

Analysis and Optimization of Massive Access to the IoT Relying on Multi-Pair Two-Way Massive MIMO Relay Systems

Zhangjie Peng, Xianzhe Chen, Wei Xu, *Senior Member, IEEE*, Cunhua Pan,
Li-Chun Wang, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—We investigate massive access in the Internet-of-Things (IoT) relying on multi-pair two-way amplify-and-forward (AF) relay systems using massive multiple-input multiple-output (MIMO). We utilize the approximate message passing (AMP) algorithm for joint device activity detection and channel estimation. Furthermore, we analyze the achievable rates for multiple pairs of active devices and derive the closed-form expressions for both maximum-ratio combining/maximum-ratio transmission (MRC/MRT) and zero-forcing reception/zero-forcing transmission (ZFR/ZFT)-based beamforming schemes adopted at the relay. Moreover, to improve the achievable sum rates, we propose a low-complexity algorithm for optimizing the pilot length L . Our simulation results verify the accuracy of the closed-form expressions of the MRC/MRT and ZFR/ZFT scenarios. Finally, the proposed pilot-length optimization algorithm performs well in both the MRC/MRT and ZFR/ZFT scenarios.

Index Terms—Massive access, Internet-of-Things (IoT), massive multiple-input multiple-output (MIMO), amplify-and-forward relay, two-way, beamforming.

This work was supported in part by the NSFC under Grant 61701307, the open research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2018D14, and the Natural Science Foundation of Shanghai under Grant 18ZR1428000. The work of Wei Xu was supported by the National Natural Science Foundation of China under Grants 62022026 and 61871109 as well as by the Natural Science Foundation of Jiangsu Province under Grants BK20190012. The work of Li-Chun Wang was supported by the Ministry of Science and Technology under the Grants MOST 110-2634-F-009-021 and 108-2911-I-009-517 through Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan. The work of Lajos Hanzo was supported by the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as by the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). (*Corresponding author: Wei Xu*)

Z. Peng is with the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, Shanghai Engineering Research Center of Intelligent Education and Bigdata, Shanghai Normal University, Shanghai 200234, China (e-mail: pengzhangjie@shnu.edu.cn).

X. Chen is with the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, (China e-mail: 1000479050@smail.shnu.edu.cn).

W. Xu is with the National Mobile Communications Research Lab, Southeast University, Nanjing 210096, China, and also with Henan Joint International Research Laboratory of Intelligent Networking and Data Analysis, Zhengzhou University, Zhengzhou, 450001 China (e-mail: wxu@seu.edu.cn).

C. Pan is with the School of Electronic Engineering and Computer Science at Queen Mary University of London, London E1 4NS, U.K. (e-mail: c.pan@qmul.ac.uk).

L.-C. Wang is with the Department of Electrical Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan (e-mail: lichun@cc.nctu.edu.tw).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: l.h@ecs.soton.ac.uk).

I. INTRODUCTION

In massive machine-type communications (mMTC) in the Internet-of-Things (IoT), each cell usually supports numerous devices, but in a coherence time interval, only few of them are active [1]–[3]. In these scenarios, device activity detection (DAD) and channel estimation (CE) constitute challenging issues. When supporting numerous devices, the assignment of long orthogonal device-specific pilot sequences would impose an excessive pilot-overhead. Hence, researchers have put forward the non-orthogonal multiple access (NOMA) to make full use of the available bandwidth. In these grant-free NOMA schemes, all devices are assigned non-orthogonal pilot sequences. Then, the active devices simultaneously transmit their pilot signals to the base station, which performs joint DAD and CE.

Due to the sparse nature of access in the IoT, DAD and CE can be transformed into a compressed sensing problem. In recent years, researchers have studied many algorithms for DAD and CE, such as those based on mean-field message passing [4] or Gaussian message passing [5], [6]. However, these algorithms do not have a rigorous performance analysis, which is usually required by researchers to study the system performance. To solve this problem, Liu and Yu [7], [8] utilized the approximate message passing (AMP) algorithm for joint DAD and CE in massive multiple-input multiple-output (MIMO) systems, and they analyzed the statistical characteristics of the system. The AMP algorithm has been studied by many researchers [9]–[15]. These studies proved that the AMP algorithm has a remarkable capability of balancing its performance by complexity trade-off.

Given the high number of potential devices in massive access, there is an urgent need to improve the system's capacity and expand its coverage. Relay-aided transmissions have been shown to be a promising technique of extending coverage area and increasing the network capacity [16]–[18]. Recently, researchers have devoted considerable efforts to integrating massive MIMO solutions into multi-pair relaying for improving the system performance [19]–[21]. Gao et al. [22] investigated a one-way multi-pair massive MIMO relay system. In a multi-pair relay system using massive MIMO, multiple pairs of users exchange their information through a relay station, which is equipped with numerous antennas. Furthermore, researchers also conceived a two-way relay system, where two communication nodes engage in simultane-

TABLE I
A BRIEF COMPARISON OF THE RELATED LITERATURE

	[7]	[8]	[15]	[27-29]	Our work
Massive access	✓	✓	✓		✓
Massive MIMO	✓	✓	✓	✓	✓
Multi-pair				✓	✓
Two-way relaying				✓	✓
Achievable rate analysis		✓	✓	✓	✓
Pilot length optimization		✓	✓		✓

ous bidirectional data transmission, which can theoretically double the bandwidth efficiency (BE) compared to one-way relaying [23]–[26]. The authors of [27] and [28] studied the power-scaling laws of the multi-pair two-way massive MIMO (MTMM) relay systems, and the authors of [29] put forward a beneficial power allocation scheme to maximize the BE. The above studies show that the MTMM relay substantially increases the coverage and improves the system capacity. In a nutshell, Table I boldly and explicitly contrasts our contributions to the existing literature.

Hence, we are inspired to use an MTMM relay for massive access in the IoT, where numerous devices are supported, but only a small fraction of them are active during a coherence time. Active devices at both sides of the relay establish stable bidirectional communications through a relay station employing numerous antennas. A two-phase transmission protocol is adopted. We utilize the AMP algorithm to perform joint DAD and CE in the first phase. Then, in the second phase, the active devices communicate with their paired devices through the relay. Furthermore, we also study the pilot length optimization problem. The main contributions of this paper are summarized as follows:

- We firstly study massive access in the IoT assisted by the MTMM relay adopting both MRC/MRT and ZFR/ZFT beamforming schemes. Furthermore, we use the AMP algorithm for joint DAD and CE in this system.
- We derive the corresponding closed-form achievable rate expressions for the above linear processing techniques for multiple pairs of active devices.
- We propose a low-complexity optimization algorithm to obtain the optimal pilot lengths for joint DAD and CE by maximizing the achievable rates for both MRC/MRT and ZFR/ZFT beamforming schemes. We prove that these pair of optimization problems can be approximated by convex optimization problems. Then, we propose a binary search algorithm for solving these problems, and show that this optimization algorithm has a low complexity. Our simulation results verify that the proposed optimization algorithm performs well for these two linear processing techniques.

The rest of this paper is organized as follows. Section II briefly introduces the MTMM relay system applied in a massive access scenario. Section III derives the closed-form expressions for the achievable rates in this access scenario, when the relay employs both MRC/MRT and ZFR/ZFT processing techniques. Section IV solves pilot length optimization

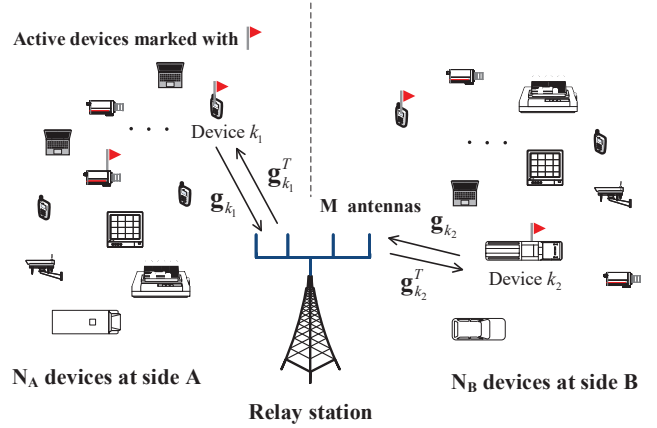


Fig. 1. System Model

problems, and proposes an algorithm for solving these problems. The main results of this paper are verified by simulation results in Section V. Finally, brief conclusions are provided in Section VI.

Notations: In this paper, matrices are denoted by bold upper case letters, vectors are denoted by bold lower case letters, and scalars are denoted by lower case letters. The conjugate, transpose, conjugate-transpose, and matrix inverse operations are respectively denoted by the superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$. Euclidean 2-norm and trace operators are denoted by $\|\cdot\|$ and $\text{tr}(\cdot)$, respectively. We use $\text{E}\{\cdot\}$ and $\text{Var}\{\cdot\}$ to denote the expectation and variance operators, respectively. And $[\mathbf{A}]_{ij}$ is used to denote the (i, j) th element of matrix \mathbf{A} . The matrix \mathbf{I}_N denotes an $N \times N$ identity matrix. In addition, the distribution of the circularly symmetric complex Gaussian vector with zero mean and covariance Σ is denoted by $\mathcal{CN}(\mathbf{0}, \Sigma)$.

II. SYSTEM MODEL

We investigate an MTMM relay system in the massive access scenario shown in Fig. 1, where N_A IoT devices distributed at side A of the relay (\mathcal{R}) want to communicate with N_B IoT devices distributed at the other side, i.e., side B. The total number of the IoT devices is $N = N_A + N_B$. We assume that the single-antenna devices at sides A and B are too far to establish direct links. Therefore, the devices at sides A and B can only exchange information via \mathcal{R} , which has M antennas. Furthermore, the relay operates in the half-duplex mode and under the TDD protocol. The vector $\mathbf{g}_n \in \mathbb{C}^{M \times 1}$ denotes the channel spanning from device n to \mathcal{R} , and \mathbf{g}_n obeys the i.i.d. $\mathcal{CN}(\mathbf{0}, \beta_n \mathbf{I}_M)$ with β_n being the path loss. We assume TDD-based channel reciprocity. Then, the downlink (DL) channel from \mathcal{R} to device n (\mathcal{D}_n) can be expressed as \mathbf{g}_n^T [29].

During the coherence interval, the active devices make up only a very small part of the total number of devices, since the IoT devices generate bursts of information [7]. For convenience, we introduce α_n as the device activity factor with $\alpha_n = 1$, if \mathcal{D}_n is active, otherwise, $\alpha_n = 0$. Then for $\forall n = 1, 2, \dots, N$, we have

$$\begin{cases} \Pr(\alpha_n = 1) = \theta, \mathcal{D}_n \text{ is active,} \\ \Pr(\alpha_n = 0) = 1 - \theta, \text{ otherwise,} \end{cases} \quad (1)$$

where θ stands for the activity probability of each device. Furthermore, in a coherence time, we use the set $\mathcal{K} = \{n : \alpha_n = 1\}$ to represent the collection of all active devices, and assume the number of active devices is $|\mathcal{K}| = \bar{K}$. Additionally, the sets of active devices at sides A and B are respectively denoted by \mathcal{K}_A and \mathcal{K}_B .

In this paper, we divide each coherence interval of T symbol durations into two phases, one for DAD and CE, while the other for data transmission.

A. Device Activity Detection and Channel Estimation [7]

For our massive access scenario, we adopt a grant-free NOMA scheme, since the number of potential devices is larger than the pilot length, i.e., $N > L$. Hence it is not feasible to have unique device-specific orthogonal pilots. For the DAD and CE phase, different pilot sequences are assigned to each device n , which is denoted as $\mathbf{a}_n = [a_{n,1}, a_{n,2}, \dots, a_{n,L}]^T \in \mathbb{C}^{L \times 1}$ with entries generated from the i.i.d. $\mathcal{CN}(0, \frac{1}{L})$. At the start of this phase, the active devices simultaneously transmit their pilot sequences to \mathcal{R} . Then, the received pilot signal $\mathbf{Y} \in \mathbb{C}^{L \times M}$ at \mathcal{R} is given by

$$\mathbf{Y} = \sum_{n=1}^N \alpha_n \sqrt{\xi_n} \mathbf{a}_n \mathbf{g}_n^T + \mathbf{Z} \triangleq \mathbf{A}\mathbf{X} + \mathbf{Z}, \quad (2)$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] \in \mathbb{C}^{L \times N}$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T \in \mathbb{C}^{N \times M}$ with $\mathbf{x}_n = \alpha_n \sqrt{\xi_n} \mathbf{g}_n \in \mathbb{C}^{M \times 1}$, $\xi_n = L p_{p,n}$ is the transmit energy with pilot power $p_{p,n}$, and \mathbf{Z} is the additive white Gaussian noise (AWGN) with entries obeying the i.i.d. $\mathcal{CN}(0, \sigma_R^2)$. Because very few of the devices are active, the row vector \mathbf{x}_n is sparse, and it obeys a Bernoulli Gaussian distribution:

$$p_{\mathbf{x}_n} = (1 - \theta) \delta_0 + \theta p_n, \quad (3)$$

where p_n denotes the distribution of vector $\sqrt{\xi_n} \mathbf{g}_n \sim \mathcal{CN}(\mathbf{0}, \xi_n \beta_n \mathbf{I}_M)$, and β_n is the path loss exponent defined before. Additionally, δ_0 denotes the point mass measure at zero.

Then, to carry out DAD and CE, we have to recover the sparse signal \mathbf{X} based on the received pilot signal \mathbf{Y} , which can be modeled as a compressed sensing problem. For this kind of problems, the AMP algorithm has been proved to be an efficient solution [7], [13], [30]. According to [7], [30], when the entries of the sensing matrix \mathbf{A} are generated according to the i.i.d. Gaussian distribution, the performance of the AMP algorithm can be predicted by the state evolution in the asymptotic regime, where we have $L, \bar{K}, N \rightarrow \infty$ with their ratios converging to some fixed positive values.

We adjust the AMP algorithm of [7] to perform joint DAD and CE. When the device k is detected to be active, we can obtain its corresponding channel estimate at the same time as

$$\hat{\mathbf{g}}_k = \frac{\mathbf{x}_k^t}{\sqrt{\xi_k}}, \quad (4)$$

where \mathbf{x}_k^t is the estimate of \mathbf{x}_1 at iteration t of the AMP algorithm. Then the channel estimation error is given by $\mathbf{e}_k = \mathbf{g}_k - \hat{\mathbf{g}}_k$. Moreover, as M becomes large, the covariance matrices of $\hat{\mathbf{g}}_k$ and \mathbf{e}_k can be respectively formulated as [8],

[15]

$$\text{cov}(\hat{\mathbf{g}}_k, \hat{\mathbf{g}}_k) = \frac{\beta_k^2 \xi_k}{\beta_k \xi_k + \tau_\infty^2} \mathbf{I}_M, \quad (5)$$

$$\text{cov}(\mathbf{e}_k, \mathbf{e}_k) = \frac{\beta_k \tau_\infty^2}{\beta_k \xi_k + \tau_\infty^2} \mathbf{I}_M, \quad (6)$$

where $\text{cov}(\cdot, \cdot)$ denotes the covariance operator. This means that the vectors $\hat{\mathbf{g}}_k$ and \mathbf{e}_k respectively obey $\hat{\mathbf{g}}_k \sim \mathcal{CN}(\mathbf{0}, \frac{\beta_k^2 \xi_k}{\beta_k \xi_k + \tau_\infty^2} \mathbf{I}_M)$ and $\mathbf{e}_k \sim \mathcal{CN}(\mathbf{0}, \frac{\beta_k \tau_\infty^2}{\beta_k \xi_k + \tau_\infty^2} \mathbf{I}_M)$, as M becomes large. In (5) and (6), the scalar τ_∞^2 is the solution converged to the state evolution of the AMP algorithm. In the high signal-to-noise ratio (SNR) regions, i.e., $\frac{L p_{p,k} \beta}{\sigma_R^2} \rightarrow \infty$, when $L > \bar{K}$, τ_∞^2 obeys [8], [15]

$$\tau_\infty^2 \rightarrow \frac{L \sigma_R^2}{L - \bar{K}}. \quad (7)$$

B. Data Transmission

In this scenario, we assume that there are $K_A = |\mathcal{K}_A|$ active devices at side A as well as $K_B = |\mathcal{K}_B|$ active devices at side B in a coherence time. We consider the case where $K = \min(K_A, K_B)$ pairs of active devices establish communication with each other via the relay. When $K_A \neq K_B$, we can use some user scheduling schemes to select the same number of active devices at each side to communicate at each time [8], [31]. Therefore, we assume $K_A = K_B = K$ for convenience, and we have $\bar{K} = 2K$ in our following paper.

In the data transmission phase, there are basically two stages. In the first stage, the K pairs of active devices simultaneously transmit their signals to \mathcal{R} . The signal received by \mathcal{R} can be expressed as

$$\mathbf{y}_r = \sum_{n \in \mathcal{K}} \mathbf{g}_n \sqrt{p_{d,n}} s_n + \mathbf{n}_R, \quad (8)$$

where the vector $\mathbf{n}_R \sim \mathcal{CN}(\mathbf{0}, \sigma_R^2 \mathbf{I}_M)$ is the AWGN at \mathcal{R} , s_n represents the signal transmitted from device n with unit norm, and $p_{d,n}$ is its corresponding transmit power.

In the second stage, the received signal is processed by the linear processing technique at \mathcal{R} , and then it is transmitted to K pairs of active devices. The signal transmitted from \mathcal{R} can be expressed as

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_r = \rho \mathbf{W} \left(\sum_{n \in \mathcal{K}} \mathbf{g}_n \sqrt{p_{d,n}} s_n + \mathbf{n}_R \right), \quad (9)$$

where \mathbf{W} is the beamforming matrix at \mathcal{R} , which is a function of the channel estimate, and ρ is chosen to satisfy the transmit power constraint $\mathbb{E} \{ \|\mathbf{y}_t\|^2 \} = p_r$ at \mathcal{R} . Therefore, the signal received by the active device k can be expressed as

$$\begin{aligned} y_k &= \underbrace{\rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \sqrt{p_{d,k'}} s_{k'}}_{\text{desired signal}} + \underbrace{\rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_k \sqrt{p_{d,k}} s_k}_{\text{self-interference}} \\ &+ \underbrace{\sum_{n \in \mathcal{K}/\{k, k'\}} \rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_n \sqrt{p_{d,n}} s_n}_{\text{inter-pair interference}} + \underbrace{\rho \mathbf{g}_k^T \mathbf{W} \mathbf{n}_R}_{\text{noise from the relay}} + n_k, \end{aligned} \quad (10)$$

where (k, k') represents a pair of active devices, and $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the AWGN noise at the device k . From (10),

we note that the received signal y_k consists of several terms. The desired signal is the signal transmitted from the active device k' . The self-interference represents the signal which is first transmitted from the active device k to the relay, then retransmitted from the relay to the active device k itself. The inter-pair interference represents the signals transmitted from other active devices pairs. The noise contributed by the relay is generated when the relay receives signals and it is amplified, when the relay retransmits the received signals.

Based on the channel estimate, the relay transmits the self-interference cancellation (SIC) coefficient $\hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_k$ for the active device k . Therefore, the signal received at the active device k after the SIC is written as

$$\begin{aligned} \tilde{y}_k = & \underbrace{\rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \sqrt{p_{d,k'}} s_{k'}}_{\text{desired signal}} + \underbrace{\rho \lambda_k \sqrt{p_{d,k}} s_k}_{\text{residual self-interference}} \\ & + \underbrace{\sum_{n \in \mathcal{K}/\{k,k'\}} \rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_n \sqrt{p_{d,n}} s_n}_{\text{inter-pair interference}} + \underbrace{\rho \mathbf{g}_k^T \mathbf{W} \mathbf{n}_R}_{\text{noise from the relay}} + n_k, \end{aligned} \quad (11)$$

where $\lambda_k = \mathbf{g}_k^T \mathbf{W} \mathbf{g}_k - \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_k$.

C. Linear Processing Techniques

We consider a pair of linear processing techniques at the relay, i.e., MRC/MRT and ZFR/ZFT, similar to those in [29]. Apart from these two techniques, LMMSE processing technique is an effective scheme, which has been studied by many researchers [32]. However, our system is a more complicated one. In fact, there is even no work on applying LMMSE scheme to an MTMM relay system. Hence, it is challenging to apply this scheme in the MTMM relay system. We will consider that in the future study.

1) *MRC/MRT Processing*: The MRC/MRT beamforming matrix is given by

$$\mathbf{W} = \hat{\mathbf{G}}^* \mathbf{P} \hat{\mathbf{G}}^H, \quad (12)$$

where $\hat{\mathbf{G}} \in \mathbb{C}^{M \times 2K}$ is the matrix of channel estimates. The first K columns of $\hat{\mathbf{G}}$ represent the channel estimates of the K active devices at side A of Fig. 1, while the last K columns of that represent those of the corresponding K active devices at side B, and $\mathbf{P} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_K \\ \mathbf{I}_K & \mathbf{0} \end{bmatrix}$ is the mapping matrix. Then, the scalar ρ of the MRC/MRT processing technique is expressed as

$$\rho = \sqrt{\frac{p_r}{\sum_{n \in \mathcal{K}} p_{d,n} \mathbb{E} \left\{ \|\mathbf{W} \mathbf{g}_n\|^2 \right\} + \mathbb{E} \left\{ \|\mathbf{W} \mathbf{n}_R\|^2 \right\}}} \triangleq \rho_{\text{mr}}. \quad (13)$$

Lemma 1: The closed-form expression of ρ_{mr} is formulated

as

$$\rho_{\text{mr}} = \sqrt{\frac{p_r}{\sum_{n \in \mathcal{K}} p_{d,n} \left(\frac{M(M+1)^2 \beta_n^4 \beta_n'^2 \xi_n^2 \xi_n'}{(\beta_n \xi_n + \tau_\infty)^2 (\beta_n' \xi_n' + \tau_\infty)} + \beta_n \eta \right) + \sigma_R^2 \eta}}, \quad (14)$$

where we have $\eta = \sum_{j \in \mathcal{K}_A} \frac{2M(M+1) \beta_j^2 \beta_j'^2 \xi_j \xi_j'}{(\beta_j \xi_j + \tau_\infty)(\beta_j' \xi_j' + \tau_\infty)}$.

Proof: See Appendix A. ■

2) *ZFR/ZFT Processing*: The ZFR/ZFT beamforming matrix is given by

$$\mathbf{W} = \hat{\mathbf{G}}^* \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \hat{\mathbf{G}}^H. \quad (15)$$

For convenience, we introduce the matrices $\bar{\mathbf{G}}$, $\bar{\mathbf{P}}$ and the vector $\bar{\mathbf{s}}$, where $\bar{\mathbf{G}} \in \mathbb{C}^{M \times 2K}$ represents the real channel matrix of the active devices corresponding to $\hat{\mathbf{G}}$. Furthermore, the matrix $\bar{\mathbf{P}} \in \mathbb{C}^{2K \times 2K}$ is a diagonal matrix, and its diagonal elements represent the square roots of the transmission power of the active devices. Finally, the elements of the vector $\bar{\mathbf{s}} \in \mathbb{C}^{2K \times 1}$ represent the unit-norm signals of the active devices. Substituting (15) into (9) and combining it with the transmit power constraint, the scalar ρ of the ZFR/ZFT processing technique is formulated as

$$\rho = \sqrt{\frac{p_r}{\mathbb{E} \left\{ \|\mathbf{W} \bar{\mathbf{G}} \bar{\mathbf{P}} \bar{\mathbf{s}}\|^2 \right\} + \mathbb{E} \left\{ \|\mathbf{W} \mathbf{n}_R\|^2 \right\}}} \triangleq \rho_{\text{zf}}. \quad (16)$$

Lemma 2: The closed-form expression of ρ_{zf} is given by

$$\rho_{\text{zf}} = \sqrt{\frac{p_r}{\sum_{n \in \mathcal{K}} \left(\frac{p_{d,n'} (\beta_n \xi_n + \tau_\infty)}{(M-2K-1) \beta_n^2 \xi_n} + \frac{\gamma p_{d,n} \beta_n \tau_\infty^2}{\beta_n \xi_n + \tau_\infty} \right) + \sigma_R^2 \gamma}}, \quad (17)$$

where we have $\gamma = \sum_{j \in \mathcal{K}} \frac{2(\beta_j \xi_j + \tau_\infty)(\beta_j' \xi_j' + \tau_\infty)}{(M-2K)(M-2K-3) \beta_j^2 \beta_j'^2 \xi_j \xi_j'}$.

Proof: See Appendix B. ■

III. RATE ANALYSIS FOR LARGE-SCALE IOT ACCESS

In this section, we aim for investigating the achievable rates of the active devices. Explicitly, we derive the closed-form expressions of the achievable rates for MRC/MRT and ZFR/ZFT processing adopted at the relay.

Given (11), the achievable rate of the active device k can be formulated as (18) at the bottom of this page. Then, we use a similar method to that in [33], [34] for deriving the lower bound of the achievable rate at the active device k for further insightful analysis. Therefore, \tilde{y}_k in (11) can be re-written as

$$\tilde{y}_k = \rho \mathbb{E} \left\{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right\} \sqrt{p_{d,k'}} s_{k'} + \underbrace{\tilde{n}_k}_{\text{effective noise}}, \quad (19)$$

$$\bar{R}_k = \frac{T-L}{2T} \mathbb{E} \left\{ \log_2 \left(1 + \frac{p_{d,k'} |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'}|^2}{p_{d,k} |\lambda_k|^2 + \sum_{n \in \mathcal{K}/\{k,k'\}} p_{d,n} |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_n|^2 + |\mathbf{g}_k^T \mathbf{W} \mathbf{n}_R|^2 + \frac{1}{\rho^2} |n_k|^2} \right) \right\}. \quad (18)$$

where the effective noise \tilde{n}_k is

$$\begin{aligned} \tilde{n}_k &= \rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \sqrt{p_{d,k'} s_{k'}} - \rho \mathbf{E} \{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \} \sqrt{p_{d,k'} s_{k'}} \\ &+ \rho \lambda_k \sqrt{p_{d,k} s_k} + \sum_{n \in \mathcal{K}/\{k,k'\}} \rho \mathbf{g}_k^T \mathbf{W} \mathbf{g}_n \sqrt{p_{d,n} s_n} \\ &+ \rho \mathbf{g}_k^T \mathbf{W} \mathbf{n}_R + n_k. \end{aligned} \quad (20)$$

Based on (19) and (20), the lower bound of the achievable rate of the active device k can be obtained as

$$R_k = \frac{T-L}{2T} \log_2 \left(1 + \frac{DS_k}{GU_k + SI_k + IP_k + NR_k + ND_k} \right), \quad (21)$$

with DS_k , GU_k , SI_k , IP_k , NR_k , and ND_k representing respectively the desired signal, the gain uncertainty, the residual self-interference after SIC, the inter-pair interference, the amplified noise emanating from the relay and the noise at the device k , which are expressed as

$$DS_k = p_{d,k'} \mathbf{E} \{ |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'}|^2 \}, \quad (22)$$

$$GU_k = p_{d,k'} \text{Var} \{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \}, \quad (23)$$

$$SI_k = p_{d,k} \mathbf{E} \{ |\lambda_k|^2 \}, \quad (24)$$

$$IP_k = \sum_{n \in \mathcal{K}/\{k,k'\}} p_{d,n} \mathbf{E} \{ |\mathbf{g}_k^T \mathbf{W} \mathbf{g}_n|^2 \}, \quad (25)$$

$$NR_k = \mathbf{E} \{ |\mathbf{g}_k^T \mathbf{W} \mathbf{n}_R|^2 \}, \quad (26)$$

$$ND_k = \frac{1}{\rho^2} \mathbf{E} \{ |n_k|^2 \}. \quad (27)$$

A. MRC/MRT Processing

When the MRC/MRT beamforming matrix is adopted at the relay, based on (21), we derive a closed-form expression for the achievable rate of the active device k in the following theorem.

Theorem 1: In the multi-pair massive access scenario using MRC/MRT processing, when $L > 2K$, the SNR is high and M tends to infinity, the achievable rate of the active device k is lower bounded by

$$R_k^{\text{mr}} = \frac{T-L}{2T} \log_2 \left(1 + \frac{DS_k^{\text{mr}}}{GU_k^{\text{mr}} + SI_k^{\text{mr}} + IP_k^{\text{mr}} + NR_k^{\text{mr}} + ND_k^{\text{mr}}} \right), \quad (28)$$

where the terms DS_k^{mr} , GU_k^{mr} , SI_k^{mr} , IP_k^{mr} , NR_k^{mr} , and ND_k^{mr} are expressed in the following equations. And in (29)-(33), the value of the term τ_∞^2 equals to $\frac{L\sigma_R^2}{L-2K}$, i.e., $\tau_\infty^2 = \frac{L\sigma_R^2}{L-2K}$.

$$DS_k^{\text{mr}} = \frac{p_{d,k'} M^2 (M+1)^2 \beta_k^4 \beta_{k'}^4 \xi_k^2 \xi_{k'}^2}{(\beta_k \xi_k + \tau_\infty^2)^2 (\beta_{k'} \xi_{k'} + \tau_\infty^2)^2}, \quad (29)$$

$$\begin{aligned} GU_k^{\text{mr}} &= p_{d,k'} \beta_k \beta_{k'} \eta + \frac{p_{d,k'} M(M+1)^2 \beta_k^3 \beta_{k'}^3 \xi_k \xi_{k'}}{(\beta_k \xi_k + \tau_\infty^2) (\beta_{k'} \xi_{k'} + \tau_\infty^2)} \\ &\times \left(\frac{\beta_k \xi_k}{\beta_k \xi_k + \tau_\infty^2} + \frac{\beta_{k'} \xi_{k'}}{\beta_{k'} \xi_{k'} + \tau_\infty^2} \right), \end{aligned} \quad (30)$$

$$\begin{aligned} SI_k^{\text{mr}} &= \frac{p_{d,k} \beta_k \tau_\infty^2}{\beta_k \xi_k + \tau_\infty^2} \left(\frac{4M(M+1)^2 \beta_k^4 \beta_{k'}^2 \xi_k^2 \xi_{k'}}{(\beta_k \xi_k + \tau_\infty^2)^2 (\beta_{k'} \xi_{k'} + \tau_\infty^2)} \right. \\ &\left. + \frac{2\beta_k \tau_\infty^2 \eta}{\beta_k \xi_k + \tau_\infty^2} + \frac{4\beta_k^2 \xi_k \eta}{\beta_k \xi_k + \tau_\infty^2} \right), \end{aligned} \quad (31)$$

$$\begin{aligned} IP_k^{\text{mr}} &= \sum_{n \in \mathcal{K}/\{k,k'\}} \left(\frac{M(M+1)^2 p_{d,n} \beta_n \beta_k^4 \beta_{k'}^2 \xi_k^2 \xi_{k'}}{(\beta_k \xi_k + \tau_\infty^2)^2 (\beta_{k'} \xi_{k'} + \tau_\infty^2)} \right. \\ &\left. + \frac{M(M+1)^2 p_{d,n} \beta_k \beta_n^4 \beta_{k'}^2 \xi_n^2 \xi_{k'}}{(\beta_n \xi_n + \tau_\infty^2)^2 (\beta_{k'} \xi_{k'} + \tau_\infty^2)} + p_{d,n} \beta_n \beta_k \eta \right), \end{aligned} \quad (32)$$

$$NR_k^{\text{mr}} = \frac{M(M+1)^2 \beta_k^4 \beta_{k'}^2 \xi_k^2 \xi_{k'} \sigma_R^2}{(\beta_k \xi_k + \tau_\infty^2)^2 (\beta_{k'} \xi_{k'} + \tau_\infty^2)} + \sigma_R^2 \beta_k \eta, \quad (33)$$

$$ND_k^{\text{mr}} = \frac{1}{\rho_{\text{mr}}^2} \sigma_k^2. \quad (34)$$

Proof: See Appendix C. ■

B. ZFR/ZFT Processing

For ZFR/ZFT processing, we derive a closed-form expression for the achievable rate of the active device k in the following theorem.

Theorem 2: In the multi-pair massive access scenario using ZFR/ZFT processing, when $L > 2K$, the SNR is high and $M > 2K$ tends to infinity, the achievable rate of the active device k is lower bounded by

$$R_k^{\text{zf}} = \frac{T-L}{2T} \log_2 \left(1 + \frac{DS_k^{\text{zf}}}{GU_k^{\text{zf}} + SI_k^{\text{zf}} + IP_k^{\text{zf}} + NR_k^{\text{zf}} + ND_k^{\text{zf}}} \right), \quad (35)$$

where the terms DS_k^{zf} , GU_k^{zf} , SI_k^{zf} , IP_k^{zf} , NR_k^{zf} , and ND_k^{zf} are respectively given by the following equations. And in (37)-(40), we have $\tau_\infty^2 = \frac{L\sigma_R^2}{L-2K}$.

$$DS_k^{\text{zf}} = p_{d,k'}, \quad (36)$$

$$\begin{aligned} GU_k^{\text{zf}} &= \frac{p_{d,k'} \tau_\infty^2}{M-2K-1} \left(\frac{1}{\beta_{k'} \xi_{k'}} + \frac{1}{\beta_k \xi_k} \right) \\ &+ \frac{p_{d,k'} \beta_k \beta_{k'} \tau_\infty^4 \gamma}{(\beta_k \xi_k + \tau_\infty^2) (\beta_{k'} \xi_{k'} + \tau_\infty^2)}, \end{aligned} \quad (37)$$

$$SI_k^{\text{zf}} = \frac{p_{d,k} \beta_k^2 \tau_\infty^4 \gamma}{(\beta_k \xi_k + \tau_\infty^2)^2} + \frac{4p_{d,k} (\beta_{k'} \xi_{k'} + \tau_\infty^2) \beta_k \tau_\infty^2}{(M-2K-1) (\beta_k \xi_k + \tau_\infty^2) \beta_{k'}^2 \xi_{k'}}, \quad (38)$$

$$\begin{aligned} IP_k^{\text{zf}} &= \sum_{n \in \mathcal{K}/\{k,k'\}} \left(\frac{p_{n,d} \beta_k \beta_n \tau_\infty^4 \gamma}{(\beta_k \xi_k + \tau_\infty^2) (\beta_n \xi_n + \tau_\infty^2)} + \frac{p_{n,d} \tau_\infty^2}{(M-2K-1)} \right. \\ &\left. \times \left(\frac{(\beta_{k'} \xi_{k'} + \tau_\infty^2) \beta_n}{(\beta_n \xi_n + \tau_\infty^2) \beta_{k'}^2 \xi_{k'}} + \frac{(\beta_n \xi_n + \tau_\infty^2) \beta_k}{(\beta_k \xi_k + \tau_\infty^2) \beta_n^2 \xi_n} \right) \right), \end{aligned} \quad (39)$$

$$NR_k^{\text{zf}} = \frac{\sigma_R^2 (\beta_{k'} \xi_{k'} + \tau_\infty^2)}{(M-2K-1) \beta_{k'}^2 \xi_{k'}} + \frac{\beta_k \tau_\infty^2 \sigma_R^2 \gamma}{\beta_k \xi_k + \tau_\infty^2}, \quad (40)$$

$$ND_k^{\text{zf}} = \frac{1}{\rho_{\text{zf}}^2} \sigma_k^2, \quad (41)$$

Proof: See Appendix D. ■

For these two massive IoT access scenarios, *Theorem 1* and *Theorem 2* indicate that system performance mainly depends

on the system parameters: L , $2K$, M , $p_{p,j}$, p_r , $p_{d,j}$, and β_j , where $j \in \mathcal{K}$. It can be readily seen that the achievable rate of the active device k increases with p_r and $p_{d,k}$. By contrast, it decreases with $p_{d,j}$ when $j \neq k$, since the self-interference and the inter-pair interference increase with $p_{d,j}$. When $p_{d,j}$, $j \in \mathcal{K}$ increases, the channel estimation error decreases, so the achievable rate increases. It can also be concluded from *Theorem 1* and *Theorem 2* that the achievable rate increases with M for both MRC/MRT and ZFR/ZFT processing. Additionally, the achievable rate of the active device k apparently decreases with K . Finally, the achievable rate is a complex function of L , which will be further discussed in the next section.

IV. OPTIMIZATION OF PILOT LENGTH

When $L > 2K$, the SNR is high and M goes to infinity, according to equations (5)-(7), we observe that the pilot length L has an impact on the channel estimate and hence on the estimation error. A larger L improves the accuracy of the channel estimation leading to an increased rate, but naturally, it reduces the useful data transmission duration within a coherence time, hence reducing the rate, and vice versa. Therefore, the optimization of the pilot length L is necessary to strike a balance between these two effects. For the simplicity of the ensuing analysis in this section, we assume $\beta_k = \beta$, $p_{p,k} = p_p$, $p_{d,k} = p_d$ and $\sigma_R^2 = \sigma_k^2 = 1$, for $\forall k$.

First, we consider the MRC/MRT processing case. The optimization problem for the pilot length L to get the maximal achievable rate can be expressed as

$$\begin{aligned} \max_L \sum_{k \in \mathcal{K}} R_k^{\text{mr}} \\ \text{s.t. } 2K < L < T, \quad L \text{ is an integer.} \end{aligned} \quad (42)$$

Theorem 3: In Problem (42), the objective function can be approximated as a concave function over L in the range of $2K < L < T$, when L is a real variable.

Proof: To deal with this problem, we reformulate the expression of

$$\text{SNR}_k^{\text{mr}} \triangleq \frac{\text{DS}_k^{\text{mr}}}{\text{GU}_k^{\text{mr}} + \text{SI}_k^{\text{mr}} + \text{IP}_k^{\text{mr}} + \text{NR}_k^{\text{mr}} + \text{ND}_k^{\text{mr}}} \quad (43)$$

into a more tractable form. After some simplifications, we arrive at

$$\text{SNR}_k^{\text{mr}} = \frac{a_3}{a_2 \chi_{\text{mr}}^2 + a_1 \chi_{\text{mr}} + a_0}, \quad (44)$$

where $\chi_{\text{mr}} = \frac{1}{(L-2K)p_p\beta}$. In (44), $a_0, a_1, a_2, a_3 > 0$ represent the coefficients that are independent of the pilot length L , which are given by

$$\begin{aligned} a_0 = & \left(4\beta^2 + \frac{4\beta}{p_r}\right) K^2 + \left(6\beta^2 + \frac{2\beta}{p_d} + \frac{2}{p_r p_d} + \frac{2\beta}{p_r}\right) K \\ & + \left(4K\beta^2 + \frac{2K\beta}{p_r} + 2\beta^2 + \frac{\beta}{p_d}\right) M + 2\beta^2 + \frac{\beta}{p_d}, \end{aligned} \quad (45)$$

$$\begin{aligned} a_1 = & \left(8\beta^2 + \frac{8\beta}{p_r}\right) K^2 + \left(12\beta^2 + \frac{4\beta}{p_d} + \frac{4}{p_r p_d} + 4\beta^2 + \frac{12\beta}{p_r}\right) K \\ & + \left(4K\beta^2 + \frac{2K\beta}{p_r} + 6\beta^2 + \frac{\beta}{p_d}\right) M + 6\beta^2 + \frac{\beta}{p_d}, \end{aligned} \quad (46)$$

$$a_2 = \left(4\beta^2 + \frac{4\beta}{p_r}\right) K^2 + \left(6\beta^2 + \frac{2\beta}{p_d} + \frac{2}{p_r p_d}\right) K, \quad (47)$$

$$a_3 = M(M+1)\beta^2. \quad (48)$$

Since the objective function of Problem (42) associated with SNR_k^{mr} given in (44) is still complicated, we consider its upper bound for simplicity, which is given by

$$\text{SNR}_k^{\text{mr}} < \frac{a_3}{a_1 \chi_{\text{mr}} + a_0} = \frac{a_3}{\frac{a_1}{(L-2K)p_p\beta} + a_0}. \quad (49)$$

Since a_0 , a_1 and a_2 have values of the same order, when χ_{mr} is small enough for ensuring that χ_{mr}^2 is many orders of magnitude smaller than χ_{mr} , the term $a_2 \chi_{\text{mr}}^2$ in (44) can be ignored. Thus, the above upper bound is tight, when χ_{mr} is very small. Note that at high SNRs, where $L p_p \beta \rightarrow \infty$, the variable $\chi_{\text{mr}} = \frac{1}{(1-\frac{2K}{L})L p_p \beta}$ converges to zero, provided that L is not close to $2K$. According to [8], the cost of DAD is large when L is close to $2K$, which degrades the sum rate. This will be verified later in Section V. Hence, the optimal L is not close to $2K$. Therefore, when we are searching for an optimal L , we consider the case where χ_{mr} is small and the upper bound is tight. More explicitly, our simulation results will show that the values of the two sum rates are tight when L is not too close to $2K$. In addition, a larger pilot length L and a higher pilot power p_p leading to a higher SNR, will make the curves of the two sum rates to appear to be tighter.

Based on (49), the objective function of Problem (42) is simplified to

$$\max_{2K < L < T} \frac{T-L}{2T} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{a_3}{\frac{a_1}{(L-2K)p_p\beta} + a_0} \right). \quad (50)$$

When L is considered in the real field in the range of $2K < L < T$, we define

$$f(L) = \log_2 \left(1 + \frac{a_3}{\frac{a_1}{(L-2K)p_p\beta} + a_0} \right), \quad (51)$$

$$g(L) = \frac{T-L}{T} f(L). \quad (52)$$

The first-order and second-order derivatives of $f(L)$ are given respectively by

$$\begin{aligned} f'(L) = & \frac{a_0 a_3}{(a_0^2 + a_0 a_3) p_p \beta (L-2K)^2 + (a_1 a_0 + a_1 a_3) (L-2K) + \frac{a_1^2}{p_p \beta}}, \end{aligned} \quad (53)$$

$$\begin{aligned} f''(L) = & \frac{-a_0 a_3 (2(a_0^2 + a_0 a_3) p_p \beta (L-2K) + a_1 a_0 + a_1 a_3)}{\left((a_0^2 + a_0 a_3) p_p \beta (L-2K)^2 + (a_1 a_0 + a_1 a_3) (L-2K) + \frac{a_1^2}{p_p \beta} \right)^2}. \end{aligned} \quad (54)$$

When L is in the range of $2K < L < T$, it can be observed that $f'(L) > 0$ and $f''(L) < 0$. Next, we focus our attention

on the characteristics of the function $g(L)$. We obtain the second-order derivative of $g(L)$ in (52) as

$$g''(L) = \frac{-2f'(L) + (T-L)f''(L)}{T}. \quad (55)$$

When L is in the range of $2K < L < T$, substituting the results $f'(L) > 0$ and $f''(L) < 0$ into (55), we can then obtain that $g''(L) < 0$. Therefore, when L is in the real field, the objective function of Problem (50) is a concave function of L , provided that $2K < L < T$. ■

Next, the ZFR/ZFT processing scenario is considered. The optimization problem of the ZFR/ZFT processing scenario is to get the maximal achievable rate over the pilot length L , which can be expressed as

$$\begin{aligned} \max_L \quad & \sum_{k \in \mathcal{K}} R_k^{\text{zf}} \\ \text{s.t.} \quad & 2K < L < T, \quad L \text{ is an integer,} \end{aligned} \quad (56)$$

where R_k^{zf} is given in (35).

Theorem 4: In Problem (56), the objective function can be approximated as a concave function over L in the range of $2K < L < T$, when L is a real variable.

Proof: Following some simplifications, we arrive at

$$\text{SNR}_k^{\text{zf}} \triangleq \frac{\text{DS}_k^{\text{zf}}}{\text{GU}_k^{\text{zf}} + \text{SI}_k^{\text{zf}} + \text{IP}_k^{\text{zf}} + \text{NR}_k^{\text{zf}} + \text{ND}_k^{\text{zf}}}, \quad (57)$$

which can be reformulated as

$$\text{SNR}_k^{\text{zf}} = \frac{b_3}{b_2 \chi_{\text{zf}}^2 + b_1 \chi_{\text{zf}} + b_0}, \quad (58)$$

where $\chi_{\text{zf}} = \frac{1}{(L-2K)p_r\beta}$. Furthermore, $b_0, b_1, b_2, b_3 > 0$ represent the coefficients that are independent of L , which are given by

$$b_0 = \frac{2c_1}{p_d p_r} K + \left(\frac{2K}{p_r} + \frac{1}{p_d} \right) \beta c_2, \quad (59)$$

$$\begin{aligned} b_1 = & \left(\frac{4p_d\beta}{p_d p_r} K + \frac{4}{p_d p_r} + \frac{2\beta}{p_d} \right) K c_1 \\ & + \left(\left(\frac{2}{p_r} + 4\beta \right) K + 2\beta + \frac{1}{p_d} \right) \beta c_2, \end{aligned} \quad (60)$$

$$b_2 = \left(\left(4\beta^2 + \frac{4p_d\beta}{p_d p_r} \right) K + \frac{2}{p_d p_r} + \frac{2\beta}{p_d} \right) K c_1, \quad (61)$$

$$b_3 = \beta^2, \quad (62)$$

where $c_1 = \frac{1}{(M-2K)(M-2K-3)}$ and $c_2 = \frac{1}{(M-2K-1)}$.

Similarly, we also consider its upper bound, which is given by

$$\text{SNR}_k^{\text{zf}} < \frac{b_3}{b_1 \chi_{\text{zf}} + b_0} = \frac{b_3}{\frac{b_1}{(L-2K)p_r\beta} + b_0}. \quad (63)$$

We will verify in Section V that the above upper bound is tight, when we are searching for the optimal L . Then, using a procedure similar to that we used for the MRC/MRT case, we obtain *Theorem 4*. ■

Theorem 3 and *Theorem 4* show the optimizations of the pilot length L . To elaborate, maximizing the achievable rates of the MRC/MRT and ZFR/ZFT processing techniques can be

Algorithm 1 Binary Search for The Optimal Pilot Length

Input: Objective function: $u(L)$.

Variable: The pilot length L .

Output: The optimal pilot length L_{opt} .

- 1: Initialize the search range $[L_{\min}, L_{\max}]$, where $L_{\max} = T - 1$ and $L_{\min} = 2K + 1$.
 - 2: **loop**
 - 3: **if** $L_{\max} - L_{\min} > 2$ **then**
 - 4: Set the feasible solution $L_{\text{opt}} = \lfloor (L_{\max} + L_{\min})/2 \rfloor$.
 - 5: Calculate the values of the objective function in $u(L_{\text{opt}} - 1)$, $u(L_{\text{opt}})$ and $u(L_{\text{opt}} + 1)$.
 - 6: **if** $u(L_{\text{opt}} - 1) > u(L_{\text{opt}})$ **then**
 - 7: Set $L_{\max} = L_{\text{opt}} - 1$.
 - 8: **else if** $u(L_{\text{opt}} + 1) > u(L_{\text{opt}})$ **then**
 - 9: Set $L_{\min} = L_{\text{opt}} + 1$.
 - 10: **else**
 - 11: **break.**
 - 12: **end if**
 - 13: Update the search range $[L_{\min}, L_{\max}]$.
 - 14: **else**
 - 15: Find the L_{opt} that maximizes $u(L)$ by exhaustive search method in the range $[L_{\min}, L_{\max}]$.
 - 16: **break.**
 - 17: **end if**
 - 18: **end loop**
-

approximated as convex optimization problems. We propose the low-complexity procedure of *Algorithm 1* to solve these two optimization problems. For the MRC/MRT processing scenario, the value of the term $u(L)$ in *Algorithm 1* is set to $u(L) = \sum_{k \in \mathcal{K}} R_k^{\text{mr}}$, where R_k^{mr} is expressed as in (28). By contrast, for ZFR/ZFT processing, the value of the term $u(L)$ in *Algorithm 1* is set to $u(L) = \sum_{k \in \mathcal{K}} R_k^{\text{zf}}$, where R_k^{zf} is defined in (35).

The corresponding computational complexities of both the exhaustive search algorithm and of *Algorithm 1* (namely our proposed binary search algorithm) are respectively shown as follows:

- **The exhaustive search algorithm:** Since L is in the range of $[2K, T]$, the computational complexity is on the order of $\mathcal{O}(T - 2K)$.
- **The proposed binary search algorithm (*Algorithm 1*):** Since a binary search is used, the computational complexity is on the order of $\mathcal{O}(\log_2(T - 2K))$.

Thus, *Algorithm 1* (our proposed binary search algorithm) has a moderate complexity. Furthermore, we test the running time of both the exhaustive search algorithm and of *Algorithm 1* in simulation. Table II compares the running times of these two algorithms. It further shows that our algorithm has a low complexity.

V. NUMERICAL RESULTS

In this section, we present our numerical results. We assume that there are $N = 1000$ devices at each side of Fig. 1, and the activity probability is $\theta = 0.05$. Hence the number of

TABLE II
A BRIEF COMPARISON OF THE RUNNING TIME FOR THE ALGORITHMS

The number of simulations	MR running time (s)		ZF running time (s)	
	Exhaustive algorithm	Algorithm 1	Exhaustive algorithm	Algorithm 1
100	15.473	0.633	16.328	0.709
500	77.470	3.024	81.435	3.124
1000	153.137	6.055	162.323	6.108
2000	327.793	13.924	346.955	12.478

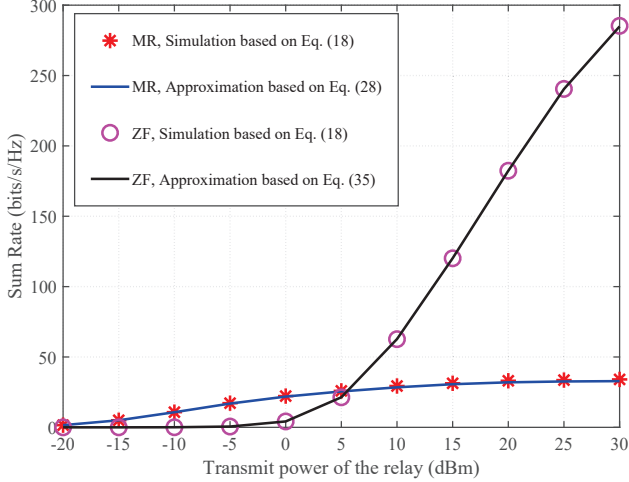


Fig. 2. The sum rate versus p_r with $N = 1000$, $\theta = 0.05$, $L = 150$, and $M = 512$.

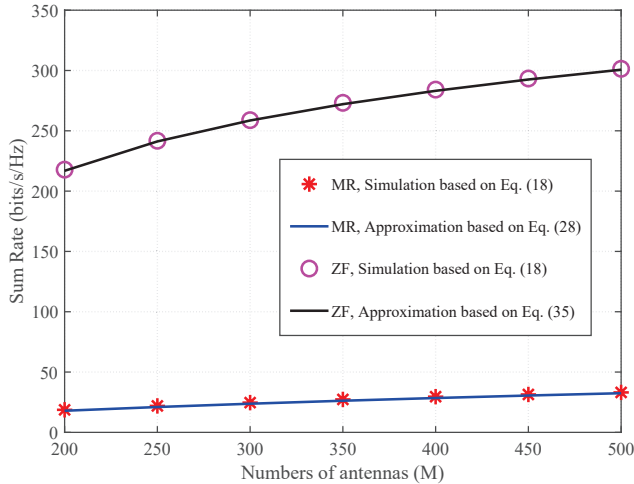


Fig. 3. The sum rate versus M with $N = 1000$, $\theta = 0.05$, $L = 150$, and $p_r = 23$ dBm.

active devices at each side is around $K = 50$. We assume each coherence interval has $T = 1000$ symbols [7], [15]. The path loss model of the wireless channel for device n is $\beta_n = -128.1 - 36.7 \log(d_n)$ in dB, $\forall n$, where d_n denote the distance between device n and the relay, which is in the range of $[0.1\text{km}, 1\text{km}]$. Additionally, we set $\sigma_R^2 = \sigma_k^2 = -109$ dBm, $p_d = \frac{p_r}{2K}$, and $p_p = 13$ dBm. Furthermore, the numerical results in this section are obtained by averaging over 1000 channel realizations.

In Fig. 2 and Fig. 3, we portray the MR and ZF simu-

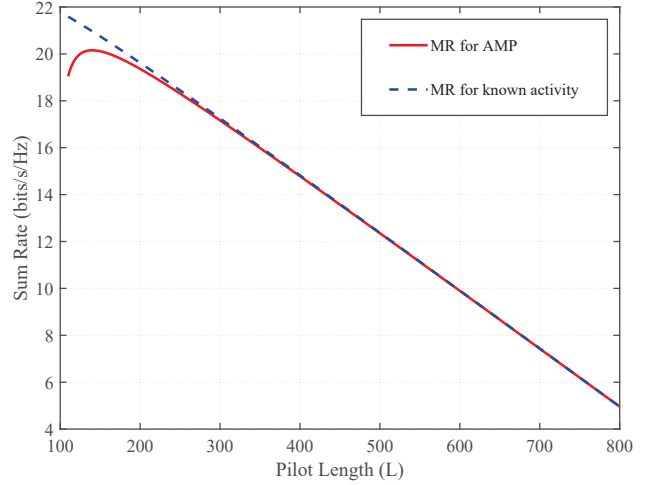


Fig. 4. The sum rate versus L in the MRC/MRT case with $N = 1000$, $\theta = 0.05$, and $M = 256$.

lation curves by substituting the MRC/MRT and ZFR/ZFT beamforming matrices into (18), respectively. Additionally, the MR and ZF approximation curves are respectively obtained according to *Theorem 1* and *Theorem 2*. In Fig. 2, we set $L = 150$ and $M = 512$. It is readily seen that the sum rate in Fig. 2 varies with p_r and p_d , as expected. We see that in both the MRC/MRT and ZFR/ZFT scenarios, the sum rates increase with p_r and p_d , when p_r and p_d are low. When p_r and p_d become higher, the ZFR/ZFT scenario still performs well, but the MRC/MRT scheme exhibits saturation. This is because the MRC/MRT processing neglects the inter-pair interference effect, which becomes excessive with high p_r and p_d . In Fig. 3, we set $L = 150$ and $p_r = 23$ dBm, where the sum rate varies with M . We find that in both the MRC/MRT and ZFR/ZFT scenarios, the sum rates increase with M , as expected. Additionally, the sum rate of the ZFR/ZFT processing is higher than that of the MRC/MRT processing. This is because the ZFR/ZFT processing focuses on eliminating the multi-user interference, which is large especially in the massive access scenario, while the ZFR/ZFT processing focuses on maximizing the desired part of the received signal.

In the massive access scenario, DAD will impose a sum rate erosion, which is the cost of DAD. Herein, the sum rates obtained by the MR and ZF schemes are compared to the sum rates obtained in the perfectly known activity scenarios, where the estimated channel and the channel estimation error of device k can be assumed respectively as

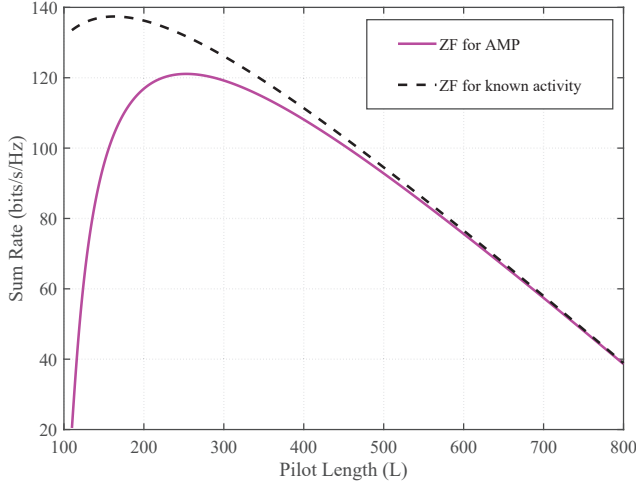


Fig. 5. The sum rate versus L in the ZFR/ZFT case with $N = 1000$, $\theta = 0.05$, and $M = 256$.

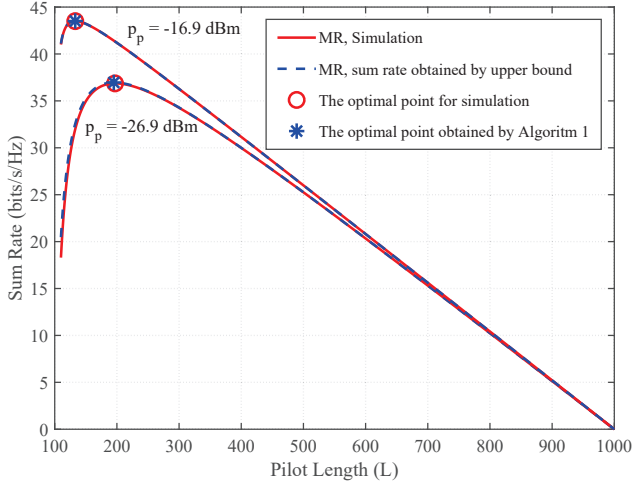


Fig. 6. Pilot length optimization in the MRC/MRT case with background settings $N = 1000$, $\theta = 0.05$, $M = 256$ and $p_r = 23$ dBm

$\hat{\mathbf{g}}_k \sim \mathcal{CN}\left(\mathbf{0}, \frac{\beta_k^2 L p_p}{\beta_k L p_p + \sigma_R^2} \mathbf{I}\right)$ and $\mathbf{e}_k \sim \mathcal{CN}\left(\mathbf{0}, \frac{\beta_k \sigma_R^2}{\beta_k L p_p + \sigma_R^2} \mathbf{I}\right)$. In Fig. 4 and Fig. 5, we set the powers as $p_r = 23$ dBm, $p_d = \frac{p_r}{2K}$ and $p_p = p_d$ to gain further insights. Observe for both the MRC/MRT and ZFR/ZFT cases that the cost of DAD is evident when L is close to $2K$, especially in the ZFR/ZFT case. Additionally, the cost of DAD is falling fast when L increases, as expected. Specifically, when $L = 120$, the DAD rates have a 7.29% drop for MR scheme and a 62.24% drop for ZF scheme, compared to the known activity rates. However, when $L = 200$, the rate drops are 1.33% for MR scheme and 14.17% for ZF scheme. We also find that the maximal sum rates of MRC/MRT processing for the cases with and without known device activity at the relay are closer than that of ZFR/ZFT processing. This means that the cost of realistic imperfect DAD in the MRC/MRT case is lower than that in the ZFR/ZFT case. However, in the ZFR/ZFT case in Fig. 5, the achievable sum rate is dramatically improved compared to the MRC/MRT case in Fig. 4.

In Section IV, we utilize upper bounds for the sum rates to solve the problems in (42) and (56). Herein, the sum rates

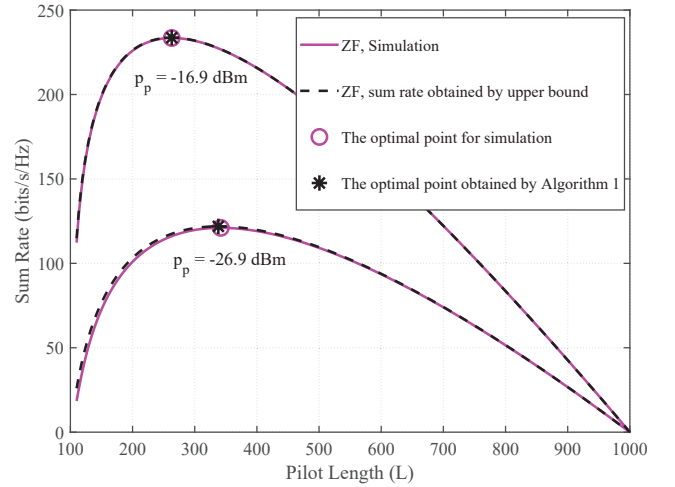


Fig. 7. Pilot length optimization in the ZFR/ZFT case with background settings $N = 1000$, $\theta = 0.05$, $M = 256$ and $p_r = 23$ dBm

obtained by the proposed upper bound are compared to the sum rates obtained by MR and ZF simulation. Fig. 6 shows that the upper bound in the MRC/MRT scenario is close to the simulation result. Additionally, the gap between the two sum rates becomes smaller, when L becomes larger and p_p becomes higher, as expected. Observe in Fig. 7 that the situation is similar for the ZFR/ZFT case. Furthermore, we use *Algorithm 1* to obtain optimal points for the sum rates. As is seen in Fig. 6 and Fig. 7, *Algorithm 1* performs well. The optimal points obtained by *Algorithm 1* for both MRC/MRT and ZFR/ZFT processing are close to the optimal points in the simulation results, when $p_p = -26.9$ dBm. Furthermore, the optimal points become the same, when $p_p = -16.9$ dBm.

VI. CONCLUSION

In this paper, we first applied an MTMM AF relaying system in a single-cell massive access scenario. Using the AMP algorithm for DAD and CE, we derived the closed-form ergodic achievable rate expressions for both MRC/MRT and ZFR/ZFT processing, which depend on the system parameters: L , $2K$, M , $p_{p,j}$, p_r , $p_{d,j}$, and β_j , where $j \in \mathcal{K}$. Additionally, we found that in the MRC/MRT case, the cost of DAD is lower, while in the ZFR/ZFT case, the sum rate is much higher. Furthermore, for both MRC/MRT and ZFR/ZFT cases, we approximated the non-convex pilot length optimization problems of maximizing the sum rate by tractable convex optimization problems. Finally, we solved them by the proposed algorithm.

APPENDIX A PROOF OF LEMMA 1

First we briefly review some known results concerning random vectors [35], [36] that will be useful in the following derivations. Given mutually independent random vectors $\mathbf{x} \in \mathbb{C}^{M \times 1}$ and $\mathbf{y} \in \mathbb{C}^{M \times 1}$ with the distributions $\mathcal{CN}(\mathbf{0}, \sigma_x^2 \mathbf{I}_M)$ and $\mathcal{CN}(\mathbf{0}, \sigma_y^2 \mathbf{I}_M)$ respectively, it can be concluded that

$$\mathbb{E} \left\{ |\mathbf{x}^H \mathbf{y}|^2 \right\} = M \sigma_x^2 \sigma_y^2, \quad (64)$$

$$\mathbb{E}\{\mathbf{xx}^H\mathbf{xx}^H\} = (M+1)\sigma_x^4\mathbf{I}_M. \quad (65)$$

To simplify the expressions for the following procedures of calculations, from (5) and (6), we define the terms $\sigma_{g,k}^2$ and $\sigma_{e,k}^2$ as

$$\sigma_{g,k}^2 \triangleq \frac{\beta_k^2 \xi_k}{\beta_k \xi_k + \tau_\infty^2}, \quad (66)$$

$$\sigma_{e,k}^2 \triangleq \frac{\beta_k \tau_\infty^2}{\beta_k \xi_k + \tau_\infty^2}. \quad (67)$$

Then, we commence by calculating ρ_{mr} in (14). Note that we have to calculate two terms $\mathbb{E}\{\|\mathbf{W}\mathbf{g}_n\|^2\}$ and $\mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\}$ to complete the proof. Firstly, we rewrite the MRC/MRT beamforming matrix based on (12) as follows:

$$\mathbf{W} = \sum_{j \in \mathcal{K}_A} (\hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H). \quad (68)$$

Next, we focus on calculating the term $\mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\}$ shown below

$$\begin{aligned} \mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\} &= \sigma_R^2 \text{tr}(\mathbb{E}\{\mathbf{W}^H \mathbf{W}\}) \\ &= \sigma_R^2 \text{tr}\left(\mathbb{E}\left\{\sum_{i \in \mathcal{K}_A} (\hat{\mathbf{g}}_{i'} \hat{\mathbf{g}}_i^T + \hat{\mathbf{g}}_i \hat{\mathbf{g}}_{i'}^T) \sum_{j \in \mathcal{K}_A} (\hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H)\right\}\right) \\ &\stackrel{(a)}{=} \sigma_R^2 \sum_{j \in \mathcal{K}_A} \text{tr}(\mathbb{E}\{(\hat{\mathbf{g}}_{j'} \hat{\mathbf{g}}_j^T + \hat{\mathbf{g}}_j \hat{\mathbf{g}}_{j'}^T) (\hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H)\}). \end{aligned} \quad (69)$$

Step (a) is justified, because when $i \neq j$, the expectation equals to zero. Based on (64), and exploiting the facts that $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ and $(\mathbf{x}^T \mathbf{y})^T = \mathbf{x}^T \mathbf{y}$, we obtain

$$\begin{aligned} &\text{tr}(\mathbb{E}\{(\hat{\mathbf{g}}_{j'} \hat{\mathbf{g}}_j^T + \hat{\mathbf{g}}_j \hat{\mathbf{g}}_{j'}^T) (\hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H)\}) \\ &= \text{tr}(\mathbb{E}\{\hat{\mathbf{g}}_{j'} \hat{\mathbf{g}}_j^T \hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_j^H\}) + \text{tr}(\mathbb{E}\{\hat{\mathbf{g}}_j \hat{\mathbf{g}}_{j'}^T \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H\}) \\ &+ \text{tr}(\mathbb{E}\{\hat{\mathbf{g}}_j \hat{\mathbf{g}}_{j'}^T \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j^H\}) + \text{tr}(\mathbb{E}\{\hat{\mathbf{g}}_{j'} \hat{\mathbf{g}}_j^T \hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_{j'}^H\}) \\ &= 2M(M+1)\sigma_{g,j}^2 \sigma_{g,j'}^2. \end{aligned} \quad (70)$$

Therefore, we have

$$\mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\} = \sigma_R^2 2M(M+1) \sum_{j \in \mathcal{K}_A} \sigma_{g,j}^2 \sigma_{g,j'}^2 \triangleq \sigma_R^2 \eta_1. \quad (71)$$

Finally, we calculate the other term $\mathbb{E}\{\|\mathbf{W}\mathbf{g}_n\|^2\}$ as follows:

$$\mathbb{E}\{\|\mathbf{W}\mathbf{g}_n\|^2\} = \mathbb{E}\{\hat{\mathbf{g}}_n^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{g}}_n\} + \mathbb{E}\{\mathbf{e}_n^H \mathbf{W}^H \mathbf{W} \mathbf{e}_n\}. \quad (72)$$

The term $\mathbb{E}\{\hat{\mathbf{g}}_n^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{g}}_n\}$ in (72) is expanded as

$$\begin{aligned} &\mathbb{E}\{\hat{\mathbf{g}}_n^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{g}}_n\} \\ &= \mathbb{E}\{\hat{\mathbf{g}}_n^T (\hat{\mathbf{g}}_{n'} \hat{\mathbf{g}}_n^T + \hat{\mathbf{g}}_n \hat{\mathbf{g}}_{n'}^T) (\hat{\mathbf{g}}_{n'}^* \hat{\mathbf{g}}_n^H + \hat{\mathbf{g}}_n^* \hat{\mathbf{g}}_{n'}^H) \hat{\mathbf{g}}_n^*\} \\ &+ \sum_{j, j' \neq n} \mathbb{E}\{\hat{\mathbf{g}}_n^T (\hat{\mathbf{g}}_{j'} \hat{\mathbf{g}}_j^T + \hat{\mathbf{g}}_j \hat{\mathbf{g}}_{j'}^T) (\hat{\mathbf{g}}_{j'}^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H) \hat{\mathbf{g}}_n^*\}, \end{aligned} \quad (73)$$

which can be calculated in a way similar to (70). Additionally, the second term $\mathbb{E}\{\mathbf{e}_n^H \mathbf{W}^H \mathbf{W} \mathbf{e}_n\}$ can be obtained similarly to (71). Thus, we arrive at

$$\mathbb{E}\{\|\mathbf{W}\mathbf{g}_n\|^2\} = M(M+1)^2 \sigma_{g,n}^4 \sigma_{g,n'}^2 + \beta_n \eta_1. \quad (74)$$

Based on (66), (67), (71), and (74), we obtain (14), which

completes this proof.

APPENDIX B PROOF OF LEMMA 2

The derivation of obtaining ρ_{zf} in (17) includes the calculation of the terms $\mathbb{E}\{\|\mathbf{W}\hat{\mathbf{G}}\bar{\mathbf{P}}\bar{\mathbf{s}}\|^2\}$ and $\mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\}$. We will use (66) and (67) to simplify the expressions for the following procedures of calculations. We first calculate the second term as

$$\begin{aligned} \mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\} &= \sigma_R^2 \text{tr}(\mathbb{E}\{\mathbf{W}^H \mathbf{W}\}) \\ &= \sigma_R^2 \text{tr}\left(\mathbb{E}\left\{\left(\hat{\mathbf{G}}^H \hat{\mathbf{G}}\right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*\right)^{-1} \mathbf{P}\right\}\right) \\ &= \sigma_R^2 \text{tr}\left(\mathbb{E}\{\hat{\mathbf{A}} \hat{\mathbf{P}} \hat{\mathbf{A}}^* \mathbf{P}\}\right) = \sigma_R^2 \sum_{j \in \mathcal{K}} \sum_{i \in \mathcal{K}} \mathbb{E}\{\hat{a}_{j,i} \hat{a}_{i,j}^*\}, \end{aligned} \quad (75)$$

where we define $\hat{\mathbf{A}} \triangleq \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}}\right)^{-1}$ and $[\hat{\mathbf{A}}]_{i,j} \triangleq \hat{a}_{i,j}$. Note that $\hat{\mathbf{G}}^H \hat{\mathbf{G}}$ is a center Wishart matrix obeying the distribution $\mathcal{W}_{2K}(M, \hat{\Sigma})$, where $\hat{\Sigma}$ is the covariance matrix of the row vectors in the estimated channel matrix $\hat{\mathbf{G}}$. According to the properties of the Wishart matrix given in [36], we have

$$\mathbb{E}\left\{\left(\hat{\mathbf{G}}^H \hat{\mathbf{G}}\right)^{-1}\right\} = \frac{\hat{\Sigma}^{-1}}{M-2K-1}. \quad (76)$$

Furthermore, the matrix $\hat{\mathbf{A}} = \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}}\right)^{-1}$ follows an inverse Wishart distribution denoted by $\mathcal{W}_{2K}^{-1}(M, \hat{\Sigma}^{-1})$. Given the definition $[\hat{\Sigma}^{-1}]_{i,j} \triangleq \varsigma_{i,j}$, an essential property [29], [37], [38] of the inverse Wishart matrix $\hat{\mathbf{A}}$ can be formulated as

$$\text{cov}(\hat{a}_{i,j}, \hat{a}_{k,l}) = \frac{2\varsigma_{i,j}\varsigma_{k,l} + (M-2K-1)(\varsigma_{i,k}\varsigma_{j,l} + \varsigma_{i,l}\varsigma_{j,k})}{(M-2K)(M-2K-1)^2(M-2K-3)}. \quad (77)$$

Since $\hat{a}_{i,j}$ is real and $\varsigma_{i,j}$ satisfies that $\varsigma_{i,j} = \sigma_{g,j}^{-2}$ if $i = j$ and $\varsigma_{i,j} = 0$ if $i \neq j$, the terms $\mathbb{E}\{\hat{a}_{j,i} \hat{a}_{i,j}^*\}$ are divided into three groups. When $i = j$, we obtain

$$\begin{aligned} \mathbb{E}\{|\hat{a}_{j,j'}|^2\} &= \text{cov}(\hat{a}_{j,j'}, \hat{a}_{j,j'}) + |\mathbb{E}\{\hat{a}_{j,j'}\}|^2 \\ &= \frac{\sigma_{g,j}^{-2} \sigma_{g,j'}^{-2}}{(M-2K)(M-2K-1)(M-2K-3)}, \end{aligned} \quad (78)$$

When $i = j'$, we have

$$\begin{aligned} \mathbb{E}\{\hat{a}_{j,j} \hat{a}_{j',j'}^*\} &= \text{cov}(\hat{a}_{j,j}, \hat{a}_{j',j'}) + \mathbb{E}\{\hat{a}_{j,j}\} \mathbb{E}\{\hat{a}_{j',j'}^*\} \\ &= \frac{2\sigma_{g,j}^{-2} \sigma_{g,j'}^{-2}}{(M-2K)(M-2K-1)^2(M-2K-3)} + \frac{\sigma_{g,j}^{-2} \sigma_{g,j'}^{-2}}{(M-2K-1)^2}. \end{aligned} \quad (79)$$

Otherwise, it is noted from (77) that

$$\mathbb{E}\{\hat{a}_{j,i} \hat{a}_{i,j'}^*\} = 0, \quad \text{if } i \neq j \text{ and } i \neq j'. \quad (80)$$

Therefore, substituting (78), (79) and (80) into (75), we arrive at

$$\mathbb{E}\{\|\mathbf{W}\mathbf{n}_R\|^2\} = \sigma_R^2 \sum_{j \in \mathcal{K}} \frac{2}{(M-2K)(M-2K-3)\sigma_{g,j}^2 \sigma_{g,j'}^2} \triangleq \sigma_R^2 \gamma. \quad (81)$$

Second, the other term $E \left\{ \|\mathbf{W}\bar{\mathbf{G}}\bar{\mathbf{P}}\bar{\mathbf{s}}\|^2 \right\}$ is calculated as

$$\begin{aligned} & E \left\{ \|\mathbf{W}\bar{\mathbf{G}}\bar{\mathbf{P}}\bar{\mathbf{s}}\|^2 \right\} \\ &= \text{tr} \left(E \left\{ \bar{\mathbf{P}}\hat{\mathbf{G}}^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{G}} \bar{\mathbf{P}} \right\} + E \left\{ \bar{\mathbf{P}}\Xi^H \mathbf{W}^H \mathbf{W} \Xi \bar{\mathbf{P}} \right\} \right) \\ &= \text{tr} \left(E \left\{ \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{P} \bar{\mathbf{P}} \bar{\mathbf{P}} \mathbf{P} \right\} \right) + \\ & \text{tr} \left(E \left\{ \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \hat{\mathbf{G}}^H \Xi \bar{\mathbf{P}} \bar{\mathbf{P}} \Xi^H \right\} \right), \end{aligned} \quad (82)$$

where Ξ is the channel estimation error matrix defined by $\Xi = \bar{\mathbf{G}} - \hat{\mathbf{G}}$. Therefore, the first term in the second line of (82) is derived as

$$\begin{aligned} \text{tr} \left(E \left\{ \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{P} \bar{\mathbf{P}} \bar{\mathbf{P}} \mathbf{P} \right\} \right) &= \text{tr} \left(E \left\{ \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \right\} E \left\{ \mathbf{P} \bar{\mathbf{P}} \bar{\mathbf{P}} \mathbf{P} \right\} \right) \\ &= \sum_{n \in \mathcal{K}} \frac{p_{d,n'}}{(M-2K-1) \sigma_{g,n}^2}. \end{aligned} \quad (83)$$

Furthermore, the second term is obtained as

$$\begin{aligned} & \text{tr} \left(E \left\{ \hat{\mathbf{G}} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \hat{\mathbf{G}}^H \Xi \bar{\mathbf{P}} \bar{\mathbf{P}} \Xi^H \right\} \right) \\ &= \text{tr} \left(E \left\{ \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \mathbf{P} \left(\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* \right)^{-1} \mathbf{P} \right\} \right) \sum_{n \in \mathcal{K}} p_{d,n} \sigma_{e,n}^2 \\ &= \gamma \sum_{n \in \mathcal{K}} p_{d,n} \sigma_{e,n}^2. \end{aligned} \quad (84)$$

According to (16) and (82), we can obtain (17) by substituting (66), (67), (81), (83), and (84) into (16). Thus, this proof is completed.

APPENDIX C PROOF OF Theorem 1

In order to obtain (28), we should derive the closed-form expressions of the terms DS_k , GU_k , SI_k , IP_k , NR_k , and ND_k , when the relay utilizes the MRC/MRT processing technique. We will use (66) and (67) to simplify the expressions for the following procedures.

a) *The term DS_k^{mr}* : We start with the following calculation:

$$\begin{aligned} E \{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \} &= E \left\{ \left(\hat{\mathbf{g}}_k^T + \mathbf{e}_k^T \right) \left[\sum_{j \in \mathcal{K}_A} \left(\hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j^H \right) \right] \left(\hat{\mathbf{g}}_{k'} + \mathbf{e}_{k'} \right) \right\} \\ &= E \left\{ \hat{\mathbf{g}}_k^T \left(\hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H + \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H \right) \hat{\mathbf{g}}_{k'} \right\} \\ &= E \left\{ \left| \hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'} \right|^2 \right\} + E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \right\} E \left\{ \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_k \right\} \\ &= M \sigma_{g,k}^2 \sigma_{g,k'}^2 + M^2 \sigma_{g,k}^2 \sigma_{g,k'}^2 = M(M+1) \sigma_{g,k}^2 \sigma_{g,k'}^2. \end{aligned} \quad (85)$$

Thus we have

$$\text{DS}_k^{\text{mr}} = p_{d,k'} |E \{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \}|^2 = p_{d,k'} M^2 (M+1)^2 \sigma_{g,k}^4 \sigma_{g,k'}^4. \quad (86)$$

b) *This term GU_k^{mr}* : The term can be expressed as

$$\begin{aligned} \text{GU}_k^{\text{mr}} &= p_{d,k'} \text{Var} \left\{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right\} \\ &= p_{d,k'} E \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\} - p_{d,k'} |E \{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \}|^2. \end{aligned} \quad (87)$$

Therefore, we only have to derive the expression for $E \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\}$, which is given by

$$\begin{aligned} & E \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\} \\ &= E \left\{ \left(\hat{\mathbf{g}}_k^T + \mathbf{e}_k^T \right) \mathbf{W} \left(\hat{\mathbf{g}}_{k'} + \mathbf{e}_{k'} \right) \left(\hat{\mathbf{g}}_{k'}^H + \mathbf{e}_{k'}^H \right) \mathbf{W}^H \left(\hat{\mathbf{g}}_k^* + \mathbf{e}_k^* \right) \right\} \\ &\stackrel{(a)}{=} E \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} + E \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \mathbf{e}_{k'} \mathbf{e}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} \\ & \quad + E \left\{ \mathbf{e}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \mathbf{e}_k^* \right\} + E \left\{ \mathbf{e}_k^T \mathbf{W} \mathbf{e}_{k'} \mathbf{e}_{k'}^H \mathbf{W}^H \mathbf{e}_k^* \right\}. \end{aligned} \quad (88)$$

Since the random vectors $\hat{\mathbf{g}}_k$, $\hat{\mathbf{g}}_{k'}$, $\hat{\mathbf{e}}_k$ and $\hat{\mathbf{e}}_{k'}$ are mutually independent, step (a) can be carried out, where the term $E \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\}$ in (88) is expressed as

$$\begin{aligned} & E \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} \\ &= E \left\{ \hat{\mathbf{g}}_k^T \left(\hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H + \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H \right) \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \left(\hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^T + \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^T \right) \hat{\mathbf{g}}_k^* \right\} \\ & \quad + \sum_{j,j' \neq k} E \left\{ \hat{\mathbf{g}}_k^T \left(\hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j^H \right) \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \left(\hat{\mathbf{g}}_{j'} \hat{\mathbf{g}}_{j'}^T + \hat{\mathbf{g}}_j \hat{\mathbf{g}}_j^T \right) \hat{\mathbf{g}}_k^* \right\}. \end{aligned} \quad (89)$$

The calculation of the first term at the right hand side of (89) can be divided into four parts

$$\begin{aligned} & E \left\{ \hat{\mathbf{g}}_k^T \left(\hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H + \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H \right) \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \left(\hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^T + \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^T \right) \hat{\mathbf{g}}_k^* \right\} \\ &= E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_k^* \right\} \\ & \quad + E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\ & \quad + E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\ & \quad + E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\}. \end{aligned} \quad (90)$$

The term $E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\}$ in (90) can be derived as

$$\begin{aligned} & E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\ &= E \left\{ \left| \frac{\hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'}}{\|\hat{\mathbf{g}}_{k'}\|} \right|^4 \|\hat{\mathbf{g}}_{k'}\|^4 \right\} = E \left\{ \left| \frac{\hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'}}{\|\hat{\mathbf{g}}_{k'}\|} \right|^4 \|\hat{\mathbf{g}}_{k'}\|^4 \right\} E \left\{ \|\hat{\mathbf{g}}_{k'}\|^4 \right\} \\ &\stackrel{(a)}{=} 2M(M+1) \sigma_{g,k}^4 \sigma_{g,k'}^4, \end{aligned} \quad (91)$$

where $\frac{\hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'}}{\|\hat{\mathbf{g}}_{k'}\|} \hat{\mathbf{g}}_{k'}$ can be regarded as a complex-valued Gaussian random variable with the distribution $\mathcal{CN} \left(0, \sigma_{g,k}^2 \right)$. Step (a) is obtained by using (65) and the fact that if a complex Gaussian random variable ϖ follows the distribution $\mathcal{CN} \left(0, \sigma^2 \right)$, then $E \left\{ |\varpi|^4 \right\} = 2\sigma^4$. Then, we calculate the last three terms at the right hand side of (90). Based on (65) and the facts that $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ and $x = x^T$ when x is a scalar, the second term can be expressed as

$$\begin{aligned} & E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\ &= E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \left(\hat{\mathbf{g}}_k^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \right)^T \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\ &= E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\ &= \text{tr} \left(E \left\{ \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \right) \\ &= \text{tr} \left(E \left\{ \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \right\} \cdot E \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \right) \\ &= \text{tr} \left((M+1) \sigma_{g,k'}^4 \mathbf{I}_M \cdot (M+1) \sigma_{g,k}^4 \mathbf{I}_M \right) \\ &= M(M+1)^2 \sigma_{g,k}^4 \sigma_{g,k'}^4. \end{aligned} \quad (92)$$

In a similar way, the third and the fourth terms can be formulated respectively as

$$\begin{aligned}
& \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* (\hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^*) \right\} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \right\} \\
&= \text{tr} \left(\mathbb{E} \left\{ \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \right\} \right) \\
&= \text{tr} \left(\mathbb{E} \left\{ \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \right\} \cdot \mathbb{E} \left\{ \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \hat{\mathbf{g}}_{k'}^* \hat{\mathbf{g}}_{k'}^T \right\} \right) \\
&= \text{tr} \left((M+1) \sigma_{g,k}^4 \mathbf{I}_M \cdot (M+1) \sigma_{g,k'}^4 \mathbf{I}_M \right) \\
&= M(M+1)^2 \sigma_{g,k}^4 \sigma_{g,k'}^4, \tag{93}
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \hat{\mathbf{g}}_{k'} \right\} \cdot \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k^* \right\} \\
&= M^2 (M+1)^2 \sigma_{g,k}^4 \sigma_{g,k'}^4. \tag{94}
\end{aligned}$$

Thus, we obtain (90) as

$$\begin{aligned}
& \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T (\hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_k^H + \hat{\mathbf{g}}_k^* \hat{\mathbf{g}}_{k'}^H) \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H (\hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_k^T + \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^T) \hat{\mathbf{g}}_k^* \right\} \\
&= M(M+1) (M^2 + 3M + 4) \sigma_{g,k}^4 \sigma_{g,k'}^4. \tag{95}
\end{aligned}$$

Furthermore, similar to (70), the second term at the right side of (89) can be obtained as

$$\begin{aligned}
& \sum_{j,j' \neq k} \mathbb{E} \left\{ \hat{\mathbf{g}}_j^T (\hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_j^H + \hat{\mathbf{g}}_j^* \hat{\mathbf{g}}_{j'}^H) \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H (\hat{\mathbf{g}}_j \hat{\mathbf{g}}_j^T + \hat{\mathbf{g}}_j \hat{\mathbf{g}}_{j'}^T) \hat{\mathbf{g}}_k^* \right\} \\
&= \sum_{j,j' \neq k} 2M(M+1) \sigma_{g,k}^2 \sigma_{g,k'}^2 \sigma_{g,j}^2 \sigma_{g,j'}^2. \tag{96}
\end{aligned}$$

When substituting (95) and (96) into (89), we arrive at

$$\begin{aligned}
& \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} = \sigma_{g,k}^2 \sigma_{g,k'}^2 \eta \\
&+ M(M+1) (M^2 + 3M + 2) \sigma_{g,k}^4 \sigma_{g,k'}^4. \tag{97}
\end{aligned}$$

The other three terms at the right hand side of (88) can be obtained by using similar steps to those used for the derivation of the first term. Therefore, the term $\mathbb{E} \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\}$ is obtained as

$$\begin{aligned}
& \mathbb{E} \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\} = \beta_k \beta_{k'} \eta + M(M+1)^2 \sigma_{g,k}^2 \sigma_{g,k'}^2 \\
&\times ((M+2) \sigma_{g,k}^2 \sigma_{g,k'}^2 + \sigma_{g,k}^2 \sigma_{e,k'}^2 + \sigma_{g,k'}^2 \sigma_{e,k}^2). \tag{98}
\end{aligned}$$

After substituting (98) and (85) into (87), we obtain GU_k^{mr} in

(30).

c) *The terms SI_k^{mr} , IP_k^{mr} and NR_k^{mr}* : The derivation of these terms can be obtained by using similar steps used for deriving the expression of GU_k^{mr} .

d) *The term ND_k^{mr}* : Since n_k is the AWGN at device k , (34) can be readily obtained.

Then, (66) and (67) is substituted into above equations, and we can obtain (28). Accordingly, we complete this proof.

APPENDIX D PROOF OF Theorem 2

Similarly, for the ZFR/ZFT processing scenario, we have to derive the equations (36)-(41) in order to prove (35). We will use (66) and (67) to simplify the expressions for the following procedures. Note that for the ZFR/ZFT beamforming matrix \mathbf{W} given in (15), we have

$$\hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} = \hat{\mathbf{g}}_{k'}^T \mathbf{W} \hat{\mathbf{g}}_k = 1, \tag{99}$$

$$\hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_k = \hat{\mathbf{g}}_{k'}^T \mathbf{W} \hat{\mathbf{g}}_{k'} = 0. \tag{100}$$

Then, we give the detailed derivation in the following:

a) *The term DS_k^{zf}* : We first calculate

$$\begin{aligned}
& \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \mathbf{g}_{k'} \right\} = \mathbb{E} \left\{ (\hat{\mathbf{g}}_k^T + \mathbf{e}_k^T) \mathbf{W} (\hat{\mathbf{g}}_{k'} + \mathbf{e}_{k'}) \right\} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \right\} + \mathbb{E} \left\{ \mathbf{e}_k^T \mathbf{W} \mathbf{e}_{k'} \right\} = 1. \tag{101}
\end{aligned}$$

Then, the term DS_k^{zf} can be obtained as

$$\text{DS}_k^{\text{zf}} = p_{d,k'} \left| \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \mathbf{g}_{k'} \right\} \right|^2 = p_{d,k'}. \tag{102}$$

b) *The term GU_k^{zf}* : The term can be expressed as

$$\begin{aligned}
& \text{GU}_k^{\text{zf}} = p_{d,k'} \text{Var} \left\{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right\} \\
&= p_{d,k'} \mathbb{E} \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\} - p_{d,k'} \left| \mathbb{E} \left\{ \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right\} \right|^2, \tag{103}
\end{aligned}$$

where the term $\mathbb{E} \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\}$ in (103) can be derived as (104) at the bottom of this page. In step (a), the vectors \mathbf{i}_k and $\mathbf{i}_{k'}$ are the column vectors of \mathbf{I}_{2K} in the same location as the vectors $\hat{\mathbf{g}}_k$ and $\hat{\mathbf{g}}_{k'}$ in matrix $\hat{\mathbf{G}}$, respectively. This is due to the fact that $\hat{\mathbf{G}}^T \hat{\mathbf{G}}^* (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} = (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^H \hat{\mathbf{G}} = \mathbf{I}_{2K}$. Step (b) is obtained by (75) and by using some further simplifications, while step (c) is based on (76).

$$\begin{aligned}
& \mathbb{E} \left\{ \left| \mathbf{g}_k^T \mathbf{W} \mathbf{g}_{k'} \right|^2 \right\} = \mathbb{E} \left\{ (\hat{\mathbf{g}}_k^T + \mathbf{e}_k^T) \mathbf{W} (\hat{\mathbf{g}}_{k'} + \mathbf{e}_{k'}) (\hat{\mathbf{g}}_k^H + \mathbf{e}_k^H) \mathbf{W}^H (\hat{\mathbf{g}}_k^* + \mathbf{e}_k^*) \right\} \\
&= \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} + \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \mathbf{e}_{k'} \mathbf{e}_{k'}^H \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} + \mathbb{E} \left\{ \mathbf{e}_k^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \mathbf{e}_k^* \right\} + \mathbb{E} \left\{ \mathbf{e}_k^T \mathbf{W} \mathbf{e}_{k'} \mathbf{e}_{k'}^H \mathbf{W}^H \mathbf{e}_k^* \right\} \\
&= 1 + \sigma_{e,k}^2 \mathbb{E} \left\{ \hat{\mathbf{g}}_k^T \mathbf{W} \mathbf{W}^H \hat{\mathbf{g}}_k^* \right\} + \sigma_{e,k}^2 \mathbb{E} \left\{ \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \mathbf{W} \hat{\mathbf{g}}_{k'} \right\} + \sigma_{e,k}^2 \sigma_{e,k'}^2 \text{tr} \left(\mathbb{E} \left\{ \mathbf{W} \mathbf{W}^H \right\} \right) \\
&\stackrel{(a)}{=} 1 + \sigma_{e,k}^2 \mathbb{E} \left\{ \mathbf{i}_k^T \mathbf{P} (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \mathbf{P} \mathbf{i}_k \right\} + \sigma_{e,k}^2 \mathbb{E} \left\{ \mathbf{i}_{k'}^T \mathbf{P} (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} \mathbf{P} \mathbf{i}_{k'} \right\} + \sigma_{e,k}^2 \sigma_{e,k'}^2 \text{tr} \left(\mathbb{E} \left\{ (\hat{\mathbf{G}}^H \hat{\mathbf{G}})^{-1} \mathbf{P} (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} \mathbf{P} \right\} \right) \\
&\stackrel{(b)}{=} 1 + \sigma_{e,k}^2 \left[\mathbb{E} \left\{ (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} \right\} \right]_{k',k'} + \sigma_{e,k}^2 \left[\mathbb{E} \left\{ (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} \right\} \right]_{k,k} + \sigma_{e,k}^2 \sigma_{e,k'}^2 \sum_{j \in \mathcal{K}} \frac{2}{(M-2K)(M-2K-3) \sigma_{g,j}^2 \sigma_{g,j'}} \\
&\stackrel{(c)}{=} 1 + \frac{\sigma_{e,k'}}{(M-2K-1) \sigma_{g,k'}} + \frac{\sigma_{e,k}}{(M-2K-1) \sigma_{g,k}} + \sigma_{e,k}^2 \sigma_{e,k'}^2 \gamma. \tag{104}
\end{aligned}$$

c) The terms SI_k^{zf} , IP_k^{zf} and NR_k^{zf} : Similar to the steps we use for deriving GU_k^{zf} , the expressions in (38), (39) and (40) can be obtained.

d) The term ND_k^{zf} : The expression in (41) is valid because n_k is the AWGN at device k .

Finally, we substitute (66) and (67) into above equations, hence accordingly, we obtain (35). Thus, this proof is completed.

REFERENCES

- [1] L. Liu and W. Yu, "Massive device connectivity with massive MIMO," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1072–1076.
- [2] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.
- [3] J. Wang, Z. Zhang, and L. Hanzo, "Joint active user detection and channel estimation in massive access systems exploiting reedmuller sequences," *IEEE J. Sel. Top. Sign. Process.*, vol. 13, no. 3, pp. 739–752, Mar. 2019.
- [4] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block sparse bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9631–9640, Oct. 2018.
- [5] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "DNN-aided block sparse bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12000–12012, Dec. 2019.
- [6] Z. Zhang, Y. Li, C. Huang, Q. Guo, L. Liu, C. Yuen, and Y. L. Guan, "User activity detection and channel estimation for grant-free random access in LEO satellite-enabled internet of things," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8811–8825, Sep. 2020.
- [7] L. Liu and W. Yu, "Massive connectivity with massive MIMO - Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.
- [8] —, "Massive connectivity with massive MIMO - Part II: Achievable rate characterization," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2947–2959, Jun. 2018.
- [9] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
- [10] —, "Message passing algorithms for compressed sensing: II. Analysis and validation," in *Proc. IEEE Inf. Theory Workshop Inf. Theory (ITW)*, Jan. 2010, pp. 1–5.
- [11] J. Kim, W. Chang, B. Jung, D. Baron, and J. C. Ye, "Belief propagation for joint sparse recovery," *Computing Research Repository - CORR*, Feb. 2011.
- [12] C. Wei, H. Liu, Z. Zhang, J. Dang, and L. Wu, "Approximate message passing-based joint user activity and data detection for NOMA," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 640–643, Mar. 2017.
- [13] Z. Chen, F. Sahrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.
- [14] L. Liu, C. Liang, J. Ma, and L. Ping, "Capacity optimality of AMP in coded systems." [Online]. Available: <https://arxiv.org/abs/1901.09559>
- [15] X. Shao, X. Chen, C. Zhong, J. Zhao, and Z. Zhang, "A unified design of massive access for cellular Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3934–3947, Apr. 2019.
- [16] K. Chen and R. Prasad, *Cooperative Communications and Networks*. Wiley, 2009. [Online]. Available: <https://ieeexplore.ieee.org/document/8042398>
- [17] W. Xu, X. Dong, and W. Lu, "MIMO relaying broadcast channels with linear precoding and quantized channel state information feedback," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5233–5245, Oct. 2010.
- [18] Z. Peng, W. Xu, L. Wang, and C. Zhao, "Achievable rate analysis and feedback design for multiuser MIMO relay with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 780–793, Feb. 2014.
- [19] H. A. Suraweera, H. Q. Ngo, T. Q. Duong, C. Yuen, and E. G. Larsson, "Multi-pair amplify-and-forward relaying with very large antenna arrays," in *Proc. IEEE ICC*, Jun. 2013, pp. 4635–4640.
- [20] W. Xu, J. Liu, S. Jin, and X. Dong, "Spectral and energy efficiency of multi-pair massive MIMO relay network with hybrid processing," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3794–3809, Sep. 2017.
- [21] Q. Wang and Y. Jing, "Performance analysis and scaling law of MR-C/MRT relaying with CSI error in multi-pair massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5882–5896, Sep. 2017.
- [22] H. Gao, T. Lv, X. Su, H. Yang, and J. M. Cioffi, "Energy-efficient resource allocation for massive MIMO amplify-and-forward relay systems," *IEEE Access*, vol. 4, pp. 2771–2787, 2016.
- [23] R. Zhang, Y. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.
- [24] K. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.
- [25] G. Amarasuriya, C. Tellambura, and M. Ardakani, "Two-way amplify-and-forward multiple-input multiple-output relay networks with antenna selection," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1513–1529, Sep. 2012.
- [26] R. Vaze and R. W. Heath, "On the capacity and diversity-multiplexing tradeoff of the two-way relay channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4219–4234, Jul. 2011.
- [27] S. Jin, X. Liang, K. Wong, X. Gao, and Q. Zhu, "Ergodic rate analysis for multipair massive MIMO two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1480–1491, Mar. 2015.
- [28] H. Cui, L. Song, and B. Jiao, "Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2636–2645, May 2014.
- [29] Y. Dai and X. Dong, "Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5932–5946, Sep. 2016.
- [30] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [31] X. Meng, Y. Chang, Y. Wang, and J. Wu, "Multi-user grouping based scheduling algorithm in massive MIMO uplink networks," in *Proc. IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2018, pp. 409–413.
- [32] L. Liu, Y. Chi, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving MIMO-NOMA: Iterative LMMSE detection," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1758–1773, Feb. 2019.
- [33] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [34] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [35] H. Cramer, *Random Variables and Probability Distributions*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1970.
- [36] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Now, 2004. [Online]. Available: <https://ieeexplore.ieee.org/document/8187216>
- [37] "Inverse-Wishart distribution," Website, 2019, https://en.wikipedia.org/wiki/Inverse-Wishart_distribution.
- [38] S. J. Press, *Applied Multivariate Analysis*. Dover Publications, 2005.



Zhangjie Peng received the Ph.D. degree in Communication and Information Engineering from Southeast University, Nanjing, China, in 2016. He is currently an Associate Professor at the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China.

His research interests include reconfigurable intelligent surface (RIS), cooperative communications, information theory, physical layer security, and machine learning for wireless communications.



Xianzhe Chen received the B.E. degree from the College of Information Science & Electronic Engineering, Zhejiang University, Zhejiang, China, in 2019. He is currently pursuing the M.E. degree at the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China.

His major research interests include massive multiple-input multiple-output systems and relaying communications.



Li-Chun Wang (M'96 – SM'06 – F'11) received the Ph. D. degree from the Georgia Institute of Technology, GA, USA, in 1996. From 1996 to 2000, he was with AT&T Laboratories, NJ, USA, where he was a Senior Technical Staff Member in the Wireless Communications Research Department. Since 2000, he has been with the Department of Electrical and Computer Engineering of National Chiao Tung University, Hsinchu, Taiwan. He holds 26 US patents, and has published over 200 journal and conference papers, and co-edited a book, “*Key Technologies for 5G Wireless Systems*,” (Cambridge University Press 2017). Dr. Wang received Distinguished Research Awards of the National Science Council, Taiwan, in 2012 and 2017, and is a co-recipient of the IEEE Communications Society Asia-Pacific Board Best Award in 2015, the Y. Z. Hsu Scientific Paper Award in 2013, and the IEEE Jack Neubauer Best Paper Award in 1997. His current research interests include software-defined mobile networks, heterogeneous networks, and data-driven intelligent wireless communications.



Wei Xu (Senior Member, IEEE) received the Ph.D. degree from Southeast University, Nanjing, China, in 2009. From 2009 to 2010, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. He is currently a Professor with the National Mobile Communications Research Laboratory, Southeast University. He has coauthored more than 100 refereed journal papers, 36 domestic patents, and four U.S. patents granted. His research interests include information theory, signal processing, and machine learning for wireless communications.

From 2012 to 2017, he was the Editor of the IEEE Communications Letters. He is currently the Editor of the IEEE Transactions on Communications and the Senior Editor of the IEEE Communications Letters. He was the recipient of the Best Paper Award from a number of prestigious IEEE conferences, including the IEEE Globecom and the IEEE ICC. He was the recipient of the Youth Science and Technology Award of China Institute of Communications in 2018.



Lajos Hanzo (Fellow, IEEE) (<http://www-mobile.ecs.soton.ac.uk>, https://en.wikipedia.org/wiki/Lajos_Hanzo) received the D.Sc. degree, holds an honorary doctorate, from the Technical University of Budapest, in 2009, and also from The University of Edinburgh, in 2015. He has published more than 1900 contributions at IEEE Xplore and 18 Wiley-IEEE Press books. He has helped the fast-track career of 119 Ph.D. students. More than 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry. He is a Fellow of Academy of Engineering, of IET, and of EURASIP. He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served as a Governor for both IEEE ComSoc and of VTS.



Cunhua Pan received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively. From 2015 to 2016, he was a Research Associate at the University of Kent, U.K. He held a post-doctoral position at Queen Mary University of London, U.K., from 2016 and 2019, where he is currently a Lecturer.

His research interests mainly include reconfigurable intelligent surfaces (RIS), intelligent reflection surface (IRS), ultra-reliable low latency communication (URLLC), machine learning, UAV, Internet of Things, and mobile edge computing. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019. He is currently an Editor of IEEE Wireless Communication Letters, IEEE Communications Letters and IEEE ACCESS. He also serves as a lead guest editor of IEEE Journal of Selected Topics in Signal Processing (JSTSP) Special Issue on Advanced Signal Processing for Reconfigurable Intelligent Surface-aided 6G Networks.