Towards energy-efficient 6G networks: uplink cell-free massive MIMO with NLD cancellation technique of hardware impairments

Asma Mabrouk, Rafik Zayani

Towards Energy-Efficient 6G Networks: Uplink Cell-free massive MIMO with NLD Cancellation Technique of Hardware Impairments

Asma Mabrouk, Rafik Zayani
CEA-Leti, Université Grenoble Alpes, 38054 Grenoble, France
asma.mabrouk@cea.fr, rafik.zayan@cea.fr

Abstract

This paper focuses on investigating uplink (UL) data transmission in cell-free massive MIMO based on orthogonal frequency division multiplexing (CF-mMIMO-OFDM) systems, taking into account the effects of hardware impairments (HWIs). Specifically, the HWIs arise from nonlinear distortions (NLD) caused by power amplifiers (PAs) at user equipment (UEs). These NLDs have a significant impact on both channel estimation and data transmission in UL CF-mMIMO-OFDM. To mitigate NLDs while maintaining a good power-efficiency, we propose a successive NLD approach that is adequate for CF-mMIMO-OFDM. Specifically, a novel frequency-domain channel estimation method is introduced that incorporates NLD cancellation. This method aims to accurately estimate the channel despite the presence of NLDs. Additionally, a successive combining-aware NLD cancellation is proposed to mitigate the NLD impact on data detection. Not that three combinng schemes are adopted, namely maximum-ratio (MR), zero-forcing (FZF), and partial-FZF (PFZF). Most-importantly, the proposed techniques are designed to be implemented in a distributed and scalable manner, highlighting the advantages of CF-mMIMO-OFDM systems. The performance of the proposed techniques are evaluated with simulations when considereing the the combining schemes. Results show the capability of our proposed NLD cancellation approach to improve both channel estimation and data detection, especially when levreaging the good features of PFZF combing scheme. For objective comparison purpose, we derived closed-form expressions on UL spectral-efficiency (SE) performance of an UL CF-mMIMO-OFDM system in presence of ideal and nonlinear PA.

I. Introduction

Cell-Free massive MIMO (CF-mMIMO) has been shown to be the most promising key technology in the development of beyond 5G networks that are expected to bring new requirements for wireless communication systems, such as higher data rates, ultra-low latency, and massive connectivity [1]–[3]. To this end, a new network architecture is needed that can provide a scalable and flexible infrastructure able to support a wide range of applications and services. Purposely, CF-mMIMO systems do not rely on a traditional cellular network architecture, with the system divided into cells [4]. Instead, a large number of distributed access points (APs) connected to a central processing unit (CPU) are used to provide seamless wireless coverage and capacity across the entire coverage area. Each AP serves all user equipments (UEs) in the same time-frequency resource block. As such, UEs can potentially have access to the same amount of resources regardless of their location. With a high density of APs and UEs, the amount of signaling and data exchanged between nodes increases leading to a substantial increase in computational complexity and resource requirements. Therefore, scalability-related issues should be adressed in CF-mMIMO systems. In particular, user-centric approach enables efficient and dynamic AP selection by UEs to optimize resource allocation and alleviate the challenges posed by a large number of APs and UEs [2], [5]. Based on time-division duplex (TDD) transmission mode, the transmissions in CF-mMIMO are carried out in coherence blocks, each divided into three phases : uplink (UL) training, UL payload transmission and downlink (DL) payload transmission. Obtaining accurate channel estimates is necessary to achieve the full potential of UL/DL transmission in CF-mMIMO by optimizing the design of the combining/precoding schemes at serving APs. Combining schemes are designed to enhance the UL performance efficiency by reducing the inter-user interference and/or enhancing the desired signal of each user. Two levels of cooperation

between APs and the CPU are proposed for UL combining in CF-mMIMO: centralized combining and distributed combining [6]. It is shown that, by leveraging distributed combining, CF-mMIMO systems can achieve better scalability [7].

While CF-mMIMO presents promising potential for enhancing wireless communications, its implementation in real-world scenarios poses practical challenges [3]. Notably, the densification feature of CF-mMIMO can indeed result in higher system efficiency at the cost of a substantial increase in energy consumption and hardware costs. This is especially true if both the APs and UEs are equipped with high-quality hardware components. Nevertheless, while low-cost hardware may be attractive due to its affordability, it often comes with the downside of reduced system efficiency. Hence, it is essential to achieve a balance between the cost of the hardware and the system's performance [8]. The aforementioned works rely on simplifying assumptions such as ideal hardware, which is unlikely to hold in practical implementation. Power amplifier (PA) non-linearity (NL) is considered one of the major HWIs in the analog transmission chain of modern communication systems. Indeed, non-linear distortion (NLD) of non-linear power amplifier (NL-PA) deteriorates both channel estimates and the useful received data.

A. Related works

The impact of potential HWIs on the performance of CF-mMIMO systems has been explored in [9]–[16]. In [9], authors showed that HWIs can significantly degrade the performance of CF-mMIMO systems in terms of spectral efficiency (SE) and energy efficiency (EE). According to the authors, although increasing the number of APs can alleviate the impact of HWIs at APs, the system performance still suffer from HWIs at UEs. The same hardware scaling law was reveled in [10] where closed-form of SE and EE expressions were derived for both UL and DL transmissions. It is also shown that max-min power control algorithm improves the performance of CF-mMIMO systems under HWIs at both APs and UEs. The impact of HWIs on UL CF-mMIMO systems with four low-complexity receiver cooperation levels has been investigated in [11]. The obtained results show that HWIs have a larger impact on the performance of large scale fading decoding (LSFD) combining receiver. In [12], authors studied the effect of HWIs on the UL and DL sum-rate of CF-mMIMO-OFDM systems and revealed that the UL sum-rate is more susceptible to HWIs at UEs than the DL sum-rate. The impact of HWIs, including phase noise, quantization errors, and PA distortion, on the physical layer security of CF-mMIMO systems in the presence of pilot spoofing attack has been presented in [13]. It is shown that the hardware scaling law is almost not applicable to secure CF-mMIMO systems under active attack. The impacts of both phase drifts and distortion noise on CF-mMIMO systems have been investigated in [14]. The achievable rate of CF-mMIMO systems with low resolution analog-to-digital converters at both the APs and users has been investigated in [15]. In [16], authors highlited the performance of fronthal-constrained CF-mMIMO under transceiver HWIs.

B. Contributions

Existing studies have predominantly focused on examining the impact of HWIs at APs and/or UEs on the performance of CF-mMIMO systems. Building upon the findings presented in [10], which indicate that the UL SE is primarily affected by HWIs at the UEs, we investigate the performance of UL OFDM-based CF-mMIMO systems under NL-PA model at UEs. Our work specifically considers a realistic NL-PA model. It is worthnoting that NLD could be compensated either at the UE side or the AP side. The former case may require additional computational complexity, potentially affecting battery life and overall UE performance. This can motivate a receiver cancellation technique which is typically the case in UL transmission where the APs have more resources in terms of power and computational complexity. Therefore, the main contributions of this work are as follows

- We propose a frequency-domain channel estimation approach that incorporates a NLD cancellation technique to improve the efficiency and accuracy of channel estimation by mitigating the impact of nonlinear distortion.

- A new receiver technique, that is able to cancel the NLD caused by NL-PAs, is developed. The proposed technique combines UL detection and NLD cancellation techniques. When applied in a CF-mMIMO-OFDM system, the NLD cancellation technique is evaluated in a realistic scenario, which includes channel estimation error and multi-user interference. Simulation results are provided to demonstrate that the proposed techniques help to improve the overall performance of the CF-mMIMO-OFDM system by reducing the impact of HWIs.
- We analyze the global SE of a CF-mMIMO-OFDM system when the proposed approaches are implemented.
- Unlike [11] and [10], instead of considering the MR combining scheme, we investigate the system performance under HWIs based on different combining techniques: MR, full-pilot zero-forcing (FZF) and partial-FZF (PFZF). The achievable SE of the system is derived for the three combining schemes. We show the ability of our NLD cancellation approach especially when levreaging the good features of PFZF comining scheme that further improve the spectral efficiency by providing an adaptable trade-off between NLD cancelation and boosting of the desired signal
- In addition, we discuss the computational complexity analysis for the proposed schemes. Specifically, we show that the technique converges within a few iterations.

The remainder of this paper is summarized in the following order. Section I describes the considered hardware-constrained CF-mMIMO transmission model and Section II presents the system performance where closed-form of the achievable SE expressions with different combining schemes are derived. The proposed techniques are detailed in Section III. Section IV exhibits simulation results and discussions. At last, Section V concludes this paper.

II. System model

We consider the UL transmission in a TDD-based CF-mMIMO-OFDM system consisting of L APs and K UEs, which are randomly distributed within a given area. UEs and APs are equipped with a single transmit antenna and N receive antennas, respectively. We assume that APs are connected to a CPU via a high-capacity and error-free fronthaul network. Operating in TDD mode, the training phase, UL and DL transmissions fit into the channel coherence block of duration τ_c, as illustrated in Fig. 1. We assume a block-fading channel model where channel coefficients remains unchanged during the coherence block. Let $\mathbf{H}_{k,l} = [\mathbf{h}_{k,l}^0, \mathbf{h}_{k,l}^1, \ldots, \mathbf{h}_{k,l}^I] \in \mathbb{C}^{N \times I}$ be the time-domain channel response between AP$_l$ and UE$_k$ that is modeled as a finite impulse response filter with I equally-spaced channel taps. The i-th channel tap is given by

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

where $\beta_{k,l}$ and $\mathbf{g}_{k,l}^i$ are the large and small scale fading components, respectively. It is assumed that $\mathbf{g}_{k,l}^i \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$ are independent and identically distributed (i.i.d.). We use OFDM modulation to divide the frequency band into multiple subcarriers that are used by UEs in an orthogonal way to transmit data simultaneously. In the frequency-domain, the channel response on the m-th subcarrier between AP$_l$ and UE$_k$ is denoted by $\mathbf{g}_{k,l,m}, m = 1, \ldots, N_{sub}$ where N_{sub} is the number of OFDM subcarriers. Following 5G-NR, based on TDD mode, the UL and DL transmissions share the same frequency band but are separated in time. A frame is typically represented as a grid in the time-frequency plane, where the frequency domain is divided into N_{sub} subcarriers spaced by Δ_f and the time domain is divided into N_c OFDM symbols. Moreover, the data transmission is organized into N_{rb} resource blocks (RB), each consisting of N_{sc} contiguous subcarriers and N_c symbols. Therefore, each RB comprises $N_c N_{sc}$ resource units (RUs), representing the smallest time-frequency resource of one subcarrier and one symbol. To ensure accurate channel estimation, pilot symbols are inserted into RUs of each RB. We assume that the number of OFDM symbols, N_c, covers the coherence time of all UEs. In each RB, a portion of RUs is used for scheduling UL and DL transmissions, where the remaining portion is used for pilot symbols. Specifically, $\xi(\tau_c - \tau_p)$ and $(1-\xi)(\tau_c - \tau_p)$ RUs are used for scheduling UL and DL transmissions, respectively, where $0 < \xi < 1$ [17].

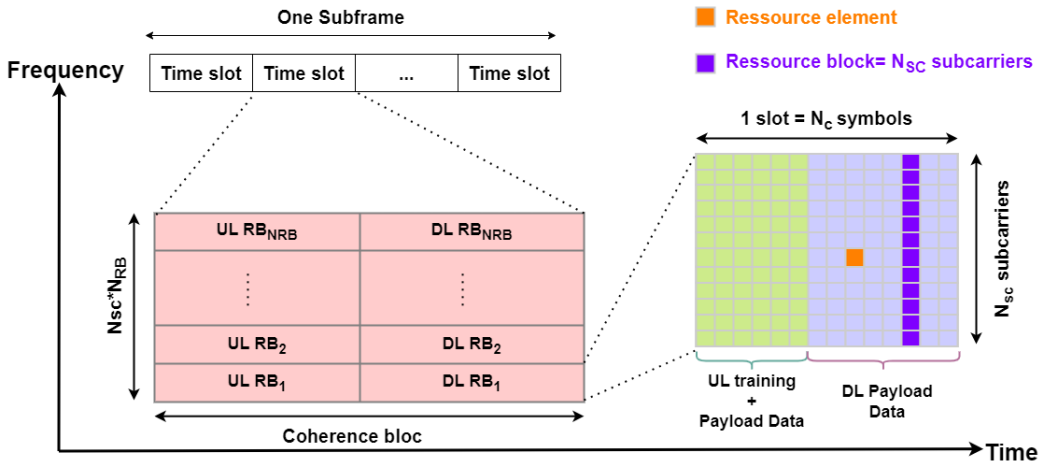


Fig. 1: OFDM TDD frame structure

A. *Hardware impairment model*

In wireless communication systems, HWIs can arise from various components. However, it is worth noting that NL-PAs are particularly prominent and significant sources of impairments. The nonlinearity encountered in a PA can be described through two types of conversion: amplitude to amplitude (AM/AM) and amplitude to phase (AM/PM). The relation between the input, r_{in}, and output, r_{out}, signals of a PA is described by its transfer function, $f(.)$, which is given by

$$r_{out} = f(r_{in}) = \Omega(\alpha\rho)\exp^{j(\psi+\Psi(\alpha\rho))} \tag{2}$$

where $\Omega(.)$ and $\Psi(.)$ denote the AM/AM and AM/PM conversions, respectively, and α is the multiplicative coefficient that is applied to the input of PA to achieve the desired PA operating point based on the specified input back-off (IBO). The IBO refers to the amount of power reduction at the input of the PA relative to its maximum rated power level. Moreover, ρ and ψ are the magnitude and the phase of the PA's input sample. To accurately model the behavior of NL-PAs, we use the memoryless modified Rapp model based on which the AM/AM and AM/PM conversions are given by [18]

$$\Omega(\rho) = \frac{G\rho}{\left(1+\left|\frac{G\rho}{V_{sat}}\right|^{2p}\right)^{\frac{1}{2p}}}, \qquad \Psi(\rho) = \frac{A\rho^q}{\left(1+\left(\frac{\rho}{B}\right)^q\right)} \tag{3}$$

where G is the small signal gain, V_{sat} is the saturation level, p is the smoothness factor and A, B and q are fitting parameters. Note that $\alpha = \frac{V_{sat}}{G\sqrt{p_{in}}}10^{\frac{-IBO[dB]}{10}}$, where p_{in} is the signal average power. Based on the Bussgang's theorem [19], the output of a NL-PA can be expressed in the following form

$$r_{out} = \kappa_0 r_{in} + d \tag{4}$$

where κ_0 denotes a complex gain and d is the added zero-mean distortion noise with variance σ_d^2. Note that, d is uncorrelated with the input signal, r_{in}. It is worth noticing that the Bussgang theorem holds only for Gaussian input signals. This condition is verified in the proposed scheme, since an OFDM modulated signal is amplified at each UE. In addition, it is mentionning that d is not Gaussian at the output of the PA but it becomes Gaussian at the receiver side (AP) after the OFDM demodulation.

B. *Uplink training phase*

In wireless CF-mMIMO-OFDM systems, pilot-based channel estimation is a common technique used to estimate the users' at the APs. For simplicity, we assume that the channel coefficients are the same

for all RUs within a RB. Therefore, channel estimation can be performed once per RB. During the UL training phase, UEs transmit orthogonal frequency-domain sequences of length τ_p through the first τ_p time-domain OFDM symbols to the APs. It is important to highlight that orthogonal RBs are assigned to each user in order to enable simultaneous transmission of pilots without causing interference with other UEs. This approach helps to mitigate pilot contamination and ensure accurate channel estimation.

Let $\phi_{k,m} \in \mathbb{C}^{\tau_p \times 1}$ denotes the pilot sequence assigned to the k-th user, where $\phi_{k,m}\phi_{k,m} = \tau_p$ and $\phi_{k,m}\phi_{k',m} = 0$, $\forall k \neq k'$. Considering NL-PA at UEs, the recieved signal, $\mathbf{Y}^p_{l,m} \in \mathbb{C}^{N \times \tau_p}$, at the l-th AP is given by

$$\mathbf{Y}^p_{l,m} = \mathbf{H}_{l,m}\left(\mathbf{P}\mathbf{\Phi}^H_m + \mathbf{D}_m\right) + \mathbf{N}_{l,m}, \tag{5}$$

where $\mathbf{P} = \mathrm{diag}(\kappa_0\sqrt{\mathrm{p}_1}, \ldots, \kappa_0\sqrt{\mathrm{p}_K})$ is the transmit power matrix. Note that the output power of the NL-PA circuit of the k-th is a linear scale of the input power p_k. $\mathbf{N}_{l,m} \in \mathbb{C}^{N \times \tau_p}$ is the AWGN matrix with variance σ^2 and $\mathbf{D}_m = [\mathbf{d}_{1,m}, \ldots, \mathbf{d}_{K,m}]$ is the matrix that contains the NLD vector of each user on the m-th subcarrier. This fisrt step to estimate the channel $\mathbf{H}_{l,m}$ is performing a de-spreading of the received pilot signal as follows [20]

$$\overline{\mathbf{H}}_{l,m} = \frac{1}{\sqrt{\tau_p}}\mathbf{Y}^p_{l,m}\mathbf{\Phi}_m \in \mathbb{C}^{N \times \tau_p}, \tag{6}$$

where $\mathbf{\Phi}_m = [\phi_{1,m}, \ldots, \phi_{\tau_p,m}] \in \mathbb{C}^{\tau_p \times \tau_p}$ represents the full-rank matrix of the channel estimates. Then, based on the estimation theory in [21], by adopting the minimum mean square error (MMSE) estimation of the channel response, $\mathbf{h}_{k,l,m}$ can be expressed as

$$\hat{\mathbf{h}}_{k,l,m} = \mathrm{c}_{k,l}\overline{\mathbf{H}}_{l,m}\mathbf{e}_k, \tag{7}$$

where

$$\mathrm{c}_{k,l} = \frac{\sqrt{\mathrm{p}_k\tau_p}\kappa_0\beta_{k,l}}{\mathrm{p}_k\tau_p|\kappa_0|^2\beta_{k,l} + \sum_{t=1}^{K}\sigma_d^2\beta_{t,l} + \sigma^2}, \tag{8}$$

represents the full-rank matrix of the channel estimates. The channel estimate, $\hat{\mathbf{h}}_{k,l,m}$, and the channel estimation error, $\tilde{\mathbf{h}}_{kl} = \mathbf{h}_{k,l,m} - \hat{\mathbf{h}}_{k,l,m}$, are independent with distributions $\mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \gamma_{k,l}\mathbf{I}_N\right)$ and $\mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, (\beta_{k,l} - \gamma_{k,l})\mathbf{I}_N\right)$, respectively, where

$$\gamma_{k,l} = \sqrt{\mathrm{p}_k\tau_p}\kappa_0^*\beta_{k,l}\mathrm{c}_{k,l} = \frac{\mathrm{p}_k\tau_p|\kappa_0|^2\beta_{k,l}^2}{\mathrm{p}_k\tau_p|\kappa_0|^2\beta_{k,l} + \sum_{t=1}^{K}\sigma_d^2\beta_{t,l} + \sigma^2} \tag{9}$$

We note form (9) the estimated channel of each UE is affected by NLD caused by both itself and other users in the network. Moreover, the level of NLD caused by other UEs varies based on their channel conditions.

III. Performance analysis

In this section, we evaluate the achievable SE of distributed UL transmission in CF-mMIMO system that consists of two stages: local processing and large scale fading decoding (LSFD) [6]. First, each AP independently estimates the channels and designs its combiner vector to perform data detection locally. The estimated data at all APs are then sent to the CPU that performs linear processing for joint detection. We consider LSFD technique at the CPU which relies solely on channel statistics since the channel estimates are not shared over the fronthaul links. Moreover, We study the system performance based on the following receive combining schemes: MR, FZF and PFZF.

A. Uplink data transmission

During the UL data transmission, the received data signal at the l-th AP on the m-th subcarrier is given by

$$\mathbf{y}_{l,m}^{\mathrm{ul}} = \sum_{k=1}^{K} \mathbf{h}_{k,l,m} \left(\sqrt{\mathrm{p}_k} \kappa_0 \mathrm{s}_{k,m} + \mathrm{d}_{k,m} \right) + \mathbf{n}_{l,m}, \tag{10}$$

where $\mathrm{s}_{k,m}$ is the transmitted data by the k-th user on the m-th subcarrier and $\mathbf{n}_{l,m} \in \mathbb{C}^{N \times 1}$ is the AWGN matrix with variance σ^2. The local data estimate of the k-th UE at the l-th AP based on the combining vector $\mathbf{v}_{k,l,m}$ is given by

$$\begin{aligned} \hat{s}_{k,l,m} &= \mathbf{v}_{k,l,m}^H \mathbf{y}_{l,m}^{\mathrm{ul}}, \\ &= \mathbf{v}_{k,l,m}^H \mathbf{h}_{k,l,m} (\sqrt{\mathrm{p}_k} \kappa_0 \mathrm{s}_{k,m} + \mathrm{d}_{k,m}) + \sum_{t=1, t \neq k}^{K} \sqrt{\mathrm{p}_t} \kappa_0 \mathbf{v}_{k,l,m}^H \mathbf{h}_{t,l,m} (\sqrt{\mathrm{p}_k} \kappa_0 \mathrm{s}_{t,m} + \mathrm{d}_{t,m}) + \mathbf{v}_{k,l,m}^H \mathbf{n}_{l,m} \end{aligned} \tag{11}$$

Local data estimates are then sent to the CPU to be linearly combined using the LSFD coefficients. The received signal at the CPU is given by

$$\begin{aligned} \hat{s}_{k,m} &= \sum_{l=1}^{L} a_{k,l,m}^* \hat{s}_{k,l,m}, \\ &= \underbrace{\sum_{l=1}^{L} \sqrt{\mathrm{p}_k} \kappa_0 a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{h}_{k,l,m} \times \mathrm{s}_{k,m}}_{\text{Desired signal}} + \underbrace{\sum_{l=1}^{L} a_{k,l,m}^* \sum_{t=1, t \neq k}^{K} \sqrt{\mathrm{p}_t} \kappa_0 \mathbf{v}_{k,l,m}^H \mathbf{h}_{t,l,m} \times \mathrm{s}_{t,m}}_{\text{Multi-User interference}} + \underbrace{\sum_{l=1}^{L} a_{k,l,m}^* \sum_{t=1}^{K} \mathbf{v}_{k,l,m}^H \mathbf{h}_{t,l,m} \times \mathrm{d}_{t,m}}_{\text{Transievers' Hardware Distortion}} \\ &+ \underbrace{\sum_{l=1}^{L} a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{n}_{l,m}}_{\text{AWGN}} \end{aligned} \tag{12}$$

where $a_{k,l,m}$ is the complex LSFD coefficient for l-th AP and k-th UE on the m-th subcarrier. As to maximize the SINR of UE_k, $\forall k \in \{1, \ldots, K\}$, the LSFD coefficients are computed as follows

$$\mathbf{a}_{k,m} = \left(\sum_{t=1}^{K} p_k \mathbb{E}\{\mathbf{g}_{k,t,m} \mathbf{g}_{k,t,m}^H\} + \sigma^2 \mathbf{F}_k \right)^{-1} \mathbb{E}\{\mathbf{g}_{k,k,m}\}, \tag{13}$$

where $\mathbf{g}_{k,t,m} = [\mathbf{v}_{k,1,m}^H \mathbf{h}_{t,1,m}, \ldots, \mathbf{v}_{k,L,m}^H \mathbf{h}_{t,L,m}]^T$ and $\mathbf{F}_k = \mathrm{diag}\left(\{\|\mathbf{v}_{k,1,m}\|^2\}, \ldots, \{\|\mathbf{v}_{k,L,m}\|^2\}\right)$.

B. Local receive combiners

In this subsection, we present the combining schemes considered in this investigation : MR combining, FZF combining, and PFZF combining.

1) MR combining is the simplest receiving scheme that maximizes the power of the desired signal while neglecting the inter-user interference. The MR combing vector is given by

$$\mathbf{v}_{k,l,m}^{\mathrm{MR}} = c_{k,l} \overline{\mathbf{H}}_{l,m} \mathbf{e}_k = \hat{\mathbf{h}}_{k,l,m} \tag{14}$$

Based on (14), we can note that this combining scheme has a low computational complexity since the combining vector is the corresponding local channel estimate and no additional computation is needed.

2) FZF is an efficient UL combining scheme since it enables the inter-user interference cancellation while maintaining strong desired signal powers. The FZF combiner vector is given by

$$\mathbf{v}_{k,l,m}^{\mathrm{FZF}} = c_{k,l} \theta_{k,l} \overline{\mathbf{H}}_{\mathrm{l,m}} \left(\overline{\mathbf{H}}_{\mathrm{l,m}}^{\mathrm{H}} \overline{\mathbf{H}}_{\mathrm{l,m}} \right)^{-1} \mathbf{e}_{\mathrm{k}}, \tag{15}$$

where $\theta_{k,l} = \mathbb{E}\{|\overline{\mathbf{H}}_{l,m}\mathbf{e}_k|^2\} = \frac{\gamma_{k,l}}{c_{k,l}^2}$. In this combining scheme, one combiner vector is proposed per pilot symbol. The FZF could be used only if $N \geq \tau_p$. Moreover, while FZF combining can eliminate interference, it can also lead to noise amplification.

3) PFZF is a modified version of the FZF Combining scheme that alleviates the constraint on the number of antennas at the APs versus the number of pilots. In PFZF combining, each AP divides the UEs into two disjoint subsets : $\mathcal{S}_l$ of strong users and $\mathcal{W}_l$ of weak users.

$$\mathcal{S}_l = [j_{l1}, \ldots, j_{l\nu_l}] = \underset{\downarrow k \in \{1,\ldots,K\}}{\text{argsort}} \{\beta_{k,l}\}, \tag{16}$$

$$\mathcal{W}_l = \{\beta_{k,l}\} \backslash \mathcal{S}_l. \tag{17}$$

We use the same strategy as in [22] to define the value of $\nu_l = |\mathcal{S}_l|$. The strong users are assigned to the FZF combiner, while the weak users are assigned to the MR combiner. Hence, the PFZF combining vector is given by

$$\mathbf{v}_{k,l,m}^{\text{PFZF}} = \begin{cases} \mathbf{v}_{k,l,m}^{\text{FZF}} & \text{if} \quad k \in \mathcal{S}_l \\ \mathbf{v}_{k,l,m}^{\text{MR}} & \text{if} \quad k \in \mathcal{W}_l \end{cases} \tag{18}$$

The choice of the combiner scheme depends on the specific system requirements, such as the number of APs, the channel conditions, and the desired performance metrics.

C. Uplink Spectral Efficiency

The CPU in a cell-free massive MIMO system is responsible for coordinating the transmissions from all the APs, and optimizing the overall system performance. The signal in (12) can be rewritten as

$$\hat{s}_{k,m} = \underbrace{\mathbb{E}\left\{\sum_{l=1}^{L} \sqrt{\mathrm{p}_k}\kappa_0 a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{h}_{k,l,m}\right\}}_{\text{DS}_{k,m}} \mathrm{s}_{k,m} + \underbrace{\sqrt{\mathrm{p}_k}\kappa_0 \left(\sum_{l=1}^{L} a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{h}_{k,l,m} - \mathbb{E}\left\{\sum_{l=1}^{L} a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{h}_{k,l,m}\right\}\right)}_{\text{BU}_{k,m}} \mathrm{s}_{k,m}$$

$$+ \sum_{t=1,t\neq k}^{K} \underbrace{a_{k,l,m}^* \sum_{l=1}^{L} \sqrt{\mathrm{p}_t}\kappa_0 \mathbf{v}_{k,l,m}^H \mathbf{h}_{t,l,m}}_{\text{MUI}_{k,t,m}} \mathrm{s}_{t,m} + \sum_{t=1}^{K} \underbrace{\sum_{l=1}^{L} a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{h}_{t,l,m} \times \mathrm{d}_{t,m}}_{\text{TDI}_{k,m}} + \underbrace{\sum_{l=1}^{L} a_{k,l,m}^* \mathbf{v}_{k,l,m}^H \mathbf{n}_{l,m}}_{\text{N}_{k,m}} \tag{19}$$

The detected signal at the CPU comprises five main parts : $\text{DS}_{k,m}$ and $\text{BU}_{k,m}$ represent the desired signal and the beamforming uncertainty for the k-th user, respectively. Moreover, the user-interference caused by the t-th user $(t \neq k)$ is denoted by $\text{MUI}_{k,t,m}$. The term $\text{TDI}_{k,m}$ presents the total received hardware distortion form all UEs in the network. Finally, $\text{N}_{k,m}$ represents the total received noise. Based on (19), the SINR of UE_k is given by

$$\text{SINR}_{k,m} = \frac{|\text{DS}_{k,m}|^2}{\mathbb{E}\{|\text{BU}_{k,m}|^2\} + \sum_{t=1,t\neq k}^{K} \mathbb{E}\{|\text{MUI}_{k,t,m}|^2\} + \sum_{t=1}^{K} \mathbb{E}\{|\text{TDI}_{k,m}|^2\} + \mathbb{E}\{|\text{N}_{k,m}|^2\}}, \tag{20}$$

(21)

Note that, the SINR expressions in (20) and (21) holds for any combiner used at the AP. The closed-form expression of the SINR of UE_k is given by

$$\text{SINR}_{k,m} = \frac{\left|\sum_{l\in\mathcal{A}_{k,m}} a_{k,l,m}^* \gamma_{k,l} + \sum_{l\in\mathcal{B}_{k,m}} N a_{k,l,m}^* \gamma_{k,l}\right|^2}{\text{A}_{k,m} + \text{B}_{k,m} + \text{C}_{k,m} + \text{D}_{k,m}}, \tag{22}$$

where

$$\mathrm{A}_{k,m} = \sum_{t=1}^{K} \mathrm{p}_t \left(\sum_{l \in \mathcal{A}_{k,m}} |a_{k,l,m}^*|^2 \frac{\gamma_{k,l}(\beta_{t,l} - \gamma_{t,l})}{(N - |\mathcal{S}_l|)} + \sum_{l \in \mathcal{B}_{k,m}} |a_{k,l,m}^*|^2 N \gamma_{k,l} \beta_{t,l} \right),$$

$$\mathrm{B}_{k,m} = \mathrm{p}_t \left(\sum_{l \in \mathcal{A}_{k,m}} a_{k,l,m}^* \gamma_{t,l} + \sum_{l \in \mathcal{B}_{k,m}} a_{k,l,m}^* N \gamma_{t,l} \right),$$

$$\mathrm{C}_{k,m} = \sigma_d^2 \left(\sum_{l \in \mathcal{A}_k} a_{k,l,m}^* \gamma_{k,l} + \sum_{l \in \mathcal{B}_k} a_{k,l,m}^* N \gamma_{k,l} \right) + \sum_{t=1}^{K} \sigma_d^2 \left(\sum_{l \in \mathcal{A}_k} |a_{k,l,m}^*|^2 \frac{\gamma_{k,l}(\beta_{t,l} - \gamma_{t,l})}{(N - |\mathcal{S}_l|)} + \sum_{l \in \mathcal{B}_k} |a_{k,l,m}^*|^2 N \gamma_{k,l} \beta_{t,l} \right),$$

$$\mathrm{D}_{k,m} = \sigma^2 \left(\sum_{l \in \mathcal{A}_k} |a_{k,l,m}^*|^2 \frac{\gamma_{k,l}}{(N - |\mathcal{S}_l|)} + \sum_{l \in \mathcal{B}_k} |a_{k,l,m}^*|^2 N \gamma_{k,l} \right),$$

$$\mathcal{A}_k = \{l = 1, \ldots, L, k \in \mathcal{S}_l\},$$

$$\mathcal{B}_k = \{l = 1, \ldots, L, k \in \mathcal{W}_l\}.$$

From the users' perspective, $\mathcal{A}_k$ denotes the set of APs considering UE_k as a strong user and $\mathcal{B}_k$ refers to the set of APs considering UE_k as a weak user. Therefore, $|\mathcal{A}_k| + |\mathcal{B}_k| = L$. We note the $\mathcal{A}_k = \emptyset$ when MR combining scheme is used. However, when the FZF receive combining is used when $\mathcal{B}_k = \emptyset$.

The achievable SE of the k-th user for the UL CF-mMIMO system is given by

$$\mathrm{SE}_k = \xi \Delta_f N_{sc} \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + \mathrm{SINR}_{\mathrm{k,m}}\right), \tag{23}$$

where Δ_f is the subcarrier spacing (SCS) and ξ denotes the portion of coherence time interval dedicated to the UL data transmission phase.

IV. Distributed NLD cancellation schemes

NL-PA at UEs impacts both channel estimation and UL transmission. We propose two NLD cancellation algorithms that help to improve the accuracy of the channel estimation and reduce the NLD impact on the UL transmission phase, leading to improved system performance. As shown in Fig. 2, the NLD cancellation is be performed locally. In fact, performing NLD cancellation at the APs can simplify the overall system architecture, as both the UEs and the CPU do not need to perform any additional processing for NLD cancellation. This can reduce the overall complexity of the system and make it easier to implement and maintain.

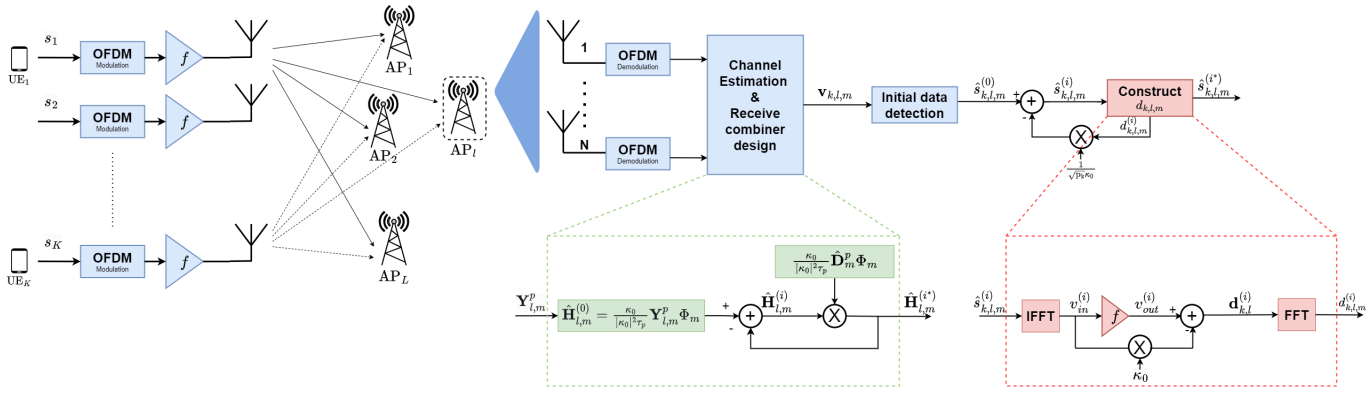


Fig. 2: CF-mMIMO-OFDM system with NLD cancellation algorithms. The algorithm is implemented in a distributed manner, i.e, it is executed locally at each AP.

A. Disributed channel estimation with NLD cancellation

As the presence of NLD can cause errors in the estimation of the channel state information, we propose a NLD cancellation technique where each AP estimates the distortion of each user separately. This is because the non-linear distortion introduced by each user's PA is different, and needs to be estimated and removed separately to mitigate its impact on their received signals. We note that channel estimation is performed in the frequency domain. The proposed method is applied at the AP side after the initial MMSE estimation and is depicted in Algorithm 1 and is summarized in the following steps:

- Distortion estimate (step 3): The AP knows the PA transfer function and the assigned pilot sequence to each user. Then the distortion introduced by the k-th user can be computed, based on the recovered time-domain pilot sequences, and transformed to the frequency domain as follows

$$\hat{\mathbf{D}}^{\mathrm{p}} = \mathrm{FFT}(f(\mathrm{IFFT}(\mathbf{\Phi})) - \kappa_0 \mathrm{IFFT}(\mathbf{\Phi})) \tag{24}$$

  This estimated distortion will be used in the iterative process to remove the received distortion from the k-th user which is influenced by the characteristics of the wireless channel.
- Initial channel estimates (step 5): The cancellation process is based on the initial channel estimates obtained from the received pilot signal in (6)

$$\hat{\mathbf{H}}_{l,m}^{(0)} = \mathbf{W}\overline{\mathbf{H}}_{l,m} = \mathbf{H}_{l,m}\mathbf{P}' + \mathbf{W}\mathbf{H}_{\mathrm{l,m}}\mathbf{D}_m\mathbf{\Phi}_m + \mathbf{W}\mathbf{N}_{l,m}\mathbf{\Phi}_m,$$

  where $\mathbf{W} = \frac{\kappa_0}{|\kappa_0|^2\tau_p}\mathbf{I}_K$ and $\mathbf{P}' = \mathrm{diag}(\sqrt{\mathrm{p}_1}, \ldots, \sqrt{\mathrm{p}_K})$. For a given user UE_k, the first estimated channel at AP_l is given by

$$\hat{\mathbf{h}}_{k,l,m}^{(0)} = \sqrt{\mathrm{p}_k}\mathbf{h}_{k,l,m} + \underbrace{\frac{\kappa_0}{|\kappa_0|^2\tau_p}\mathbf{h}_{k,l,m}\mathbf{d}_{k,m}\phi_{k,m}}_{\text{received individual distortion}} + \frac{\kappa_0}{|\kappa_0|^2\tau_p}\sum_{t=1,t\neq k}^{K}\mathbf{h}_{t,l,m}\mathbf{d}_{t,m}\phi_{k,m} + \frac{\kappa_0\mathbf{N}_{l,m}\phi_{k,m}}{|\kappa_0|^2\sqrt{\tau_p}}. \tag{25}$$

  The received distortion is correlated with the channel of the user, which means that the distortion estimation process needs to take the channel information into account.
- Distortion removal (step 10): The AP then uses an iterative cancellation process to remove the received distortion from the received signal. In the i-th iteration, it estimates the total NLD term by multiplying the estimated distortion by the estimated channel in the (i-1)th iteration.
- Stop criterion (step 11): The iterative cancellation process is repeated until either a maximum number of iterations is reached, $i^* = N_{itr}$ or until no further cancellation is done between two successive iterations, $i^* < N_{itr}$.

Algorithm 1 Channel Estimation correction at the l-th AP

1: Input: $\mathbf{Y}_l^p, \Phi, N_{itr}$
2: Output: Estimated channels with distortion cancellation $\hat{\mathbf{H}}_{l,m}^{(i^*)}$
3: $\hat{\mathbf{D}}^{\mathrm{p}} = \mathrm{FFT}(f(\mathrm{IFFT}(\mathbf{\Phi})) - \kappa_0 \mathrm{IFFT}(\mathbf{\Phi}))$
4: **for** $m \leftarrow 1$ to N_{sc} **do**
5: $\quad\hat{\mathbf{H}}_{l,m}^{(0)} = \frac{\kappa_0}{|\kappa_0|^2\tau_p}\mathbf{Y}_{l,m}^p\mathbf{\Phi}_m$
6: $\quad$**for** $i \leftarrow 1$ to $(N_{itr}-1)$ **do**
7: $\quad\quad\hat{\mathbf{H}}_{l,m}^{(i)} = \hat{\mathbf{H}}_{l,m}^{(0)} - \frac{\kappa_0}{|\kappa_0|^2\tau_p}\hat{\mathbf{H}}_{l,m}^{(i-1)}\hat{\mathbf{D}}_m^{\mathrm{p}}\mathbf{\Phi}_m$
8: $\quad\quad$**if** $\|\hat{\mathbf{H}}_{l,m}^{(i)} - \hat{\mathbf{H}}_{l,m}^{(i-1)}\|^2 <= 10^{-10}$ **then**
9: $\quad\quad\quad$**break**
10: $\quad\quad$**end if**
11: $\quad$**end for**
12: **end for**

B. Disributed UL data detection with NLD cancellation

The distributed NLD cancellation is illustrated in **Algorithm. 2**. The block diagram of the proposed algorithm is shown in Fig. 2. In the UL data transmission, after receiving the signals from the user equipment (UE), each AP performs receive combining to obtain a combined signal that contains contributions from multiple UEs. The combined signal is then processed to estimate and remove the NLD introduced by the non-linear power amplifiers (PAs) of the UEs.

After combining the received signals from all users, the resulting signal at each AP contains the NLD introduced by the power amplifiers (PAs) of the users. The goal of the NLD cancellation process is to estimate and remove this distortion, in order to improve the quality of the received signal. The process of NLD cancellation technique involves two main parts: individual distortion estimation and total distortion removal. The first step is to estimate the nonlinear distortion introduced by each user. This estimation is performed using an iterative algorithm. Since the detected signal of each user is affected by the NLD received from all users, once the distortion introduced by each user has been estimated, the second step aims to cancel out the NLD effects caused by all users' transmission. This process is carried out at each AP for each user for all sub-carriers per-RB.

Algorithm 2 NLD cancellation at the l-th AP

1: **Input :** $\hat{\mathbf{s}}_{k,l}, N_{itr}$
2: **Output :** Estimated data with NLD cancellation $\hat{\mathbf{s}}_{k,l}^{(i^*)}$
3: **Initialization :** $\hat{s}_{k,l,m}^{(0)} = \frac{\hat{s}_{k,l,m}}{\kappa_0\sqrt{\mathrm{p}_k}\gamma_{k,l}\mathbb{1}_{k\in\mathcal{G}_l}+\kappa_0\sqrt{\mathrm{p}_k}\hat{\mathbf{h}}_{k,l,m}^H\hat{\mathbf{h}}_{k,l,m}\mathbb{1}_{k\in\mathcal{P}_l}}$, $\forall k\ \forall m$
4: **for** $k \leftarrow 1$ to K **do**
5: **for** $i \leftarrow 0$ to $(N_{itr}-1)$ **do**
6: $v_{in}^{(i)} = \mathrm{IFFT}\left(\hat{\mathbf{s}}_{k,l}^{(i)}\right)$
7: $\hat{\mathbf{d}}_{k,l}^{(i)} = \mathrm{FFT}\left(f(v_{in}^{(i)}) - \kappa_0 v_{in}^{(i)}\right)$
8: **for** $m \leftarrow 1$ to N_{sub} **do**
9: $\hat{\mathbf{s}}_{k,l,m}^{(i+1)} = \hat{\mathbf{s}}_{k,l,m}^{(0)} - \frac{\hat{\mathrm{d}}_{k,l,m}^{(i)}}{\sqrt{\mathrm{p}_k}\kappa_0}$
10: **end for**
11: **if** $\|\hat{\mathbf{s}}_{k,l}^{(i+1)} - \hat{\mathbf{s}}_{k,l}^{(i)}\|^2 <= 10^{-10}$ **then**
12: **break**
13: **end if**
14: **end for**
15: **end for**
16: $i^* = i$
17: **for** $k \leftarrow 1$ to K **do**
18: **for** $m \leftarrow 1$ to N_{sc} **do**
19: $\hat{s}_{k,l,m}^{(i^*)} = \hat{s}_{k,l,m} - \sum_{t=1}^{K} \mathbf{v}_{k,l,m}^H \hat{\mathbf{h}}_{t,l,m} \times \hat{\mathrm{d}}_{t,l,m}^{(i^*)}$
20: **end for**
21: **end for**

Individual NLD estimation

Total NLD cancellation

In the following, we detail the NLD cancellation process for a given user. Before starting the iterative process, it is important to note that a specific normalization step that depends on the used combining scheme is needed (step 3). The aim here is to separate the signal and distortion components of the user of interest in its initial detected signal from other components (i.e, multi-user interference, received NLD

from other users and the thermal noise). The resulting signal after this step is given by

$$\hat{s}^{(0)}_{k,l,m} = \frac{\hat{s}_{k,l,m}}{\eta_{k,l,m}} = \sum_{t=1}^{K} \frac{\mathbf{v}^H_{k,l,m}\mathbf{h}_{t,l,m}}{\eta_{k,l,m}} \left(\sqrt{\mathrm{p}_t}\kappa_0 \mathrm{s}_{t,m} + \mathrm{d}_{t,m}\right) + \frac{\mathbf{v}^H_{k,l,m}\mathbf{n}_{l,m}}{\eta_{k,l,m}}, \tag{26}$$

$$= \mathrm{s}_{k,m} + \frac{\mathrm{d}_{k,m}}{\sqrt{\mathrm{p}_k}\kappa_0} + \sum_{t=1,t\neq k}^{K} \sqrt{\mathrm{p}_t}\kappa_0 \frac{\mathbf{v}^H_{k,l,m}\mathbf{h}_{t,l,m}}{\eta_{k,l,m}} \times \mathrm{s}_{t,m} + \sum_{t=1,t\neq k}^{K} \frac{\mathbf{v}^H_{k,l,m}\mathbf{h}_{t,l,m}}{\eta_{k,l,m}} \times \mathrm{d}_{t,m} + \frac{\mathbf{v}^H_{k,l,m}\mathbf{n}_{l,m}}{\eta_{k,l,m}},$$

where

$$\eta_{k,l,m} = \sqrt{\mathrm{p}_k}\kappa_0 \mathbf{v}^H_{k,l,m}\mathbf{h}_{k,l,m} = \begin{cases} \sqrt{\mathrm{p}_k}\kappa_0 \hat{\mathbf{h}}^H_{k,l,m}\hat{\mathbf{h}}_{k,l,m} & \text{MR} \\ \sqrt{\mathrm{p}_k}\kappa_0 \gamma_{k,l} & \text{FZF} \\ \kappa_0\sqrt{\mathrm{p}_k}\gamma_{k,l}\mathbb{1}_{k\in\mathcal{S}_l} + \kappa_0\sqrt{\mathrm{p}_k}\hat{\mathbf{h}}^H_{k,l,m}\hat{\mathbf{h}}_{k,l,m}\mathbb{1}_{k\in\mathcal{W}_l} & \text{PFZF} \end{cases} \tag{27}$$

The above signal is the input signal of the iterative algorithm. At each iteration, first, the distortion introduced the k-th user, $d_{k,l,m}$, is estimated by imitating the behavior of PAs and then the term $\frac{d_{k,l,m}}{\sqrt{\mathrm{p}_t}\kappa_0}$ is subtracted from the initial detected signal, $\hat{s}^{(0)}_{k,l,m}$. The iterative process of the algorithm is basically based on three steps :

- (step 6) : The input signal is OFDM modulated into time-domain by appliying the IFFT operation.
- (step 7) : Resulting signal from (step 6) is amplified using the known PA model and then subtracted from its original version. Thus, we are constructing the NLD part introduced by the k-th user in the time-domain then reproduce its frequency-domain representation via the FFT operation.
- (step 9) : Finally, the estimated NLD of the k-th user, $\hat{\mathbf{d}}^{(i)}_{k,l}$, is subtracted from the initial normalized detected signal in (26) for all subcarriers. As such, its NLD is decreased in an iterative manner.

In this way, if the AP was able to estimate the distortion, the outcome of i-th iteration is less-distorted compared to that of the (i-1)-th iteration and the algorithm tends to converge to the nondistorted data after a number of iterations. The iterative process ends either when the maximum number of iterations is reached, $i^* = N_{itr}$, or when almost the same corrected signals are achieved in two consecutive iterations (step 11), $i^* < N_{itr}$. Once the NLD introduced by each user has been estimated, the second step is to subtract this distortion from the initial detected signal of each user for all subcarriers based on channel estimates (step 19).

We can observe from (26) that the effectiveness of NLD cancellation in UL CF-mMIMO is indeed influenced by the Signal-to-Interference-plus-Noise Ratio (SINR) of the received signal. Particularly, a higher SINR generally allows for better estimation of the NLD introduced by each user. In fact, when the SINR is high, it means that the desired user's signal is strong compared to the interference and noise levels. In this case, the system has a clearer and more distinguishable signal from the desired user, making it easier to estimate introduced by that user and and more effective total non-linear cancellation.

C. Complexity Analysis of the proposed techniques

In this section, the computational complexity analysis is provided for the proposed schemes compared with the conventional UL channel estimation and data detection schemes. In terms of complexity analysis, we consider the number of complex operations needed for the proposed algorithms. The added complexity of the proposed schemes are given as follows

1) Distributed channel estimation with NLD cancellation : The added computational complexity to channel estimation process arises from two main operations in the algorithm: the IFFT/FFT (Steps 3) in the initial NLD estimation and the iterative NLD subtraction (Step 7). The complexity of the IFFT/FFT is $\mathcal{O}(N_{sc}\log N_{sc})$, while the NLD substruction operation has a complexity of $\mathcal{O}(N_{itr}(NK\tau_p + N^2\tau_p))$.
2) Distributed UL data detection with NLD cancellation : Regarding the proposed individual NLD estimation algorithm, it is evident that the additional computational, compared to data detection

in the traditional CF-mMIMO combining scheme, mainly stems from two key operations: the IFFT/FFT (Steps 6 and 7) and the NLD substruction (step 9). The complexity of the IFFT/FFT is $\mathcal{O}(N_{sc} \log N_{sc})$, while the NLD substruction operation has a complexity of $\mathcal{O}(N_{sc})$. Thus the total additional computational complexity per iteration is given by $\mathcal{O}(N_{sc} + 2N_{sc} \log(N_{sc}))$. The second part of the algorithm, total NLD cancellation, the complexity is $\mathcal{O}(N^2)$ for each user at each subcarrier. The parts of the algorithm are completely independent, then the computational complexity of the overall data NLD cancellation process is the sum of their individual complexities.

D. Convergence Analysis

In this section, the convergence analysis of the proposed iterative process in Algorithm 1 and Algorithm 2 are provided. For the sake of illustration, we consider the system configuration illustrated in Fig. 3 with 100 APs and 15 UEs randomly distributed within an area of size 1km×1km.

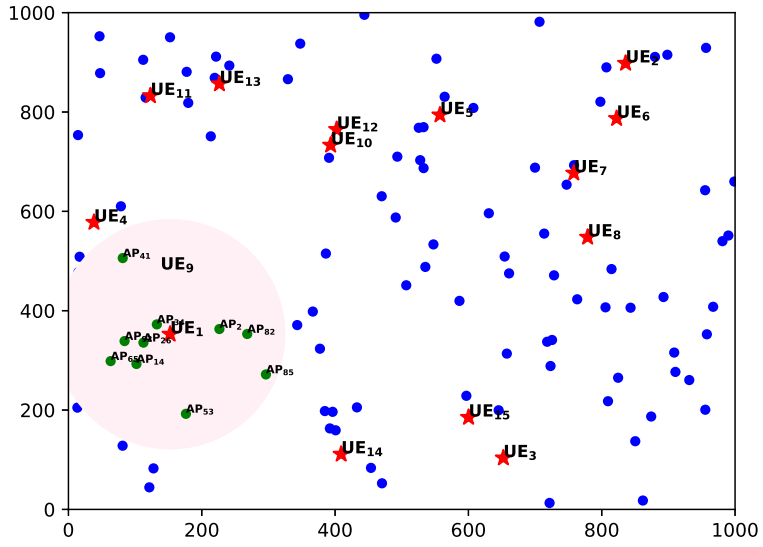


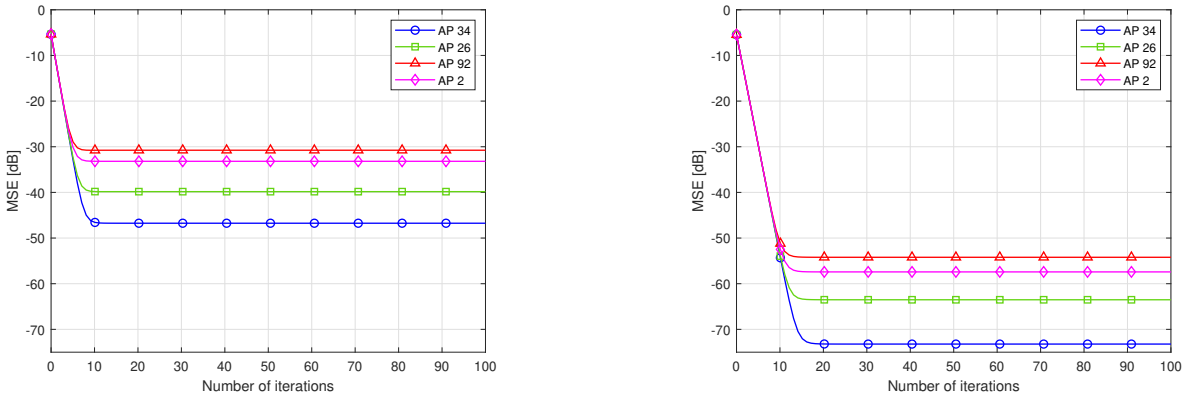
Fig. 3: Random network system topology of a CF-mMIMO system where 15 UEs and 100 APs whith N=16 are randomly distributed inside the 1km x 1km square

We analyze the NLD estimation performance at different AP locations within the CF-mMIMO system. This will involve comparing the convergence behavior and accuracy of the NLD estimation algorithm for different AP locations and based on different combining schemes. Since the NLD estimation is performed per user, we assume that UE_1 is the user of interest. For the two proposed distortion estimation methods, we present the mean square error (MSE) between the original data of UE_1 and its estimated data after cancellation of its own estimated NLD at the following APs: AP_{34}, AP_{26}, AP_{92}, AP_2 versus the number of iterations. The algorithm convergence is analyzed for two noise powers : σ^2=-93 dBm and σ^2=-120 dB.

Starting with NLD cancellation for channel estimation, in Fig. 4, we show the number of iterations needed for the convergence of the proposed algorithm for two noise powers: (a) -93 dBm and (b) -120 dBm. It is evident that the MSE of all APs decreases as the noise power decreases. Specifically, we observe that at a noise power level of σ^2 = -120dBm, the algorithm requires 20 iterations for convergence at AP_{34}, resulting in an MSE value of -73dB. In contrast, at the same AP but with a noise power level of σ^2 = -93dB, only 12 iterations are needed, yielding to an MSE of -46dB. Let's consider AP_{92}, which is located farther from UE_1. At a noise power level of σ^2 = -120dBm, the algorithm converges after 23

iterations, achieving an MSE of -54dB. Conversely, at a noise power level of σ^2 = -93dB, convergence is reached in 10 iterations, with an MSE value of -30dB.

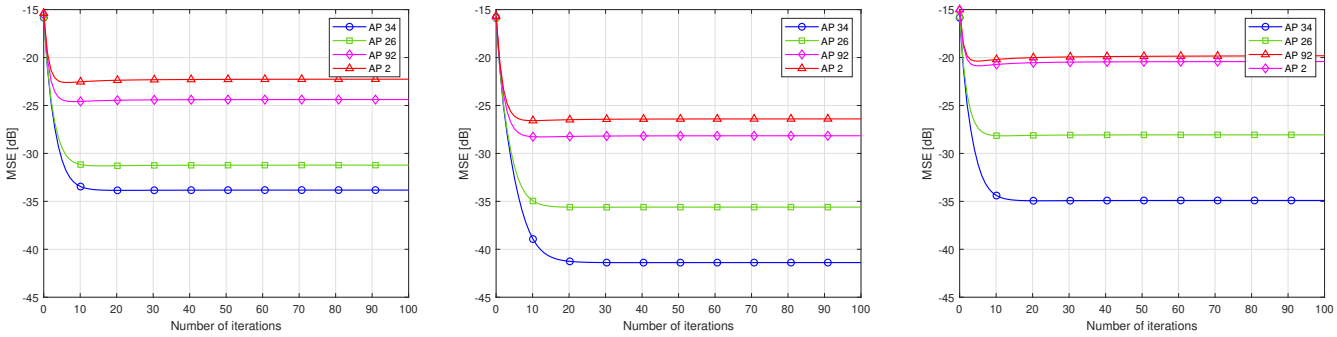
Clearly, a decrease in noise power leads to more accurate distortion estimation and improved cancellation performance, resulting in a lower MSE in the distortion estimation. However, it should be noted that as the noise level decreases, the number of iterations required for convergence may increase. This is due to the fact that with reduced noise, the cancellation process becomes more efficient, necessitating additional iterations to achieve the desired level of distortion suppression. Moreover, when considering a fixed noise power, the achievable MSE in NLD estimation can vary among different APs. Specifically, APs situated closer to UE_1 generally possess the potential to attain a lower MSE. Nevertheless, achieving a low MSE value may necessitate a higher number of iterations in the estimation process. This is primarily because APs in closer proximity to the user typically experience stronger received signal power compared to those located farther away. The higher signal power results in an improved SNR, thereby enhancing the accuracy of the estimation process.



(a) $\sigma^2 = -93$ dBm (b) $\sigma^2 = -120$ dBm

Fig. 4: MSE versus the number of iterations N_{itr} for channel estimation when $N = 16$, $K = \tau_p = 15$, $L = 100$ and IBO$= 3$

In Fig.5 and 6, we plot the MSE performance of NLD cancellation in data detection based on MR, PFZF and FZF combining for the same configuration when σ^2=-93dBm and σ^2=-120dBm, respectively. First, we can observe from the two figures that, regardless the noise level, the effectiveness of NLD cancellation depends on the used combining scheme at the AP.

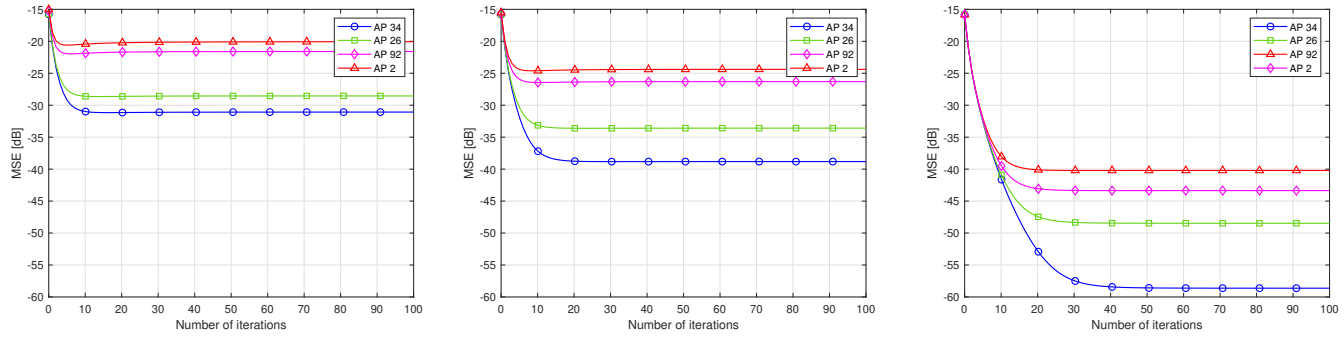


(a) MR (b) PFZF (c) FZF

Fig. 5: MSE versus the number of iterations N_{itr} for Algorithm1 when $N = 16$, $K = \tau_p = 15$, $L = 100$, IBO$= 3$ and $\sigma^2 = -93$dBm

As depicted in Figure 5, we observe that NLD estimation is less effective when the noise power level is σ^2 = -93 dBm. This is particularly evident for APs located farther away from the user UE1, such as AP_{92} and AP_2. In fact, in scenarios with high noise levels, the presence of noise significantly deteriorates the quality of the received signal, posing a greater challenge for NLD estimation, especially for APs with weaker channel conditions towards UE_1. Furthermore, it is worth noting that the FZF combining scheme exhibits a relatively high error floor value (-28 dB) at AP_{26} compared to MR and PFZF combining schemes, which achieve -31 dB and -35 dB, respectively. This is due to the fact that the noise amplification associated to FZF combining limits its effectiveness in accurately estimating and canceling NLD, particularly when encountering unfavorable channel conditions. On the other hand, MR combining may be susceptible to multi-user interference. However, being a compromise between MR and FZF, PFZF combining provides a better NLD estimation capability.

Moreover, it is worth noting that both MR and FZF combining schemes converge in fewer iterations (10 iterations), whereas PFZF requires 20 iterations. This arises from the fact that additional cancellation attempts may not significantly improve the distortion estimation due to the amplified noise level in the case of FZF and the interference level in the case of MR. In fact, in low noise conditions, FZF can



(a) MR (b) PFZF (c) FZF

Fig. 6: MSE versus the number of iterations N_{itr} for Algorithm1 when $N = 16$, $K = \tau_p = 15$, $L = 100$, IBO$= 3$ and $\sigma^2 = -120$dBm

effectively suppress interference and amplify the desired signal with minimal noise amplification. This allows for accurate estimation and cancellation of NLD. On the other hand, MR combining may be more susceptible to interference dominance, which can hinder the accurate estimation of NLD in low noise conditions. On the other hand, despite reducing noise amplification compared to FZF combining but not achieving the same level of interference suppression.

Remark: The convergence speed of the iterative NLD estimation algorithm depends on various factors, including the used combining scheme and the quality of initial detected signal of each user. The cancellation algorithm needs to be carefully designed to estimate and cancel total NLD in the presence of multi-user interference and considering the distributed nature of the APs.

V. Simulation Results

This section presents numerical results that show the effectiveness of the proposed scheme. The system being considered is a CF-mMIMO, consisting of L APs and K single-antenna users, operating in a 1km$\times$1km square-shaped area. The locations of both APs and UEs are randomly and uniformly distributed. The large-scale channel coefficient between UE_k and AP_l is modeled as follows

$$\beta_{k,l} = \mathrm{PL}_{k,l} 10^{\frac{\sigma_{sh} z_{k,l}}{10}}, \tag{28}$$

where $\sigma_{sh} = 4$ dB is the standard deviation, $z_{k,l} \sim \mathcal{N}(0,1)$ and $\mathrm{PL}_{k,l}$ is the path loss from UE_k to AP_l. Considering the micro cell path loss models of the 3GPP model which is given by [6]

$$\mathrm{PL}_{k,l}[\mathrm{dB}] = -36.7\log_{10}\left(\frac{\mathrm{Dist}_{k,l}}{1\mathrm{m}}\right) - 22.7 - 26\log_{10}\left(\frac{\mathrm{F_c}}{1\mathrm{GHz}}\right), \tag{29}$$

where $\mathrm{Dist}_{k,l}$ is the distance between UE_k and AP_l and $\mathrm{F_c}$=3.5GHz is the carrier frequency. The key simulation parameters are summarized in Table I. The parameters of PAs are set as follows G = 16, Vsat = 1.9, p = 1.1, A = -345, B = 0.17 and q = 4. Hereafter, "Lin PA" refers to the ideal case where there

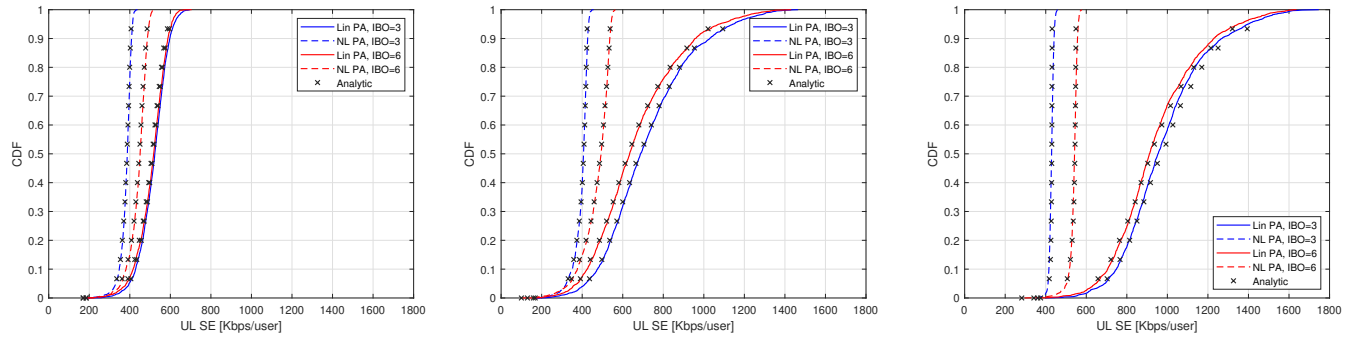
Parameter	Value	Parameter	Value
T_c	1ms	p_U	0.1 W
Δ_f	15 kHz	p_A	0.825 W
N_{sc}^{rb}	12	p_{BT}	1 W
N_c	14	C_l^f	100Mbps
τ_c	168	μ_{max}	0.785

TABLE I: Simulation parameters

are no HWIs and "NL PA" denotes the case under HWIs and without NLD cancellation.

A. *Impact of HWI on the uplink SE*

The impact of the NLD on the CDF of the uplink SE of CF-mMIMO-OFDM system is shown in Fig. 7, for L=100, N=16 and K=15. First, It can be seen that the results obtained from closed-form expressions



(a) MR (b) FZF (c) PFZF

Fig. 7: CDF of uplink SE per-RB and per-user when L=100, N=16, K=τ_p=15 and σ^2=-93dBm for two values of IBO=[3, 6]dB

(black stars) are close to the ones given by Monte-Carlo simulations for linear (solid curves) and non-linear cases (dashed curves). Moreover, regardless the used combiner, we note that the impact of NLD reduces when the IBO increases. However, it may also be noted that the system achieves better SE for "Lin PA" case as the IBO factor decreases. In fact, increasing the IBO reduces the NLD effects at the cost of the power efficiency. It should be noted that the impairments effect becomes more severe for better users with better channel gains (high percentiles). In fact, the total NLD is proportional to users' channel quality. On the other hand, the achieved SE of UEs with poor channel conditions is dominated by noise and inter-user interference and hence the impact of PA distortions is not clearly seen. As illustrated, at CDF of 95% for IBO=3dB, NLD causes a loss of 33% 63% and 69% in SE of the considered system for MR, FZF and PFZF, respectively. While the MR scheme is less affected due to the presence of noise and multiuser interference, the FZF and PFZF schemes are more susceptible to non-linear distortion, leading to a sever degradation of the system SE.

Fig. 8 shows the comparison of the median average SE of the system based on different combining schemes with or without NLD, while varying the number of antennas per AP. First, for the ”Lin PA” case, we observe that increasing N leads to higher average spectral efficiency for all combining schemes. By increasing the number of antennas, the system benefits from increased spatial degrees of freedom, improved beamforming, and enhanced interference suppression capabilities. Second, for the ”Lin PA” case, we note that the PFZF combining scheme tends to provide the highest average SE, followed by FZF, while MR offers the lowest spectral efficiency among the considered combining schemes. In the MR combining scheme, the received signals are already affected by noise and multiuser interference. Therefore, the impact of NLD on the MR combining scheme may be relatively less prominent compared to PFZF and FZF. However, the overall performance of MR can still be affected by the presence of NLD. The PFZF combining scheme balances the trade-off between complexity and interference suppression.

On the other hand, we can observe that, for the ”NL PA” case, the MR combining scheme still performs worse than FZF and PFZF. However, the achievable average SE is almost the same for FZF and PFZF combining schemes. In fact, the distortion can degrade the performance of interference suppression and compromise the benefits of these schemes.

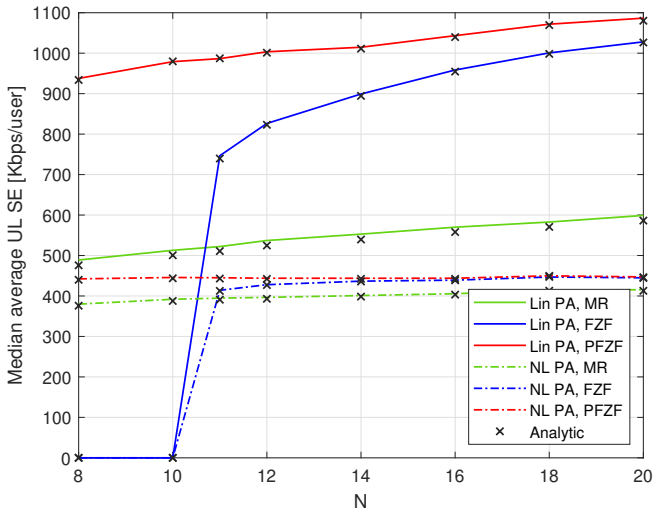
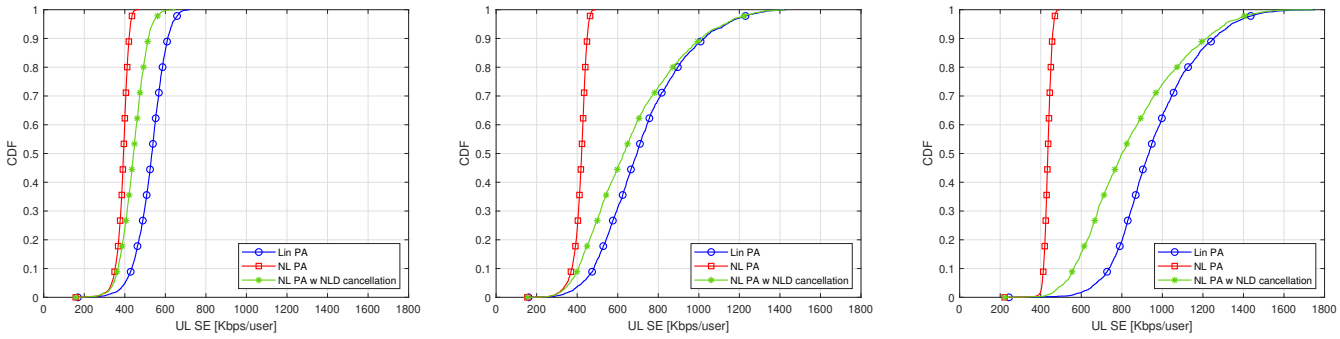


Fig. 8: Median of average uplink SE per-RB when L=100, K=τ_p=10, IBO= 3 dB and σ^2=-93 dBm

B. NLD cancellation/Spectral efficiency

Fig. 11 shows the CDFs of the SEs achieved by the CF-mMIMO-OFDM adopting, respectively, the MR, FZF and PFZF combining schemes. For the sake of comparaison, the figure depicts the achievable SE of the proposed NLD cancellation technique and ”Lin PA” and ”NL PA” cases. We can clearly note that, for all combining schemes, the NLD cancellation technique is generally more effective for high percentiles of users with good channel conditions, rather than users with poor channel conditions. It is also worth mentioning that the NLD cancellation is less effective for MR combining.

We can also see that the effectiveness of the proposed scheme is limited when the MR combining scheme is used. Particularly, at CDF of 95%, the achievable SE is improved from 428 to 536 kbps/user (gain of 25%). For low percentiles, at CDF of 20%, the provided gain decreases to 6%. When using the FZF combining scheme, the achievable gain after NLD cancelation is only 16.7%, at CDF of 2%. At CDF of 95%, the achievable SE through FZF combining is 2.5 times better than the achieved SE in the ”NL PA” case. Moreover, the gain in the achievable SE, when the PFZF combining scheme is used, is about 49% at CDF of 2%. At CDF of 95%, the achievable SE through PFZF combining is 2.8 times better than the achieved SE in the ”NL PA” case. While the MR combining scheme is limited by noise and multi-user interference, the FZF combining scheme faces challenges related to noise amplification and channel estimation errors that can affect the effectiveness of non-linear distortion cancellation.



(a) MR (b) FZF (c) PFZF

Fig. 9: CDF of uplink SE per-RB and per-user when L=100, N=16, K=τ_p=15, σ^2=-93dBm and IBO=3dB

Fig. 10 shows the median average SE with K = τ_p = 10, N = 8 versus N for the three used combiners. The figure also compares the median average SE of "Lin PA" and "NL PA w SDC" cases. We can note that, for all the combining schemes, the average SE increases when increasing the number of antennas at the APs. Moreover, the PFZF still provides the highest SE while MR gives the lowest SE. However, it does not directly impact the received non-linear distortion cancellation process.

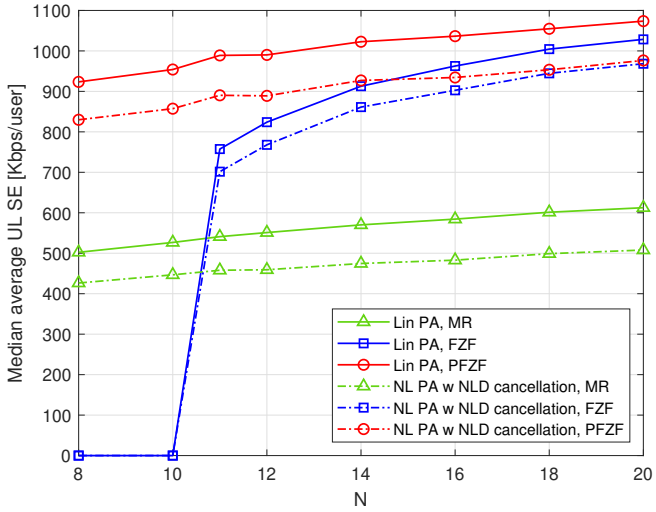
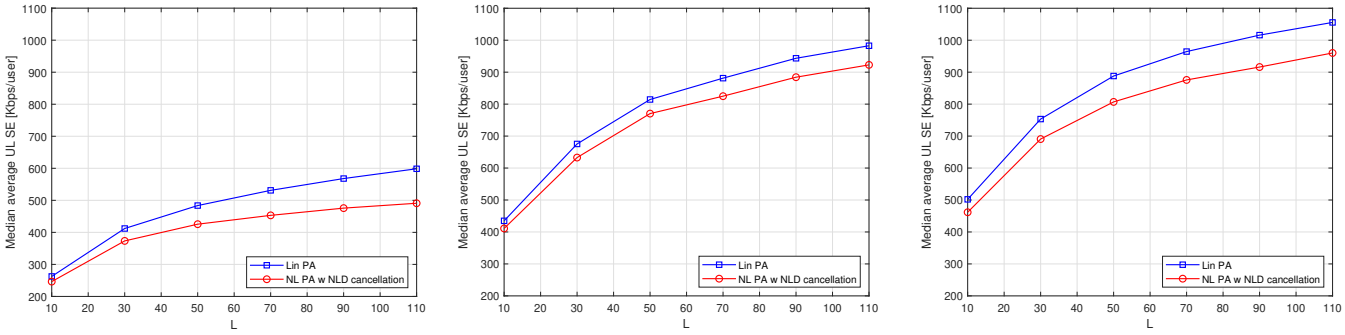


Fig. 10: Median of average uplink SE per-RB versus the number of antennas per AP when L=100, K=τ_p=10, IBO= 3 dB and σ^2=-93 dBm

VI. Conclusion

In this study, we investigated the performance of a CF-mMIMO system in the presence of HWIs at UEs. We derived the achievable SE for an UL CF-mMIMO-OFDM system affected by non-linear power amplifier (NL PA) distortion based on three combining schemes : MR, FZF and PFZF. Our simulation results aligned closely with the analytical derivation. We proposed two distributed non-linear distortion (NLD) cancellation techniques to effectively address NLD effects on both channel estimation and data detection. The goal was to mitigate the impact of NLD and improve the overall system performance. Importantly, we found that the NLD cancellation schemes depends on channel conditions between the users and APs. Moreover, data correction depends also on the specific combining scheme used (MR, FZF, or PFZF). Numerical results demonstrated the effectiveness of the proposed NLD cancellation algorithm.



(a) MR (b) FZF (c) PFZF

Fig. 11: Median of average uplink SE per-RB versus the number of APs when N=16, K=τ_p=10, IBO= 3 dB and σ^2=-93 dBm

ACKNOWLEDGMENT

This work was supported by the framework of the ALEX6 Project receiving fund by the French CARNOT-ANR under Carnot-SCIENCE-ALEX6.

REFERENCES

[1] Ian F. Akyildiz, Ahan Kak, and Shuai Nie. 6G and beyond: The future of wireless communications systems. *IEEE Access*, 8:133995–134030, 2020.

[2] Shuaifei Chen, Jiayi Zhang, Jing Zhang, Emil Björnson, and Bo Ai. A survey on user-centric cell-free massive mimo systems. *Digital Communications and Networks*, (5):695–719, 2022.

[3] Muntadher Alsabah, Marwah Abdulrazzaq Naser, Basheera M. Mahmmod, Sadiq H. Abdulhussain, Mohammad R. Eissa, Ahmed Al-Baidhani, Nor K. Noordin, Sadiq M. Sait, Khaled A. Al-Utaibi, and Fazirul Hashim. 6g wireless communications networks: A comprehensive survey. *IEEE Access*, 9:148191–148243, 2021.

[4] Hien Quoc Ngo, Alexei Ashikhmin, Hong Yang, Erik G. Larsson, and Thomas L. Marzetta. Cell-free massive mimo versus small cells. *IEEE Transactions on Wireless Communications*, 16(3):1834–1850, 2017.

[5] Özlem Tugfe Demir, Emil Björnson, and Luca Sanguinetti. 2021.

[6] Emil Björnson and Luca Sanguinetti. Making cell-free massive mimo competitive with mmse processing and centralized implementation. *IEEE Transactions on Wireless Communications*, 19(1):77–90, 2020.

[7] Emil Björnson and Luca Sanguinetti. Scalable cell-free massive mimo systems. *IEEE Transactions on Communications*, 68(7):4247–4261, 2020.

[8] Qingqing Wu, Geoffrey Ye Li, Wen Chen, Derrick Wing Kwan Ng, and Robert Schober. An overview of sustainable green 5g networks. *IEEE Wireless Communications*, 24(4):72–80, 2017.

[9] Jiayi Zhang, Yinghua Wei, Emil Björnson, Yu Han, and Xu Li. Spectral and energy efficiency of cell-free massive mimo systems with hardware impairments. In *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6, 2017.

[10] Jiayi Zhang, Yinghua Wei, Emil Björnson, Yu Han, and Shi Jin. Performance analysis and power control of cell-free massive mimo systems with hardware impairments. *IEEE Access*, 6:55302–55314, 2018.

[11] Jiakang Zheng, Jiayi Zhang, Luming Zhang, Xiaodan Zhang, and Bo Ai. Efficient receiver design for uplink cell-free massive mimo with hardware impairments. *IEEE Transactions on Vehicular Technology*, 69(4):4537–4541, 2020.

[12] Zahra Mokhtari and Rui Dinis. Sum-rate of cell free massive mimo systems with power amplifier non-linearity. *IEEE Access*, 9:141927–141937, 2021.

[13] Xianyu Zhang, Daoxing Guo, Kang An, and Bangning Zhang. Secure communications over cell-free massive mimo networks with hardware impairments. *IEEE Systems Journal*, 14(2):1909–1920, 2020.

[14] Yao Zhang, Qi Zhang, Han Hu, Longxiang Yang, and Hongbo Zhu. Cell-free massive mimo systems with non-ideal hardware: Phase drifts and distortion noise. *IEEE Transactions on Vehicular Technology*, 70(11):11604–11618, 2021.

[15] Xiaoling Hu, Caijun Zhong, Xiaoming Chen, Weiqiang Xu, Hai Lin, and Zhaoyang Zhang. Cell-free massive mimo systems with low resolution adcs. *IEEE Transactions on Communications*, 67(10):6844–6857, 2019.

[16] Hamed Masoumi and Mohammad Javad Emadi. Performance analysis of cell-free massive mimo system with limited fronthaul capacity and hardware impairments. *IEEE Transactions on Wireless Communications*, 19(2):1038–1053, 2020.

[17] Rafik Zayani, Jean-Baptiste Doré, Benoit Miscopein, and David Demmer. Local papr-aware precoding for energy-efficient cell-free massive mimo-ofdm systems. *IEEE Transactions on Green Communications and Networking*, pages 1–1, 2023.

[18] Nokia. Realistic Power Amplifier Model for the New Radio Evaluation. *document R4-163314, 3GPP TSG-RAN WG4 Meeting 79*, 2016.

[19] R. Price. A useful theorem for nonlinear devices having gaussian inputs. *IRE Transactions on Information Theory*, 4(2):69–72, 1958.

[20] M Thomas, EG Larsson, and H Yang. Hien quoc ngoet, ; fundamentals of massive mimo, 2016.
[21] Steven M Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
[22] Jiayi Zhang, Jing Zhang, Emil Björnson, and Bo Ai. Local partial zero-forcing combining for cell-free massive mimo systems. *IEEE Transactions on Communications*, 69(12):8459–8473, 2021.