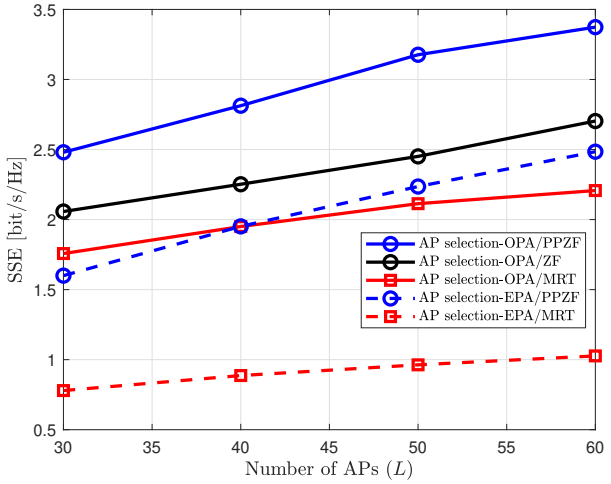


Fig. 4: SSE with AP selection and the proposed power allocation versus L for PPZF, ZF, and MRT precoding schemes. Here, $M = 4$ and $K = 10$, and $r = 100$ m with AP selection scheme, EPA, and OPA.

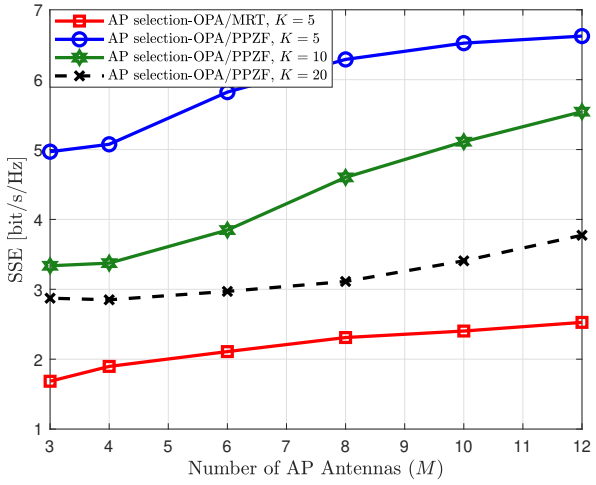


Fig. 5: SSE with AP selection and the proposed power allocation versus M, when $LM = 240$ and $r = 100$ m.

A. Performance Evaluation

1) Performance of the Proposed AP Selection: First, we investigate the performance of the proposed AP selection in Algorithm 1 for different precoding schemes. Figure 3 shows the SSE versus the number of APs, L. From this figure, we can draw the following conclusions:

- The proposed AP selection provides a noticeable secrecy improvement. More specifically, it provides performance gains of up to 220% and 730% for the CF-mMIMO system with PPZF and MRT precoding schemes, respectively. This performance gain is reasonable: i) primarily due to the fact that in a CF-mMIMO system with distributed APs, there are some APs which are in the close vicinity of the eavesdropper and hence serving the attacked user 1 by these APs may result in high values for the overheard SINR; ii) secondly, there are some APs which are located far away from user 1 and, thus, will not contribute significantly to its overall SE.
- The PPZF scheme provides a better SSE performance than the MRT scheme due to its ability to cancel the inter-user interference. This result highlights the importance of precoding scheme for secure CF-mMIMO systems.
- The gap between PPZF and MRT is always noticeable. Interestingly, PPZF along with the proposed AP selection provides around 2-fold improvement in the SSE compared to the MRT scheme for scenarios with a varying number of APs. In principle, this is reasonable since compared to MRT, PPZF offers an excellent balance between interference cancellation and the boosting of the desired signal. More importantly, it can provide high interference cancellation over a wide range of the numbers of antennas/active APs. Therefore, with PPZF, the benefits of the AP selection to enhance the SSE are more pronounced.

2) Performance of the Proposed Power Optimization: Figure 4 shows the advantage of the proposed power allocation as a function of the number of APs for CF-mMIMO system with the proposed AP selection. In this figure, "EPA" corresponds to the case with equal power allocation, i.e., $\rho_{l,k} = \rho_{\max}/K$ for all l and k, while "OPA" corresponds to our proposed optimized power allocation algorithm (i.e. Algorithm 2). Numerical results lead to the following observations:

- The CF-mMIMO system with the proposed power allocation remarkably outperforms the CF-mMIMO system with the EPA scheme. More specifically, the proposed power allocation in Algorithm 2 provides additional performance gain of around 55% and 100% over equal power allocation for a CF-mMIMO system relying on PPZF and MRT precoding scheme, respectively, which demonstrates the significance of our power allocation solution.
- Increasing the number of APs along with the proposed power allocation scheme amplifies the gain of PPZF over MRT.
- In Fig. 4, we also present the SSE performance of the CF-mMIMO system with ZF precoding, where the interference towards all the users is suppressed. Compared to ZF design, PPZF offers better SSE performance (around 30%). This is reasonable, since ZF avails of a modest array gain of $M - K$ as almost all the DoFs are used to mitigate the interference.

We would like to highlight that the proposed AP selection and power optimization algorithms can be applied to different channel models. More precisely, we can use the analytical expressions in (25) and (30), instead of the closed-form expressions in (26) and (31), respectively, to numerically implement our proposed AP selection Algorithm 1 and power optimization Algorithm 2. For example, by using channel realizations from the 3GPP 38.901 channel model implemented in QuaDRiGa [48], developed by the Fraunhofer Heinrich Hertz Institute, we can obtain the result in Fig. 7, which verifies

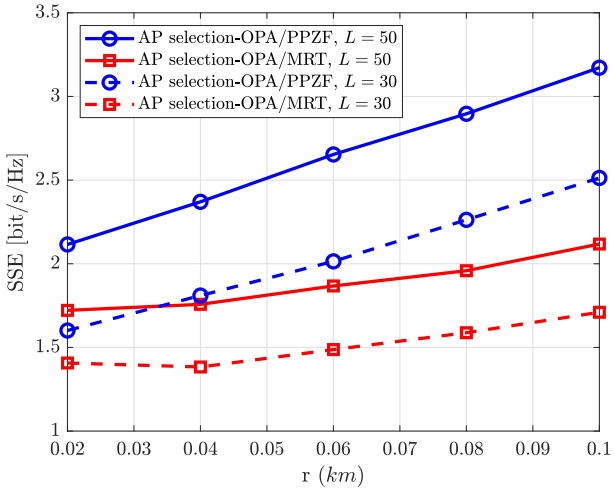


Fig. 6: SSE with AP selection and optimized power allocation versus r for $L = 30$ and $L = 50$, and $K = 20$.

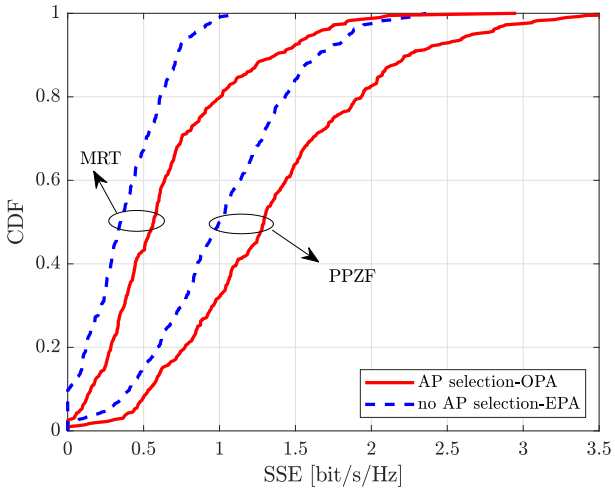


Fig. 7: CDF of the SSE with AP selection and optimized power allocation for the QuaDRiGa channel model. Here, $L = 30$, $K = 10$, and $r = 400$ m.

the benefit of our proposed scheme. More specifically, the proposed AP selection and power allocation yield a median SSE performance gain of around 65% and 30% for the CF-mMIMO system with the QuaDRiGa channel model relying on PMRT and PPZF precoding designs, respectively.

3) Effect of the Number of Antennas Per AP: In Fig. 5, we examine the SSE of CF-mMIMO as a function of the number of antennas per AP, M, when the total number of service antennas is fixed, i.e., $LM = 240$. For a fixed LM, increasing the number of antennas per AP (reducing the number of APs) has two effects on the SSE, namely, (i) higher array gain (a positive effect) and (ii) lower degree of macro-diversity and higher path losses (a negative effect). The positive effect becomes dominant, which enhances the SSE performance. In addition, we observe that by increasing the number of users, K, the secrecy performance of the CF-mMIMO system deteriorates. Nevertheless, the CF-mMIMO system comprising a high number of users and employing the proposed AP selection and power optimization still yields an excellent SSE performance. Interestingly, the CF-mMIMO system with 20 users and PPZF scheme offers around 60% SSE gain compared to the system comprising 5 users with MRT scheme. This is reasonable because, PPZF has the ability to cancel the inter-user interference.

4) Effect of the Eavesdropper Location: Figure 6 shows the SSE of CF-mMIMO system as a function of the position of the eavesdropper, the radius of a circle around user 1, r. As expected, the SSE improves with the increase of r. We can also see that the secrecy enhancement obtained by using a higher number of APs is more pronounced when the eavesdropper is located closer to the legitimate user 1.

B. Multiple-Antenna Eavesdropper

In the previous sections, we studied the performance of secure CF-mMIMO under active eavesdropping when the eavesdropper has a single antenna. We now examine how the SSE performance changes as eavesdropper is equipped with multiple antennas. We consider an eavesdropper with N_E antennas and then rewrite the eavesdropper's receive signals in (27) as follows:

$$\mathbf{z}_E = \sum_{l=1}^{L} \sqrt{\rho_{l,1}} \mathbf{H}_{l,E}^{\mathrm{H}} \mathbf{w}_{l,1} s_1 + \sum_{t \neq 1}^{K} \sum_{l=1}^{L} \sqrt{\rho_{l,t}} \mathbf{H}_{l,E}^{\mathrm{H}} \mathbf{w}_{l,t} s_t + \mathbf{n}_E, \tag{60}$$

where $\mathbf{z}_E \in \mathcal{C}^{N_E \times 1}$ represents the eavesdropper's received signals across her N_E antennas, $\mathbf{H}_{l,E} \in \mathcal{C}^{M \times N_E}$ is the channel matrix from the l-th AP to the eavesdropper, and $\mathbf{n}_E \in \mathcal{C}^{M \times N_E}$ is the additive zero-mean Gaussian noise vector each with variance equal to one. To detect the signal transmitted to user 1, the most effective and low-complexity strategy for the eavesdropper is to employ the maximum-ratio combining (MRC) scheme. Therefore, the SE at the eavesdropper can be written as

$$R_E^{\mathrm{MRC}} = \log_2 \left(1 + \mathrm{SINR}_E^{\mathrm{MRC}}\right), \tag{61}$$

where

$$\mathrm{SINR}_E^{\mathrm{MRC}} \approx \sum_{n=1}^{N_E} \frac{\mathbb{E}\{|\mathrm{BU}_{E,1}^n|^2\}}{\sum\limits_{t \neq 1}^{K} \mathbb{E}\{|\mathrm{UI}_{\mathrm{E},t}^n|^2\} + 1} = \sum_{n=1}^{N_E} \mathrm{SINR}_E^n, \tag{62}$$

where SINR_E^n is the received SINR at the n-th antenna of the eavesdropper given in (31), while Eq. (62) follows from the well known MRC result that the collective SNR is the summation of SNRs at each element.

Figure 8 shows the SSE as a function of the number of antennas at eavesdropper, N_E. The findings show that increasing the number of eavesdropper antennas has a significant impact on the secrecy performance of CF-mMIMO for both PPZF and MRT precoding schemes. Furthermore, we can infer that the reduction in the SSE attributed to the increased number of eavesdropper antennas is more pronounced when employing our proposed AP selection and power optimization approaches

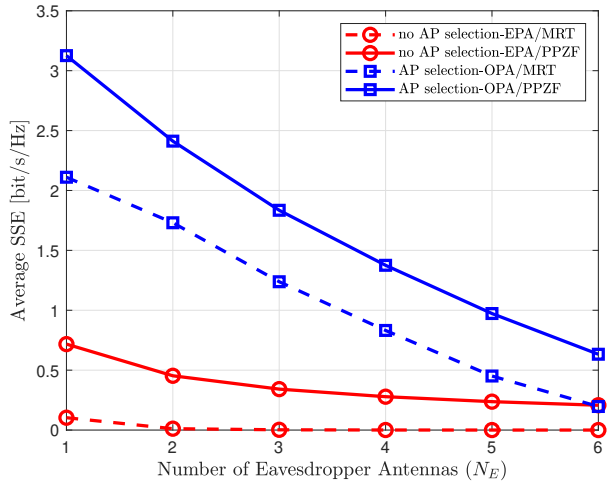


Fig. 8: SSE with AP selection and the proposed power allocation versus the number of eavesdropper antennas N_E for PPZF and MRT precoding schemes. Here, $L = 50$, $M = 4$, $K = 10$, and $r = 100$ m.

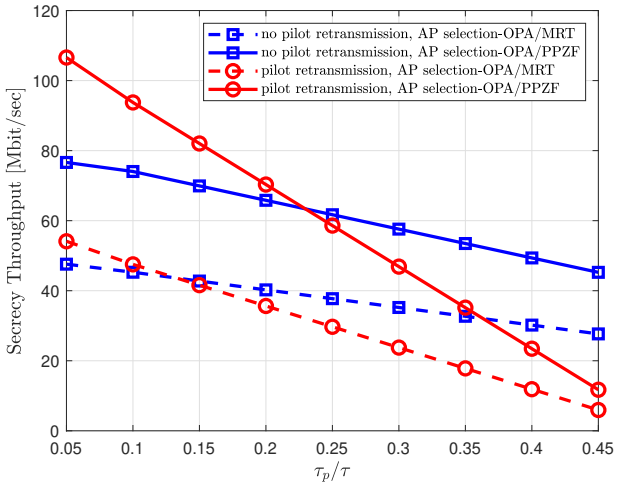


Fig. 9: Secrecy throughput with AP selection and the proposed power allocation versus τ_p/τ for PPZF and MRT precoding schemes. Here, $L = 50$, $M = 4$, $K = 10$, $B = 20$ MHz, and $r = 200$ m.

compared to the scenario where all APs are utilized with equal power allocation.

C. Mitigation of Eavesdropping Attack

As discussed in Section II, the pilot spoofing attack during the uplink training phase degrades considerably the SSE performance of the system. Therefore, in this subsection we develop a counter strategy to suppress the effect of pilot spoofing attack. We propose a *pilot re-transmission scheme* where the pilot sequences will be re-transmitted when the eavesdropping attack has been detected using the proposed scheme in Section V. We consider the worst case where the eavesdropper has prior knowledge of the pilot sequences and tries to launch a pilot spoofing attack using those training sequences. In this case, the CF-mMIMO system can adapt the training sequences based on the knowledge of the current pilot transmission rather than merely transmitting them randomly. Our proposed pilot re-transmission scheme is outlined as follows:

- **Initialization**: Choose the value of pilot length τ_p and set the number of the pilot re-transmission rounds as $N_{\rm pr} = 1$.
- **Step 1**: Each user k sends pilot sequence $\boldsymbol{\phi}_k$.
- **Step 2**: The eavesdropping attack detection scheme in Section V is performed by the APs for all the users.
- **Step 3**: If there is no eavesdropper attacking Stop. Otherwise, go to step 4.
- **Step 4**: If we conclude that the system is overheard by an eavesdropper, and one user (say user 1) is specifically targeted, then the system will find $\boldsymbol{\phi}_1^*$ such that the eavesdropping attack detection scheme results that there is no eavesdropper present in the system attacking user 1. Then, user 1 will re-transmit this new pilot $\boldsymbol{\phi}_1^*$ and other users will re-transmit the new pilots from the remaining orthogonal pilot set. Set $N_{\rm pr} = 2$.

Accordingly, the achievable SE expression for user 1 can be obtained as

$$R_1^{\rm pr} = \log_2\left(1 + {\rm SINR}_1^{\rm pr}\right), \tag{63}$$

where

$${\rm SINR}_1^{\rm pr} = \frac{\left(\sum\limits_{l=1}^{L}\sqrt{\left(M - |\mathcal{S}_l|\right)\rho_{l,1}\gamma_{l,1}^{\rm pr}}\right)^2}{\sum\limits_{t=1}^{K}\sum\limits_{l=1}^{L}\rho_{l,t}\left(\beta_{l,1} - \delta_{l,1}\gamma_{l,1}^{\rm pr}\right) + 1}, \tag{64}$$

with

$$\gamma_{l,1}^{\rm pr} = \begin{cases} \frac{\tau_p\rho_u\beta_{l,1}^2}{\tau_p\rho_u\beta_{l,1}+1}, & N_{\rm pr} = 2, \\ \frac{\tau_p\rho_u\beta_{l,1}^2}{\tau_p\rho_u\beta_{l,1}+1}, & N_{\rm pr} = 1 \text{ and } \mathcal{H}_{1,0}, \\ \frac{\tau_p\rho_u\beta_{l,1}^2}{\tau_p\rho_u\beta_{l,1}+\tau_p\rho_E\beta_{l,E}+1}, & N_{\rm pr} = 1 \text{ and } \mathcal{H}_{1,1}, \end{cases} \tag{65}$$

and the achievable SE of the eavesdropper can be obtained as $R_E^{\rm pr} = \log_2(1 + {\rm SINR}_E^{\rm pr})$ where

$${\rm SINR}_E^{\rm pr} = \frac{\sum\limits_{l=1}^{L}\rho_{l,1}\beta_{l,E}}{\sum\limits_{t\neq 1}^{K}\sum\limits_{l=1}^{L}\rho_{l,t}\beta_{l,E} + 1}. \tag{66}$$

We highlight that the proposed pilot re-transmission scheme might consume more uplink training resources. Therefore, the cost of the pilot re-transmission scheme entails a loss in the SE of the users which highly depends on the application scenarios. For example, this cost would be non-negligible when the length of the coherence interval is small (e.g., in high mobility environments or/and the number of users is large) or/and the length of the uplink training phase is large. Therefore, to ensure a fair comparison between conventional CF-mMIMO (with no pilot re-transmission) and secure CF-mMIMO with pilot re-transmission scheme, we consider the

per-user (eavesdropper) throughputs which take into account the pilot training overhead and are defined as

$$S_a = B\left(1 - \frac{\tau_p}{\tau}\right) R_a, \tag{67}$$

and

$$S_a^{\rm pr} = B\left(1 - \frac{N_{\rm pr}\tau_p}{\tau}\right) R_a^{\rm pr}, \tag{68}$$

where $a \in \{1, E\}$, τ is the total length of the coherence interval in samples, and B is the spectral bandwidth. The term τ_p/τ signifies the fact that, for each coherence interval of length τ samples and with no pilot re-transmission, we consume τ_p samples for the uplink training phase, while for the case of pilot re-transmission, we consume $N_{\rm pr}\tau_p$ samples.

Figure 9 compares the performance of the CF-mMIMO with pilot re-transmission and with no pilot re-transmission schemes as a function of τ_p/τ for PPZF and MRT precoding designs. The results indicate that for PPZF precoding, pilot re-transmission outperforms no pilot re-transmission when $\tau_p/\tau < 0.23$, and exhibits an inferior performance when $\tau_p/\tau \geq 0.23$. This result is in accordance with the above discussion and implies that the lengths of τ and τ_p are the key factors determining the extent to which CF-mMIMO with pilot re-transmission outperforms conventional CF-mMIMO (with no pilot re-transmission).

VII. Conclusion

We investigated the SSE of CF-mMIMO systems with multi-antenna APs and PPZF precoding under an active eavesdropping attack and the presence of channel estimation errors. We derived closed-form expressions for the SE at the legitimate user and eavesdropper, and thereby the SSE. We proposed a large-scale-based AP selection approach to improve the SSE and a large-scale-based power optimization approach to maximize the received SINR at the legitimate user, subject to a maximum allowable SINR at the eavesdropper and individual QoS requirements and transmit power constraints. We showed that our proposed AP selection and power optimization approaches provide significant SE gains. Our results also confirm that for a CF-mMIMO system experiencing active eavesdropping, the PPZF precoding scheme can achieve a noticeable SSE gain compared to the MRT precoding scheme. Insights from the simulation results show that the number of APs is a dominating factor of the SSE performance. When the number of APs is large, a secure CF-mMIMO system relying on the proposed approaches and PPZF scheme can offer excellent performance even for higher user loads. Note that the proposed AP selection and power optimization algorithms are not limited to Rayleigh fading channels. They are applicable to various channel models by using numerical methods, as evidenced in Fig. 7. Finally, interesting research topics for future research include: (i) investigating CF-mMIMO systems under multiple-eavesdropper attacks; (ii) developing joint AP selection and power optimization schemes to maximize the SSE; and (iii) developing secure CF-mMIMO systems with both APs and users equipped with multiple antennas over correlated fading channels.

Appendix A
Proof of Proposition 1

We first calculate the numerator of (25) as

$$\begin{aligned}
&\Bigg| \sum\nolimits_{l\in\mathcal{Z}_k} \sqrt{\rho_{l,k}}\mathbb{E}\left\{(\hat{\mathbf{h}}_{l,k} + \tilde{\mathbf{h}}_{l,k})^{\rm H}\mathbf{w}_{l,k}^{\rm PZF}\right\} \\
&+ \sum\nolimits_{p\in\mathcal{M}_k} \sqrt{\rho_{p,k}}\mathbb{E}\left\{(\hat{\mathbf{h}}_{p,k} + \tilde{\mathbf{h}}_{p,k})^{\rm H}\mathbf{w}_{p,j}^{\rm PMRT}\right\} \Bigg|^2 \\
&= \left(\sum_{l\in\mathcal{Z}_k}\sqrt{(M-|\mathcal{S}_l|)\rho_{l,k}\gamma_{l,k}} + \sum_{p\in\mathcal{M}_k}\sqrt{(M-|\mathcal{S}_p|)\rho_{p,k}\gamma_{p,k}}\right)^2, \\
&= \left(\sum\nolimits_{l=1}^{L}\sqrt{(M-|\mathcal{S}_l|)\rho_{l,k}\gamma_{l,k}}\right)^2.
\end{aligned} \tag{69}$$

The first term in the denominator of (25) is given by

$$\begin{aligned}
&\sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF} + \sum_{p\in\mathcal{M}_t}\sqrt{\rho_{p,l}}\mathbf{h}_{p,k}^{\rm H}\mathbf{w}_{p,t}^{\rm PMRT}\right|^2\right\} \\
&= \sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF}\right|^2\right\} \\
&+ 2\sum_{t=1}^{K}{\rm Re}\left\{\sum_{l\in\mathcal{Z}_t}\sum_{p\in\mathcal{M}_t}\sqrt{\rho_{l,t}\rho_{p,t}}\mathbb{E}\left\{\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF}\left(\mathbf{w}_{p,t}^{\rm PMRT}\right)^{\rm H}\mathbf{h}_{p,k}\right\}\right\} \\
&+ \sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{p\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\mathbf{h}_{p,k}^{\rm H}\mathbf{w}_{p,t}^{\rm PMRT}\right|^2\right\}.
\end{aligned} \tag{70}$$

Now, we compute the first term of (70) as follows:

If $k \in \mathcal{S}_l$, then the term $\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF}$ can be calculated based on (12). Thus, we have

$$\begin{aligned}
&\sum_{t=1}^{K}\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF}\right|^2\right\} \\
&= \sum_{t=1}^{K}\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}(\alpha_{l,k,t}^{\rm PZF} + \tilde{\mathbf{h}}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF})\right|^2\right\}, \\
&= \left(\sum_{l\in\mathcal{Z}_k}\sqrt{\rho_{l,k}}\alpha_{l,k,k}^{\rm PZF}\right)^2 + \sum_{t=1}^{K}\sum_{l\in\mathcal{Z}_t}\rho_{l,t}(\beta_{l,k} - \gamma_{l,k}), \\
&= \left(\sum_{l\in\mathcal{Z}_k}\sqrt{\rho_{l,k}(M-|\mathcal{S}_l|)\gamma_{l,k}}\right)^2 + \sum_{t=1}^{K}\sum_{l\in\mathcal{Z}_t}\rho_{l,t}(\beta_{l,k} - \gamma_{l,k}).
\end{aligned} \tag{71}$$

If $k \notin \mathcal{S}_l$, then $k \in \mathcal{W}_l$ and $\mathbf{h}_{l,k}$ is independent of $\mathbf{w}_{l,t}^{\rm PZF}$ $\forall t \neq k$. Hence,

$$\begin{aligned}
&\sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF}\right|^2\right\} \\
&= \sum\nolimits_{t=1}^{K}\sum\nolimits_{l\in\mathcal{Z}_t}\rho_{l,t}\,\mathbb{E}\left\{\left|\mathbf{h}_{l,k}^{\rm H}\mathbf{w}_{l,t}^{\rm PZF}\right|^2\right\}, \\
&= \sum\nolimits_{t=1}^{K}\sum\nolimits_{l\in\mathcal{Z}_t}\rho_{l,t}\mathbb{E}\left\{\mathbf{h}_{l,k}^{\rm H}\mathbb{E}\left\{\mathbf{w}_{l,t}^{\rm PZF}\left(\mathbf{w}_{l,t}^{\rm PZF}\right)^{\rm H}\right\}\mathbf{h}_{l,k}\right\}, \\
&= \sum\nolimits_{t=1}^{K}\sum\nolimits_{l\in\mathcal{Z}_t}\rho_{l,t}\beta_{l,k}.
\end{aligned} \tag{72}$$

Now, we combine (71) and (72) to obtain

$$\sum_{t=1}^{K}\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}}\right|^2\right\}$$
$$=\left(\sum_{l\in\mathcal{Z}_k}\sqrt{\rho_{l,k}(M-|\mathcal{S}_l|)\gamma_{l,k}}\right)^2+\sum_{t=1}^{K}\sum_{l\in\mathcal{Z}_t}\rho_{l,t}(\beta_{l,k}-\delta_{l,k}\gamma_{l,k}). \tag{73}$$

Similarly, the third term of the right-hand-side in (70) can be computed as follows:

If $k\in\mathcal{W}_p$, then $\mathbf{h}_{p,k}$ is independent of $\mathbf{w}_{p,t}^{\mathrm{PMRT}}$ $\forall t\neq k$. Thus, we obtain

$$\sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{p\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right|^2\right\}$$
$$=\sum\nolimits_{t=1}^{K}\sum\nolimits_{p\in\mathcal{M}_t}\rho_{p,t}\,\mathbb{E}\left\{\left|\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right|^2\right\}$$
$$=\left(\sum\nolimits_{l\in\mathcal{Z}_k}\sqrt{\rho_{l,k}(M-|\mathcal{S}_l|)\gamma_{l,k}}\right)^2$$
$$+\sum_{t=1}^{K}\sum_{p\in\mathcal{M}_t}\rho_{p,t}\mathbb{E}\left\{\mathbf{h}_{p,k}^{\mathrm{H}}\mathbb{E}\left\{\mathbf{w}_{p,t}^{\mathrm{PMRT}}\left(\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right)^{\mathrm{H}}\right\}\mathbf{h}_{p,k}\right\}$$
$$=\left(\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{l,k}(M-|\mathcal{S}_p|)\gamma_{l,k}}\right)^2+\sum_{t=1}^{K}\sum_{p\in\mathcal{M}_t}\rho_{p,t}\beta_{p,k}. \tag{74}$$

If $k\in\mathcal{S}_p$, then $\hat{\mathbf{h}}_{p,k}\mathbf{w}_{p,t}^{\mathrm{PMRT}}=0$, and we have

$$\sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{p\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right|^2\right\}$$
$$=\sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{p\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\tilde{\mathbf{h}}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right|^2\right\},$$
$$=\sum\nolimits_{t=1}^{K}\sum\nolimits_{p\in\mathcal{M}_t}\rho_{p,t}(\beta_{p,k}-\gamma_{p,k}). \tag{75}$$

Adding (74) to (75), we obtain

$$\sum\nolimits_{t=1}^{K}\mathbb{E}\left\{\left|\sum\nolimits_{\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right|^2\right\}$$
$$=\left(\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{p,k}(M-|\mathcal{S}_p|)\gamma_{p,k}}\right)^2+\sum_{t=1}^{K}\sum_{p\in\mathcal{M}_t}\rho_{p,t}(\beta_{p,k}-\delta_{p,k}\gamma_{p,k}). \tag{76}$$

The second term of (70) is equal to 0 $\forall t\neq k$. Moreover, for $t=k$, we obtain

$$2\sum_{t=1}^{K}\mathrm{Re}\left\{\sum_{l\in\mathcal{Z}_k}\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{l,k}\rho_{p,k}}\mathbb{E}\left\{\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,k}^{\mathtt{PZF}}\left(\mathbf{w}_{p,k}^{\mathrm{PMRT}}\right)^{\mathrm{H}}\mathbf{h}_{p,k}\right\}\right\}$$
$$=2\sum_{l\in\mathcal{Z}_k}\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{l,k}\rho_{p,k}}\mathbb{E}\left\{\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,k}^{\mathtt{PZF}}\right\}\mathbb{E}\left\{\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,k}^{\mathrm{PMRT}}\right\},$$
$$=2\sum_{l\in\mathcal{Z}_k}\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{l,k}\rho_{p,k}(M-|\mathcal{S}_l|)(M-|\mathcal{S}_p|)\gamma_{l,k}\gamma_{p,k}}. \tag{77}$$

Now, using (73), (76), and (77), we can rewrite (70) as

$$\sum_{t=1}^{K}\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,k}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}}+\sum_{p\in\mathcal{M}_t}\sqrt{\rho_{p,l}}\mathbf{h}_{p,k}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathrm{PMRT}}\right|^2\right\}$$
$$=\left(\sum_{l\in\mathcal{Z}_k}\sqrt{\rho_{l,k}(M-|\mathcal{S}_l|)\gamma_{l,k}}\right)^2+\sum_{t=1}^{K}\sum_{l\in\mathcal{Z}_t}\rho_{l,t}(\beta_{l,k}-\delta_{l,k}\gamma_{l,k})$$
$$+\left(\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{p,k}(M-|\mathcal{S}_p|)\gamma_{p,k}}\right)^2+\sum_{t=1}^{K}\sum_{p\in\mathcal{M}_t}\rho_{p,t}(\beta_{p,k}-\delta_{p,k}\gamma_{p,k})$$
$$+2\sum_{l\in\mathcal{Z}_k}\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{l,k}\rho_{p,k}(M-|\mathcal{S}_l|)(M-|\mathcal{S}_p|)\gamma_{l,k}\gamma_{p,k}},$$
$$=\left(\sum_{l\in\mathcal{Z}_k}\sqrt{\rho_{l,k}(M-|\mathcal{S}_l|)\gamma_{l,k}}+\sum_{p\in\mathcal{M}_k}\sqrt{\rho_{p,k}(M-|\mathcal{S}_p|)\gamma_{p,k}}\right)^2$$
$$+\sum_{t=1}^{K}\sum_{l=1}^{L}\rho_{l,t}(\beta_{l,k}-\delta_{l,k}\gamma_{l,k}),$$
$$=\left(\sum_{l=1}^{L}\sqrt{\rho_{l,k}(M-|\mathcal{S}_l|)\gamma_{l,k}}\right)^2+\sum_{t=1}^{K}\sum_{l=1}^{L}\rho_{l,t}(\beta_{l,k}-\delta_{l,k}\gamma_{l,k}). \tag{78}$$

Substituting (69) and (78) in (25), $\mathrm{SINR}_k^{\mathrm{PPZF}}$ in (26) is obtained. We next provide a closed-form expression for the SE for the eavesdropper.

Appendix B
Proof of Proposition 2

Based on (30), the received SINR at the eavesdropper can be written as

$$\mathrm{SINR}_E=\frac{\mathbb{E}\left\{|\mathrm{BU}_{E,1}|^2\right\}}{\sum_{t\neq1}^{K}\mathbb{E}\left\{|\mathrm{UI}_{\mathrm{E},t}|^2\right\}+1}. \tag{79}$$

We first calculate $\mathbb{E}\left\{|\mathrm{BU}_{E,1}|^2\right\}$. By denoting $\hat{\mathbf{h}}_{p,E}=\sqrt{\alpha_{p,1}}\mathbf{h}_{p,1}$, where $\alpha_{p,1}=\left(\rho_E\beta_{p,E}^2\right)/\left(\rho_u\beta_{p,1}^2\right)$, and $\gamma_{p,E}=\alpha_{p,1}\gamma_{p,1}$, we have

$$\mathbb{E}\left\{|\mathrm{BU}_{E,1}|^2\right\}$$
$$=\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_1}\sqrt{\rho_{l,1}}\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,1}^{\mathtt{PZF}}+\sum_{p\in\mathcal{M}_1}\sqrt{\rho_{p,1}}\mathbf{h}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right|^2\right\}$$
$$=\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_1}\sqrt{\rho_{l,1}}\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,1}^{\mathtt{PZF}}\right|^2\right\}+\mathbb{E}\left\{\left|\sum_{p\in\mathcal{M}_1}\sqrt{\rho_{p,1}}\mathbf{h}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right|^2\right\}$$
$$+2\mathrm{Re}\left\{\sum_{l\in\mathcal{Z}_1}\sum_{p\in\mathcal{M}_1}\sqrt{\rho_{l,1}\rho_{p,1}}\mathbb{E}\left\{\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,1}^{\mathtt{PZF}}\left(\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right)^{\mathrm{H}}\mathbf{h}_{p,E}\right\}\right\},$$
$$=\left(\sum_{l\in\mathcal{Z}_1}\sqrt{\rho_{l,1}(M-|\mathcal{S}_l|)\gamma_{l,E}}\right)^2+\sum_{l\in\mathcal{Z}_1}\rho_{l,1}(\beta_{l,E}-\gamma_{l,E})$$
$$+\sum_{p\in\mathcal{M}_1}\rho_{p,1}\mathbb{E}\left\{\left|\hat{\mathbf{h}}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right|^2\right\}$$
$$+\sum_{p\in\mathcal{M}_1}\sum_{q\in\mathcal{M}_1,q\neq p}\sqrt{\rho_{p,1}\rho_{q,1}}\mathbb{E}\left\{\hat{\mathbf{h}}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right\}\mathbb{E}\left\{(\mathbf{w}_{q,1}^{\mathtt{PMRT}})^{\mathrm{H}}\hat{\mathbf{h}}_{q,E}\right\}$$
$$+\sum\nolimits_{p\in\mathcal{M}_1}\rho_{p,1}(\beta_{p,E}-\gamma_{p,E})$$
$$+2\sum_{l\in\mathcal{Z}_1}\sum_{p\in\mathcal{M}_1}\sqrt{\rho_{l,1}\rho_{p,1}(M-|\mathcal{S}_l|)(M-|\mathcal{S}_p|)\gamma_{l,E}\gamma_{p,E}}, \tag{80}$$

where we have used $\gamma_{l,E} = \alpha_{l,1}\gamma_{l,1}$, and the fact that $\tilde{\mathbf{h}}_{p,E} = \mathbf{h}_{p,E} - \hat{\mathbf{h}}_{p,E}$ is independent of $\hat{\mathbf{h}}_{p,E}$ and has zero mean. Since $\hat{\mathbf{h}}_{p,E}^{\mathrm{H}} = \sqrt{\alpha_{p,1}}\hat{\mathbf{h}}_{p,1}^{\mathrm{H}}$, the third term of (80) can be re-written as

$$\begin{aligned}
&\sum\nolimits_{p\in\mathcal{M}_1} \rho_{p,1}\mathbb{E}\left\{\left|\hat{\mathbf{h}}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right|^2\right\} \\
&= \sum\nolimits_{p\in\mathcal{M}_1} \frac{\rho_{p,1}\alpha_{p,1}}{(M-|\mathcal{S}_p|)\gamma_{p,1}}\mathbb{E}\left\{\left|\hat{\mathbf{h}}_{p,1}^{\mathrm{H}}\mathbf{B}_p\hat{\mathbf{h}}_{p,1}\right|^2\right\}, \\
&= \sum\nolimits_{p\in\mathcal{M}_1} \rho_{p,1}\gamma_{p,E}(M-|\mathcal{S}_p|+1). \qquad (81)
\end{aligned}$$

Moreover, by using (13) and the fact that $\mathbb{E}\left\{\hat{\mathbf{h}}_{p,1}^{\mathrm{H}}\mathbf{B}_p\hat{\mathbf{h}}_{p,1}\right\} = \mathbb{E}\left\{\mathrm{tr}(\mathbf{B}_p\hat{\mathbf{h}}_{p,1}\hat{\mathbf{h}}_{p,1}^{\mathrm{H}})\right\}$, we have

$$\begin{aligned}
&\sum\nolimits_{p\in\mathcal{M}_1} \sqrt{\rho_{p,1}}\mathbb{E}\left\{\hat{\mathbf{h}}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right\} \\
&= \sum\nolimits_{p\in\mathcal{M}_1} \frac{\sqrt{\rho_{p,1}\alpha_{p,1}}}{\sqrt{(M-|\mathcal{S}_p|)\gamma_{p,1}}}\mathbb{E}\left\{\hat{\mathbf{h}}_{p,1}^{\mathrm{H}}\mathbf{B}_p\hat{\mathbf{h}}_{p,1}\right\}, \\
&= \sum\nolimits_{p\in\mathcal{M}_1} \frac{\sqrt{\rho_{p,1}\alpha_{p,1}}}{\sqrt{(M-|\mathcal{S}_p|)\gamma_{p,1}}}\mathbb{E}\left\{\mathrm{tr}(\mathbf{B}_p\hat{\mathbf{h}}_{p,1}\hat{\mathbf{h}}_{p,1}^{\mathrm{H}})\right\}, \\
&= \sum\nolimits_{p\in\mathcal{M}_1} \frac{\sqrt{\rho_{p,1}\alpha_{p,1}}}{\sqrt{(M-|\mathcal{S}_p|)\gamma_{p,1}}}\gamma_{p,1}\mathbb{E}\left\{\mathrm{tr}(\mathbf{B}_p)\right\}, \\
&= \sum\nolimits_{p\in\mathcal{M}_1} \sqrt{\rho_{p,1}(M-|\mathcal{S}_p|)\gamma_{p,E}}. \qquad (82)
\end{aligned}$$

Thus, we can rewrite the fourth term of (80) as

$$\begin{aligned}
&\sum_{p\in\mathcal{M}_1}\sum_{q\in\mathcal{M}_1} \sqrt{\rho_{p,1}\rho_{q,1}}\mathbb{E}\left\{\hat{\mathbf{h}}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,1}^{\mathtt{PMRT}}\right\}\mathbb{E}\left\{(\mathbf{w}_{q,1}^{\mathtt{PMRT}})^{\mathrm{H}}\hat{\mathbf{h}}_{q,E}\right\} \\
&= \sum_{p\in\mathcal{M}_1}\sum_{q\in\mathcal{M}_1} \sqrt{\rho_{p,1}\rho_{q,1}(M-|\mathcal{S}_p|)(M-|\mathcal{S}_q|)\gamma_{p,E}\gamma_{q,E}}. \qquad (83)
\end{aligned}$$

To this end, by substituting (81) and (83) into (80), we obtain

$$\begin{aligned}
&\mathbb{E}\left\{|\mathrm{BU}_{E,1}|^2\right\} = \left(\sum_{l=1}^{L}\sqrt{\rho_{l,1}(M-|\mathcal{S}_l|)\gamma_{l,E}}\right)^2 \\
&+\sum_{l\in\mathcal{Z}_1}\rho_{l,1}(\beta_{l,E}-\gamma_{l,E})+\sum_{p\in\mathcal{M}_1}\rho_{p,1}\beta_{p,E}, \\
&= \left(\sum_{l=1}^{L}\sqrt{\rho_{l,1}(M-|\mathcal{S}_l|)\gamma_{l,E}}\right)^2+\sum_{l=1}^{L}\rho_{l,1}\beta_{l,E}-\sum_{l\in\mathcal{Z}_1}\rho_{l,1}\gamma_{l,E}. \qquad (84)
\end{aligned}$$

Similarly, for $t \neq 1$, we calculate $\mathbb{E}\left\{|\mathrm{UI}_{E,t}|^2\right\}$ as

$$\begin{aligned}
&\mathbb{E}\left\{|\mathrm{UI}_{E,t}|^2\right\} = \mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}}+\sum_{p\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\mathbf{h}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathtt{PMRT}}\right|^2\right\}, \\
&= \mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}}\right|^2\right\}+\mathbb{E}\left\{\left|\sum_{p\in\mathcal{M}_t}\sqrt{\rho_{p,t}}\mathbf{h}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathtt{PMRT}}\right|^2\right\} \\
&+2\,\mathrm{Re}\left\{\sum_{l\in\mathcal{Z}_t}\sum_{p\in\mathcal{M}_t}\sqrt{\rho_{l,t}\rho_{p,t}}\mathbb{E}\left\{\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}}\left(\mathbf{w}_{p,t}^{\mathtt{PMRT}}\right)^{\mathrm{H}}\mathbf{h}_{p,E}\right\}\right\}. \qquad (85)
\end{aligned}$$

The first term of (85) is computed for two conditions: 1) when user $1 \in \mathcal{S}_l$, then $\hat{\mathbf{h}}_{l,1}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}} = 0$ and $\hat{\mathbf{h}}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}} = 0$ ($\hat{\mathbf{h}}_{l,E} = \sqrt{\alpha_{l,1}}\hat{\mathbf{h}}_{l,1}$); 2) when user $1 \notin \mathcal{S}_l$ and hence $\hat{\mathbf{h}}_{l,E}$ is independent of $\mathbf{w}_{l,t}^{\mathtt{PZF}}$. Then, we obtain

$$\mathbb{E}\left\{\left|\sum_{l\in\mathcal{Z}_t}\sqrt{\rho_{l,t}}\mathbf{h}_{l,E}^{\mathrm{H}}\mathbf{w}_{l,t}^{\mathtt{PZF}}\right|^2\right\} = \sum_{l\in\mathcal{Z}_t}\rho_{l,t}(\beta_{l,E}-\delta_{l,1}\gamma_{l,E}). \quad (86)$$

Similarly, for the the second term of (85), $\hat{\mathbf{h}}_{p,1}^{\mathrm{H}}\mathbf{B}_p = 0$ and $\hat{\mathbf{h}}_{p,E}^{\mathrm{H}}\mathbf{B}_p = 0$ when user $1 \in \mathcal{S}_p$. Moreover, $\mathbf{h}_{p,1}$ is independent of $\mathbf{w}_{p,t}^{\mathtt{PMRT}}$ when user $1 \notin \mathcal{S}_p$. Then, we obtain

$$\mathbb{E}\left\{\left|\sum\nolimits_{p\in\mathcal{M}_t^{\prime}}\sqrt{\rho_{p,t}}\mathbf{h}_{p,E}^{\mathrm{H}}\mathbf{w}_{p,t}^{\mathtt{PMRT}}\right|^2\right\} = \sum\nolimits_{p\in\mathcal{M}_t^{\prime}}\rho_{p,t}(\beta_{p,E}-\delta_{p,1}\gamma_{p,E}). \quad (87)$$

The third term of (85) is equal to 0 as $\mathbf{h}_{l,E}$ and $\mathbf{h}_{p,E}$ are independent of both $\mathbf{w}_{l,t}^{\mathtt{PZF}}$ and $\mathbf{w}_{p,t}^{\mathtt{PMRT}}$, respectively. By substituting (86) and (87) into (85), we obtain

$$\begin{aligned}
\mathbb{E}\left\{|\mathrm{UI}_{\mathrm{E,t}}|^2\right\} &= \sum\nolimits_{l\in\mathcal{Z}_t}\rho_{l,t}(\beta_{l,E}-\delta_{l,1}\gamma_{l,E}) \\
&+\sum\nolimits_{p\in\mathcal{M}_t}\rho_{p,t}(\beta_{p,E}-\delta_{p,1}\gamma_{p,E}), \\
&= \sum\nolimits_{l=1}^{L}\rho_{p,t}(\beta_{l,E}-\delta_{l,1}\gamma_{l,E}). \qquad (88)
\end{aligned}$$

By plugging (81) and (88) into (79), SINR_E in (31) can be obtained.

REFERENCES

[1] Y. S. Atiya, Z. Mobini, H. Q. Ngo, and M. Matthaiou, "Cell-free massive MIMO with protective partial zero-forcing and active eavesdropping," in *Proc. IEEE VTC*, Jun. 2023, pp. 1–5.

[2] M. Matthaiou, O. Yurduseven, H. Q. Ngo, D. Morales-Jimenez, S. L. Cotton, and V. F. Fusco, "The road to 6G: Ten physical layer challenges for communications engineers," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 64–69, Jan. 2021.

[3] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw*, vol. 2019, no. 1, pp. 1–13, Dec. 2019.

[4] M. Mohammadi, Z. Mobini, H. Q. Ngo, and M. Matthaiou, "Next generation multiple access with cell-free massive MIMO," *Proc. IEEE*, 2024.

[5] Z. Mobini, H. Q. Ngo, M. Matthaiou, and L. Hanzo, "Cell-free massive MIMO surveillance of multiple untrusted communication links," *IEEE Internet Things J.*, pp. 1–1, 2024.

[6] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, Jun. 2015.

[7] Z. Mobini, M. Mohammadi, and C. Tellambura, "Wireless-powered full-duplex relay and friendly jamming for secure cooperative communications," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 621–634, Mar. 2019.

[8] X. Zhou, B. Maham, and A. Hjørungnes, "Pilot contamination for active eavesdropping," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 903–907, Mar. 2012.

[9] Q. Xiong, Y.-C. Liang, K. H. Li, and Y. Gong, "An energy-ratio-based approach for detecting pilot spoofing attack in multiple-antenna systems," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 932–940, May 2015.

[10] J. Xie, Y.-C. Liang, J. Fang, and X. Kang, "Two-stage uplink training for pilot spoofing attack detection and secure transmission," in *Proc. IEEE ICC*, May 2017, pp. 1–6.

[11] X. Zhang, D. Guo, K. An, Z. Ding, and B. Zhang, "Secrecy analysis and active pilot spoofing attack detection for multigroup multicasting cell-free massive MIMO systems," *IEEE Access*, vol. 7, pp. 57 332–57 340, Apr. 2019.

[12] N. Li, Y. Gao, K. Xu, M. Guo, N. Sha, X. Wang, and G. Wang, "Spatial sparsity-based pilot attack detection and transmission countermeasure for cell-free massive MIMO system," *IEEE Syst. J.*, vol. 17, no. 2, pp. 2065–2076, Jun. 2023.

[13] Y. Fan, X. Wang, and X. Liao, "On the secure degrees of freedom for two-user MIMO interference channel with a cooperative jammer," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5390–5402, Aug. 2019.
[14] J. Zhu, R. Schober, and V. K. Bhargava, "Secure transmission in multicell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4766–4781, Sep. 2014.
[15] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inf. Forensics Security*, vol. 62, no. 7, pp. 3880–3900, Jul. 2016.
[16] N.-P. Nguyen, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and K. Tourki, "Secure massive MIMO with the artificial noise-aided downlink training," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 802–816, Apr. 2018.
[17] W. Xu, B. Li, L. Tao, and W. Xiang, "Artificial noise assisted secure transmission for uplink of massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6750–6762, Jul. 2021.
[18] J. Chen, X. Chen, W. H. Gerstacker, and D. W. K. Ng, "Resource allocation for a massive MIMO relay aided secure communication," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1700–1711, Aug. 2016.
[19] K. Guo, Y. Guo, and G. Ascheid, "Security-constrained power allocation in MU-massive-MIMO with distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8139–8153, Dec. 2016.
[20] M. Li, G. Ti, and Q. Liu, "Secure beamformer designs in MU-MIMO systems with multiuser interference exploitation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8288–8301, Sep. 2018.
[21] F. Zhu, F. Gao, H. Lin, S. Jin, J. Zhao, and G. Qian, "Robust beamforming for physical layer security in BDMA massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 775–787, Apr. 2018.
[22] Z. Lin, M. Lin, B. Champagne, W.-P. Zhu, and N. Al-Dhahir, "Secure and energy efficient transmission for RSMA-based cognitive satellite-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 251–255, Feb. 2021.
[23] S. Timilsina, D. Kudathanthirige, and G. Amarasuriya, "Physical layer security in cell-free massive MIMO," in *Proc. IEEE GLOBECOM*, Dec. 2018, pp. 1–7.
[24] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive MIMO networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724–4737, Oct. 2018.
[25] X. Zhang, D. Guo, K. An, and B. Zhang, "Secure communications over cell-free massive MIMO networks with hardware impairments," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1909–1920, Jun. 2020.
[26] M. Alageli, A. Ikhlef, F. Alsifiany, M. A. M. Abdullah, G. Chen, and J. Chambers, "Optimal downlink transmission for cell-free SWIPT massive MIMO systems with active eavesdropping," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1983–1998, Nov. 2019.
[27] S. Elhoushy and W. Hamouda, "Nearest APs-based downlink pilot transmission for high secrecy rates in cell-free massive MIMO," in *Proc. IEEE GLOBECOM*, Dec. 2020, pp. 1–6.
[28] X. Zhang, T. Liang, K. An, G. Zheng, and S. Chatzinotas, "Secure transmission in cell-free massive MIMO with RF impairments and low-resolution ADCs/DACs," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8937–8949, Sep. 2021.
[29] Y. Zhang, W. Xia, G. Zheng, H. Zhao, L. Yang, and H. Zhu, "Secure transmission in cell-free massive MIMO with low-resolution DACs over Rician fading channels," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2606–2621, Apr. 2022.
[30] H. Q. Ngo, H. Tataria, M. Matthaiou, S. Jin, and E. G. Larsson, "On the performance of cell-free massive MIMO in Ricean fading," in *Proc. IEEE Asilomar Conf. Signals, Systems, and Computers*, Nov. 2018, pp. 980–984.
[31] H. A. Ammar and R. Adve, "Power delay profile in coordinated distributed networks: User-centric v/s disjoint clustering," in *Proc. IEEE GlobalSIP*, Nov. 2019, pp. 1–5.
[32] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.
[33] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, "Structured massive access for scalable cell-free massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1086–1100, Apr. 2021.
[34] 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.
[35] T. Marzetta, E. Larsson, Y. Hong, and H. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[36] M. Mohammadi, Z. Mobini, D. Galappaththige, and C. Tellambura, "A comprehensive survey on full-duplex communication: Current solutions, future trends, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2190–2244, Fourthquarter 2023.
[37] R. Miller and W. Trappe, "On the vulnerabilities of csi in $mimo$ wireless communication systems," *IEEE Trans. Mob. Comput.*, vol. 11, no. 8, pp. 1386–1398, Aug. 2012.
[38] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.
[39] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Secrecy rate beamforming for multicell networks with information and energy harvesting," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 677–689, Feb. 2017.
[40] H. H. M. Tam, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 872–884, Feb. 2017.
[41] X. Qiu, T. Jiang, S. Wu, and M. Hayes, "Physical layer authentication enhancement using a gaussian mixture model," *IEEE Access*, vol. 6, pp. 53 583–53 592, Sep. 2018.
[42] X. Tian, M. Li, and Q. Liu, "Random-training-assisted pilot spoofing detection and security enhancement," *IEEE Access*, vol. 5, pp. 27 384–27 399, Dec. 2017.
[43] Q. Xiong, Y.-C. Liang, K. H. Li, Y. Gong, and S. Han, "Secure transmission against pilot spoofing attack: A two-way training-based scheme," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 5, pp. 1017–1026, May 2016.
[44] S. Im, H. Jeon, J. Choi, and J. Ha, "Secret key agreement with large antenna arrays under the pilot contamination attack," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6579–6594, Dec. 2015.
[45] X. Liu, Y. Tao, C. Zhao, and Z. Sun, "Detect pilot spoofing attack for intelligent reflecting surface assisted systems," *IEEE Access*, vol. 9, pp. 19 228–19 237, Feb. 2021.
[46] W. Xu, S. Xu, and B. Li, "Detection of pilot spoofing attack in massive MIMO systems," in *Proc. IEEE ICC*, Jul. 2019, pp. 1–6.
[47] D. Kapetanović, G. Zheng, K.-K. Wong, and B. Ottersten, "Detection of pilot contamination attack using random training and massive MIMO," in *Proc. IEEE PIMRC*, Sep. 2013, pp. 13–18.
[48] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "QuaDRiGa - quasi deterministic radio channel generator, user manual and documentation," Fraunhofer Heinrich Hertz Institute., Tech. Rep., 2016.

Yasseen Sadoon Atiya received the B.Sc degree in electrical engineering from the University of Babylon, Babylon, in 2007 and the M.Sc degree in electronics and communication from the University of Baghdad, Baghdad, in 2012. Currently, he is pursuing his Ph.D. at the Centre for Wireless Innovation (CWI), Queen's University Belfast. His research focuses on physical-layer security and cell-free massive MIMO systems.

Zahra Mobini received the B.S. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2006, and the M.S and Ph.D. degrees, both in electrical engineering, from the M. A. University of Technology and K. N. Toosi University of Technology, Tehran, Iran, respectively. From November 2010 to November 2011, she was a Visiting Researcher at the Research School of Engineering, Australian National University, Canberra, ACT, Australia. She is currently a Post-Doctoral Research Fellow at the Centre for Wireless Innovation (CWI), Queen's University Belfast (QUB). Before joining QUB, she was an Assistant and then Associate Professor with the Faculty of Engineering, Shahrekord University, Shahrekord, Iran (2015-2021). Her research interests include physical-layer security, massive MIMO, cell-free massive MIMO, full-duplex communications, and resource management and optimization. She has co-authored many research papers in wireless communications. She has actively served as the reviewer for a variety of IEEE journals, such as TWC, TCOM, and TVT.

Hien Quoc Ngo is currently a Reader with Queen's University Belfast, U.K. His main research interests include massive MIMO systems, cell-free massive MIMO, reconfigurable intelligent surfaces, physical layer security, and cooperative communications. He has co-authored many research papers in wireless communications and co-authored the Cambridge University Press textbook *Fundamentals of Massive MIMO* (2016).

He received the IEEE ComSoc Stephen O. Rice Prize in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, the Best Ph.D. Award from EURASIP in 2018, and the IEEE CTTC Early Achievement Award in 2023. He also received the IEEE Sweden VT-COM-IT Joint Chapter Best Student Journal Paper Award in 2015. He was awarded the UKRI Future Leaders Fellowship in 2019. He serves as the Editor for the IEEE Transactions on Wireless Communications, IEEE Transactions on Communications, the Digital Signal Processing, and the Physical Communication (Elsevier). He was a Guest Editor of IET Communications, and a Guest Editor of IEEE ACCESS in 2017.

Michail Matthaiou (Fellow, IEEE) obtained his Ph.D. degree from the University of Edinburgh, U.K. in 2008. He is currently a Professor of Communications Engineering and Signal Processing and Deputy Director of the Centre for Wireless Innovation (CWI) at Queen's University Belfast, U.K. He has also held research/faculty positions at Munich University of Technology (TUM), Germany and Chalmers University of Technology, Sweden. His research interests span signal processing for wireless communications, beyond massive MIMO, reflecting intelligent surfaces, mm-wave/THz systems and AI-empowered communications.

Dr. Matthaiou and his coauthors received the IEEE Communications Society (ComSoc) Leonard G. Abraham Prize in 2017. He currently holds the ERC Consolidator Grant BEATRICE (2021-2026) focused on the interface between information and electromagnetic theories. To date, he has received the prestigious 2023 Argo Network Innovation Award, the 2019 EURASIP Early Career Award and the 2018/2019 Royal Academy of Engineering/The Leverhulme Trust Senior Research Fellowship. His team was also the Grand Winner of the 2019 Mobile World Congress Challenge. He was the recipient of the 2011 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East and Africa Region and a co-recipient of the 2006 IEEE Communications Chapter Project Prize for the best M.Sc. dissertation in the area of communications. He has co-authored papers that received best paper awards at the 2018 IEEE WCSP and 2014 IEEE ICC. In 2014, he received the Research Fund for International Young Scientists from the National Natural Science Foundation of China. He is currently the Editor-in-Chief of Elsevier Physical Communication, a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS and IEEE SIGNAL PROCESSING MAGAZINE, and an Area Editor for IEEE TRANSACTIONS ON COMMUNICATIONS. He is an IEEE and AAIA Fellow.