

Linear Interference Suppression With Covariance Mismatches in MIMO-OFDM Systems

Chien-Chun Cheng, *Student Member, IEEE*, Serdar Sezginer, *Senior Member, IEEE*, Hikmet Sari, *Fellow, IEEE*, and Yu T. Su, *Senior Member, IEEE*

Abstract—Interference cancellation is a key design concern for next-generation communication systems. One practical approach is the interference rejection combining (IRC) scheme, which is already being considered as the base receiver for next-generation 3GPP specifications. IRC treats interference as a stationary Gaussian process to facilitate simple interference suppression. However, this stationary assumption does not hold in practice; for example, the variances of the pilot signal and of the data signal from the interfering base station (eNB) are quite different. This considerably impacts the possible improvements of an interference-aware receiver. In this paper, novel interference suppression schemes are proposed, which handle separately the interfering pilot and the interfering data signals. The proposed schemes take into account channel estimation errors and the errors in estimating the covariance of interference plus noise. The performance of the proposed schemes is validated by using realistic channel models and asynchronous scenarios for MIMO-OFDM systems.

Index Terms—Multi-user interference, MIMO-OFDM, imperfect CSI.

I. INTRODUCTION

WITH the explosive growth of wireless data traffic, there is a pressing need to densify access nodes to improve the wireless network capacity. For both conventional cell-splitting and pico/small cells in heterogeneous networks (HetNets) [1], the cell density is severely limited by inter-cell interference. Therefore, intercell interference cancellation and avoidance techniques play a critical role in future-generation mobile networks [2]. Current LTE standardization efforts focus on transmit strategies, e.g., fractional frequency reuse (FFR), almost blank subframe (ABS), and multi-user MIMO (MU-MIMO) [3]. However, it is equally important that user equip-

ment (UE) can detect, report and suppress interference in a distributed manner.

The successive interference cancellation (SIC) scheme estimates co-channel interference (CCI) and sequentially cancels CCI from the received waveform according to their signal strengths [4], [5]. For such a scheme to work properly, a receiver has to be able to obtain accurate estimates of the channel response, frequency offsets, time delays, and multiple interferers' magnitudes. In practice, accurate estimation is very difficult to obtain and may require modification/coordination of pilots used from different cells [6]. Although the corresponding theory is well understood, the high implementation cost has hindered its application in UEs.

Another approach for UE is interference suppression via whitening or equalization [7], [8]. This technique estimates the statistics of the interference plus noise to suppress interference and is known as interference rejection combining (IRC). The main idea is to mitigate interference by exploiting the spatial diversity with multiple antennas. The IRC scheme is attractive for many reasons. First, it is feasible even when interference statistics are not known, and it significantly improves the throughput performance of the cell-edge users compared to the minimum mean square error (MMSE) receiver that treats interference as additive white Gaussian noise (AWGN) [9]. It is also suitable for dealing with asynchronous interference [10] and does not involve high complexity or accurate synchronization with the interfering base station (eNB).

In this paper, we consider a MIMO-OFDM system in an interference limited scenario. The identically distributed interference assumption is not valid for standardized cellular systems like LTE as the pilot and data symbols' time-frequency locations (TFLs) are different from cell to cell. The latter TFL offsets between desired and interfering LTE signals result in different cross-correlation statistics in different parts of a signal frame. The main focus of this paper is to take into account the mismatches between covariance matrices of the interfering pilot and data waveforms and propose robust interference suppression receivers. Based on the frame format of the LTE standard, we propose a simplified one-dimensional model for the performance analysis and algorithm developments. We derive three novel receiver structures based on MMSE and SIC criteria. The performance of the proposed schemes is assessed using the triply selective channel and their robustness is confirmed by the simulated behaviors in different asynchronous cases.

The system model we consider is related to the two-user interference channel (two-user IC) [11]–[13] and the MIMO X channel [14], [15]. However, there are some differences:

Manuscript received December 20, 2013; revised May 28, 2014 and August 25, 2014; accepted September 30, 2014. Date of publication October 15, 2014; date of current version December 8, 2014. The associate editor coordinating the review of this paper and approving it for publication was G. Yue.

C.-C. Cheng is with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu 30056, Taiwan, and also with the Department of Telecommunications, SUPÉLEC, Plateau de Moulon, Gif sur Yvette 91190, France (e-mail: cccheng.cm98g@nctu.edu.tw; chien-chun.cheng@supelec.fr).

S. Sezginer is with SEQUANS Communications, Les Portes de La Défense, Colombes 92700, France (e-mail: serdar@sequans.com).

H. Sari is with the Department of Telecommunications, SUPÉLEC, Plateau de Moulon, Gif sur Yvette 91190, France, and also with SEQUANS Communications, Les Portes de La Défense, Colombes 92700, France (e-mail: hikmet.sari@supelec.fr; hikmet@sequans.com).

Y. T. Su is with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu 30056, Taiwan (e-mail: ytsu@mail.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2014.2363097

1) we deal with a more realistic model by using a frame structure similar to that of the LTE specifications, 2) channel estimation errors and covariance estimation errors are considered and compensated in the symbol decoding, 3) the asynchronous case is taken into account which is often neglected in studying the two-user IC, and 4) we employ no interference alignment (IA) schemes which may require more ideal assumptions at eNBs but our results can be extended to the case when IA is in place for further improvements.

Notations: We use upper and lower case boldface letters for matrices and vectors. The symbol $(\cdot)^T$ denotes the transpose, $(\cdot)^*$ yields the complex conjugate, $(\cdot)^H$ is the Hermitian conjugate. The symbol $\|\cdot\|$ denotes the Frobenius norm, $\det(\cdot)$ is the determinant, $|\cdot|$ represents the absolute value, and $\text{vec}(\cdot)$ yields the column vectorization. The (i, j) -th entry of a matrix \mathbf{A} is denoted as $[\mathbf{A}]_{ij}$. The symbol \otimes represents the Kronecker product and \sim denotes “distributed as”. The expectation is defined as $\mathbb{E}[\cdot]$, and the variance operation is $\text{var}[\cdot]$. The distribution of a circularly symmetric complex Gaussian random vector \mathbf{z} with mean $\boldsymbol{\mu}_z$ and covariance $\boldsymbol{\Sigma}_z$ is denoted by $\mathbf{z} \sim \mathcal{N}_c(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$.

II. SYSTEM MODEL

We consider a two-user IC for MIMO-OFDM systems. The received waveform of the UE can be expressed as

$$\mathbf{y}^{tk} = \mathbf{H}^{tk} \mathbf{p} x^{tk} + \alpha \bar{\mathbf{H}}^{tk} \bar{\mathbf{p}} \bar{x}^{tk} + \mathbf{z}^{tk} \in \mathcal{C}^{N_R}, \quad (1)$$

where at each time-frequency index (t, k) , \mathbf{H}^{tk} , $\bar{\mathbf{H}}^{tk}$ are the serving channel matrix and the interfering channel matrix of size $N_R \times N_T$, respectively, \mathbf{y}^{tk} is the N_R dimensional received signal vector, x^{tk} , \bar{x}^{tk} are the transmitted symbols, \mathbf{p} , $\bar{\mathbf{p}}$ are the N_T dimensional precoding vectors, \mathbf{z}^{tk} is the N_R dimensional AWGN vector, and α is a positive real number related to the signal-to-interference ratio (SIR). To simplify our discussion, each node is equipped with two antennas, i.e., $N_R = 2$ and $N_T = 2$.¹ We further assume that the entry of \mathbf{z} is i.i.d. Gaussian of the form $\mathcal{N}_c(0, N_0)$. Probability density functions (pdfs) of channel matrices \mathbf{H}^{tk} , $\bar{\mathbf{H}}^{tk}$ are assumed to be $\text{vec}(\mathbf{H}) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I})$ and $\text{vec}(\bar{\mathbf{H}}) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I})$, respectively. The precoding vectors \mathbf{p} and $\bar{\mathbf{p}}$ are chosen from the LTE codebook set \mathcal{P} , e.g.,

$$\mathcal{P} = \frac{1}{\sqrt{2}} \cdot \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ i \end{bmatrix}, \begin{bmatrix} 1 \\ -i \end{bmatrix} \right\}. \quad (2)$$

The data-pilot format of the LTE specification is shown in Fig. 1, where two resource blocks (RBs) with non-overlapping pilots are considered between the serving eNB and the interfering eNB.

A. One-Dimensional Model

Assume the channel is static over the time and frequency spans within one RB, the pilot number is $N_p = 16$ and the number of data symbols is $N_d = 152$. Since the pilot symbols

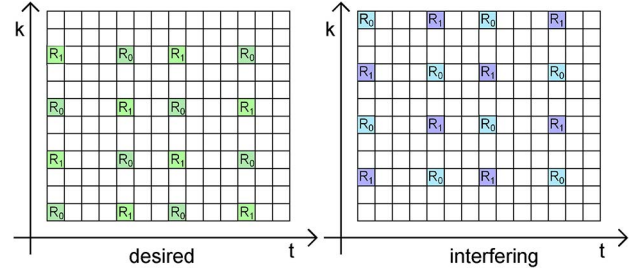


Fig. 1. Mapping of the serving and interfering pilot symbols, where R_0 and R_1 indicate the pilot positions of the 1st and 2nd antenna, respectively.

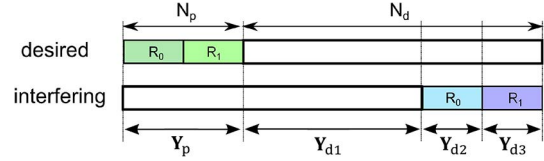


Fig. 2. The one-dimensional model which arranges pilot and data symbols.

from two eNBs are not placed at the same time-frequency locations, the received signal is subjected to interference with different statistics. Figs. 1 and 2 depict a typical relative desired and interfering signal pattern where, without loss of generality, we place the serving pilot symbols at the head of the frame and the interfering pilot symbols at the end. These two figures suggest that we divide the received samples in a typical frame into four intervals, giving

$$\mathbf{Y}_p = \mathbf{H}\mathbf{X}_p + \alpha \bar{\mathbf{H}}\bar{\mathbf{p}}\bar{\mathbf{X}}_{d1} + \mathbf{Z}_p \in \mathcal{C}^{2 \times 16} \quad (3)$$

$$\mathbf{Y}_{d1} = \mathbf{H}\mathbf{p}\mathbf{X}_1 + \alpha \bar{\mathbf{H}}\bar{\mathbf{p}}\bar{\mathbf{X}}_{d2} + \mathbf{Z}_{d1} \in \mathcal{C}^{2 \times (152-16)} \quad (4)$$

$$\mathbf{Y}_{d2} = \mathbf{H}\mathbf{p}\mathbf{X}_2 + \alpha \bar{\mathbf{H}}\bar{\mathbf{X}}_{p1} + \mathbf{Z}_{d2} \in \mathcal{C}^{2 \times 8} \quad (5)$$

$$\mathbf{Y}_{d3} = \mathbf{H}\mathbf{p}\mathbf{X}_3 + \alpha \bar{\mathbf{H}}\bar{\mathbf{X}}_{p2} + \mathbf{Z}_{d3} \in \mathcal{C}^{2 \times 8} \quad (6)$$

where $\mathbf{Y}_p, \mathbf{Y}_{d1}, \mathbf{Y}_{d2}, \mathbf{Y}_{d3}$ are the received signal matrices, \mathbf{X}_p is the serving pilot matrix, $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ are the serving data matrices, $\bar{\mathbf{X}}_{d1}, \bar{\mathbf{X}}_{d2}$ are the interfering data matrices, $\bar{\mathbf{X}}_{p1}, \bar{\mathbf{X}}_{p2}$ are the interfering pilot matrices while $\mathbf{Z}_p, \mathbf{Z}_{d1}, \mathbf{Z}_{d2}, \mathbf{Z}_{d3}$ represent AWGN. Note that during the two intervals specified by (3) and (4), the covariance of interference plus noise is the same. The above model assumes a synchronous network scenario, i.e., all UEs are synchronized with a universal clock such that both the desired and interfering downlink signals arrive at the desired UE simultaneously. We show in Section IV-C that the proposed receivers work in the asynchronous case as well. In LTE, pilots from a multi-antenna transmitter are transmitted in a time division manner, i.e., when a pilot signal is transmitted by one of the antennas, the other antennas remain silent to avoid self-interference. We further assume that the pilot signal is not spatially precoded so that the matrix products of the pilot signal are

$$\bar{\mathbf{X}}_{p1} \bar{\mathbf{X}}_{p1}^H = \frac{\mathcal{E}_p N_p}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (7)$$

$$\bar{\mathbf{X}}_{p2} \bar{\mathbf{X}}_{p2}^H = \frac{\mathcal{E}_p N_p}{2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad (8)$$

where $\bar{\mathbf{X}}_{p1}$ is the transmitted pilot signal from the 1st antenna and $\bar{\mathbf{X}}_{p2}$ is from the 2nd antenna. The pilot energy is denoted

¹Extensions to different numbers of antennas are straightforward.

by \mathcal{E}_p and the data energy is \mathcal{E}_d . We further define the signal-to-noise ratio (SNR) and SIR as $\text{SNR} = 1/N_0$ and $\text{SIR} = \alpha^{-2}$ under the unit energy assumption, e.g., $\mathcal{E}_d = \mathcal{E}_p = 1$.

B. Problem Formulation

The problem is finding the optimal linear combination of the received signal for decoding the data of interest ($\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$). Assuming the CSI from the serving eNB and the interfering eNB are given, the covariance matrices of the received signal $\mathbf{Y}_{d1}, \mathbf{Y}_{d2}, \mathbf{Y}_{d3}$ are as follows:

$$\Sigma_{\mathbf{Y}_{d1}} = \mathbb{E}[\mathbf{Y}_{d1} | \mathbf{h}, \bar{\mathbf{h}}, \bar{\mathbf{X}}_{p1}, \bar{\mathbf{X}}_{p2}] \quad (9)$$

$$= \mathbf{h}\mathbf{h}^H + \bar{\mathbf{h}}\bar{\mathbf{h}}^H + N_0\mathbf{I} \quad (10)$$

$$\Sigma_{\mathbf{Y}_{d2}} = \mathbf{h}\mathbf{h}^H + \bar{\Psi}_{p1} + N_0\mathbf{I} \quad (11)$$

$$\Sigma_{\mathbf{Y}_{d3}} = \mathbf{h}\mathbf{h}^H + \bar{\Psi}_{p2} + N_0\mathbf{I}, \quad (12)$$

where $\mathbf{h} = \mathbf{H}\mathbf{p}$ and $\bar{\mathbf{h}} = \alpha\bar{\mathbf{H}}\bar{\mathbf{p}}$ are the equivalent channel vectors. The covariance matrices of size 2×2 are given by

$$\bar{\Psi}_{p1} = \alpha^2 \bar{\mathbf{H}} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \bar{\mathbf{H}}^H \quad (13)$$

$$\bar{\Psi}_{p2} = \alpha^2 \bar{\mathbf{H}} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \bar{\mathbf{H}}^H. \quad (14)$$

In this case, the conventional IRC (which treats interference as a single Gaussian process) naturally leads to limited system performances. In this paper, we propose new approaches based on two basic concepts: exploiting the covariance estimation with MMSE and regenerating the pilot signal with SIC.

III. PROPOSED SCHEMES

A. Diagonal Loading (IRC-DL)

Treating interference as Gaussian noise, the optimal linear combining based on LMMSE criterion has the following form,

$$\begin{cases} \hat{\mathbf{X}}_1 = \mathbf{h}^H (\Sigma_{\mathbf{Y}_{d1}})^{-1} \mathbf{Y}_{d1} \\ \hat{\mathbf{X}}_2 = \mathbf{h}^H (\Sigma_{\mathbf{Y}_{d2}})^{-1} \mathbf{Y}_{d2} \\ \hat{\mathbf{X}}_3 = \mathbf{h}^H (\Sigma_{\mathbf{Y}_{d3}})^{-1} \mathbf{Y}_{d3} \end{cases} \quad (15)$$

where $\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3$ are the estimates of the desired symbols $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$. In practice, we substitute the channel and covariance estimations into this theoretical form. The serving channel is assumed to be estimated by the least squares (LS):

$$\hat{\mathbf{H}} = \frac{1}{N} \mathbf{Y}_p \mathbf{X}_p^H, \quad (16)$$

and the estimated channel vector $\hat{\mathbf{h}} = \hat{\mathbf{H}}\mathbf{p}$ is obtained by the given precoding vector \mathbf{p} . The covariance matrices are estimated by the sample covariance

$$\hat{\Sigma}_{\mathbf{Y}_{d1}} = \hat{\mathbf{h}}\hat{\mathbf{h}}^H + \hat{\mathbf{V}}\hat{\mathbf{V}}^H/N_p \quad (17)$$

$$\hat{\Sigma}_{\mathbf{Y}_{d2}} = \mathbf{Y}_{d2} \mathbf{Y}_{d2}^H / (N_p/2) \quad (18)$$

$$\hat{\Sigma}_{\mathbf{Y}_{d3}} = \mathbf{Y}_{d3} \mathbf{Y}_{d3}^H / (N_p/2) \quad (19)$$

where $\hat{\mathbf{V}} = \mathbf{Y}_p - \hat{\mathbf{H}}\mathbf{X}_p$ is the residual matrix which is a function of noise, interference and channel estimation error. We use $\hat{\mathbf{V}}$ instead of the received signal matrix \mathbf{Y}_{d1} because the former has smaller variances (see Appendix A). Substituting estimated results, therefore, the resulting output of IRC-DL is given by

$$\begin{cases} \hat{\mathbf{X}}_{1,d1} = \hat{\mathbf{h}}^H (\hat{\Sigma}_{\mathbf{Y}_{d1}})^{-1} \mathbf{Y}_{d1} \\ \hat{\mathbf{X}}_{2,d2} = \hat{\mathbf{h}}^H (\hat{\Sigma}_{\mathbf{Y}_{d2}} + \nu\mathbf{I})^{-1} \mathbf{Y}_{d2} \\ \hat{\mathbf{X}}_{3,d3} = \hat{\mathbf{h}}^H (\hat{\Sigma}_{\mathbf{Y}_{d3}} + \nu\mathbf{I})^{-1} \mathbf{Y}_{d3}, \end{cases} \quad (20)$$

where ν is the diagonal loading (DL) value to compensate sample covariance errors. This method is often used in the robust beamforming design for its simplicity [16]. However, finding the optimal DL value is nontrivial, but it is basically related to the sample covariance errors (see Appendix B). One common solution is to use the inverse condition number (ratio of the largest to the smallest singular value) of the covariance estimate [6]. [17] suggests an alternate method by using the estimation error of the estimated covariance matrix.

B. LS With Compensation (LS-C)

If the interfering pilot signal is perfectly known at the UE receiver, we can cancel interference to further improve the system performance [18]. We rewrite that part of the received signal with interfering pilot, $\mathbf{Y}_{d2}, \mathbf{Y}_{d3}$, as

$$\mathbf{Y}'_p = \mathbf{h}\mathbf{X}_{d1} + \bar{\mathbf{h}}\bar{\mathbf{X}}_p + \mathbf{Z}'_p \in \mathcal{C}^{2 \times N_p}, \quad (21)$$

where the received matrix $\mathbf{Y}'_p = [\mathbf{Y}_{d2}, \mathbf{Y}_{d3}]$ relates to the interfering pilot symbols, $\mathbf{X}_{d1} = [\mathbf{X}_2, \mathbf{X}_3]$ denotes the serving data matrix, $\bar{\mathbf{X}}_p = [\bar{\mathbf{X}}_{p1}, \bar{\mathbf{X}}_{p2}]$ denotes the interfering pilot matrix, and $\mathbf{Z}'_p = [\mathbf{Z}_{d2}, \mathbf{Z}_{d3}]$ is the AWGN matrix. A simple LS channel estimator for both serving and interfering channels is thus given by

$$\hat{\mathbf{H}} = \mathbf{Y}_p \mathbf{X}_p^H / N_p = \mathbf{H} + \mathbf{E} \quad (22)$$

$$\hat{\mathbf{H}}_b = \mathbf{Y}'_p \bar{\mathbf{X}}_p^H / N_p = \alpha\bar{\mathbf{H}} + \bar{\mathbf{E}} \quad (23)$$

where $\hat{\mathbf{H}}$ is the estimate of the serving channel matrix, $\hat{\mathbf{H}}_b$ is the estimate of the interfering channel matrix, and the channel estimation errors are given by

$$\mathbf{E} = \bar{\mathbf{h}}\bar{\mathbf{X}}_{d1} \mathbf{X}_p^H / N_p + \mathbf{Z}_p \mathbf{X}_p^H / N_p \quad (24)$$

$$\bar{\mathbf{E}} = \mathbf{h}\mathbf{X}_{d1} \bar{\mathbf{X}}_p^H / N_p + \mathbf{Z}'_p \bar{\mathbf{X}}_p^H / N_p. \quad (25)$$

Once the interfering channel estimation is available, we regenerate the received pilot signal via the product of the channel estimate and the given pilot signal as $\hat{\mathbf{H}}_b \bar{\mathbf{X}}_p$. Therefore, the residual matrix of pilot cancellation is obtained as follows:

$$\hat{\mathbf{V}}_b = \mathbf{Y}'_p - \hat{\mathbf{H}}_b \bar{\mathbf{X}}_p \quad (26)$$

$$= \mathbf{h}\mathbf{X}_{d1} - \bar{\mathbf{E}}\bar{\mathbf{X}}_p + \mathbf{Z}'_p \in \mathcal{C}^{2 \times N_p}. \quad (27)$$

Considering the channel estimation errors on both the serving and interfering channels, we derive the decoding metrics of LS-C under the conditional LMMSE criterion as

$$\hat{\mathbf{X}}_{d1,ls} = \mathbf{w}_{ls}^H \hat{\mathbf{V}}_b \quad (28)$$

where

$$\mathbf{w}_{ls}^H = \arg \min_{\mathbf{w}^H} \mathbb{E} \left[\|\mathbf{X}_{d1} - \mathbf{w}^H \hat{\mathbf{V}}_b\|^2 | \hat{\mathbf{h}}, \hat{\mathbf{h}}_b \right] \quad (29)$$

$$= \mathbb{E} \left[\mathbf{X}_{d1} \hat{\mathbf{V}}_b^H | \hat{\mathbf{h}}, \hat{\mathbf{h}}_b \right] \mathbb{E} \left[\hat{\mathbf{V}}_b \hat{\mathbf{V}}_b^H | \hat{\mathbf{h}}, \hat{\mathbf{h}}_b \right]^{-1} \quad (30)$$

$$= \hat{\mathbf{h}}^H \left(\frac{N_p^2 (\hat{\mathbf{h}} \hat{\mathbf{h}}^H)}{N_p^2 - 1} + \frac{N_p (\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H)}{N_p^2 - 1} + N_0 \mathbf{I} \right)^{-1} \quad (31)$$

where $\hat{\mathbf{h}}_b = \bar{\mathbf{p}} \hat{\mathbf{H}}_b$ is the equivalent interfering channel vector. More details can be found in Appendix C. This result is more complicated than the conventional solution that ignores channel estimation errors. Especially, the LS-C requires the precoding vector $\bar{\mathbf{p}}$ to obtain the term of $\hat{\mathbf{h}}_b$, which is difficult to obtain in practice. Therefore, an alternative way is derived by using the covariance of the residual matrix. To see this, we first notice that the conditional covariance matrix of $\hat{\mathbf{V}}$ is

$$\Sigma_v \triangleq \mathbb{E}[\hat{\mathbf{V}} \hat{\mathbf{V}}^H | \hat{\mathbf{h}}, \hat{\mathbf{h}}_b] \quad (32)$$

$$= \frac{N_p^2}{N_p^2 - 1} \hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H + \frac{N_p}{N_p^2 - 1} \hat{\mathbf{h}} \hat{\mathbf{h}}^H + N_0 \mathbf{I}. \quad (33)$$

This covariance matrix is dominated by $\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H$, which is the term we are interested in. Substituting the covariance matrix Σ_v for the term of $\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H$, we have

$$\mathbf{w}_{ls}^H = \hat{\mathbf{h}}^H \left(\hat{\mathbf{h}} \hat{\mathbf{h}}^H + \frac{1}{N_p} \Sigma_v + \frac{N_p - 1}{N_p} N_0 \mathbf{I} \right)^{-1}, \quad (34)$$

which does not depend on the precoding vector $\bar{\mathbf{p}}$. As a result, it is much more practical due to using less information from interference. Substituting the sample covariance estimation for the true covariance,

$$\hat{\Sigma}_v = \frac{1}{N_p} \hat{\mathbf{V}} \hat{\mathbf{V}}^H, \quad (35)$$

the resulting output of LS-C becomes

$$\begin{cases} \hat{\mathbf{X}}_{1,ls} = \hat{\mathbf{h}}^H \left(\hat{\Sigma}_{\mathbf{Y}_{d1}} \right)^{-1} \mathbf{Y}_{d1} \\ \hat{\mathbf{X}}_{d1,ls} = \hat{\mathbf{w}}_{ls}^H \hat{\mathbf{V}}_b. \end{cases} \quad (36)$$

where $\hat{\mathbf{w}}_{ls}$ is obtained by using the sample covariance estimation $\hat{\Sigma}_v$ in the theoretical combining vector \mathbf{w}_{ls} . Note that the LS-C is a mixed approach. We suppress the interfering data signal by its statistics and suppress the interfering pilot signal by regeneration and successive cancellation.

C. LMMSE With Compensation (LMMSE-C)

Regenerating the interfering pilot signal is not an easy task, especially in the presence of interfering channel estimation

error even with perfect information about interfering pilot. However, if the perfect SIR is available, we can significantly improve the receiver performance by using the LMMSE channel estimator

$$\text{vec}(\hat{\mathbf{H}}_{b,lm}) = \bar{\mathbf{X}}_{pc}^H (\alpha^2 \bar{\mathbf{X}}_{pc} \bar{\mathbf{X}}_{pc}^H + \hat{\Sigma}_{vbc})^{-1} \mathbf{y}'_p, \quad (37)$$

where $\mathbf{y}'_p = \text{vec}(\mathbf{Y}'_p)$ is the vectorized form of the received matrix, $\hat{\Sigma}_{vbc} = \mathbf{I} \otimes (\hat{\mathbf{V}}_b \hat{\mathbf{V}}_b^H / N_p)$ is the residual matrix, and $\bar{\mathbf{X}}_{pc} = \bar{\mathbf{X}}_p^T \otimes \mathbf{I}$ is the interfering pilot matrix. Replacing the LS channel estimates by the LMMSE channel estimates, we have the resulting output of LMMSE-C

$$\begin{cases} \hat{\mathbf{X}}_{1,lm} = \hat{\mathbf{h}}^H (\hat{\Sigma}_{\mathbf{Y}_{d1}})^{-1} \mathbf{Y}_{d1} \\ \hat{\mathbf{X}}_{d1,lm} = \hat{\mathbf{w}}_{ls}^H \hat{\mathbf{V}}_{b,lm}, \end{cases} \quad (38)$$

where $\hat{\mathbf{V}}_{b,lm} = \mathbf{Y}'_p - \hat{\mathbf{H}}_{b,lm} \bar{\mathbf{X}}_p$ is the residual matrix of the LMMSE estimate. Note that the only difference between LS-C and LMMSE-C is the residual matrix $\hat{\mathbf{V}}_{b,lm}$. We only modify the residual matrix due to its substantial impact on the system performance. This benefit is verified in the numerical results section.

IV. ANALYSIS

In this section, we try to answer three basic questions: 1) what is the optimum receiver? 2) what is the performance loss of a conventional IRC receiver? and 3) what is the impact of different propagation delays of serving and interfering eNBs?

A. Optimal Receiver (OPT)

We begin by answering the first question of optimal detection of the desired signal $x \in \mathbf{X}_1$ in the presence of an interfering signal from a neighbor eNB. The optimal MAP receiver given that the pilot symbols, the received vector $\mathbf{y} \in \mathbf{Y}_{d1}$, precoding vectors, SNR and SIR are known, is

$$\hat{x} = \arg \max_x \log P(x | \mathbf{y}, \mathbf{Y}_p, \mathbf{Y}'_p, \mathbf{X}_p, \bar{\mathbf{X}}_p, \alpha, \mathbf{p}, \bar{\mathbf{p}}) \quad (39)$$

$$= \arg \max_x \log \sum_{\mathbf{X}_{d1}} \sum_{\bar{\mathbf{x}}} \frac{\exp(\mathbf{b}^H \Sigma^{-1} \mathbf{A}^{-1} \mathbf{b})}{\det(\mathbf{A})} \quad (40)$$

where we apply the Gaussian integral [19] on the channel matrices and define the parameters as follows:

$$\mathbf{A} = \mathbf{I}_{2N_T N_R} + \Sigma \left((\mathcal{X} \mathcal{X}^H / N_0)^T \otimes \mathbf{I}_{N_R} \right) \quad (41)$$

$$\mathbf{b} = \Sigma \text{vec}(\mathcal{Y} \mathcal{X}^H / N_0) \quad (42)$$

$$\mathcal{X} = \begin{bmatrix} \mathbf{X}_p & \mathbf{X}_{d1} & \mathbf{p}x \\ \mathbf{p} \bar{\mathbf{X}}_{d1} & \bar{\mathbf{p}} \bar{\mathbf{X}}_p & \bar{\mathbf{p}} \bar{x} \end{bmatrix} \quad (43)$$

$$\mathcal{Y} = (\mathbf{Y}_p, \mathbf{Y}'_p, \mathbf{y}) \quad (44)$$

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & \alpha^2 \end{bmatrix} \otimes \mathbf{I}_{N_T N_R}. \quad (45)$$

The computational load of the OPT is quite heavy due to the multiple summations over all candidate transmitted signals \mathbf{X}_{d1} , $\bar{\mathbf{X}}_{d1}$, and $\bar{\mathbf{x}}$. This formulation implies that we softly decode (compute all candidates) $\bar{\mathbf{X}}_{d1}$ and \mathbf{X}_{d1} for better channel

state information (CSI) estimates; and then softly decode \bar{x} to achieve maximum likelihood detection.

For comparison purpose, we compute the OPT performance bound when the interfering symbols $\bar{\mathbf{X}}_{d1}$ and the serving symbols \mathbf{X}_{d1} can be decoded perfectly, i.e., both serving and interfering pilots are interference free. In this case, the resulting detector reduces to

$$\hat{x}_{opt} = \arg \max_x \log \sum_{\bar{x}} \frac{\exp(\mathbf{b}^H \boldsymbol{\Sigma}^{-1} \mathbf{A}^{-1} \mathbf{b})}{\det(\mathbf{A})}. \quad (46)$$

The associated performance will be used as the benchmark for assessing the performance of the proposed algorithms. Note that the log-sum-exp operation can result in a computing problem, which can be solved by the log-sum-exp trick [20].

The OPT not only provides a performance benchmark but also helps our understanding of the two-user interference channel. We summarize our observations as follows:

- All interfering symbols $\bar{\mathbf{X}}_{d1}$, $\bar{\mathbf{X}}_{d2}$, and $\bar{\mathbf{X}}_p$ should be decoded and subtracted in a decoding metric. Therefore, both the serving channel \mathbf{H} and the interfering channel $\alpha \bar{\mathbf{H}}$ need to be estimated.
- Good quality of CSI is difficult to obtain because both the pilot symbols \mathbf{X}_p and $\bar{\mathbf{X}}_p$ interfere with the data symbols. These data symbols cannot be decoded and subtracted easily. Therefore, one has to compensate for the channel estimation errors associated with \mathbf{H} and $\bar{\mathbf{H}}$.

Moreover, it is worth noting that all information of interference is useful, but only some of it can be obtained easily, i.e., 1) the strength of interference (SIR) and 2) the positions and the sequences of the interfering pilot symbols. Both of them are available from the cell ID sequences of the interfering eNB.

B. SINR Analysis of IRC

We first consider the conventional IRC scheme [7] which treats interference as a stationary Gaussian process. The IRC detector is

$$\hat{\mathbf{X}}_{irc} = \hat{\mathbf{h}}^H (\hat{\mathbf{h}} \hat{\mathbf{h}}^H + \hat{\mathbf{V}} \hat{\mathbf{V}}^H / N_p)^{-1} \mathbf{Y}_d, \quad (47)$$

where $\hat{\mathbf{X}}_{irc}$ is the estimate of transmitted data matrices ($\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$) and $\mathbf{Y}_d = (\mathbf{Y}_{d1}, \mathbf{Y}_{d2}, \mathbf{Y}_{d3})$ is the corresponding received signal matrix. Unlike the proposed detectors, this scheme decodes \mathbf{X}_2 and \mathbf{X}_3 with incorrect covariance matrices. The covariance mismatch leads to

$$\text{SINR}_{x_1} = \beta^2 / (\beta(1 - \beta)) \quad (48)$$

$$\text{SINR}_{x_2} = \beta^2 / (\beta(1 - \beta) + \psi_1) \quad (49)$$

$$\text{SINR}_{x_3} = \beta^2 / (\beta(1 - \beta) + \psi_2). \quad (50)$$

where β^2 is the desired signal power and ψ_1, ψ_2 are residual interference terms due to the covariance mismatch (see Appendix F). If $\mathcal{E}_d \gg N_0$, one can show that $\psi_1 > 0$, $\psi_2 > 0$, and

$$\psi_1, \psi_2 \propto (\text{SNR})^2 \cdot \alpha^2 \cdot \frac{\mathcal{E}_p}{\mathcal{E}_d}. \quad (51)$$

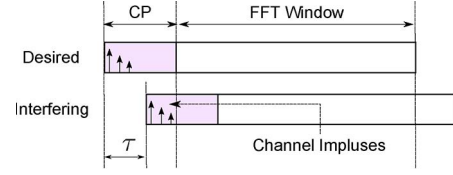


Fig. 3. An example of an asynchronous interference.

Hence, the IRC scheme suffers from mismatch-induced performance degradation when 1) SNR is high, 2) a strong interference is present, and 3) the pilot-to-data power ratio is high, i.e., $\frac{\mathcal{E}_p}{\mathcal{E}_d} \gg 1$. It is worth mentioning the special case when the interfering eNB transmits no data but pilot symbols only. In this case, we have $\mathcal{E}_d = 0$ such that $\frac{\mathcal{E}_p}{\mathcal{E}_d} = \infty$. This happens in several Inter-Cell Interference Coordination (ICIC) approaches [2] and causes significant performance degradation of an IRC receiver.

C. The Asynchronous Case

Suppose the length of the Discrete Fourier Transform (DFT) window is N_F , the length of the cyclic prefix (CP) is N_{cp} , the maximum channel delay spread is N_L , and the timing difference between two received signals is denoted by τ as shown in Fig. 3. All of these parameters are given in terms of the number of OFDM samples. We should consider two ranges for the propagation delay τ : 1) those causing no inter-symbol interference (ISI) and 2) those which lead to ISI, which are shown in the following.

1) *Case Without ISI*: The first subset of delays $\tau \in \{0, \dots, N_{cp} - N_L\}$ does not destroy orthogonality and only introduces a phase rotation in each subcarrier [21]. Assuming the channels are non-selective, the received signal at (tk) th slot is

$$\mathbf{y} = \mathbf{h}x + e^{-j\frac{2\pi k\tau}{N_F}} \bar{\mathbf{h}}\bar{x} + \mathbf{z}, \quad (52)$$

where $0 < \tau \leq N_{cp} - N_L$. Note that the phase rotation $\exp(-j2\pi k\tau/N_F)$ is unknown and degrades the interfering channel estimate. On the other hand, since the phase rotation is a function of subcarrier index k instead of time and space, there is no impact on the spatial covariance matrix, i.e., the covariance matrix of \mathbf{y} is the same as in the synchronous case, that is

$$\boldsymbol{\Sigma}_y = \mathbf{h}\mathbf{h}^H + \bar{\mathbf{h}}\bar{\mathbf{h}}^H + N_0\mathbf{I}. \quad (53)$$

Hence, we can expect IRC-DL will be more robust than LS-C and LMMSE-C in the ISI-free case.

2) *Case With ISI*: If the delay τ is outside the above range, orthogonality among the subcarriers will be destroyed. Mathematically, we have

$$\mathbf{y} = \mathbf{h}x + \kappa e^{-j\frac{2\pi k\tau}{N_F}} \bar{\mathbf{h}}\bar{x} + \alpha \mathcal{I}_\tau^{tk} + \mathbf{z},$$

for $\tau < 0$ or $\tau > N_{cp} - N_L$, where $\kappa \approx (N_F - \Delta(\tau))/N_F$, $\Delta(\tau)$ is a positive value related to delay τ , and \mathcal{I}_τ^{tk} is the ISI

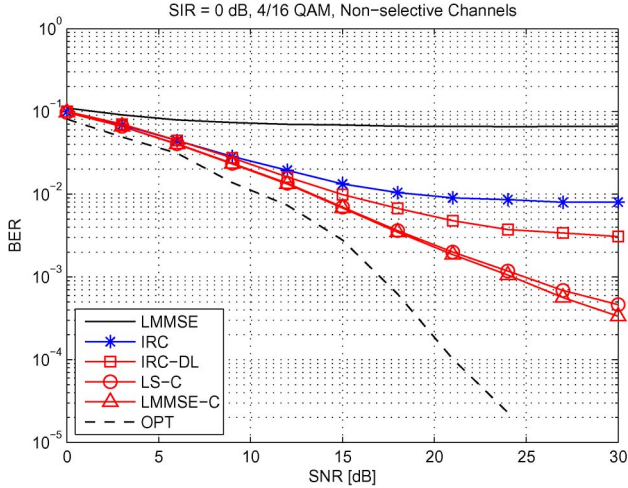


Fig. 4. BER vs. SNR in the strong interference region, SIR = 0 dB.

term. The impact on the covariance matrices is shown as:

$$\Sigma_{\mathbf{y}} = \mathbf{h}\mathbf{h}^H + \kappa^2 \bar{\mathbf{h}}\bar{\mathbf{h}}^H + \alpha^2 \text{var}(\mathcal{I}_{\tau}^{tk}) + N_0 \mathbf{I} \quad (54)$$

$$\approx \mathbf{h}\mathbf{h}^H + \left(1 - \frac{\Delta(\tau)}{N_F}\right)^2 \bar{\mathbf{h}}\bar{\mathbf{h}}^H + \alpha^2 \left(2 \frac{\Delta(\tau)}{N_F} - \left(\frac{\Delta(\tau)}{N_F}\right)^2\right) \mathbf{I} + N_0 \mathbf{I}. \quad (55)$$

Here, the ISI term \mathcal{I}_{τ}^{tk} is approximated by Gaussian noise as in [22]. With this approximation, the covariance matrix is not a function of time or frequency, i.e., without the index t and the index k . Therefore, we can make a similar conclusion that different delays have no impact on the covariance estimation. However, since our proposed scheme is based on the accurate pilot positions, the index shift due to a large propagation delay needs to be avoided by proper coordination between eNBs.

V. NUMERICAL RESULTS

A. Non-Selective Channels

To evaluate the proposed schemes, we start with a constant channel which is time invaring and frequency non-selective during $N_p + N_d$ intervals (within one RB). The data-pilot format follows the LTE specification illustrated in Figs. 1 and 2 where we define $N_p = 16$, and $N_d = 152$. The data of interest ($\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$) are modulated with Gray-coded 4-QAM. The interfering data matrices ($\bar{\mathbf{X}}_{d1}, \bar{\mathbf{X}}_{d2}$) are modulated with Gray-coded 16-QAM. In what follows, we compare the performance of the conventional schemes: LMMSE (which treats interference as AWGN), IRC, and the proposed approaches: IRC-DL, LS-C, LMMSE-C, and OPT.

Fig. 4 shows the BER performance of different SNRs for SIR = 0 dB. These curves show substantial improvements of the proposed schemes over the conventional solutions. The curve of IRC-DL shows the performance gain by interference suppression with proper grouping. The curves of LS-C and LMMSE-C show the benefit of interfering pilot cancellation. In this case, interfering pilot cancellation can get the largest performance gain. Moreover, OPT shows the possible further

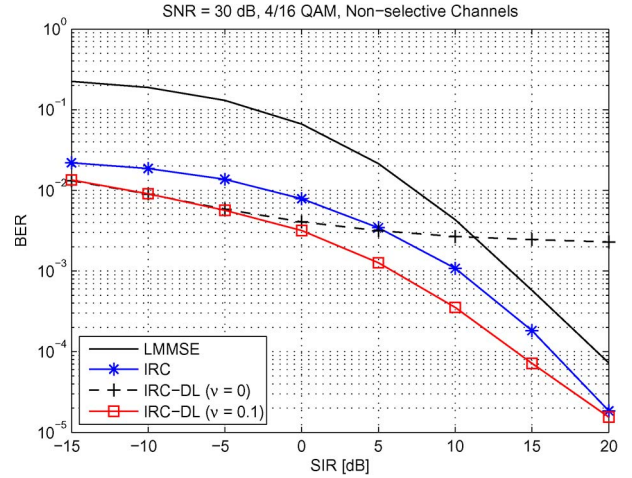


Fig. 5. An illustration of the impact of DL values for SNR = 30 dB.

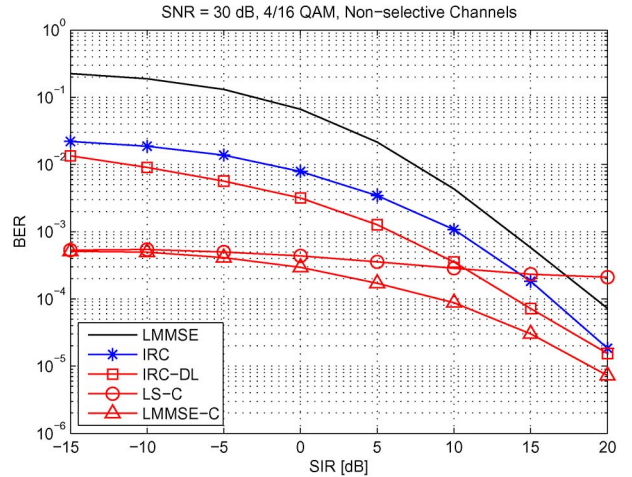


Fig. 6. BER vs. SIR in the high SNR region, SNR = 30 dB.

improvement if we can decode the interfering data symbols and further improve the channel estimation on both serving and interfering channels.

Fig. 5 shows the impact of using the DL value when IRC-DL is applied. We compare two cases: 1) without compensation ($\nu = 0$) and 2) given a fixed DL value ($\nu = 0.1$). These curves illustrate a huge performance loss at medium and low interference levels when we ignore the sample covariance error. Therefore, it is essential to apply some compensation schemes. Even the fixed-value DL scheme, which is extremely simple and suboptimal, provides substantial improvements.

Fig. 6 shows the BER performance of different SIRs for SNR = 30 dB. These curves demonstrate the advantage of using LMMSE-C. Note that the error floor of LS-C occurs for SIR ≥ -10 dB. This is because accurate CSI of interference is difficult to obtain under the weak interference assumption. This performance degradation can be highly reduced by using LMMSE-C.

Fig. 7 shows a special case in which the interfering eNB only transmits the pilot symbols. We set the power of interfering data symbols equal to zero $\mathcal{E}_d = 0$ to show the case of $\mathcal{E}_p/\mathcal{E}_d \rightarrow \infty$. One can see that the conventional IRC performs poorly as the LMMSE. This result supports our SINR analysis in which the

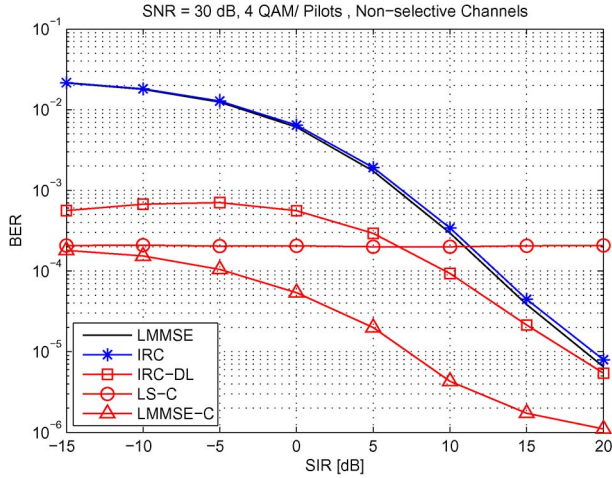


Fig. 7. BER vs. SIR for SNR = 30 dB; no interfering data symbols.

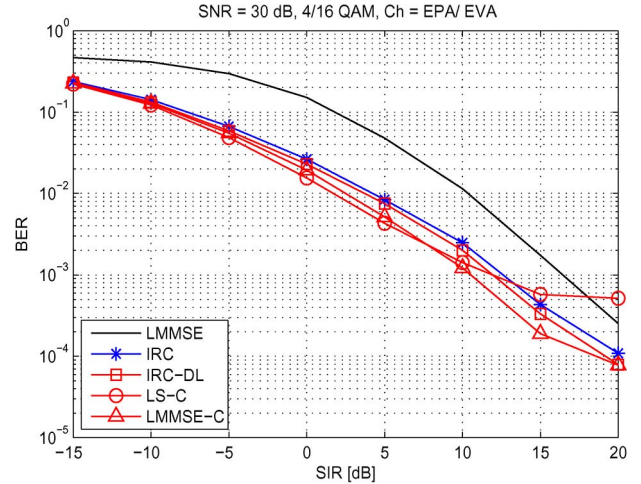


Fig. 9. BER vs. SIR under SNR = 30 dB with triply selective channels: $f_d = 10/10$ Hz, EPA/EVA.

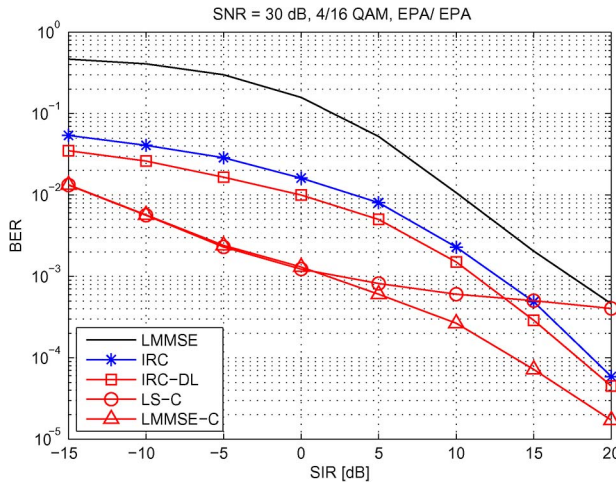


Fig. 8. BER vs. SIR under SNR = 30 dB with triply selective channels: $f_d = 10/10$ Hz, EPA/EVA.

performance loss of IRC is proportional to the pilot-data power ratio $\mathcal{E}_p/\mathcal{E}_d$. Moreover, this also implies that the proposed schemes can be applied to some ICIC schemes when intended symbols interfere with the pilot symbols or control symbols only.

B. Triply Selective Channels

For fading channels, we assume the serving channel distribution information (CDI) is perfectly known but the interfering CDI is unknown. The system parameters are given as follows: normal CP $N_{CP} = 72$, DFT-size $N_F = 1024$, and 2 GHz carrier frequency. All of them are in terms of OFDM samples. The discrete-time MIMO fading channel is generated using the triply selective channel [23]. In what follows, we denote CDI on both channels by (serving CDI/interfering CDI) and the UE receiver performs the data detection by the proposed scheme but replacing the serving channel estimator by the 2D-MMSE channel estimator instead of the LS channel estimator.

Fig. 8 shows the BER performance for SNR = 30 dB on triply selective channels. CDI of both channels is given by: maximum Doppler frequency $f_d = 10$ Hz, the power delay

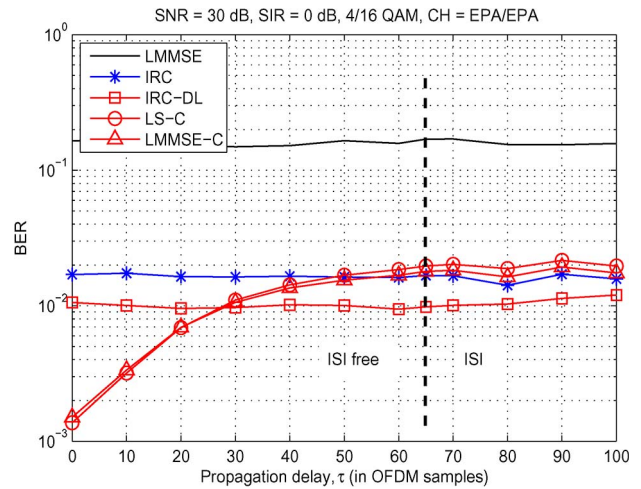


Fig. 10. BER vs. propagation delay, τ with SNR = 30 dB and SIR = 0 dB on triply selective channels: $f_d = 10/10$ Hz, EPA/EVA.

profiles C_{ISI} is Extended Pedestrian A (EPA) model defined in LTE specification, and no spatial correlation for Ψ_{Tx} and Ψ_{Rx} . Under these slow fading and slightly selective assumptions, the results are basically the same as those corresponding to the constant channels.

Fig. 9 shows the BER performance on the more selective interfering channel. We use the same CDI as in Fig. 8 but substituting Extended Vehicular A (EVA) for EPA on the interfering channel, i.e., the serving CDI is ($f_d = 10$ Hz, $C_{ISI} = \text{EPA}$, $\Psi_{Rx} = \Psi_{Tx} = \mathbf{I}$) and the interfering CDI is ($f_d = 10$ Hz, $C_{ISI} = \text{EVA}$, $\Psi_{Rx} = \Psi_{Tx} = \mathbf{I}$). The curves illustrate an approximately 3 dB gain of LMMSE-C with respect to the IRC at BER = 10^{-3} . However, we lose some performance gains because the interfering channel changes too much within one RB. This can be further improved if we choose a smaller cluster such as a half RB.

Fig. 10 illustrates the asynchronous case on triply selective channels. CDI of both channels is given as: 10 Hz of the Doppler frequency, EPA model as the power delay profile, and no spatial correlation between antennas. According to EPA, The maximal channel delay spread is given by $N_L = 7$ OFDM

samples. Therefore, the maximal delay of ISI-free region will be $\tau = 65$ OFDM samples shown by the dashed line. First of all, we show that our schemes provide substantial improvements on triply selective channels without propagation delay $\tau = 0$. Secondly, when the timing delay increases, we show how robust the IRC-DL is. In fact, the performance degradation of IRC-DL only happens after a very large delay (200 OFDM samples) and its performance converges to that of the IRC.

VI. CONCLUSION

In this paper, we have proposed several inter-cell interference suppression schemes for MIMO-OFDM systems. As the covariance between the desired and interfering signals varies in different parts of a frame due to pilot location offset between the two OFDM signals. By taking into account such distinct statistical properties within a frame, we derived novel relatively low-complexity LMMSE and SIC detection schemes. Numerical results confirm that separate treatments of the interfering pilot and data signals do result in substantial performance improvements, which also clearly reveal the importance of judicious exploitation of the interfering signals' (such as those from neighboring cells) structure and statistics. The simulations include flat and triply selective channels with/without timing delays of arrival signals. The most significant performance gain is achieved when the interference is strong, SNR is high and there is no timing delay. This happens because the interfering pilot signal can be cancelled accurately in this case. In the asynchronous case, we find that the proposed IRC-DL based on the LMMSE criterion provides reliable and stable performance.

APPENDIX A

COMPARISON OF COVARIANCE ESTIMATES

We compare two different ways to estimate the covariance of \mathbf{Y}_{d1} : 1) using the residual matrix $\hat{\mathbf{V}}$ and 2) using the received signal \mathbf{Y}_{d1} . Assume that \mathbf{Y}_{d1} and $\hat{\mathbf{V}}$ have the same symbol length, i.e., $N_p = N_d - N_p \triangleq N_o$, and the columns of \mathbf{Y}_{d1} and $\hat{\mathbf{V}}$ are i.i.d. complex Gaussian vectors. Ignoring channel estimation errors, we conclude that

$$\hat{\mathbf{V}}\hat{\mathbf{V}}^H \sim \mathcal{W}_{N_R}(\boldsymbol{\Sigma}_{\mathbf{V}}, N_o)$$

$$\mathbf{Y}_{d1}\mathbf{Y}_{d1}^H \sim \mathcal{W}_{N_R}(\boldsymbol{\Sigma}_{\mathbf{Y}_{d1}}, N_o),$$

where \mathcal{W}_{N_R} denotes the complex Wishart matrix [24] with N_R degrees of freedom, $\boldsymbol{\Sigma}_{\mathbf{V}} = \mathbf{h}\mathbf{h}^H + N_o\mathbf{I}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}_{d1}} = \mathbf{h}\mathbf{h}^H + \boldsymbol{\Sigma}_{\mathbf{V}}$ are the true covariance matrices. The variances of their diagonal entries are then given by

$$\text{var}([\hat{\mathbf{V}}\hat{\mathbf{V}}^H]_{mm}) = \frac{2}{N_o^2}[\boldsymbol{\Sigma}_{\mathbf{V}}]_{mm}^2$$

$$\text{var}([\mathbf{Y}_{d1}\mathbf{Y}_{d1}^H]_{mm}) = \frac{2}{N_o^2}[\boldsymbol{\Sigma}_{\mathbf{Y}_{d1}}]_{mm}^2$$

which are proportional to the covariance estimation errors. Since $[\boldsymbol{\Sigma}_{\mathbf{V}}]_{mm}$ is always smaller than $[\boldsymbol{\Sigma}_{\mathbf{Y}_{d1}}]_{mm}$, the estimation errors of $\hat{\mathbf{V}}$ are smaller than that of \mathbf{Y}_{d1} in general.

APPENDIX B
DIAGONAL LOADING

Let \mathbf{w} is the linear combination of the received matrix \mathbf{Y}_{d2} to decode the intended signal, i.e., $\hat{\mathbf{X}}_{2,d1} = \mathbf{w}^H\mathbf{Y}_{d2}$. Suppose the perfect CSI and the perfect covariance estimation are given, the SINR of \mathbf{Y}_{d2} is obtained by

$$\text{SINR}_{\mathbf{Y}_{d2}} = \frac{|\mathbf{w}^H\mathbf{h}|^2}{\mathbf{w}^H(\hat{\boldsymbol{\Psi}}_{p1} + N_o\mathbf{I})\mathbf{w}}$$

The problem can be rewritten as the following form

$$\max_{\mathbf{w}} \frac{\text{SINR}_{\mathbf{Y}_{d2}}}{1 + \text{SINR}_{\mathbf{Y}_{d2}}} = \max_{\mathbf{w}} \frac{|\mathbf{w}^H\mathbf{h}|^2}{\mathbf{w}^H\boldsymbol{\Sigma}_{\mathbf{Y}_{d2}}\mathbf{w}}$$

Using the sample covariance $\boldsymbol{\Sigma}_{\mathbf{Y}_{d2}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{d2}} + \boldsymbol{\Delta}$ and assuming the sample covariance errors $\boldsymbol{\Delta}$ can be upper bounded, the max-min SINR of \mathbf{Y}_{d2} is equivalent to

$$\arg \min_{\mathbf{w}} \max_{\boldsymbol{\Delta}} \mathbf{w}^H(\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{d2}} + \boldsymbol{\Delta})\mathbf{w}$$

subject to $\mathbf{w}^H\mathbf{h} = 1, \|\boldsymbol{\Delta}\|_F \leq \nu$

Since $\boldsymbol{\Delta}$ is the Hermitian matrix due to the sample covariance estimate, we have the following property:

$$\max_{\|\boldsymbol{\Delta}\|_F \leq \nu} \mathbf{w}^H\boldsymbol{\Delta}\mathbf{w} = \nu \cdot \mathbf{w}^H\mathbf{w}.$$

Thus, defining $\mathbf{R}^* = \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{d2}} + \nu\mathbf{I}$, we rewrite

$$\arg \min_{\mathbf{w}} \mathbf{w}^H\mathbf{R}^*\mathbf{w}, \text{ subject to } \mathbf{w}^H\mathbf{h} = 1$$

The Lagrangian $L: \mathcal{C}^{N_R} \times \mathcal{C} \rightarrow \mathcal{R}$ is defined as

$$L(\mathbf{w}, \lambda) = \mathbf{w}^H\mathbf{R}^*\mathbf{w} + \frac{\lambda^H}{2}(\mathbf{w}^H\mathbf{h} - 1) + \frac{\lambda}{2}(\mathbf{h}^H\mathbf{w} - 1)$$

We differentiate $L(\mathbf{w}, \lambda)$ with respect to \mathbf{w}^H and λ , and set the partial derivatives equal to zero. We have: $\mathbf{w}^* = -\mathbf{R}^{*-1}\mathbf{h}\frac{\lambda}{2}$. Imposing the constraint $\mathbf{h}^H\mathbf{w}^* = 1 = -\mathbf{h}^H\mathbf{R}^{*-1}\mathbf{h}\frac{\lambda}{2}$, then $\lambda^* = -2(\mathbf{h}^H\mathbf{R}^{*-1}\mathbf{h})^{-1}$. Therefore, we have the optimal solution of \mathbf{w} as:

$$\mathbf{w}^* = \mathbf{R}^{*-1}\mathbf{h} \cdot (\mathbf{h}^H\mathbf{R}^{*-1}\mathbf{h})^{-1} \approx \mathbf{R}^{*-1}\mathbf{h}$$

$$= \left(\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{d2}} + \nu\mathbf{I}\right)^{-1} \mathbf{h}.$$

Here we omit the immaterial constant. Note that this result is equivalent to the diagonal-loading form. Therefore, the diagonal loading value is equal to the upper-bound of the Frobenius norm of the sample covariance errors.

APPENDIX C
DERIVATION OF LS-C

For notational brevity, we ignore the conditional part of a conditional expectation operator, e.g., $\mathbb{E}[\mathbf{X}|\mathbf{h}] = \mathbb{E}[\mathbf{X}]$. The

weighting vector of the LS-C is given by

$$\begin{aligned} \mathbf{w}_{ls}^H &\stackrel{(a)}{=} \mathbb{E} \left[\mathbf{X}_{d1} \hat{\mathbf{V}}_b^H \right] \mathbb{E} \left[\hat{\mathbf{V}}_b \hat{\mathbf{V}}_b^H \right]^{-1} \\ &\stackrel{(b)}{=} \mathbb{E}[\mathbf{h}^H] \left(\mathbb{E}[\mathbf{h}\mathbf{h}^H] + N_0 \mathbf{I} \right)^{-1} \\ &= \hat{\mathbf{h}}^H \left(\frac{N_p^2 (\hat{\mathbf{h}}\hat{\mathbf{h}}^H)}{N_p^2 - 1} + \frac{N_p (\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H)}{N_p^2 - 1} + N_0 \mathbf{I} \right)^{-1}. \end{aligned}$$

In (a), the two conditional expectations on the RHS are

$$\begin{aligned} \mathbb{E} \left[\mathbf{X}_{d1} \hat{\mathbf{V}}_b^H \right] &= \mathbb{E} \left[\mathbf{X}_{d1} (\mathbf{h}\mathbf{X}_{d1} - \bar{\mathbf{E}}\bar{\mathbf{X}}_p + \mathbf{Z}'_p)^H \right] \\ &= N_p \cdot \mathbb{E}[\mathbf{h}^H] - \mathbb{E} \left[\mathbf{X}_{d1} \bar{\mathbf{X}}_p^H \bar{\mathbf{E}}^H \right] \\ &= N_p \cdot \mathbb{E}[\mathbf{h}^H] - \mathbb{E} \left[\mathbf{X}_{d1} \bar{\mathbf{X}}_p^H \left(\frac{1}{N_p} \bar{\mathbf{X}}_p \mathbf{X}_{d1}^H \mathbf{h}^H \right) \right] \\ &= \left(N_p - \frac{1}{N_p} \text{Tr} (\bar{\mathbf{X}}_p^H \bar{\mathbf{X}}_p \mathbb{E} [\mathbf{X}_{d1}^H \mathbf{X}_{d1}]) \right) \mathbb{E}[\mathbf{h}^H] \\ &= (N_p - 2) \cdot \mathbb{E}[\mathbf{h}^H] \\ \text{var}[\hat{\mathbf{V}}_b] &= \mathbb{E} \left[N_p \mathbf{h}\mathbf{h}^H + \bar{\mathbf{E}}\bar{\mathbf{X}}_p \bar{\mathbf{X}}_p^H \bar{\mathbf{E}}^H \right] + N_0 N_p \mathbf{I} \\ &\quad - 2 \cdot \mathbb{E} \left[\bar{\mathbf{E}}\bar{\mathbf{X}}_p \mathbf{X}_{d1}^H \mathbf{h}^H + \bar{\mathbf{E}}\bar{\mathbf{X}}_p \mathbf{Z}'_p^H \right] \\ &= (N_p - 2) \cdot \left(\mathbb{E}[\mathbf{h}\mathbf{h}^H] + N_0 \mathbf{I} \right), \end{aligned}$$

where we have used the following identities,

$$\begin{aligned} \mathbb{E} \left[\bar{\mathbf{E}}\bar{\mathbf{X}}_p \bar{\mathbf{X}}_p^H \bar{\mathbf{E}}^H \right] &= \frac{\text{var} (\mathbf{h}\mathbf{X}_{d1} \bar{\mathbf{X}}_p^H \bar{\mathbf{X}}_p) + \text{var} (\mathbf{Z}'_p \bar{\mathbf{X}}_p^H \bar{\mathbf{X}}_p)}{N_p^2} \\ &= 2 \left(\mathbb{E}[\mathbf{h}\mathbf{h}^H] + N_0 \mathbf{I} \right) \\ \mathbb{E} \left[\bar{\mathbf{E}}\bar{\mathbf{X}}_p \mathbf{X}_{d1}^H \mathbf{h}^H \right] &= \frac{1}{N_p} \mathbb{E} \left[\mathbf{h}\mathbf{X}_{d1} \bar{\mathbf{X}}_p^H \bar{\mathbf{X}}_p \mathbf{X}_{d1}^H \mathbf{h}^H \right] \\ &= \frac{1}{N_p} \text{Tr} (\bar{\mathbf{X}}_p^H \bar{\mathbf{X}}_p \cdot \text{var} (\mathbf{X}_{d1}^H)) \text{var} (\mathbf{h}^H) \\ &= 2 \cdot \mathbb{E}[\mathbf{h}\mathbf{h}^H] \\ \mathbb{E} \left[\bar{\mathbf{E}}\bar{\mathbf{X}}_p \mathbf{Z}'_p^H \right] &= \frac{1}{N_p} \text{var} (\mathbf{Z}'_p \bar{\mathbf{X}}_p^H) = 2N_0 \mathbf{I}. \end{aligned}$$

In (b), since the entries of the channel estimation error matrix are zero mean, we have $\mathbb{E}[\mathbf{h}] = \hat{\mathbf{h}}$. Invoking the relations

$$\begin{aligned} \text{var}(\mathbf{E}\mathbf{p}) &= \frac{\text{var} (\mathbf{h}_b \bar{\mathbf{X}}_{d1} \mathbf{X}_p^H \mathbf{p}) + \text{var} (\mathbf{Z}_p \mathbf{X}_p^H \mathbf{p})}{N_p^2} \\ &= \frac{\text{Tr} (\mathbf{X}_p^H \mathbf{p}\mathbf{p}^H \mathbf{X}_p) \cdot \text{Cov}(\bar{\mathbf{h}}) + N_p N_0 \mathbf{I}}{N_p^2} \\ &= \frac{1}{N_p} \left(\mathbb{E}[\bar{\mathbf{h}}\bar{\mathbf{h}}^H] + N_0 \mathbf{I} \right) \\ \mathbb{E}[\bar{\mathbf{h}}\bar{\mathbf{h}}^H] &= \hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H + \text{var}(\bar{\mathbf{E}}\bar{\mathbf{p}}^H) \\ &= \hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H + \frac{1}{N_p} \left(\mathbb{E}[\mathbf{h}\mathbf{h}^H] + N_0 \mathbf{I}_{N_R} \right), \end{aligned}$$

We immediately obtain

$$\begin{aligned} \mathbb{E}[\mathbf{h}\mathbf{h}^H] &= \hat{\mathbf{h}}\hat{\mathbf{h}}^H + \text{var}(\mathbf{E}\mathbf{p}^H) \\ &= \hat{\mathbf{h}}\hat{\mathbf{h}}^H + \frac{1}{N_p} \left(\mathbb{E}[\bar{\mathbf{h}}\bar{\mathbf{h}}^H] + N_0 \mathbf{I} \right) \\ &= \hat{\mathbf{h}}\hat{\mathbf{h}}^H + \frac{N_0}{N_p} \mathbf{I} \\ &\quad + \frac{1}{N_p} \left(\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H + \frac{1}{N_p} \left(\mathbb{E}[\mathbf{h}\mathbf{h}^H] + N_0 \mathbf{I}_{N_R} \right) \right) \\ &= \left(1 + \frac{1}{N_p^2} + \dots \right) \hat{\mathbf{h}}_1 \hat{\mathbf{h}}_1^H \\ &\quad + \left(\frac{1}{N_p} + \frac{1}{N_p^3} + \dots \right) \hat{\mathbf{h}}_2 \hat{\mathbf{h}}_2^H \\ &\quad + \left(\frac{1}{N_p} + \frac{1}{N_p^2} + \dots \right) N_0 \mathbf{I}_{N_R} \\ &= \frac{N_p^2}{N_p^2 - 1} \hat{\mathbf{h}}_1 \hat{\mathbf{h}}_1^H + \frac{N_p}{N_p^2 - 1} \hat{\mathbf{h}}_2 \hat{\mathbf{h}}_2^H + \frac{1}{N_p - 1} N_0 \mathbf{I}_{N_R}. \end{aligned}$$

Similarly, we can show that the conditional expectations related to the interfering channel and the residual matrix are given by

$$\begin{aligned} \mathbb{E}[\mathbf{h}_b \mathbf{h}_b^H] &= \frac{N_p^2}{N_p^2 - 1} \hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H + \frac{\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H}{N_p - 1} + \frac{N_0 \mathbf{I}}{N_p - 1} \\ \mathbb{E} \left[\hat{\mathbf{V}} \hat{\mathbf{V}}^H \right] &= \frac{N_p^2 (\hat{\mathbf{h}}_b \hat{\mathbf{h}}_b^H)}{N_p^2 - 1} + \frac{N_p (\hat{\mathbf{h}}\hat{\mathbf{h}}^H)}{N_p^2 - 1} + N_0 \mathbf{I}. \end{aligned}$$

APPENDIX D DERIVATION OF OPT

We first rewrite the system model into a compact form. The received symbol matrices are given by

$$\begin{cases} \mathbf{Y}_p = \mathbf{H}\mathbf{X}_p + \alpha \bar{\mathbf{H}}\bar{\mathbf{p}}\bar{\mathbf{X}}_{d1} + \mathbf{Z}_p \\ \mathbf{Y}'_p = \mathbf{H}\mathbf{p}\bar{\mathbf{X}}_{d1} + \alpha \bar{\mathbf{H}}\bar{\mathbf{p}}\bar{\mathbf{X}}_p + \mathbf{Z}'_p \\ \mathbf{y} = \mathbf{H}\mathbf{p}x + \alpha \bar{\mathbf{H}}\bar{\mathbf{p}}\bar{x} + \mathbf{z}_{d1} \end{cases}$$

where $\mathbf{y} \in \mathcal{C}^{N_R}$ is the vector of \mathbf{Y}_{d1} , x is the signal of interest from the data matrix \mathbf{X}_1 and \mathbf{z}_{d1} is the AWGN vector. We further define $\mathcal{Y} \triangleq (\mathbf{Y}_p, \mathbf{Y}'_p, \mathbf{y})$, $\mathcal{Z} = (\mathbf{Z}_p, \mathbf{Z}'_p, \mathbf{z}_{d1})$, $\mathcal{H} = (\mathbf{H}, \alpha \bar{\mathbf{H}})$ and

$$\mathcal{X} \triangleq \begin{bmatrix} \mathbf{X}_p & \mathbf{X}_{d1} & \mathbf{p}x \\ \bar{\mathbf{p}}\bar{\mathbf{X}}_{d1} & \bar{\mathbf{p}}\bar{\mathbf{X}}_p & \bar{\mathbf{p}}\bar{x} \end{bmatrix}.$$

Then we can rewrite the received matrices as:

$$\mathcal{Y} = \mathcal{H}\mathcal{X} + \mathcal{Z}.$$

Since \mathcal{Z} is still the circularly symmetric complex Gaussian, we have the conditional probability $P(\mathcal{Y}|\mathcal{X}, \mathcal{H})$ as follows:

$$P(\mathcal{Y}|\mathcal{X}, \mathcal{H}) = \frac{\exp(-\|\mathcal{Y} - \mathcal{H}\mathcal{X}\|^2 / N_0)}{(\pi N_0)^{N_R(2N_p+1)}}.$$

We define $\text{vec}(\mathcal{H}) \sim \mathcal{N}_c(\mathbf{0}, \Sigma)$ because both channel matrices \mathbf{H} and $\bar{\mathbf{H}}$ are complex Gaussian matrices. Then taking the average of \mathcal{H} by using the Gaussian integral [19], we have

$$\begin{aligned} P(\mathcal{Y}|\mathcal{X}) &= \mathbb{E}_{\mathcal{H}} [P(\mathcal{Y}|\mathcal{X}, \mathcal{H})] \\ &= \frac{\exp(\mathbf{b}^H \Sigma^{-1} \mathbf{A}^{-1} \mathbf{b} - \|\mathbf{Y}\|^2/N_0)}{(\pi N_0)^{(2N_p+1)N_R} \det(\mathbf{A})} \end{aligned}$$

where

$$\mathbf{A} = \mathbf{I}_{2N_T N_R} + \Sigma \left((\mathcal{X} \mathcal{X}^H / N_0)^T \otimes \mathbf{I}_{N_R} \right)$$

$$\mathbf{b} = \Sigma \text{vec}(\mathcal{Y} \mathcal{X}^H / N_0)$$

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & \alpha^2 \end{bmatrix} \otimes \mathbf{I}_{N_T N_R}.$$

Based on this fact, the OPT metric can be expressed as follows:

$$\begin{aligned} \hat{x} &= \arg \max_x \log P(x|\mathbf{y}, \mathbf{Y}_p, \mathbf{Y}'_p, \mathbf{X}_p, \bar{\mathbf{X}}_p, \alpha, \mathbf{p}, \bar{\mathbf{p}}) \\ &= \arg \max_x \log \mathbb{E}_{\mathbf{X}_{d1}, \bar{\mathbf{X}}_{d1}, \bar{x}} [P(\mathcal{Y}|\mathcal{X})] \\ &= \arg \max_x \log \sum_{\mathbf{X}_{d1}} \sum_{\bar{\mathbf{X}}_{d1}} \sum_{\bar{x}} \frac{\exp(\mathbf{b}^H \Sigma^{-1} \mathbf{A}^{-1} \mathbf{b})}{\det(\mathbf{A})} \end{aligned}$$

where we apply the law of iterated expectations to add the lacking random variables in the conditional set.

APPENDIX E SINR ANALYSIS

Assuming that perfect channel estimation and perfect covariance estimation are given, we derive three SINRs, one for each of the transmitted data matrices \mathbf{X}_1 , \mathbf{X}_2 , and \mathbf{X}_3 . First of all, the resulting output of the conventional IRC on \mathbf{X}_1 is

$$\begin{aligned} \hat{x}_{1,irc} &= \mathbf{w}_{irc}^H \mathbf{y}_{d1} \\ &= \mathbf{w}_{irc}^H \mathbf{h} x_1 + \mathbf{w}_{irc}^H \bar{\mathbf{h}} \bar{x}_{d2} + \mathbf{w}_{irc}^H \mathbf{z}_{d1} \\ &\triangleq \beta x_1 + \omega_1 \end{aligned}$$

where $\hat{x}_{1,irc}$, \mathbf{y}_{d1} , x_1 , \bar{x}_{d2} , and \mathbf{z}_{d1} are the column elements of $\hat{\mathbf{X}}_{1,irc}$, \mathbf{Y}_{d1} , \mathbf{X}_1 , $\bar{\mathbf{X}}_{d2}$, and \mathbf{Z}_{d1} , respectively. Next, we derive the average power of βx_1 and ω_1 by their variances,

$$\begin{aligned} \text{var}(\beta x_1) &= \beta^2 \mathcal{E}_d \\ \text{var}(\omega_1) &= (\beta - \beta^2) \mathcal{E}_d. \end{aligned}$$

Therefore, the SINR of \mathbf{X}_1 is obtained as:

$$\text{SINR}_{x_1} = \frac{\text{var}(\beta x_1)}{\text{var}(\omega_1)} = \frac{\beta^2}{\beta(1-\beta)}.$$

Following the same process, we derive the SINR of \mathbf{X}_2

$$\hat{x}_{2,irc} = \mathbf{w}_{irc}^H \mathbf{y}_{d2} \triangleq \beta x_2 + \omega_2$$

and their variances are

$$\begin{aligned} \text{var}(\beta x_2) &= \beta^2 \mathcal{E}_d \\ \text{var}(\omega_2) &= \mathbf{w}_{irc}^H (\alpha^2 \mathcal{E}_p \bar{\Psi}_{p1} + N_0 \mathbf{I}) \mathbf{w}_{irc} \\ &= (\beta - \beta^2) \mathcal{E}_d + \alpha^2 \mathcal{E}_p \mathbf{w}_{irc}^H \bar{\Psi}_{p1} \mathbf{w}_{irc} \\ &\quad - \mathcal{E}_d \mathbf{w}_{irc}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \mathbf{w}_{irc} \\ &\triangleq (\beta - \beta^2) \mathcal{E}_d + \mathcal{E}_p \phi_1 - \mathcal{E}_d \phi_2, \end{aligned}$$

where \mathbf{y}_{d2} and $\hat{x}_{2,irc}$ are the column of \mathbf{Y}_{d2} and the column of the estimate of \mathbf{X}_2 . Thus, the SINR of \mathbf{X}_2 is obtained by

$$\text{SINR}_{x_2} = \frac{\beta^2}{\beta(1-\beta) + \frac{\mathcal{E}_p}{\mathcal{E}_d} \phi_1 - \phi_2}$$

The residual interference $(\mathcal{E}_p/\mathcal{E}_d)\phi_1 - \phi_2$ is due to the covariance mismatch between the weighting vector \mathbf{w}_{irc} and the covariance matrix of \mathbf{y}_{d2} . To get more insight, we first define $\mathbf{R}^{-1} = \bar{\mathbf{h}} \bar{\mathbf{h}}^H + (N_0/\mathcal{E}_d) \mathbf{I}$ and then rewrite \mathbf{w}_{irc}^H as:

$$\mathbf{w}_{irc}^H = \mathbf{h}^H \left(\mathbf{R} - \frac{\mathbf{R} \mathbf{h} \mathbf{h}^H \mathbf{R}}{1 + \mathbf{h}^H \mathbf{R} \mathbf{h}} \right).$$

Substituting this form into ϕ_1 and ϕ_2 , we have

$$\begin{aligned} \phi_1 &= \frac{\mathbf{h}^H \mathbf{R} \Psi_{p1} \mathbf{R} \mathbf{h}}{|1 + \mathbf{h}^H \mathbf{R} \mathbf{h}|^2} = \frac{|\mathbf{h}^H \mathbf{R} \bar{\mathbf{h}}_{p1}|^2}{|1 + \mathbf{h}^H \mathbf{R} \mathbf{h}|^2} \\ &= \xi^2 \cdot \left| \mathbf{h}^H \left[\left(\frac{N_0}{\mathcal{E}_d} + \bar{\mathbf{h}}^H \bar{\mathbf{h}} \right) \mathbf{I} - \bar{\mathbf{h}} \bar{\mathbf{h}}^H \right] \bar{\mathbf{h}}_{p1} \right|^2 \\ &= \xi^2 \cdot \left| \mathbf{h}^H \bar{\mathbf{h}}_{p1} \frac{N_0}{\mathcal{E}_d} + \bar{\mathbf{h}}^H \bar{\mathbf{h}} \mathbf{h}^H \bar{\mathbf{h}}_{p1} - \mathbf{h}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \bar{\mathbf{h}}_{p1} \right|^2, \end{aligned}$$

and

$$\begin{aligned} \phi_2 &= \frac{|\mathbf{h}^H \mathbf{R} \bar{\mathbf{h}}|^2}{|1 + \mathbf{h}^H \mathbf{R} \mathbf{h}|^2} \\ &= \xi^2 \cdot \left| \mathbf{h}^H \left[\left(\frac{N_0}{\mathcal{E}_d} + \bar{\mathbf{h}}^H \bar{\mathbf{h}} \right) \mathbf{I} - \bar{\mathbf{h}} \bar{\mathbf{h}}^H \right] \bar{\mathbf{h}} \right|^2 \\ &= \xi^2 \cdot \left| \frac{N_0}{\mathcal{E}_d} \mathbf{h}^H \bar{\mathbf{h}} \right|^2, \end{aligned}$$

where $\bar{\Psi}_{p1} = \bar{\mathbf{h}}_{p1} \bar{\mathbf{h}}_{p1}^H$ by the Cholesky decomposition and

$$\begin{aligned} \xi &= |1 + \mathbf{h}^H \mathbf{R} \mathbf{h}|^{-1} \left| \frac{N_0^2}{\mathcal{E}_d^2} + \frac{N_0}{\mathcal{E}_d} \cdot \bar{\mathbf{h}}^H \bar{\mathbf{h}} \right|^{-1} \\ &= \left| \left(\frac{N_0}{\mathcal{E}_d} \right)^2 + \left(\frac{N_0}{\mathcal{E}_d} \right) (\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h}) \right|^{-1}. \end{aligned}$$

To obtain above results, we need to rewrite \mathbf{R} by the Sherman-Morrison formula, that is

$$\begin{aligned} \mathbf{R} &= \left(\bar{\mathbf{h}} \bar{\mathbf{h}}^H + \frac{N_0}{\mathcal{E}_d} \mathbf{I} \right)^{-1} = \frac{\mathcal{E}_d}{N_0} \mathbf{I} - \frac{\frac{\mathcal{E}_d}{N_0} \bar{\mathbf{h}} \bar{\mathbf{h}}^H \frac{\mathcal{E}_d}{N_0}}{1 + \bar{\mathbf{h}}^H \left(\frac{\mathcal{E}_d}{N_0} \right) \bar{\mathbf{h}}} \\ &= \left(\frac{N_0^2}{\mathcal{E}_d^2} + \frac{N_0}{\mathcal{E}_d} \bar{\mathbf{h}}^H \bar{\mathbf{h}} \right)^{-1} \left[\left(\frac{N_0}{\mathcal{E}_d} + \bar{\mathbf{h}}^H \bar{\mathbf{h}} \right) \mathbf{I} - \bar{\mathbf{h}} \bar{\mathbf{h}}^H \right]. \end{aligned}$$

Therefore, the residual interference due to the covariance mismatch is obtained as follows:

$$\begin{aligned} \psi_1 &\triangleq \frac{\mathcal{E}_p}{\mathcal{E}_d} \phi_1 - \phi_2 \\ &= \frac{\frac{\mathcal{E}_p}{\mathcal{E}_d} |\mathbf{h}^H \bar{\mathbf{h}}_{p1}|^2 - |\bar{\mathbf{h}}^H \bar{\mathbf{h}}|^2}{1 + |\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h}|^2 + 2 \frac{N_0}{\mathcal{E}_d} (\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h})} \\ &\quad + \frac{\mathcal{E}_p \mathcal{E}_d}{\mathcal{E}_d N_0} \left(\frac{2 \Re [\bar{\mathbf{h}}_{p1}^H \mathbf{h} (\bar{\mathbf{h}}^H \bar{\mathbf{h}} \mathbf{h}^H \bar{\mathbf{h}}_{p1} - \mathbf{h}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \bar{\mathbf{h}}_{p1})]}{1 + |\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h}|^2 + 2 \frac{N_0}{\mathcal{E}_d} (\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h})} \right) \\ &\quad + \frac{\mathcal{E}_p \mathcal{E}_d^2}{\mathcal{E}_d N_0^2} \left(\frac{|\bar{\mathbf{h}}^H \bar{\mathbf{h}} \mathbf{h}^H \bar{\mathbf{h}}_{p1} - \mathbf{h}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \bar{\mathbf{h}}_{p1}|^2}{1 + |\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h}|^2 + 2 \frac{N_0}{\mathcal{E}_d} (\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h})} \right). \end{aligned}$$

Suppose $\mathcal{E}_d \gg N_0$, then we can simplify further,

$$\begin{aligned} \psi_1 &\approx \frac{\mathcal{E}_p \mathcal{E}_d^2}{\mathcal{E}_d N_0^2} \left(\frac{|\bar{\mathbf{h}}^H \bar{\mathbf{h}} \mathbf{h}^H \bar{\mathbf{h}}_{p1} - \mathbf{h}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \bar{\mathbf{h}}_{p1}|^2}{1 + |\bar{\mathbf{h}}^H \bar{\mathbf{h}} + \mathbf{h}^H \mathbf{h}|^2} \right) \\ &= \frac{\mathcal{E}_p}{\mathcal{E}_d} \cdot \left(\frac{\mathcal{E}_d}{N_0} \right)^2 \cdot \frac{\alpha^6 \zeta_1}{\alpha^4 \zeta_2 + \alpha^4 \zeta_3 + \zeta_4} \geq 0 \\ &\propto \frac{\mathcal{E}_p}{\mathcal{E}_d} \cdot \text{SNR}^2 \cdot \alpha^2 \end{aligned}$$

where we have the variable α from the interfering channel vectors $\bar{\mathbf{h}}$ and $\bar{\mathbf{h}}_{p1}$, and $\{\zeta_1, \zeta_2, \zeta_3, \zeta_4\}$ are defined for the other terms we are not interested in. Finally, The SINR of \mathbf{X}_3 is obtained as follows:

$$\text{SINR}_{x_3} = \frac{\beta^2}{\beta(1-\beta) + \psi_2}$$

where ψ_2 can be obtained by using $\bar{\mathbf{h}}_{p2}$ from $\bar{\Psi}_{p2} = \bar{\mathbf{h}}_{p2} \bar{\mathbf{h}}_{p2}^H$. Following the same process as for SINR_{x_2} , we can have $\psi_2 \propto (\mathcal{E}_p / \mathcal{E}_d) \cdot \text{SNR}^2 \cdot \alpha^2$.

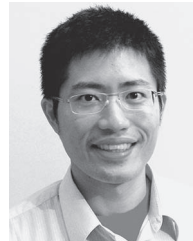
ACKNOWLEDGMENT

The present work was carried out within the framework of Celtic-Plus SHARING project.

REFERENCES

- [1] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.
- [2] D. Lopez-Perez *et al.*, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, 2011.
- [3] D. Bai *et al.*, "Lte-advanced modem design: Challenges and perspectives," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 178–186, Feb. 2012.
- [4] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
- [5] P. Alexander, M. Reed, J. Asenstorfer, and C. Schlegel, "Iterative multi-tuser interference reduction: Turbo CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1008–1014, Jul. 1999.
- [6] R. Narasimhan and S. Cheng, "Channel estimation and co-channel interference rejection for LTE-advanced MIMO uplink," in *Proc. IEEE WCNC*, 2012, pp. 416–420.
- [7] M. Escartin and P. Ranta, "Interference rejection with a small antenna array at the mobile scattering environment," in *Proc. 1st IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, Apr. 1997, pp. 165–168.
- [8] D. Asztély and B. Ottersten, "MLSE and spatio-temporal interference rejection combining with antenna arrays," in *Proc. IEEE 9th Eusipco*, 1998, pp. 1341–1344.

- [9] Y. Ohwatari, N. Miki, T. Asai, T. Abe, and H. Taoka, "Performance of advanced receiver employing interference rejection combining to suppress inter-cell interference in lte-advanced downlink," in *Proc. IEEE VTC—Fall*, 2011, pp. 1–7.
- [10] V. Kotszsch, W. Rave, and G. Fettweis, "Interference cancellation and suppression in asynchronous cooperating base station systems," in *Proc. Int. ITG WSA*, 2012, pp. 78–85.
- [11] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [12] R. H. Etkin, D. N. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
- [13] G. Bresler and D. Tse, "The two-user Gaussian interference channel: A deterministic view," *Eur. Trans. Telecommun.*, vol. 19, no. 4, pp. 333–354, 2008.
- [14] S. A. Jafar and S. Shamai, "Degrees of freedom region of the MIMO X channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 151–170, Jan. 2008.
- [15] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.
- [16] X. Mestre and M.-A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, Jan. 2006.
- [17] N. Ma and J. T. Goh, "Efficient method to determine diagonal loading value," in *Proc. IEEE ICASSP*, 2003, vol. 5, pp. V-341-1–V-341-4.
- [18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [19] G. Taricco, "Optimum receiver design and performance analysis of arbitrarily correlated rician fading mimo channels with imperfect channel state information," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1114–1134, Mar. 2010.
- [20] K. P. Murphy, "Naive Bayes classifiers," Univ. British Columbia, 2006.
- [21] H. Minn, V. Bhargava, and K. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 822–839, Jul. 2003.
- [22] M. Speth, S. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broad-band systems using OFDM. I," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668–1677, Nov. 1999.
- [23] C. Xiao, J. Wu, S.-Y. Leong, Y. Zheng, and K. Letaief, "A discrete-time model for triply selective MIMO rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1678–1688, Sep. 2004.
- [24] S. W. Nydick, "The wishart and inverse wishart distributions," *J. Stat.*, 2012.



Chien-Chun Cheng (S'11) received the B.S. degree in mechanical engineering and the M.S. degree in communication engineering from National Central University (NCU), Zhongli, Taiwan, in 2005 and 2007, respectively. He is currently working toward the joint Ph.D. degree at National Chiao Tung University (NCTU), Hsinchu, Taiwan and Ecole supérieure d'électricité (Supelec), Gif-sur-Yvette, France. His current research interests mainly lie in the areas of wireless communication and signal processing.



Serdar Sezginer (S'99–M'07–SM'11) was born in Bandirma, Turkey, in 1977. He received the B.Sc. and M.Sc. degrees in electrical and electronics engineering, both from Middle East Technical University (METU), Ankara, Turkey, in 2000 and 2003, respectively, and the Ph.D. degree from University of Paris-Sud XI, Orsay, France, in 2006. He is the recipient of the EEA Best Thesis Award of France in the area of signal and image processing. He is currently with Sequans Communications, Paris, France, where he is managing the physical layer

research and development activities such as algorithm study, architecture design and implementation. His research interests mainly lie in the areas of digital communications and statistical signal processing, including synchronization, estimation, equalization, and diversity techniques.



Hikmet Sari (S'78–M'81–SM'88–F'95) received the Engineering and Ph.D. degrees from the ENST, Paris, France, and the Post-Doctoral Habilitation degree from the University of Paris XI, Orsay Cedex, Paris. From 1980 to 2002, he held various research and management positions at Philips Research Laboratories, SAT, Alcatel, Pacific Broadband Communications, and Juniper Networks. He is currently a Professor and the Head of the Telecommunications Department of SUPÉLEC and also Chief Scientist at Sequans Communications. His distinctions include the IEEE Fellow Grade in 1995, the André Blondel Medal in 1995, the Edwin H. Armstrong Award in 2003, the Harold Sobol Award in 2012, as well as election to Academia Europaea (the Academy of Europe) and to the Science Academy of Turkey in 2012.

Dr. Sari has served as an Editor of the *IEEE TRANSACTIONS ON COMMUNICATIONS* (1987–1981), a Guest Editor of the *European Transactions on Telecommunications* (1993) and of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS* (1999), and an Associate Editor of the *IEEE COMMUNICATIONS LETTERS* (1999–2002). He served as a Distinguished Lecturer of the IEEE Communications Society in 2001–2006, as a member of the IEEE Fellow Evaluation Committee in 2002–2007, and as a member of the Awards Committee in 2005–2007. He was Chair of the Communication Theory Symposium of ICC 2002, Technical Program Chair of ICC 2004, Vice General Chair of ICC 2006, General Chair of PIMRC 2010, General Chair of WCNC 2012, Executive Chair of WCNC 2014, and General Chair of ICUWB 2014. He also chaired the Globecom and ICC Technical Content (GITC) Committee in 2010–2011. Currently, he is Vice President for Conferences of the IEEE Communications Society, and is serving as Technical Program Chair of EuCNC 2015, Executive Co-Chair of ICC 2016, and Executive Chair of ICC 2017.



Yu T. Su (SM'06) received the B.S. degree in electrical engineering from Tatung Institute of Technology, Taipei, Taiwan, in 1974, and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, CA, USA, in 1983. From 1983 to 1989, he was with LinCom Corporation, Los Angeles, CA, USA, where his last position was Corporate Scientist and was involved in various satellite communication systems design and verification. Since late 1989, he has been with the National Chiao Tung University (NCTU), Hsinchu, Taiwan, where he is currently a Professor in the Department of Electrical and Computer Engineering and chairing the NCTU's Committee for Liberal Arts Education. He was an Associative Dean of the College of Electrical and Computer Engineering from 2004 to 2007 and Head of the Communications Engineering Department from 2001 to 2003. He is also affiliated with the Microelectronic and Information Systems Research Center of the same university and served as a Deputy Director during 1997–2000. From 2005 to 2008, he was the Area Coordinator of the National Science Council's Telecommunications Programme. His main research interests include communication theory and statistical signal processing.

Dr. Su has served as an Editor of the *IEEE TRANSACTIONS ON COMMUNICATIONS* (1987–1981), a Guest Editor of the *European Transactions on Telecommunications* (1993) and of the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS* (1999), and an Associate Editor of the *IEEE COMMUNICATIONS LETTERS* (1999–2002). He served as a Distinguished Lecturer of the IEEE Communications Society in 2001–2006, as a member of the IEEE Fellow Evaluation Committee in 2002–2007, and as a member of the Awards Committee in 2005–2007. He was Chair of the Communication Theory Symposium of ICC 2002, Technical Program Chair of ICC 2004, Vice General Chair of ICC 2006, General Chair of PIMRC 2010, General Chair of WCNC 2012, Executive Chair of WCNC 2014, and General Chair of ICUWB 2014. He also chaired the Globecom and ICC Technical Content (GITC) Committee in 2010–2011. Currently, he is Vice President for Conferences of the IEEE Communications Society, and is serving as Technical Program Chair of EuCNC 2015, Executive Co-Chair of ICC 2016, and Executive Chair of ICC 2017.