Estimation and Mitigation of Channel Non-Reciprocity in Massive MIMO

Orod Raeesi, *Student Member, IEEE,* Ahmet Gokceoglu, *Member, IEEE,* and Mikko Valkama, *Senior Member, IEEE*

***Abstract*—Time-division duplex (TDD) based massive MIMO systems rely on the reciprocity of the wireless propagation channels when calculating the downlink precoders based on uplink pilots. However, the effective uplink and downlink channels incorporating the analog radio front-ends of the base station (BS) and user equipments (UEs) exhibit non-reciprocity due to non-identical behavior of the individual transmit and receive chains. When the downlink precoder is not aware of such channel non-reciprocity (NRC), system performance can be significantly degraded due to the NRC induced interference terms. In this work, we consider a general TDD-based massive MIMO system where frequency-response mismatches at both the BS and UEs, as well as the mutual coupling mismatches at the BS large-antenna system all coexist and induce channel NRC. Based on the NRC-impaired signal models, we first propose a novel iterative estimation method for acquiring both the BS and UE side NRC matrices and then also propose an efficient NRC-aware downlink precoder design which utilizes the obtained estimates. Furthermore, an efficient pilot signaling scheme between the BS and UEs is introduced in order to facilitate executing the proposed estimation method and the NRC-aware precoding technique in practical systems. Comprehensive numerical results indicate substantially improved spectral efficiency performance when the proposed NRC estimation and NRC-aware precoding methods are adopted, compared to the existing state-of-the-art methods.**

***Index Terms*—Beamforming, channel non-reciprocity, channel state information, frequency-response mismatch, linear precoding, massive MIMO, mutual coupling, time division duplexing (TDD).**

I. Introduction

MASSIVE MIMO is one of the key potential technologies for upcoming 5G systems [1] where base stations (BSs) deploy very large antenna arrays, e.g., several tens or hundreds of antenna units per array, to facilitate high beamforming and spatial multiplexing gains. In such systems, it is not feasible to transmit downlink pilots from each BS antenna in order to estimate the corresponding spatial channels at user equipments (UEs) and feedback the channel state information (CSI) to BS, as the amount of overhead in such approach is proportional to the number of antennas in the BS side [2]. Massive MIMO systems are thus envisioned to primarily deploy time-division duplex (TDD) based radio access and rely on the reciprocity of the physical uplink and downlink channels when obtaining CSI at the BS. This, in turn, requires substantially smaller pilot or reference signal overhead being only proportional to the number of UEs [3].

While it is a common assumption in TDD systems that the physical propagation channels are reciprocal within a coherence interval [2], [3], the impacts of the BS and UE side transceiver analog front-ends on the *effective downlink and uplink channels* are not reciprocal. This hardware induced phenomenon is often referred to as the channel non-reciprocity (NRC) problem [4], [5]. Typically, the mismatches in the frequency-responses (FRs) of both the BS and UE side radio front-ends between the transmit and receive modes are seen as the main cause of NRC. Another important source of NRC considered in literature is the differences in mutual coupling (MC) of BS antenna units and the associated RF transceivers under transmit and receive modes [6], [7].

The impacts of the NRC on the achievable system performance have been studied in various works in the recent literature. To this end, [5] provides downlink sum-rate analysis for a general multi-user MIMO system with zero-forcing (ZF) and eigen-beamforming types of precoding under NRC due to FR mismatch. Then, specifically focusing on massive MIMO systems, [8], [9] study achievable downlink sum-rates for maximum-ratio transmission (MRT) and ZF precoding schemes, demonstrating significant performance degradation under practical values of the NRC parameters.

There is also a large amount of work reported in the literature addressing the estimation and mitigation of NRC in TDD based MIMO systems [2], [4], [6], [10]–[17]. These studies can be divided into three main categories as follows:

i) The BS carries out "self-calibration" using a reference antenna with the help of additional circuitry [4], [6]. This method is capable of estimating the BS side NRC only.
ii) The BS carries out "self-calibration" without additional circuitry. Coupling between the BS antennas is utilized when exchanging pilot signals with the reference antenna [2], [10]–[13]. Similar to i), also this method estimates only the BS side NRC and commonly focuses only on the FR mismatch estimation, thus neglecting the mutual coupling mismatch.
iii) The BS transmits specific pilot signals to UEs and UEs send back the received signals in certain properly precoded forms to facilitate the BS side NRC parameter estimation, which is often referred to as over-the-air (OTA) approach [6], [14]–[17].

O. Raeesi, A. Gokceoglu, and M. Valkama are with the Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere 33720, Finland (e-mail: orod.raeesi@tut.fi; ahmet.gokceoglu@tut.fi; mikko.e.valkama@tut.fi).

This work was supported by the Finnish Funding Agency for Technology and Innovation (Tekes) under the project "5th Evolution Take of Wireless Communication Networks (TAKE-5)", by the Academy of Finland under the projects 284694 and 288670 and TUT Graduate School.

In this work, we focus on OTA-based estimation and mitigation of NRC in a multi-user massive MIMO system context deploying MRT or ZF precoding. The novelty and contributions of this paper can be summarized as follows:

1) We consider generalized NRC induced by coexisting FR mismatches of all associated radio transceivers at UE and BS sides as well as the mutual coupling mismatches in the BS side large-array antenna system, unlike many of the earlier works, such as [2], [4], [10]–[17], that neglect BS mutual coupling mismatch. In this respect, only [6] reports similar modeling, however, the proposed estimation and mitigation scheme in [6] is suitable mainly for small scale MIMO systems, e.g., 2-4 BS antennas due to the excessive computational complexity as well as unaffordable overhead in signaling back the required downlink channel estimates from the UEs to the BS when large antenna arrays are considered.
2) We address estimation and mitigation of the NRC sources of both the UE and the BS sides, unlike many other works that address only BS side NRC, e.g., [2], [4], [10]–[13], [15]. As shown in [18], with the popular assumption of not having downlink demodulation pilots, UE side NRC can be a major cause of performance degradation in multi-user massive MIMO systems, thus strongly motivating to incorporate such effects in the NRC estimation and mitigation processes.
3) Unlike other massive MIMO NRC mitigation works [2], [10]–[13] which all assume the availability of downlink demodulation pilots, we consider the appealing massive MIMO scenario in which there are no downlink demodulation pilots and thus UEs rely on the statistical properties of the beamformed channels to decode the received downlink signals [8], [19]–[22]. This further motivates to include both the BS and UE side NRC sources in the estimator and mitigation developments.
4) We demonstrate and evaluate the performance of our proposed scheme under imperfect uplink CSI, unlike other works which commonly rely on the perfect uplink CSI assumption [4], [11]–[13], [15].

The organization of the paper is as follows. Fundamental signal models of the considered massive MIMO system with MRT and ZF-based precoding schemes under NRC are first presented in Section II. Then, the NRC-aware downlink precoding approach is formulated for given NRC estimates. In Section III, a novel pilot signaling method between the BS and UEs is introduced which is followed by the proposed novel iterative estimation of the BS and UE side NRC matrices. Also the corresponding complexity analysis is provided. The results of empirical performance evaluations in terms of the NRC estimation mean-squared error (MSE) and the achievable system spectral efficiency are presented in Section IV, incorporating the proposed estimation-mitigation scheme together with existing state-of-the-art NRC estimation/mitigation methods for reference. Finally, conclusions are drawn in Section V.

Notations: Throughout the paper, vectors and matrices are denoted with lower and upper case bold letters, respectively, e.g., vector $\mathbf{x}$, matrix $\mathbf{Y}$. The superscripts $(.)^*$, $(.)^{\mathrm{T}}$, $(.)^{\mathrm{H}}$, and $(.)^{\dagger}$ indicate complex-conjugation, transposition, Hermitian-transpose, and Moore-Penrose pseudo inverse operations, respectively. The expectation operator is shown by $\mathbb{E}[.]$, while $\mathrm{Tr}\,(.)$ represents the trace operator. The $\mathrm{diag}\,(.)$ operator transforms a vector $\mathbf{v}$ to a diagonal matrix with the elements of $\mathbf{v}$ at its diagonal, and vice versa, reads the diagonal elements of the input matrix into a column vector. $\Re\{.\}$ and $\Im\{.\}$ work element-wise and return real and imaginary parts of complex-valued arguments, respectively. The element in the i'th row and j'th column of matrix $\mathbf{V}$ is represented by v_{ij}, whereas the i'th element on the main diagonal of a diagonal matrix $\mathbf{C}$ is shown by c_i. The complex-valued zero-mean circularly symmetric Gaussian distribution with variance σ^2 is denoted as $\mathcal{CN}\left(0,\sigma^2\right)$. Finally, $\mathbf{I}_n$ and $\mathbf{0}_n$ denote the $n \times n$ identity and all-zero matrices, respectively.

II. System Model and Problem Formulation

We consider a TDD based single-cell multi-user downlink transmission scenario where the BS with a large number of antenna units, denoted by N, transmits to K single-antenna UEs simultaneously in the same time-frequency resource, where $N \gg K$. All signal and system models are written for an arbitrary subcarrier of the underlying orthogonal frequency division multiplexing/multiple access (OFDM/OFDMA) waveform, that is, before IFFT and after FFT on the TX and RX sides, respectively.

In an ideal TDD massive MIMO system, the effective uplink and downlink channels consist of only the reciprocal physical channels. Building on that, the downlink transmission is done by beamforming the multi-user downlink data based on the estimated channels from uplink pilot sequences of length τ_u symbols [2], [3]. A fundamental assumption is that the uplink training and downlink transmission both take place clearly within one channel coherence interval, meaning that the channels are essentially static. In this work, we assume the same procedure for the uplink training and downlink transmission, however, we consider more generalized uplink and downlink effective channel models which are *non-reciprocal* due to the radio front-end mismatches and non-idealities. In this respect, the uplink model for channel estimation phase [19] and the corresponding downlink received signal model in beamformed data transmission phase under the non-reciprocal effective channels can be expressed as

$$\begin{aligned} &\text{Uplink Training :} && \mathbf{Y} = \sqrt{\rho_u \tau_u}\mathbf{G} + \mathbf{Z}_u \\ &\text{Downlink Transmission :} && \mathbf{r} = \sqrt{\rho_d}\mathbf{H}\mathbf{x} + \mathbf{z}_d, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{N\times 1}$ denotes the precoded multi-user data, whereas $\mathbf{G} \in \mathbb{C}^{N\times K}$ and $\mathbf{H} \in \mathbb{C}^{K\times N}$ are the effective non-reciprocal uplink and downlink multi-user MIMO channels, respectively, which will be elaborated in detail later in Section II-A. $\mathbf{Z}_u \in \mathbb{C}^{N\times K}$ is the processed noise matrix at the BS, while $\mathbf{z}_d \in \mathbb{C}^{K\times 1}$ denotes the UE side multi-user thermal noise vector, both assumed to consist of i.i.d. $\mathcal{CN}\left(0,1\right)$ elements. The average signal to noise ratios (SNRs) in the uplink and downlink are denoted as ρ_u and ρ_d, respectively. This basic system framework is largely based on and following the seminal work by Marzetta in [19], [23] where reciprocal channels were assumed.

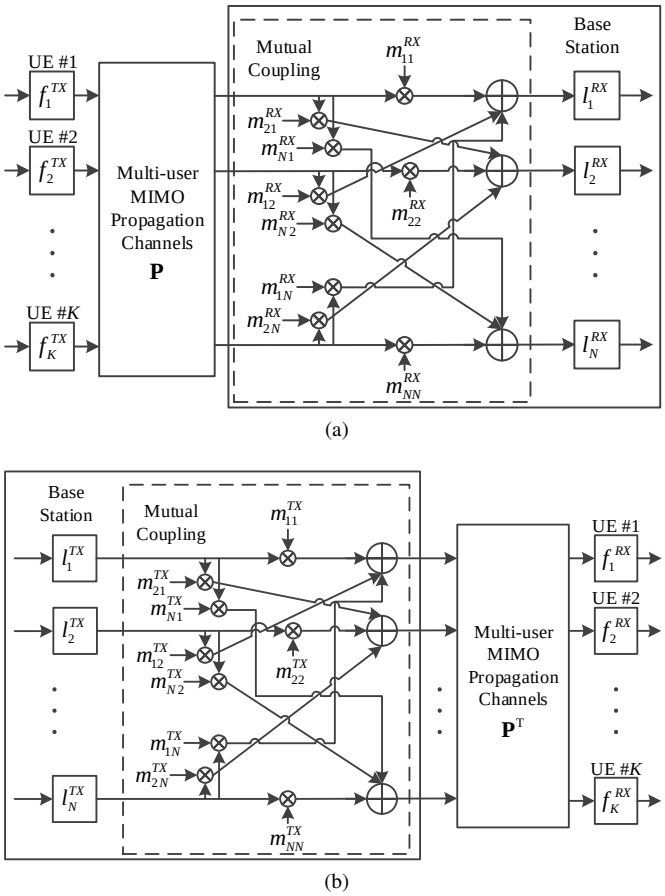


Fig. 1. Basic system models for (a) uplink and (b) downlink transmission and reception including physical propagation channels, transceiver frequency responses and antenna mutual coupling in the devices.

A. Effective and Relative Uplink and Downlink Channels

As illustrated in Fig. 1, the complete description of the uplink and downlink effective channels appearing in (1) can be expressed as

$$\left.\begin{aligned}\mathbf{G} &= \mathbf{E}_r\mathbf{P}\mathbf{F}_t\\ \mathbf{H} &= \mathbf{F}_r\mathbf{P}^{\mathrm{T}}\mathbf{E}_t\end{aligned}\right\} \Rightarrow \mathbf{H} \neq \mathbf{G}^{\mathrm{T}}, \tag{2}$$

with $\mathbf{E}_r = \mathbf{L}_r\mathbf{M}_r$ and $\mathbf{E}_t = \mathbf{M}_t\mathbf{L}_t$. In above, $\mathbf{F} \in \mathbb{C}^{K\times K}$ is the joint frequency-response matrix of the UEs, $\mathbf{L} \in \mathbb{C}^{N\times N}$ is the frequency-response matrix of the BS, $\mathbf{M} \in \mathbb{C}^{N\times N}$ is the mutual coupling matrix of the BS, $\mathbf{P} \in \mathbb{C}^{N\times K}$ is the reciprocal physical channel, while the subscripts t and r denote the transmitting and receiving modes, respectively. Note that the frequency-response matrices, $\mathbf{F}$ and $\mathbf{L}$, are diagonal, while the mutual coupling matrix $\mathbf{M}$ in general has both non-zero diagonal and off-diagonal entries.

In general, the effective channels with above assumptions and modeling are clearly non-reciprocal, i.e., $\mathbf{H} \neq \mathbf{G}^{\mathrm{T}}$, due to differences in the TX and RX modes of the radio front-end and array responses, i.e., $\mathbf{F}_t \neq \mathbf{F}_r$, $\mathbf{L}_t \neq \mathbf{L}_r$ and $\mathbf{M}_t \neq \mathbf{M}_r^{\mathrm{T}}$. Hence, the effective uplink and downlink channels can be described relative to each other as

$$\mathbf{H} = \mathbf{A}\mathbf{G}^{\mathrm{T}}\mathbf{B}, \tag{3}$$

where, $\mathbf{A} = \mathbf{F}_r\mathbf{F}_t^{-1}$ and $\mathbf{B} = \mathbf{L}_r^{-1}\left(\mathbf{M}_r^{\mathrm{T}}\right)^{-1}\mathbf{M}_t\mathbf{L}_t$.

In general, $\mathbf{A} \in \mathbb{C}^{K\times K}$ is a diagonal matrix and the k'th diagonal entry, denoted as a_k, corresponds to the frequency-response ratio of k'th UE at TX and RX modes. In the following, similar to [5], [6], [8], we will use the decomposition of the form $\mathbf{A} = \mathbf{I}_K + \mathbf{A}'$, where the diagonal matrix $\mathbf{A}' \in \mathbb{C}^{K\times K}$ measures the deviation from the ideal unity frequency-response ratio. The k'th diagonal entry of $\mathbf{A}'$ is denoted as a'_k, such that $a_k = 1 + a'_k$.

In (3), $\mathbf{B} \in \mathbb{C}^{N\times N}$ is a full matrix that incorporates both the frequency-responses and mutual coupling at the BS side in TX and RX modes. In the following, for notational convenience, we will use the decomposition $\mathbf{B} = \mathbf{I}_N + \mathbf{B}'$, where $\mathbf{B}' \in \mathbb{C}^{N\times N}$ accounts for the deviation of diagonal and off-diagonal entries from the ideal reciprocal response.

The detailed modeling of the entries of the above matrices is based on the practical physical circuits driven NRC modeling introduced in [6], in which σ_{F}^2 is denoting the variance of the diagonal elements in $\mathbf{F}_t$ and $\mathbf{F}_r$, while the corresponding variance of the diagonal elements in $\mathbf{L}_t$ and $\mathbf{L}_r$ is denoted by σ_{L}^2. In this model [6], $\mathbf{M}_t$ and $\mathbf{M}_r$ are generated as

$$\begin{aligned}\mathbf{M}_t &= \left(\mathbf{I}_N - \mathbf{\Gamma}_t\mathbf{\Phi}\right)^{-1}\\ \mathbf{M}_r &= \left(\mathbf{I}_N - \mathbf{\Gamma}_r\mathbf{\Phi}\right)^{-1},\end{aligned} \tag{4}$$

where $\mathbf{\Gamma}_t \in \mathbb{C}^{N\times N}$ and $\mathbf{\Gamma}_r \in \mathbb{C}^{N\times N}$ are diagonal matrices representing input reflection coefficients at transmitting and receiving modes, respectively, while $\mathbf{\Phi} \in \mathbb{C}^{N\times N}$ is the scattering matrix of the BS antenna system, defined as

$$\mathbf{\Phi} = \left(\frac{\mathbf{Z}}{Z_0} + \mathbf{I}_N\right)^{-1}\left(\frac{\mathbf{Z}}{Z_0} - \mathbf{I}_N\right). \tag{5}$$

In above, Z_0 denotes the reference impedance of the antenna ports and $\mathbf{Z} \in \mathbb{C}^{N\times N}$ is the BS impedance matrix whose elements depend on the antenna spacing and layout [24]. As illustrated in (4), for a chosen antenna spacing and layout which result into a certain BS antenna system scattering matrix $\mathbf{\Phi}$, the powers of the elements in $\mathbf{M}_t$ and $\mathbf{M}_r$ are controlled by the corresponding input reflection coefficient matrices $\mathbf{\Gamma}_t$ and $\mathbf{\Gamma}_r$, respectively. The variance of the diagonal elements in the input reflection coefficient matrices is denoted by σ_{rc}^2.

The characterization as given in (2) and/or (3) is generally referred to in the literature as channel non-reciprocity [5], [6]. The ideal reciprocal channel model is a special case where $\mathbf{A} = \mathbf{B} = \mathbf{I}$, i.e., $\mathbf{A}' = \mathbf{B}' = \mathbf{0}$.

B. Channel Estimation and Beamforming under NRC

First, we shortly address the influence of NRC when the downlink transmission is carried out without any processing against the NRC, i.e., NRC-blind precoding is adopted. In this respect, the required downlink channel estimate in the BS is obtained from the orthogonal uplink training signals, with the observation model given already on the first line of (1), complemented, e.g., with LMMSE channel estimator as described in [19], [23]. This yields formally

$$\hat{\mathbf{H}} = \hat{\mathbf{G}}^{\mathrm{T}}, \tag{6}$$

where $\hat{\mathbf{H}} \in \mathbb{C}^{K\times N}$ and $\hat{\mathbf{G}} \in \mathbb{C}^{N\times K}$ are the estimated downlink and uplink effective channels, respectively.

Using the estimated downlink effective channel in (6), the user data vector $\mathbf{s} = [s_1, ..., s_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ which is assumed to have element-wise power normalization of the form $\mathbb{E}\left[|s_k|^2\right] = 1$, is precoded as

$$\mathbf{x} = \beta \mathbf{U}\mathbf{s}, \tag{7}$$

where the NRC-blind linear precoding matrix $\mathbf{U} \in \mathbb{C}^{N\times K}$ reads [23]

$$\mathbf{U} = \begin{cases} \hat{\mathbf{H}}^{\mathrm{H}}, & \text{for MRT} \\ \hat{\mathbf{H}}^{\mathrm{H}} \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}}\right)^{-1}, & \text{for ZF.} \end{cases} \tag{8}$$

In above, without loss of generality, the scalar β can be chosen to satisfy unit average transmit power constraint as [19]

$$\beta = \left(\sqrt{\mathbb{E}[\mathrm{Tr}\left(\mathbf{U}^{\mathrm{H}}\mathbf{U}\right)]}\right)^{-1}. \tag{9}$$

C. Received Signal at UE under NRC

The multi-user received downlink signal vector is given by the second line of (1). Plugging the precoded symbol vector expression in (7) into (1), the received signal for k'th user corresponding to the k'th element of $\mathbf{r}$ can be written as

$$r_k = \sqrt{\rho_d}\beta \mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k s_k + \sqrt{\rho_d}\beta \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^{\mathrm{T}}\mathbf{u}_i s_i + z_{d,k}, \tag{10}$$

where $\mathbf{u}_k$ and $\mathbf{h}_k^{\mathrm{T}}$ denote the k'th column and row vectors of the precoder and the effective downlink channel matrices, respectively. Notice that by denoting the k'th column of the uplink effective channel matrix as $\mathbf{g}_k$, the effective downlink channel towards the k'th user can be expressed as

$$\mathbf{h}_k^{\mathrm{T}} = a_k \mathbf{g}_k^{\mathrm{T}}\mathbf{B}. \tag{11}$$

In general, conventional MIMO systems employ downlink pilots to acquire downlink CSI for detection purposes. However, in massive MIMO systems, as shown in [8], [19]–[22], it is commonly assumed that UEs employ only the statistical properties of the beamformed channel, namely $\mathbb{E}\left[\beta \mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right]$, as the downlink CSI to decode the received signal. This assumption is justified by the law of large numbers which implies that $\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k \rightarrow \mathbb{E}\left[\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right]$ when BS array size N increases, commonly known as the channel hardening concept [23], [25]. Utilizing such approach in acquiring downlink CSI in UEs eliminates the need for sending downlink demodulation pilots which directly reduces downlink overhead. Alternatively, blind downlink channel estimation [26] can also be pursued. Building on this and plugging (11) into (10), the received signal under NRC can be re-written in a general form as

$$r_k = \sqrt{\rho_d}\mathbb{E}\left[\beta \mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right] s_k + z_{\mathrm{SI},k} + z_{\mathrm{IUI},k} + z_{d,k}, \tag{12}$$

where the self-interference (SI), $z_{\mathrm{SI},k}$, and the inter-user-interference (IUI), $z_{\mathrm{IUI},k}$, are given by

$$\begin{aligned} z_{\mathrm{SI},k} &= \sqrt{\rho_d}\beta \left(a_k \mathbf{g}_k^{\mathrm{T}}\mathbf{B}\mathbf{u}_k - \mathbb{E}\left[\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right]\right) s_k \\ z_{\mathrm{IUI},k} &= \sqrt{\rho_d}\beta \sum_{i=1, i\neq k}^{K} a_k \mathbf{g}_k^{\mathrm{T}}\mathbf{B}\mathbf{u}_i s_i. \end{aligned} \tag{13}$$

Note that the basic received signal model in (12) and (13) holds independently of the channel hardening phenomenon. However, the channel hardening assumption is in practice needed in order to obtain good decoding performance with statistical beamformed channel based downlink CSI. In general, the statistical beamformed downlink CSI, $\mathbb{E}\left[\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right]$, can be interpreted either for given NRC realizations or averaged also over the NRC variable statistics. In this manuscript, we will take the latter approach as that is independent of the NRC realizations, and thus facilitates efficient UE data decoding without any downlink demodulation pilots. This is also well motivated by the main topic of the article which is the joint estimation and mitigation of the BS and UE side NRC at the BS. Finally, we note that static channels within uplink training and downlink transmission are assumed, and thus practical aspects such as channel aging within one uplink training and downlink transmission cycle are not considered.

Based on (13), it can be clearly observed that the NRC-blind precoder $\mathbf{u}$ which is constructed based on the estimated uplink effective channel $\hat{\mathbf{G}}$, through $\hat{\mathbf{H}} = \hat{\mathbf{G}}^{\mathrm{T}}$, cannot take into account the NRC effects from a_k and $\mathbf{B}$, which results into increased interference levels and thus reduced downlink spectral efficiency. This is illustrated through an elementary system spectral efficiency evaluation in Fig. 2, with the detailed evaluation assumptions being described in Section IV. It can be noticed that in particular in the ZF precoder case, NRC-blind precoding results into substantial performance degradation, hence strongly motivating to develop efficient NRC estimation and mitigation techniques.

D. NRC-Aware Downlink Precoding Principle

As shown in Section II-C above, if MRT and/or ZF precoders are applied naively without accounting for NRC, there are additional SI and IUI terms that can substantially degrade the quality of the received signal at the UE side. Here, we introduce an efficient NRC mitigation approach, called NRC-aware precoding, which seeks to cancel out the effects of NRC by properly modifying the precoder.

Assuming that the BS has already estimates of the NRC matrices $\mathbf{A}$ and $\mathbf{B}$, denoted by $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, the NRC-aware precoding approach transforms the basic linear precoders given in (8) as

$$\mathbf{U}_{nrc} = \hat{\mathbf{B}}^{-1}\mathbf{U}\hat{\mathbf{A}}^{-1}. \tag{14}$$

With ideal estimates of $\mathbf{A}$ and $\mathbf{B}$ and ideal uplink channel estimation, the above transformed precoder can be easily shown to diagonalize the beamformed multi-user channel matrix $\mathbf{H}\mathbf{U}_{nrc}$. This applies in an exact form with ZF precoding while is a very good approximation with MRT precoding when the BS array size N is large. Note that, in the special case where the NRC estimation method is capable of estimating the BS side NRC only, (14) reduces to $\mathbf{U}_{nrc} = \hat{\mathbf{B}}^{-1}\mathbf{U}$.

The system spectral efficiency performance with the NRC-aware precoder, assuming ideal NRC estimates, is shown in Fig. 2. As can be observed, the NRC-aware precoder achieves the ideal system performance, i.e., the performance with fully reciprocal channels. The evaluation setup and details of spectral efficiency calculations will be described in Section IV.

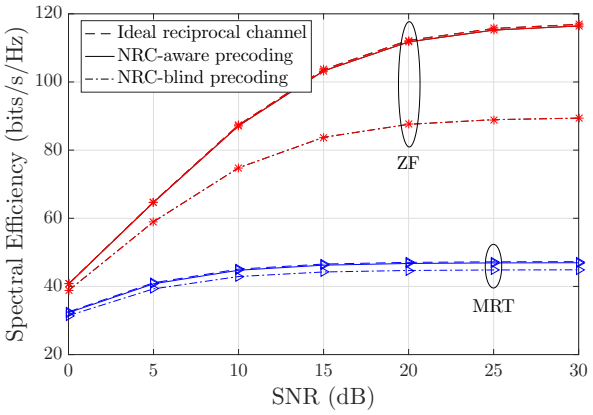


Fig. 2. Spectral efficiency vs. downlink SNR (ρ_d) for $N = 100$, $K = 20$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

It is also noted that an alternative approach to construct NRC-aware downlink precoder is to first correct the effective downlink channel estimates based on the estimated NRC matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ as $\hat{\mathbf{H}}_{cor} = \hat{\mathbf{A}}\hat{\mathbf{H}}\hat{\mathbf{B}}$, and then calculate the precoder matrix based on this corrected downlink channel estimate. However, in this work, we focus on the NRC mitigation approach described in (14) which is shown in Fig. 2 to practically reach the performance of the ideal reciprocal channel as long as accurate estimates of $\mathbf{A}$ and $\mathbf{B}$ are available.

III. Proposed Estimation of NRC Matrices

The NRC mitigation method, i.e., the NRC-aware precoder described in Section II-D requires the knowledge of the matrices $\mathbf{A}$ and $\mathbf{B}$ at the BS. The information about these matrices is not readily available, hence calling for efficient estimation approaches. Thus, in this section, we will propose a novel iterative OTA estimation framework for acquiring accurate estimates of $\mathbf{A}$ and $\mathbf{B}$, based on a novel pilot signaling concept between the BS and UEs. Furthermore, computational complexity analysis of the proposed estimation method is pursued.

In general, the NRC variances σ_{F}^2, σ_{L}^2, σ_{rc}^2, and the corresponding realizations of the elements of $\mathbf{A}$ and $\mathbf{B}$ depend on hardware characteristics and operating conditions, e.g., temperature, which vary slowly in time. Thus, the NRC characteristics and the corresponding realizations of $\mathbf{A}$ and $\mathbf{B}$ can be assumed to stay constant over many propagation channel coherence intervals [14]. Therefore, it is sufficient to perform the NRC estimation relatively infrequently, e.g, once in every 10 minutes or even more irregularly [2], [10], which makes the estimation pilot overhead negligible, when compared to other signaling and pilot overheads that commonly rise from channel estimation procedures.

A. Proposed Pilot Signaling

In order to estimate the matrices $\mathbf{A}$ and $\mathbf{B}$, we propose the following round-trip pilot signaling approach:

1) BS transmits an $N \times N$ orthonormal pilot matrix $\mathbf{X}_{nrc}$.
2) Upon reception, without decoding, UEs send back the conjugated versions of the received samples.

Based on the above scheme, the received multi-user pilot signal matrix at UE side can be written as

$$\mathbf{R}_{nrc} = \sqrt{\tilde{\rho}_d}\mathbf{H}\mathbf{X}_{nrc} + \tilde{\mathbf{Z}}_d, \tag{15}$$

where $\tilde{\rho}_d$ is the downlink SNR and $\tilde{\mathbf{Z}}_d$ is the $K \times N$ multi-user receiver noise matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. The tilde sign is used in above and what follows to distinguish these variables between the actual data transmission and pilot signaling phases. Then, the corresponding received signal at the BS with the UEs sending back the conjugated version of (15) reads

$$\begin{aligned}\mathbf{Y}_{nrc} &= \sqrt{\tilde{\rho}_u}\mathbf{G}\mathbf{R}_{nrc}^* + \tilde{\mathbf{Z}}_u \\ &= \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\mathbf{G}\mathbf{H}^*\mathbf{X}_{nrc}^* + \tilde{\mathbf{Z}}_{tot},\end{aligned} \tag{16}$$

where $\tilde{\rho}_u$ is the uplink pilot SNR and $\tilde{\mathbf{Z}}_u$ is the $N \times N$ BS receiver noise matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. The total effective noise matrix seen at the BS is denoted as $\tilde{\mathbf{Z}}_{tot} = \sqrt{\tilde{\rho}_u}\mathbf{G}\tilde{\mathbf{Z}}_d^* + \tilde{\mathbf{Z}}_u$. Note that, for notational simplicity, we have assumed all the UEs to have the same uplink pilot SNR value $\tilde{\rho}_u$, while in practice, $\tilde{\rho}_u$ is likely to be different for different UEs.

In above, the duration of the described overall NRC-related pilot signaling is $2N$ symbols where the uplink and downlink channels are assumed to be fixed. The coherence time of the physical channels is typically in the order of several hundred symbol intervals, determined mostly by the mobility of the UEs and the system center-frequency. Hence, we assume a scenario where the coherence time is at least $2N + K$ symbols, taking into account both NRC-related pilot signaling and uplink channel estimation. As mentioned in the previous section, matrices $\mathbf{A}$ and $\mathbf{B}$ are expected to change very slowly compared to channel coherence time and hence it is assumed that their values are fixed during the above pilot signaling. Fig. 3 illustrates the overall assumed radio frame or sub-frame structure of the considered massive MIMO TDD system including the proposed NRC estimation phase.

B. Overall Estimation Framework

As the initial step in estimating $\mathbf{A}$ and $\mathbf{B}$, the BS processes the received signal $\mathbf{Y}_{nrc}$ in (16) as $\mathbf{Q} = \mathbf{Y}_{nrc}^*\mathbf{X}_{nrc}^{\mathrm{H}}$. Since the pilot matrix $\mathbf{X}_{nrc}$ has the property $\mathbf{X}_{nrc}^{\mathrm{H}}\mathbf{X}_{nrc} = \mathbf{I}_N$, the processed signal can be expressed as

$$\mathbf{Q} = \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\mathbf{G}^*\mathbf{A}\mathbf{G}^{\mathrm{T}}\mathbf{B} + \mathbf{V}, \tag{17}$$

where the processed noise matrix is given by $\mathbf{V} = \tilde{\mathbf{Z}}_{tot}^*\mathbf{X}_{nrc}^{\mathrm{H}}$.

Now the target is to estimate $\mathbf{A}$ and $\mathbf{B}$ from (17) assuming that the BS has the uplink channel estimate $\hat{\mathbf{G}}$. In this respect, denoting the estimates at m'th iteration as $\hat{\mathbf{A}}(m)$ and $\hat{\mathbf{B}}(m)$, we propose the following iterative estimation framework:

1) Initialize, $\hat{\mathbf{A}}(0) = \mathbf{I}_K$, obtain the estimate $\hat{\mathbf{B}}(1)$.
2) Substitute $\hat{\mathbf{B}}(1)$ for $\mathbf{B}$ in (17) and obtain estimate $\hat{\mathbf{A}}(1)$.
3) Successively refine the estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ by fixing the current value of one and solving for the other from (17).

In above, $\mathbf{I}_K$ is used for initialization since the deviation matrix $\mathbf{A}'$ in $\mathbf{A} = \mathbf{I}_K + \mathbf{A}'$ is in practice small. Notice that

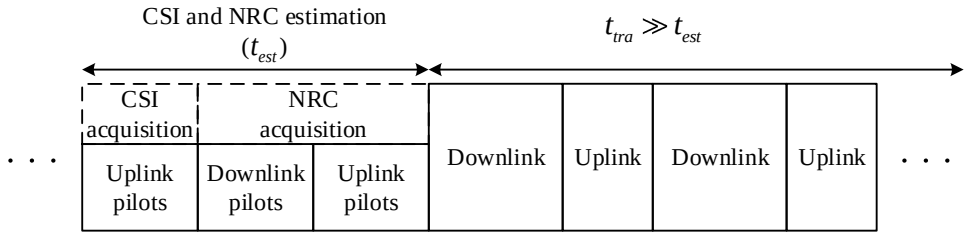


Fig. 3. Assumed radio frame or sub-frame structure incorporating CSI and NRC estimation as well as actual data transmission phases.

the processed received signal in (17) and the corresponding uplink channel estimate are available at multiple parallel sub-carriers in an OFDM/OFDMA based radio system. Hence, the above iterative estimation scheme can be carried out in a per subcarrier manner as well. Furthermore, as mentioned in [6], transceivers' baseband-to-baseband behavior can be modeled by allpass-like transfer functions, therefore it is reasonable to assume that the NRC matrices $\mathbf{A}$ and $\mathbf{B}$ are largely similar over a set of adjacent subcarriers C_{sc} where we conservatively assume $C_{sc} \leq 10$, whereas $\mathbf{G}$ is subject to variations depending on the frequency selectivity of the propagation channels. Based on these assumptions, the estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ can be obtained by averaging the per subcarrier estimates over C_{sc} neighboring subcarriers, i.e.,

$$\begin{aligned} \hat{\mathbf{A}} &= \frac{1}{C_{sc}} \sum_{l=1}^{C_{sc}} \hat{\mathbf{A}}_l \\ \hat{\mathbf{B}} &= \frac{1}{C_{sc}} \sum_{l=1}^{C_{sc}} \hat{\mathbf{B}}_l, \end{aligned} \tag{18}$$

where l denotes the relative subcarrier index. Next we will present the actual proposed methods to obtain the estimates for $\mathbf{A}$ and $\mathbf{B}$. To simplify the notation, we will drop the relative subcarrier index l. Note that the above averaging helps reducing the NRC estimation errors, mainly caused by additive receiver noise, regardless of how small the particular NRC values are, unless being below the machine precision and thus swallowed by finite word-length effects.

C. Proposed Estimation of $\mathbf{B}$

As described earlier, $\hat{\mathbf{B}}$ is iteratively refined using the current estimate of $\mathbf{A}$. The proposed estimator builds on solving the matrix equation in (17) based on minimizing the Frobenius norm criterion. With this setting, the refined estimate of $\mathbf{B}$ can be formulated as

$$\hat{\mathbf{B}}(m+1) = \underset{\mathbf{B}}{\operatorname{argmin}} \left|\left| \mathbf{Q} - \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^* \hat{\mathbf{A}}(m) \hat{\mathbf{G}}^{\mathrm{T}} \mathbf{B} \right|\right|_F^2, \tag{19}$$

where the subscript F denotes the Frobenius norm.

Next, by denoting $\mathbf{T}(m) = \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^* \hat{\mathbf{A}}(m) \hat{\mathbf{G}}^{\mathrm{T}} \in \mathbb{C}^{N \times N}$, we have the following identity

$$||\mathbf{Q} - \mathbf{T}(m)\mathbf{B}||_F^2 = \sum_{j=1}^{K} ||\mathbf{q}_j - \mathbf{T}(m)\mathbf{b}_j||^2, \tag{20}$$

where $\mathbf{q}_j \in \mathbb{C}^{N \times 1}$ and $\mathbf{b}_j \in \mathbb{C}^{N \times 1}$ denote the j'th columns of $\mathbf{Q}$ and $\mathbf{B}$, respectively. Since the j'th term in the sum depends only on $\mathbf{b}_j$, minimizing the total sum is equivalent to separately minimizing each term $||\mathbf{q}_j - \mathbf{T}(m)\mathbf{b}_j||^2$. Thus, the estimation of matrix $\mathbf{B}$ is eventually simplified to estimation of each column of $\mathbf{B}$, independently.

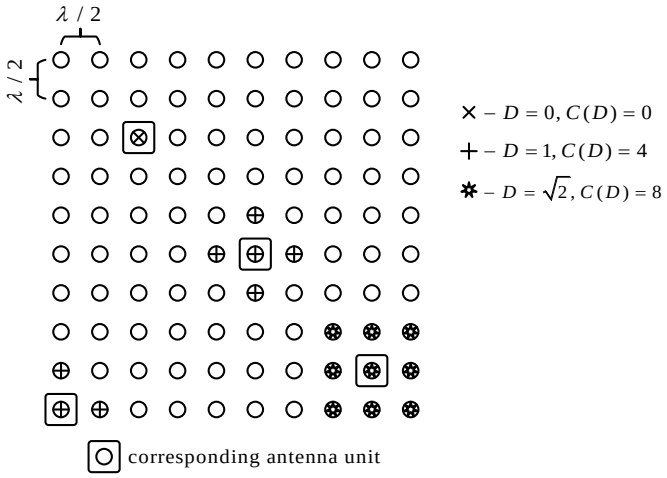


Fig. 4. Illustration of sparsity threshold D on 10×10 rectangular BS antenna grid with $\lambda/2$ antenna spacing.

As mentioned earlier, the BS NRC matrix incorporates both the frequency-responses and the mutual coupling effects at the BS side. The power of mutual coupling between two different antenna units is related to their physical distance, thus the off-diagonal elements in $\mathbf{B}$ become smaller as the distance between the two corresponding antenna units increases. Here, in estimating the BS NRC matrix $\mathbf{B}$, we treat those off-diagonal entries which are corresponding to two antennas with a distance larger than a pre-defined threshold, called sparsity threshold D, as zeros, yielding a sparse matrix structure for $\hat{\mathbf{B}}$. We also define the maximum number of coupled neighboring antenna elements as $C(D)$. In Fig. 4 an example 10×10 rectangular antenna layout with $\lambda/2$ antenna spacing between the neighboring elements is shown with 3 different values of D, namely $D = 0$, 1 and $\sqrt{2}$, measured as multiples of $\lambda/2$. When $D = 0$, it is assumed that there is no mutual coupling and $C(D) = 0$, whereas for $D = 1$ and $D = \sqrt{2}$, the central antenna elements are coupled with at most $C(D) = 4$ and $C(D) = 8$ closest neighboring antenna elements. Note that, the antenna elements close to the edges of the grid are coupled with lower number of antenna units. This is illustrated in Fig. 4 where for $D = 1$, the bottom left antenna element is assumed to be coupled with only $2 \leq C(1) = 4$ antennas.

The following column-wise estimator will build on the

assumption that $\mathbf{B}$ has a sparse structure and the number of non-zero row entries in any column j, denoted as R_j, satisfies

$$R_j \leq R\left(D\right), \tag{21}$$

where $R(D) = C(D)+1$. It is further assumed that the index of non-zero entries of $\mathbf{b}_j$ are known, which is directly determined by the antenna system architecture and geometry, and the assumed pre-defined coupling threshold discussed above. With these assumptions, we define a reduced vector of dimension $R_j \times 1$, $\mathbf{b}_j^{red}$, that contains the non-zero entries of $\mathbf{b}_j$. The indexes of non-zero entries of $\mathbf{b}_j$ are then mapped to the column numbers which we keep when constructing $\mathbf{T}_j^{red}(m) \in \mathbb{C}^{N\times R_j}$. If the j'th row is kept when constructing $\mathbf{b}_j^{red}$, then similarly, the j'th column is kept to construct $\mathbf{T}_j^{red}(m)$. Based on these, we can formulate the estimation of columns of $\hat{\mathbf{B}}(m+1)$ through a reduced system of equations as

$$\hat{\mathbf{b}}_j^{red}(m+1) = \underset{\mathbf{b}_j^{red}}{\operatorname{argmin}} \left|\left|\mathbf{q}_j - \mathbf{T}_j^{red}(m)\mathbf{b}_j^{red}\right|\right|^2. \tag{22}$$

The solution to (22) is then given by

$$\hat{\mathbf{b}}_j^{red}(m+1) = \left(\mathbf{T}_j^{red}(m)\right)^{\dagger}\mathbf{q}_j. \tag{23}$$

Once the vector $\hat{\mathbf{b}}_j^{red}(m+1)$ is solved from (23), then $\hat{\mathbf{b}}_j(m+1)$ can be obtained straightforwardly by appending zeros to the appropriate rows.

Note that, when $\mathbf{A} \approx \mathbf{I}_K$, we also have $\mathbf{G}^*\mathbf{A}\mathbf{G}^{\mathrm{T}} \approx \mathbf{G}^*\mathbf{G}^{\mathrm{T}}$, where the matrix $\mathbf{G}^*\mathbf{G}^{\mathrm{T}} = \left(\mathbf{G}\mathbf{G}^{\mathrm{H}}\right)^*$ is positive semi-definite matrix and of rank K if $\mathbf{G}$ is of rank K. The obtained $\hat{\mathbf{b}}_j^{red}$ and the corresponding minimum expression from (22) depend on the corresponding values of K and D. The column space of $\mathbf{T}_j^{red}(m)$ has higher dimensionality for larger K. Thus, when D is fixed, for larger K one can solve for $\hat{\mathbf{b}}_j^{red}$ from (23) which yields smaller values of $\left|\left|\mathbf{q}_j - \mathbf{T}_j^{red}(m)\mathbf{b}_j^{red}\right|\right|^2$.

D. Proposed Estimation of $\mathbf{A}$

Next, given $\hat{\mathbf{B}}(m)$ from (19), the (refined) estimate of $\mathbf{A}$ can be formulated based on minimizing the Frobenius norm criterion as

$$\hat{\mathbf{A}}(m) = \underset{\mathbf{A}}{\operatorname{argmin}} \left|\left|\mathbf{Q} - \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^*\mathbf{A}\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right|\right|_F^2. \tag{24}$$

For diagonal $\mathbf{A}$, the solution to (24) can be obtained as

$$\hat{\mathbf{A}}(m) = \operatorname{diag}\left(\hat{\boldsymbol{\xi}}\right). \tag{25}$$

where $\hat{\boldsymbol{\xi}} = \left[\mathbf{I}_K, i\mathbf{I}_K\right]\hat{\boldsymbol{\psi}}$ and the vector $\hat{\boldsymbol{\psi}} \in \mathbb{C}^{2K\times 1}$ is given as

$$\hat{\boldsymbol{\psi}} = \left(\sum_{j=1}^{N}\bar{\mathbf{W}}_j^{\mathrm{T}}\bar{\mathbf{W}}_j\right)^{-1}\sum_{j=1}^{N}\bar{\mathbf{W}}_j^{\mathrm{T}}\bar{\mathbf{q}}_j. \tag{26}$$

In above, $\bar{\mathbf{q}}_j = \left[\Re\{\mathbf{q}_j^{\mathrm{T}}\}, \Im\{\mathbf{q}_j^{\mathrm{T}}\}\right]^{\mathrm{T}} \in \mathbb{C}^{2N\times 1}$, and defining the $N \times K$ matrix $\mathbf{W}_j = \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^*\operatorname{diag}\left(\left[\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right]_j\right)$, with $\left[\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right]_j$ being the j'th column of $\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)$, $\bar{\mathbf{W}}_j \in \mathbb{C}^{2N\times 2K}$ is given as

$$\bar{\mathbf{W}}_j = \begin{bmatrix}\Re\{\mathbf{W}_j\}, & -\Im\{\mathbf{W}_j\} \\ \Im\{\mathbf{W}_j\}, & \Re\{\mathbf{W}_j\}\end{bmatrix}. \tag{27}$$

Proof: See Appendix.

E. Computational Complexity Analysis

The computational complexity of the proposed NRC estimation method includes the complexity of both the BS side NRC estimation and the UE side NRC estimation phases, with all estimation processing carried out at the BS side. The order of the computing complexity is here analyzed, strictly-speaking per iteration round. However, since the numerical results of Section IV show that only $2-4$ iterations are needed in practice, the below analysis reflects correctly the overall order of the estimation complexity.

For BS side, (23) should be performed to estimate j'th column of $\mathbf{B}$. The first step here is to invert $\mathbf{T}_j^{red}(m)$. The size of $\mathbf{T}_j^{red}(m)$ is $N \times R_j$, thus the complexity is $\mathcal{O}\left(NR_j^2\right)$. In the next step, there is a matrix-vector multiplication as $\left(\mathbf{T}_j^{red}(m)\right)^{\dagger}\mathbf{q}_j$ where the size of $\left(\mathbf{T}_j^{red}(m)\right)^{\dagger}$ is $R_j \times N$ and the size of $\mathbf{q}_j$ is $N \times 1$. Thus the complexity is $\mathcal{O}\left(NR_j\right)$. As mentioned, these two steps are required to be performed to construct each of the K columns of $\hat{\mathbf{B}}$. Therefore, the total complexity for BS NRC estimation is of the form

$$\mathcal{O}\left(\sum_{j=1}^{K} NR_j^2\left(1+\frac{1}{R_j}\right)\right) \approx \mathcal{O}\left(KNR\left(D\right)^2\right). \tag{28}$$

For UE side NRC estimation, the first step is to calculate $\sum_{j=1}^{N}\bar{\mathbf{W}}_j^{\mathrm{T}}\bar{\mathbf{W}}_j$ where the size of $\bar{\mathbf{W}}_j$ is $2N \times 2K$. Thus the complexity of each matrix multiplication is $\mathcal{O}\left(NK^2\right)$ and total complexity is $\mathcal{O}\left(N^2K^2\right)$. Next, the obtained $2K \times 2K$ matrix is inverted and the complexity is $\mathcal{O}\left(K^3\right)$. Then, in order to calculate $\sum_{j=1}^{N}\bar{\mathbf{W}}_j^{\mathrm{T}}\bar{\mathbf{q}}_j$, a $2K \times 2N$ by $2N \times 1$ matrix multiplication is required to be performed N times. Thus, the complexity is $\mathcal{O}\left(N^2K\right)$. Finally, a matrix-vector multiplication, $\left(\sum_{j=1}^{N}\bar{\mathbf{W}}_j^{\mathrm{T}}\bar{\mathbf{W}}_j\right)^{-1}\sum_{j=1}^{N}\bar{\mathbf{W}}_j^{\mathrm{T}}\bar{\mathbf{q}}_j$, is needed where the sizes of the matrix and the vector are $2K \times 2K$ and $2K \times 1$, respectively. Thus, the complexity is $\mathcal{O}\left(K^2\right)$. Therefore, the total complexity for UE NRC estimation is of the form

$$\mathcal{O}\left(\left(N^2K^2+K^3\right)\left(1+\frac{1}{K}\right)\right) \approx \mathcal{O}\left(N^2K^2+K^3\right). \tag{29}$$

Based on the derived complexity orders in (28) and (29), we can express the total complexity for our proposed NRC estimation method as

$$\text{Total complexity} \approx \mathcal{O}\left(KNR\left(D\right)^2 + N^2K^2 + K^3\right). \tag{30}$$

For reference, the computing complexity of the method proposed in [6] is shown in [27] to be of the order $\mathcal{O}\left(\left(N^2+K^2\right)^2\right)$. Thus, the method proposed in this manuscript is of substantially lower complexity, especially when the BS array size N is large.

IV. Numerical Evaluations and Analysis

A. Basic Evaluation Settings and Performance Measures

In this section, by using extensive computer simulations, we evaluate the performance of the proposed NRC estimation

and mitigation scheme. We also compare its performance to the performance of two other existing schemes in literature, namely the direct-path based least squares (LS) known as "Argos" [2] and the generalized neighbor LS [11]. The latter is the optimized version of the generalized LS method presented in [10] and is shown in [11] to have the best performance amongst several existing NRC estimation methods. Both LS based methods estimate the BS NRC by the means of mutual coupling between the BS antennas, while they depend on the downlink demodulation pilots to compensate the NRC in the UE side.

At communication system level, we consider the downlink spectral efficiency as the key performance metric. For the proposed method and the corresponding system which does not utilize downlink demodulation pilots, this is defined as

$$\eta_{s,1} = K\left(1-\frac{\tau_u}{T}\right)\log_2\left(1+\mathrm{SINR}_{\mathrm{eff}}\right). \tag{31}$$

In above, T refers to the total number of symbols in a channel coherence interval, while $\mathrm{SINR}_{\mathrm{eff}}$ in (31) is the effective signal to interference and noise ratio (SINR) [19], [21], which can be written, based on (10)−(13), as

$$\mathrm{SINR}_{\mathrm{eff}} = \frac{\left|\sqrt{\rho_d}\hat{\alpha}_k\right|^2}{I_{\mathrm{SI}}+I_{\mathrm{IUI}}+1}, \tag{32}$$

where $\hat{\alpha}_k = \mathbb{E}\left[\beta\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right]$ is the assumed statistical complex beamforming gain of the useful signal term available at the receiver of the k'th UE. Note that the expression in (31) is strictly-speaking a lower-bound, reflecting a worst-case scenario with uncorrelated Gaussian interference [19], [21]. In (32), the powers of the SI and IUI terms are given by

$$\begin{aligned} I_{\mathrm{SI}} &= \rho_d\mathbb{E}\left[\left|\beta\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k-\hat{\alpha}_k\right|^2\right] \\ I_{\mathrm{IUI}} &= \rho_d\beta^2\sum_{i=1,i\neq k}^{K}\mathbb{E}\left[\left|\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_i\right|^2\right], \end{aligned} \tag{33}$$

where the expectations are over different NRC variable and propagation channel realizations. Finally, we note that the expectations in (33) are evaluated numerically.

On the contrary, the other two NRC estimation methods utilize $\tau_d \geq K$ downlink pilots for downlink CSI acquisition as described in [23]. Thus, we redefine the lower-bound on spectral efficiency as [23]

$$\eta_{s,2} = K\left(1-\frac{\tau_u+\tau_d}{T}\right)\mathbb{E}\left[\log_2\left(1+\mathrm{SINR}_{\mathrm{inst}}\right)\right]. \tag{34}$$

Furthermore, the instantaneous SINR, $\mathrm{SINR}_{\mathrm{inst}}$, is defined as [23]

$$\mathrm{SINR}_{\mathrm{inst}} = \frac{\left|\sqrt{\rho_d}\hat{\alpha}_{k,\mathrm{MMSE}}\right|^2}{\sum_{i=1}^{K}\mathbb{E}\left[\left|\sqrt{\rho_d}\epsilon_i\right|^2\right]+\sum_{i=1,i\neq k}^{K}\left|\sqrt{\rho_d}\hat{\alpha}_{i,\mathrm{MMSE}}\right|^2+1}, \tag{35}$$

where $\hat{\alpha}_{i,\mathrm{MMSE}}$ is the MMSE estimate of $\beta\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_i$, while $\epsilon_i = \beta\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_i - \hat{\alpha}_{i,\mathrm{MMSE}}$ is the MMSE channel estimation error and is uncorrelated to $\hat{\alpha}_{i,\mathrm{MMSE}}$ [23]. Finally, the expectations in (34)–(35) are evaluated numerically.

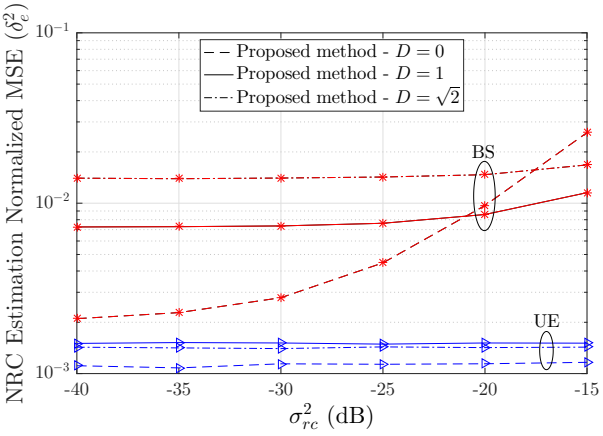
The other highly relevant performance metric is the normalized mean squared error (MSE) for NRC estimation which is defined as

$$\delta_e^2 = \begin{cases} \dfrac{\left|\left|\mathbf{B}-\hat{\mathbf{B}}\right|\right|_F^2}{||\mathbf{B}||_F^2}, & \text{for BS side} \\ \dfrac{\left|\left|\mathrm{diag}\left(\mathbf{A}\right)-\mathrm{diag}\left(\hat{\mathbf{A}}\right)\right|\right|_F^2}{||\mathrm{diag}\left(\mathbf{A}\right)||_F^2}, & \text{for UE side.} \end{cases} \tag{36}$$

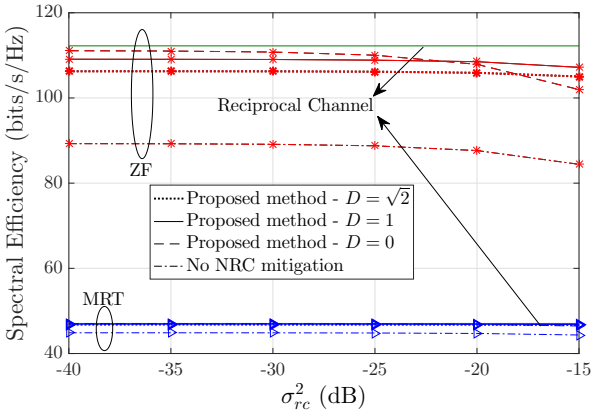
As a baseline simulation scenario, we consider a BS which is equipped with $N = 100$ infinitely thin $\lambda/2$ dipole antennas in a 10×10 square layout with $\lambda/2$ spacing as illustrated in Fig. 4. The impedance matrix $\mathbf{Z}$ in (5) is computed based on [24] for the assumed carrier-frequency of $f_c = 3.5$ GHz while $Z_0 = 50\ \Omega$ [6]. The impedances are assumed to be frequency-independent, as the modulated signal RF bandwidth is much smaller than the carrier frequency. The uplink channel matrix $\mathbf{G}$ is assumed to have i.i.d. $\mathcal{CN}(0,1)$ elements. The BS serves $K = 20$ single-antenna UEs simultaneously on the same time-frequency resource through either ZF or MRT precoding. We assume a block-fading scenario where each coherence interval contains $T = 250$ OFDM symbols. The number of uplink pilots sent by each UE in each coherence interval is equal to the number of scheduled UEs, $\tau_u = K$, and the uplink SNR in this phase is assumed to be $\rho_u = 0$ dB. As already stated in Section II-C, we assume that the statistical beamformed downlink CSI, $\mathbb{E}\left[\beta\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right]$, used by the UEs for data detection when no downlink CSI pilots are available, is computed through averaging over both channel and NRC variables. For the specific assumed example case of $\mathbf{G}$ having i.i.d. $\mathcal{CN}(0,1)$ elements, this can be shown to yield [23]

$$\mathbb{E}\left[\beta\mathbf{h}_k^{\mathrm{T}}\mathbf{u}_k\right] = \begin{cases} \sqrt{\frac{(N-K)\tau_u\rho_u}{K(\tau_u\rho_u+1)}}, & \text{for ZF} \\ \sqrt{\frac{N\tau_u\rho_u}{K(\tau_u\rho_u+1)}}, & \text{for MRT.} \end{cases} \tag{37}$$

Whereas, for the cases where UEs rely on downlink pilots for decoding purposes, i.e., direct-path based LS and generalized neighbor LS methods, the number of downlink pilots in each coherence interval is set to be $\tau_d = K$ [23], and their SNR is equal to the downlink SNR in data transmission phase which is assumed to be $\rho_d = 20$ dB. The SNR of the coupling channel between two neighboring antennas is set to be 80 dB for the two mentioned NRC mitigation reference methods [11]. The uplink and downlink SNRs for the pilot signaling in the proposed NRC estimation framework are set to be $\tilde{\rho}_u = 0$ dB and $\tilde{\rho}_d = 10$ dB, respectively. Note that while different UEs are likely to have different $\tilde{\rho}_u$ values in practice, in order to have fair comparison between the proposed and the two mentioned state-of-the-art NRC estimation methods, we have deliberately assumed a worst case scenario in which all the UEs have very low uplink SNR value of $\tilde{\rho}_u = 0$ dB. In the proposed method, the estimated NRC matrices are averaged over 10 neighboring subcarriers, $C_{sc} = 10$, over which the NRC realizations are assumed to be constant. Finally, the variances of transceivers frequency-responses in both the BS and the UE side are assumed to be -20 dB, i.e., $\sigma_{\mathrm{L}}^2 = \sigma_{\mathrm{F}}^2 =$

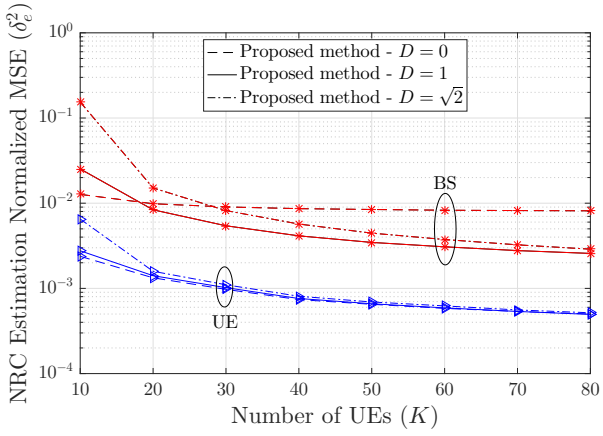


(a)

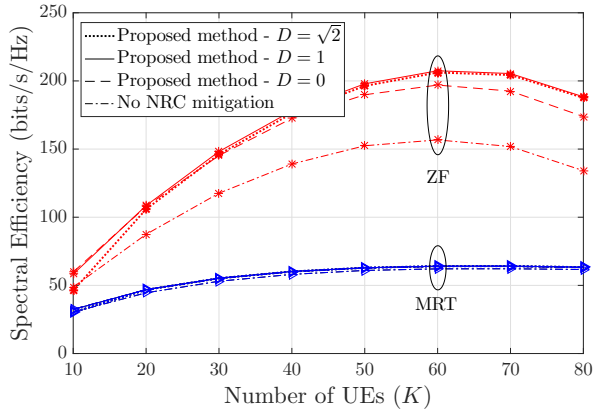


(b)

Fig. 5. (a) NRC estimation normalized MSE and (b) system spectral efficiency vs. input reflection coefficients variance (σ_{rc}^2) for different values of the sparsity threshold D with $N = 100$, $K = 20$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.



(a)



(b)

Fig. 6. (a) NRC estimation normalized MSE and (b) system spectral efficiency vs. number of UEs (K) for $N = 100$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

-20 dB. For simulation simplicity, the distributions of the NRC variables are assumed to be Gaussians. These are the baseline simulation settings, while some of the parameter values are also varied in the evaluations.

B. Numerical Results

1) Effect of the Sparsity Distance Threshold D: Here, we will study the effect of D on the normalized MSE and the system spectral efficiency. In this respect, Fig. 5a illustrates the normalized MSE of the UE and BS NRC estimation under the baseline system settings, with the value of σ_{rc}^2 being varied. It can be seen that the choice of $D = 0$, i.e., estimating only the diagonal elements of $\mathbf{B}$, yields the lowest MSE for UE NRC estimation. Note that, in the proposed NRC estimation method, the choice of D influences the UE side estimation as well since $\mathbf{A}$ and $\mathbf{B}$ are estimated iteratively as described in Section III-B. On the other hand, the highest BS NRC estimation accuracy is achieved for $D = 0$ only when $\sigma_{rc}^2 \leq -21$ dB, whereas higher estimation accuracy is obtained for $D = 1$ when $\sigma_{rc}^2 > -21$ dB. Following that, the spectral efficiencies plotted in Fig. 5b illustrate the combined effect of UE and BS NRC estimation. As can be seen, the highest spectral efficiency is achieved for $D = 0$ when $\sigma_{rc}^2 \leq -22$ dB and for $D = 1$ when $\sigma_{rc}^2 > -22$ dB.

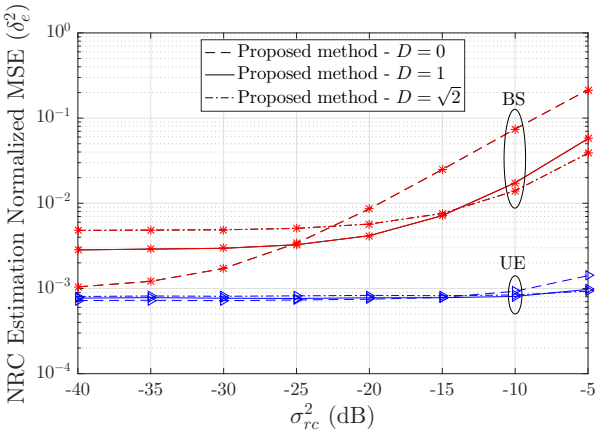


Fig. 7. NRC estimation normalized MSE vs. input reflection coefficients variance (σ_{rc}^2) for different values of the sparsity threshold D with $N = 100$, $K = 40$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

For fixed NRC characteristics of $\sigma_{\mathrm{L}}^2 = \sigma_{\mathrm{F}}^2 = \sigma_{rc}^2 = -20$ dB, Fig. 6 evaluates the normalized estimation MSE and the system spectral efficiency for different values of D and against the number of scheduled UEs K. Fig. 6a shows that higher

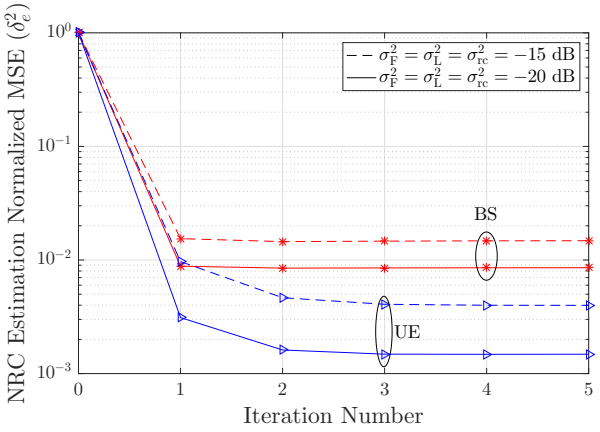


Fig. 8. NRC estimation normalized MSE vs. NRC estimation iteration number for $N = 100$, $K = 20$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

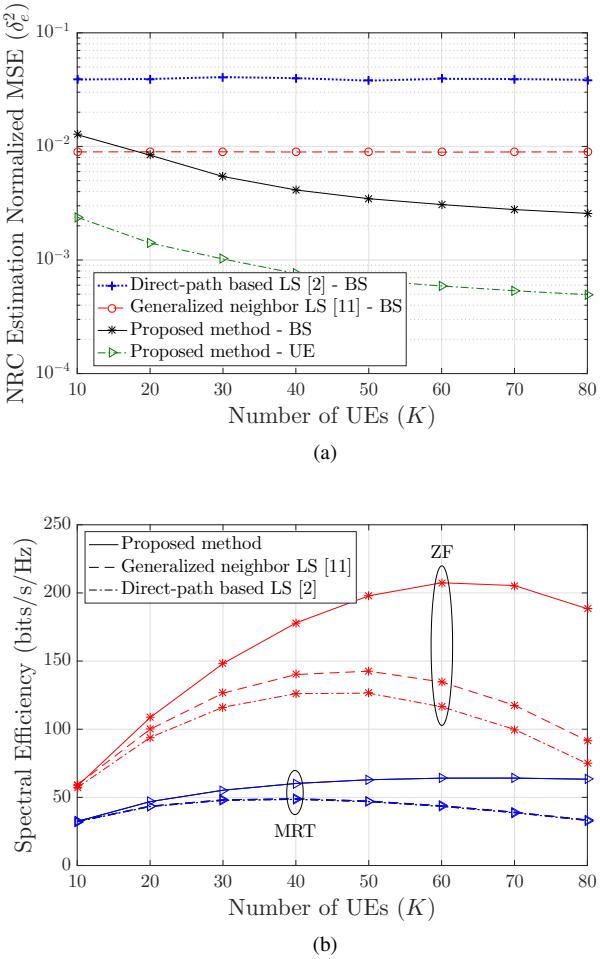


Fig. 9. (a) NRC estimation normalized MSE and (b) system spectral efficiency vs. number of UEs (K) for $N = 100$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

UE and BS NRC estimation accuracy is achieved for $D = 0$ when $K < 20$, whereas when the number of scheduled users exceeds $K \geq 20$ the choice of $D = 1$ yields the highest BS NRC estimation accuracy. For $K \geq 20$, UE NRC estimation performances are largely similar for all three choices of D. Following these, Fig. 6b illustrates that from spectral efficiency perspective, the optimum sparsity distance threshold is $D = 0$ for $K < 20$ and $D = 1$ for $K \geq 20$. Thus, except in Fig. 7 which similarly evaluates the impacts of D on the NRC estimation performance, in the continuation $D = 0$ and $D = 1$ will be used under the settings of $K < 20$ and $K \geq 20$, respectively. As discussed in the previous section, when $\mathbf{A} \approx \mathbf{I}_K$, $\mathbf{G}^*\mathbf{A}\mathbf{G}^{\rm T}$ which is used in the estimation process is of rank K. Therefore, having higher number of K increases the accuracy of the BS NRC estimation in the proposed method which facilitates the estimation of more non-diagonal elements in $\mathbf{B}$, i.e., higher values for D. In order to better illustrate this point, Fig. 7 provides additional results for higher σ_{rc}^2 values with $K = 40$, while keeping other simulation setup and parameter settings the same as in Fig. 5a. As expected, for the considered higher value of K (i.e., $K = 40$), $D = \sqrt{2}$ outperforms $D = 1$ at higher σ_{rc}^2 values.

It should be noted that for all the cases in Fig. 5, Fig. 6, and Fig. 7, the proposed iterative NRC estimator is executed over a sufficient amount of iterations such that convergence is obtained. This is commonly in the order of 4 iterations, as illustrated more specifically next.

2) Effect of the Number of Iterations: Fig. 8 illustrates the reduction in NRC estimation normalized MSE over NRC estimation iteration steps. It is shown in Fig. 8 that, even with high NRC levels of $\sigma_{\rm L}^2 = \sigma_{\rm F}^2 = \sigma_{rc}^2 = -15$ dB, having 4 iteration rounds is sufficiently good for the proposed NRC estimator to converge. Therefore, in the continuation, we set the number of iteration rounds to 4.

3) Effect of the Number of Scheduled Users K: In Fig. 9, the NRC estimation normalized MSE and the system spectral efficiency are plotted against the number of scheduled UEs K for $\sigma_{rc}^2 = -20$ dB. Fig. 9a shows that while the direct-path based LS has the worst performance, the proposed method is the best option for estimating BS NRC for $K \geq 20$ with a high accuracy where MSE is in the order of 10^{-3}. For the direct-path based LS [2] and the generalized neighbor LS [11], the normalized MSE for UE side NRC is not shown. It is mentioned in [2] and [11] that additional downlink pilot signaling per coherence interval can be used together with UE side estimation for UE side NRC acquisition. However, no detailed description is provided on the actual pilot signal structure or the actual estimation method.

The corresponding system spectral efficiency performances are evaluated and shown in Fig. 9b. The proposed NRC estimation and mitigation scheme clearly outperforms the direct-path based LS and the generalized neighbor LS methods. The difference between the performance of the proposed method and the other two methods increases as K grows. Remarkably, for $K = 70$, the difference between the proposed method and the other two methods is already in the order of 100 bits/s/Hz. Another advantage in utilizing the proposed NRC estimation scheme is that the optimum number of UEs K_{opt}, which is defined as the number of scheduled UEs which maximizes the spectral efficiency, is higher compared to the other two NRC estimation methods. For instance under ZF precoding K_{opt} is around 60 for the proposed method whereas for the LS based methods K_{opt} is between 40 and 50.

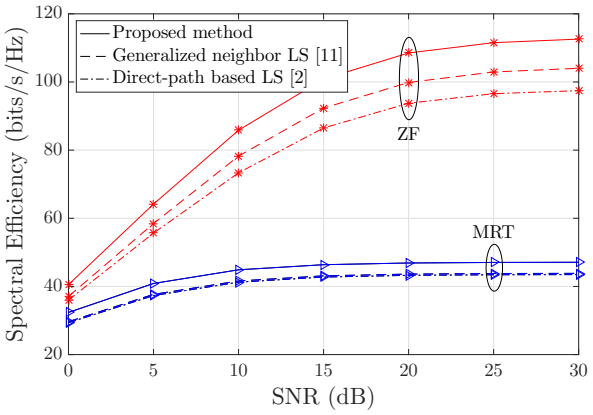


Fig. 10. Spectral efficiency vs. downlink SNR (ρ_d) for $N = 100$, $K = 20$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

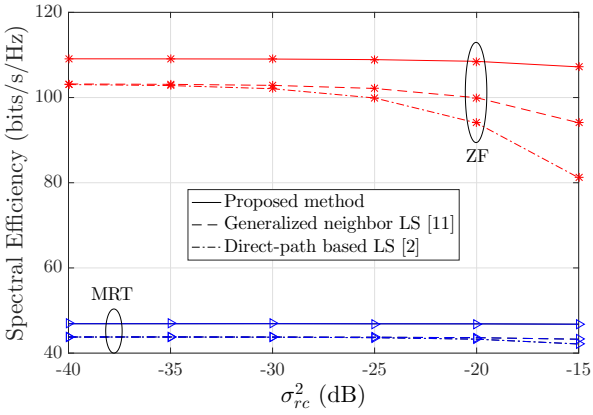


Fig. 11. Spectral efficiency vs. input reflection coefficients variance (σ^2_{rc}) for $N = 100$, $K = 20$, $\tau_u = K$, $\rho_u = 0$ dB, $T = 250$.

4) Effect of Downlink SNR: In Fig. 10, the system spectral efficiency is plotted against the downlink SNR when $\sigma^2_{rc} = -20$ dB. The results show clear advantage in employing the proposed method in estimating NRC for all SNR values. The proposed estimation and mitigation method outperforms the LS based methods for both low and high SNR regions. Especially, the performance difference is most visible for high SNR region under ZF precoding.

5) Effect of the Input Reflection Coefficient: Fig. 11 shows the impact of the variance of the input reflection coefficients on the achievable spectral efficiency. The proposed estimation and mitigation method again outperforms the other two LS based methods. The difference between the proposed method and the other two methods increases as σ^2_{rc} grows, which is due to the ability of the proposed method to estimate the non-diagonal elements in the BS NRC matrix. It should be noted that $D = 1$ is used for obtaining the results in Fig. 11, and there is still room for improving the performance of the proposed method by adaptively selecting the optimum D according to the level of σ^2_{rc} as shown already in Fig. 5b and Fig. 7.

6) Summary of the Obtained Results: Overall, as observed through the extensive numerical evaluations in various scenarios, the proposed NRC estimation method outperforms the other two state-of-the-art methods. Selected technical aspects can be summarized as follows:

- Employing the proposed NRC estimation method eliminates the need to send downlink demodulation pilots since the proposed OTA framework facilitates estimating both the BS side and UE side NRC characteristics in the base station. Therefore, more time-frequency resources can be allocated in each coherence interval for actual downlink data transmission purposes which improves the spectral efficiency.
- The proposed NRC estimation method is more and more superior over the two reference methods when the number of scheduled UEs K grows. One reason for this is that increasing K is forcing the other two NRC estimation methods to spend more time for downlink pilot transmission in each coherence interval, while a larger number of scheduled users improves the accuracy of the proposed NRC estimation method.
- Due to the ability to estimate also non-diagonal elements of the BS NRC matrix, the difference between the performance of the proposed NRC estimation method and the other two methods increases as the power of BS antenna mutual coupling mismatch grows.

V. Conclusion

In this work, we proposed an efficient NRC estimation and mitigation framework for multi-user massive MIMO TDD networks to compensate the jointly coexisting BS and UE side NRC. In general, even relatively modest NRC levels can cause significant performance loss in the achievable spectral efficiency when only standard NRC-blind MRT or ZF downlink precoding is employed. A novel OTA-based approach incorporating a dedicated round-trip pilot signaling with feasible pilot overhead together with sparsity-aided efficient iterative estimation techniques were proposed for the acquisition of the NRC matrices at the BS. Unlike the existing state-of-the-art methods, the proposed NRC estimation method acquires both the UE transceiver NRC as well as the BS transceiver and antenna system NRC, and does not rely on downlink demodulation pilots during the actual data transmission phase to compensate the NRC in the UE side. Therefore, it can be efficiently employed in massive MIMO systems that rely only on the statistical knowledge of the beamformed downlink channels at terminals for data decoding with very low system pilot overhead. The extensive computer simulations showed that for practical values of the NRC levels, SNRs and the number of spatially multiplexed users, the proposed estimation and mitigation method clearly outperforms the existing state-of-the-art methods in terms of the system spectral efficiency.

Appendix
Proof for Estimation of $\mathbf{A}$

Let

$$\mathcal{L} \triangleq \left|\left|\mathbf{Q} - \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^{*}\mathbf{A}\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right|\right|_F^2. \tag{38}$$

Then,

$$\begin{aligned}\mathcal{L} &= \sum_{j=1}^{N} \left|\left| \mathbf{q}_j - \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^* \mathbf{A} \left[\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right]_j \right|\right|^2 \\ &= \sum_{j=1}^{N} \left|\left| \mathbf{q}_j - \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^* \mathrm{diag}\left(\left[\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right]_j\right)\boldsymbol{\xi} \right|\right|^2,\end{aligned} \tag{39}$$

where $\boldsymbol{\xi} \triangleq [a_1, a_2, \cdots, a_K]^{\mathrm{T}}$.

By using $\mathbf{W}_j = \sqrt{\tilde{\rho}_u}\sqrt{\tilde{\rho}_d}\hat{\mathbf{G}}^* \mathrm{diag}\left(\left[\hat{\mathbf{G}}^{\mathrm{T}}\hat{\mathbf{B}}(m)\right]_j\right)$, (39) can be re-written as

$$\begin{aligned}\mathcal{L} &= \sum_{j=1}^{N} ||\mathbf{q}_j - \mathbf{W}_j\boldsymbol{\xi}||^2 \\ &= \sum_{j=1}^{N} \left|\left|\bar{\mathbf{q}}_j - \bar{\mathbf{W}}_j\boldsymbol{\psi}\right|\right|^2 \triangleq \bar{\mathcal{L}}\left(\boldsymbol{\psi}\right),\end{aligned} \tag{40}$$

where $\boldsymbol{\psi} \triangleq \left[\Re\{\boldsymbol{\xi}^{\mathrm{T}}\}, \Im\{\boldsymbol{\xi}^{\mathrm{T}}\}\right]^{\mathrm{T}}$. Therefore, the solution $\hat{\mathbf{A}}(m)$ can be obtained by

$$\hat{\mathbf{A}}(m) = \mathrm{diag}\left(\left[\mathbf{I}_K, i\mathbf{I}_K\right]\hat{\boldsymbol{\psi}}\right), \tag{41}$$

where

$$\hat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}\in\mathbb{R}^K}{\mathrm{argmin}}\ \bar{\mathcal{L}}\left(\boldsymbol{\psi}\right). \tag{42}$$

Since $\bar{\mathcal{L}}(\boldsymbol{\psi})$ is convex, (42) can be solved from the partial derivative equation $\frac{\partial\bar{\mathcal{L}}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}} = 0$, which finally yields the solution given by (26).

ACKNOWLEDGMENT

The estimation of $\mathbf{A}$ is building on the work presented in [28]. The proof in Appendix A is a result of discussion with the authors of [28], H. Q. Ngo and E. G. Larsson, whose technical guidance is greatly acknowledged.

REFERENCES

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[2] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 53–64.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[4] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation," in *Radio and Wireless Conference, 2003. RAWCON '03. Proceedings*, Aug 2003, pp. 183–186.

[5] Y. Zou, O. Raeesi, R. Wichman, A. Tolli, and M. Valkama, "Analysis of channel non-reciprocity due to transceiver and antenna coupling mismatches in TDD precoded multi-user MIMO-OFDM downlink," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sept 2014, pp. 1–7.

[6] M. Petermann, M. Stefer, F. Ludwig, D. Wubben, M. Schneider, S. Paul, and K. D. Kammeyer, "Multi-user pre-processing in multi-antenna OFDM TDD systems with non-reciprocal transceivers," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3781–3793, September 2013.

[7] H. Wei, D. Wang, and X. You, "Reciprocity of mutual coupling for TDD massive MIMO systems," in *Wireless Communications Signal Processing (WCSP), 2015 International Conference on*, Oct 2015, pp. 1–5.

[8] W. Zhang, H. Ren, C. Pan, M. Chen, R. C. de Lamare, B. Du, and J. Dai, "Large-scale antenna systems with UL/DL hardware mismatch: Achievable rates analysis and calibration," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1216–1229, April 2015.

[9] F. Athley, G. Durisi, and U. Gustavsson, "Analysis of massive MIMO with hardware impairments and different channel models," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, May 2015, pp. 1–5.

[10] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1815–1831, April 2014.

[11] J. Vieira, F. Rusek, and F. Tufvesson, "Reciprocity calibration methods for massive MIMO based on antenna coupling," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 3708–3712.

[12] H. Wei, D. Wang, H. Zhu, J. Wang, S. Sun, and X. You, "Mutual coupling calibration for multiuser massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 606–619, Jan 2016.

[13] H. Wei, D. Wang, J. Wang, and X. You, "TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent," *Science China Information Sciences*, vol. 59, no. 10, p. 102306, 2015.

[14] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, 2005.*, vol. 1, August 2005, pp. 403–406.

[15] Y. Zou, O. Raeesi, and M. Valkama, "Efficient estimation and compensation of transceiver non-reciprocity in precoded TDD multi-user MIMO-OFDM systems," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sept 2014, pp. 1–7.

[16] M. Guillaud and F. Kaltenberger, "Towards practical channel reciprocity exploitation: Relative calibration in the presence of frequency offset," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 2525–2530.

[17] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *2010 Future Network Mobile Summit*, June 2010, pp. 1–10.

[18] O. Raeesi, A. Gokceoglu, Y. Zou, E. Bjrnson, and M. Valkama, "Performance analysis of multi-user massive MIMO downlink under channel non-reciprocity and imperfect CSI," *IEEE Transactions on Communications*, vol. PP, no. 99, pp. 1–1, 2018.

[19] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, February 2013.

[20] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel estimation and linear precoding in multiuser multiple-antenna TDD systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2102–2116, Jun 2011.

[21] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, August 2011.

[22] E. G. Larsson and H. V. Poor, "Joint beamforming and broadcasting in massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 3058–3070, April 2016.

[23] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, Oct 2013, pp. 293–298.

[24] S. A. Schelkunoff and H. T. Friis, *Antennas Theory and Practice*. New York: John Wiley & Sons, 1952.

[25] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, Sept 2004.

[26] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2921–2935, May 2017.

[27] M. Petermann, M. Stefer, D. Wbben, M. Schneider, and K. D. Kammeyer, "Low-complexity calibration of mutually coupled non-reciprocal multi-antenna OFDM transceivers," in *2010 7th International Symposium on Wireless Communication Systems*, Sept 2010, pp. 285–289.

[28] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3249–3252.

Orod Raeesi received his M.Sc. degree (with distinction) from Tampere University of Technology (TUT), Tampere, Finland, in 2011, and is currently pursuing the Ph.D. degree at TUT. Currently, he is working as a communications system specialist with Nokia Mobile Networks, Espoo, Finland. His research interests include IEEE 802.11 MAC and PHY layer challenges, massive MIMO systems, TDD channel non-reciprocity, ultra-reliable low latency communications, system-level simulations, and 5G mobile radio networks.

Ahmet Gokceoglu received M.Sc. (2010) and Ph.D. Degrees (2014) from the Department of Electronics and Communications Engineering, Tampere University of Technology, Finland, where he also held a postdoctoral researcher position (2014-2016) together with a visiting researcher position at Linköping university (2015). His research interests are information theory, modeling and performance analysis of MIMO-OFDM systems, advanced transmitter and receiver signal processing techniques and algorithm design for MAC and physical layer. He is currently working as an algorithm specialist with Huawei, Stockholm, Sweden.

Mikko Valkama (S'00–M'01–SM'15) was born in Pirkkala, Finland, on November 27, 1975. He received the M.Sc. and Ph.D. Degrees (both with honors) in electrical engineering (EE) from Tampere University of Technology (TUT), Finland, in 2000 and 2001, respectively. In 2002, he received the Best Ph.D. Thesis -award by the Finnish Academy of Science and Letters for his dissertation entitled "Advanced I/Q signal processing for wideband receivers: Models and algorithms". In 2003, he was working as a visiting post-doc research fellow with the Communications Systems and Signal Processing Institute at SDSU, San Diego, CA. Currently, he is a Full Professor and Laboratory Head at the Laboratory of Electronics and Communications Engineering at TUT, Finland. His general research interests include radio communications, communications signal processing, estimation and detection techniques, signal processing algorithms for flexible radios, cognitive radio, full-duplex radio, radio localization, and 5G mobile radio networks.