

On the effect of imperfect timing synchronization on pilot contamination

Antonios Pitarokoilis, Emil Björnson and Erik G Larsson

The self-archived postprint version of this conference article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-142306>

N.B.: When citing this work, cite the original publication.

Pitarokoilis, A., Björnson, E., Larsson, E. G., (2017), On the effect of imperfect timing synchronization on pilot contamination, *2017 IEEE International Conference on Communications (ICC)*.

<https://doi.org/10.1109/ICC.2017.7996671>

Original publication available at:

<https://doi.org/10.1109/ICC.2017.7996671>

Copyright: IEEE

<http://www.ieee.org/>

©2017 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.



On the Effect of Imperfect Timing Synchronization on Pilot Contamination

Antonios Pitarokoilis¹, Emil Björnson², Erik G. Larsson²

¹ School of Electrical Engineering and ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden, apit@kth.se

² Dept. of Electrical Engineering (ISY), Linköping University, Linköping, Sweden, {emil.bjornson, erik.g.larsson}@liu.se

Abstract—The phenomenon of pilot contamination (PC) in multi-cell Massive MIMO systems is investigated in the presence of imperfect timing synchronization (TS). In particular, a basic setup is considered, where a base station (BS) is perfectly synchronized with the user of its cell, but there is imperfect TS between the BS and the user in another cell, possibly due to different propagation distances. A discrete-time system model is derived based on the continuous-time system model. The discrete-time system model accurately captures the phenomenon of imperfect TS in terms of the timing mismatch and the pulse shaping filter impulse responses. The derived discrete-time system model is used to study the achievable rates of a two-cell Massive MIMO uplink. It is shown that the structure imposed to the pilot contaminating signal due to the imperfect TS can be leveraged to mitigate the effect of PC. The level of PC suppression is quantified as a function of the timing mismatch and the characteristics of the transmit/receive pulse shaping filters.

I. INTRODUCTION

Massive MIMO is considered as a major candidate for future cellular wireless networks due to the unprecedentedly large gains in radiated energy and spectral efficiency that it provides over contemporary networks [1]. The promised gains are reaped by coherently combining the information signals with estimated channel impulse responses (IRs). These estimates are typically acquired at the base stations (BSs) by processing received signals that are generated by training, i.e., known and mutually orthogonal sequences transmitted by the users in predetermined time and frequency resources. The wireless channels, however, remain approximately constant only for a limited interval (*coherence interval*). Hence, the number of orthogonal training sequences is also limited.

Due to the finite number of orthogonal sequences that can fit in the coherence interval, non-orthogonal sequences must be reused between neighboring cells. Consequently, the channel estimate of a user served by a specific BS is *contaminated* by the channels from those users in neighboring cells that are assigned the same pilot sequence. The undesirable consequence is that interference is directed from the BS towards certain users in neighboring cells in the downlink and interference from certain users in neighboring cells is received by the BS in the uplink. This phenomenon is called as *pilot contamination* (PC) [2] and causes the achievable rate to remain bounded even when the number of BS antennas, M , grows unbounded.

The problem of PC has attracted significant interest by researchers and many approaches have been pursued, an

overview of which can be found in [3]. The authors in [4] propose a multi-cell MMSE technique that improves the performance of Massive MIMO in the presence of PC. In [5] a method to eliminate PC is proposed by exploiting the knowledge of the channel covariance matrices. Blind channel estimation schemes based on subspace methods are proposed in [6] and [7]. The authors in [8] propose a PC precoding scheme based on the large-scale fading coefficients.

In all the above works and in most of the multi-cell literature perfect timing synchronization (TS) is assumed. Even though this is possible to achieve between a BS and its served users, it is impossible to achieve TS throughout the multi-cellular network. The widely varying distances between the users and the BSs in other cells introduce timing mismatches in the communication. Prior work on the issue is [9], which investigates asynchronous cooperative multi-cell networks. To the authors' knowledge, this is the first treatment of imperfect TS within Massive MIMO. Further, the effect of PC is investigated under imperfect TS.

The contributions of this work can be summarized as follows. From a continuous-time model with imperfect TS a discrete-time model is derived. The discrete-time model describes clearly the effect of imperfect TS in terms of the timing mismatch, ϵ , and the characteristics of the sampled overall channel IR. The derived discrete-time system model is used to investigate the performance of an uplink channel, where a reference BS with a large number of antennas serves a reference user that is in perfect TS. A user in some other cell communicates simultaneously with its own BS using a non-orthogonal training sequence with respect to the training sequence of the reference user. The interfering user is in imperfect TS with the reference BS. Achievable rates for the reference user are derived rigorously and it is shown that PC can be efficiently mitigated by leveraging the spreading in the signal of the interfering user, which is caused due to the imperfect TS. Design considerations are investigated based on the structure of the derived achievable rates. The effects of timing mismatch, ϵ , and the sampled overall IR are reflected in the achievable rate expressions in a very intuitive way.

II. SYSTEM SETUP

In the considered setup, a single-antenna user is served (*served user*, (SU)) by a multi-antenna base station (BS) with M antenna elements. A neighboring single-antenna user is served by another BS and uses the same (or a non-orthogonal) training sequence (*pilot contaminating user*, (PCU)). The SU

This work was supported by the Swedish Research Council (VR), ELLIIT and CENIIT.

is in perfect TS with the BS, which is practically achievable using timing advance. However, there is a TS mismatch of ϵT_s seconds between the BS and the PCU, where $\epsilon \in [-0.5, 0.5)$ is a deterministic and known constant and T_s is the symbol interval. The mismatch is caused by the fact that the PCU is synchronized to another BS. The setup is depicted in Fig. 1.

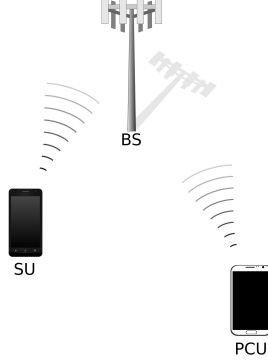


Fig. 1: The SU and the PCU are using non-orthogonal training sequences and are transmitting simultaneously. The SU is in perfect TS with the BS. There is a TS mismatch between the BS and the PCU.

Narrowband transmission is assumed and the propagation channels from the SU to the BS antenna array and from the PCU to the BS antenna array are given by

$$\mathbf{c}(t) = \sqrt{\beta} [h_1, \dots, h_M]^T \delta(t) = \sqrt{\beta} \mathbf{h} \delta(t), \quad (1)$$

$$\mathbf{c}_\epsilon(t) = \sqrt{\beta_\epsilon} [h_{\epsilon,1}, \dots, h_{\epsilon,M}]^T \delta(t) = \sqrt{\beta_\epsilon} \mathbf{h}_\epsilon \delta(t), \quad (2)$$

respectively, where $\delta(\cdot)$ is the Dirac delta function¹. A *block fading* model is assumed, according to which the small scale vectors $\mathbf{h} = [h_1, \dots, h_M]^T$ and $\mathbf{h}_\epsilon = [h_{\epsilon,1}, \dots, h_{\epsilon,M}]^T$ remain constant for a *coherence interval* of $N_c T_s$ seconds, where N_c is an integer, and subsequently change to an independent realization. The small scale vectors \mathbf{h} and \mathbf{h}_ϵ are two independent circularly symmetric complex Gaussian random vectors, i.e., $\mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$ and $\mathbf{h}_\epsilon \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$. The large scale fading coefficients β and β_ϵ are deterministic constants, which remain fixed throughout the communication.

III. DISCRETE-TIME MODEL WITH IMPERFECT SYNCHRONIZATION

In this section, a discrete-time system model is rigorously derived from the corresponding continuous-time model for the case of imperfect TS between a single-antenna transmitter and a single-antenna receiver. Let $\{\dots, x[k], \dots\}$ be a sequence of symbols to be transmitted. Assuming pulse-amplitude modulation, the corresponding transmitted continuous-time waveform, $x(t)$, is given by

$$x(t) = \sum_k x[k] p_T(t - \epsilon T_s - k T_s), \quad (3)$$

¹In this work, flat fading channels $\mathbf{c}(t)$ and $\mathbf{c}_\epsilon(t)$ are considered. More realistic channel models with a longer delay spread are left for future work.

where $p_T(\cdot)$ is the baseband-equivalent IR of the transmit pulse shaping filter. The time reference at the receiver is assumed to be the absolute time reference. There is an ϵT_s mismatch between the time reference at the transmitter and the absolute time reference, which appears as a time-shift in (3). The continuous-time received signal is given by

$$r(t) = (x \star c)(t) + n(t), \quad (4)$$

where \star is the convolution operation and $n(t)$ is a continuous-time white Gaussian noise process. The continuous-time signal after the pulse shaping filter, $p_R(\cdot)$, at the receiver is

$$y(t) = (p_R \star r)(t) + (p_R \star n)(t). \quad (5)$$

The pulse shaping filter, $p_R(\cdot)$, is assumed to be matched to $p_T(\cdot)$. The filtered continuous-time waveform in (5) is sampled at $t = iT_s$ and the i -th sample is given by

$$y[i] = y(iT_s) = (p_R \star r)(iT_s) + (p_R \star n)(iT_s). \quad (6)$$

When the propagation channel $c(t)$ is modeled as in (2), the sampled received signal in (6) is given by

$$y[i] = \sum_{l=-\infty}^{+\infty} \sqrt{\beta_\epsilon} h_\epsilon R_p(lT_s - \epsilon T_s) x[i-l] + n[i], \quad (7)$$

where $n[i] \triangleq (p_R \star n)(iT_s)$ is a white discrete-time sequence of circularly symmetric complex Gaussian random variables, i.e., $n[i] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2)$, and $R_p(t)$ is the convolution of the transmit filter and the matched receive filter,

$$R_p(t) = \int_{-\infty}^{+\infty} p_T^*(\tau) p_T(\tau + t) d\tau. \quad (8)$$

It is assumed that the pulse shaping filters $p_T(\cdot)$ and $p_R(\cdot)$ satisfy the Nyquist criterion for intersymbol interference (ISI) [10] when $R_p(t)$ is sampled with perfect TS ($\epsilon = 0$) it holds $R_p(lT_s) = 0, \forall l \neq 0$ and $R_p(0) = 1$.

Since the pulses are approximately time-limited [11], $R_p(t)$ will have negligible values as $|t|$ grows large. With $r_{\epsilon,l} \triangleq R_p(lT_s - \epsilon T_s)$ it is possible to truncate the discretized IR in (7) so that $100(1-\epsilon)\%$ of the energy is contained in the truncated IR, where ϵ is arbitrarily small. The truncated discrete-time system model is given by

$$y[i] = \sqrt{\beta_\epsilon} h_\epsilon \sum_{l=-L_1}^{L_2} r_{\epsilon,l} x[i-l] + n[i]. \quad (9)$$

Observation 1: The discrete-time IRs in (7) and (9) are non-causal, even though the propagation channel is strictly causal. The non-causality is due to the transmit and receive filtering. However, the system in (9) is still realizable. In practice, the transmit and receive filters are time-shifted and the non-causal part is truncated so that minimum distortion is introduced to the resulting filters. This results in a causal overall IR and, hence, a realizable system.

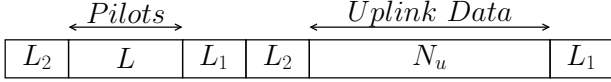


Fig. 2: Transmission Scheduling with cyclic pre- and postfix. The assumption that pilots and uplink data are separated is not fundamental and is made merely to simplify the analysis.

IV. DISCRETE-TIME MULTI-USER SYSTEM MODEL

Based on the discrete-time model described in Section III, the signal received at the m -th BS antenna element at time i , $y_m[i]$, for the setup considered in Section II is given by

$$y_m[i] = \sqrt{\rho\beta}h_mx[i] + \sqrt{\rho\beta_\epsilon}h_{\epsilon,m} \sum_{l=-L_1}^{L_2} r_{\epsilon,l}x_\epsilon[i-l] + n_m[i], \quad (10)$$

where $n_m[i] \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ and ρ is the average signal-to-noise ratio (SNR) at the m -th BS antenna element when only the SU transmits information symbols with $\mathbb{E}\{|x[i]|^2\} = 1$ and $\beta = 1$. One symbol is transmitted per channel use by SU and PCU, $x[i]$ and $x_\epsilon[i]$, respectively, and N_c is the number of channel uses per coherence interval. A per channel use average power constraint is assumed for both users, SU and PCU, i.e., $\mathbb{E}\{|x[i]|^2\} = 1$ and $\mathbb{E}\{|x_\epsilon[i]|^2\} = 1$. The received vector is given by

$$\mathbf{y}[i] = \sqrt{\rho\beta}\mathbf{h}x[i] + \sqrt{\rho\beta_\epsilon}\mathbf{h}_\epsilon \sum_{l=-L_1}^{L_2} r_{\epsilon,l}x_\epsilon[i-l] + \mathbf{n}[i]. \quad (11)$$

The first term is the signal from the SU, which is in perfect TS with the BS and hence the effective channel is flat fading. The second term is the signal from the PCU, which is in TS mismatch and hence the effective channel contains ISI.

A. Transmission Scheduling

The coherence time of N_c channel uses is split into two parts, one for uplink pilot transmission of $L = L_1 + L_2 + 1$ channel uses and one for uplink data transmission of N_u channel uses. Both transmission intervals are padded with cyclic pre- and postfixes, i.e., a cyclic prefix of L_2 symbols consists of the last L_2 symbols of a block of symbols and is padded before the block. A cyclic postfix of L_1 symbols consists of the first L_1 symbols of a block of symbols and is padded after the block, as shown in Fig. 2.

B. Uplink Channel Estimation and Data Transmission

A sequence of L channel uses is spent for channel training. The desired user transmits an arbitrary sequence, $\mathbf{b} = [b[0], \dots, b[L-1]]^T$, with energy $\|\mathbf{b}\|^2 = L$ and the PCU transmits a similar arbitrary sequence, $\mathbf{b}_\epsilon = [b_\epsilon[0], \dots, b_\epsilon[L-1]]^T$, with energy $\|\mathbf{b}_\epsilon\|^2 = L$. The analysis holds both for the case of orthogonal sequences, which corresponds to the scenario without PC for $\epsilon = 0$, and for non-orthogonal sequences, which corresponds to the scenario with PC when $\epsilon = 0$.

At the i -th channel use of the training interval, after the removal of the pre- and postfixes, the received vector at the BS is given by

$$\mathbf{y}_p[i] = \sqrt{\rho\beta}\mathbf{b}[i]\mathbf{h} + \sqrt{\rho\beta_\epsilon}\mathbf{r}_{p,i}^T\mathbf{b}_\epsilon\mathbf{h}_\epsilon + \mathbf{n}_p[i], \quad (12)$$

where $\mathbf{r}_p^T \triangleq [r_{\epsilon,0}, \dots, r_{\epsilon,-L_1}, r_{\epsilon,L_2}, \dots, r_{\epsilon,1}]$ and $\mathbf{r}_{p,i}^T$ is the cyclic shift of \mathbf{r}_p^T to the right for i positions. The received vector during the whole training interval is given by

$$\mathbf{y}_p = \sqrt{\rho}(\mathbf{A} \otimes \mathbf{I}_M)\mathbf{h}_p + \mathbf{n}_p, \quad (13)$$

where $\mathbf{y}_p = [\mathbf{y}_p[0], \dots, \mathbf{y}_p[L_1 + L_2]]$, $\mathbf{h}_p = [\mathbf{h}^T \mathbf{h}_\epsilon^T]^T$, $\mathbf{n}_p = [\mathbf{n}_p[0], \dots, \mathbf{n}_p[L_1 + L_2]]$, $\mathbf{A} = [\sqrt{\beta}\mathbf{b} \quad \sqrt{\beta_\epsilon}\mathbf{R}_p\mathbf{b}_\epsilon]$ with $\mathbf{r}_{p,i}^T$ being the i -th row of \mathbf{R}_p , $i = 0, \dots, L_1 + L_2$ and \otimes is the Kronecker product. The MMSE estimate for the vector \mathbf{h}_p given the observation \mathbf{y}_p is

$$\begin{aligned} \hat{\mathbf{h}}_p &= \mathbb{E}\{\mathbf{h}_p\mathbf{y}_p^H\} (\mathbb{E}\{\mathbf{y}_p\mathbf{y}_p^H\})^{-1} \mathbf{y}_p \\ &= \left(\left(\sqrt{\rho} (\rho\mathbf{A}^H\mathbf{A} + \mathbf{I}_2)^{-1} \mathbf{A}^H \right) \otimes \mathbf{I}_M \right) \mathbf{y}_p. \end{aligned} \quad (14)$$

The MMSE estimate $\hat{\mathbf{h}}_p$ is distributed as $\hat{\mathbf{h}}_p \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \hat{\mathbf{C}}_p)$, with $\hat{\mathbf{C}}_p = \left(\rho\mathbf{A}^H (\rho\mathbf{A}\mathbf{A}^H + \mathbf{I}_L)^{-1} \mathbf{A} \right) \otimes \mathbf{I}_M$ and the associated estimation error $\tilde{\mathbf{h}}_p = \mathbf{h}_p - \hat{\mathbf{h}}_p$ is statistically independent with $\hat{\mathbf{h}}_p$ and is distributed as $\tilde{\mathbf{h}}_p \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_{2M} - \hat{\mathbf{C}}_p)$.

Similarly to \mathbf{h}_p , the estimate $\hat{\mathbf{h}}_p$ is partitioned as $\hat{\mathbf{h}}_p = [\hat{\mathbf{h}}^T \hat{\mathbf{h}}_\epsilon^T]^T$. The corresponding covariance matrices are $\hat{\mathbf{C}} \triangleq \mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = \hat{c}\mathbf{I}_M$ and $\hat{\mathbf{C}}_\epsilon \triangleq \mathbb{E}\{\hat{\mathbf{h}}_\epsilon\hat{\mathbf{h}}_\epsilon^H\} = \hat{c}_\epsilon\mathbf{I}_M$, where $\hat{c} = \rho\beta\mathbf{b}^H (\rho\mathbf{A}\mathbf{A}^H + \mathbf{I}_L)^{-1} \mathbf{b}$ and $\hat{c}_\epsilon = \rho\beta_\epsilon\mathbf{b}_\epsilon^H \mathbf{R}_p (\rho\mathbf{A}\mathbf{A}^H + \mathbf{I}_L)^{-1} \mathbf{R}_p\mathbf{b}_\epsilon$. The estimates $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}_\epsilon$ are correlated with cross-covariance $\hat{\mathbf{C}}_{x_\epsilon} = \mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}_\epsilon^H\} = \hat{c}_{x_\epsilon}\mathbf{I}_M$, where $\hat{c}_{x_\epsilon} \triangleq \rho\sqrt{\beta\beta_\epsilon}\mathbf{b}^H (\rho\mathbf{A}\mathbf{A}^H + \mathbf{I}_L)^{-1} \mathbf{R}_p\mathbf{b}_\epsilon$.

The received vector at the i -th channel use of the data transmission interval after the removal of the cyclic pre- and postfixes is given by (11) or, for short, when a block of N_u symbols is sent

$$\mathbf{y}[i] = \sqrt{\rho\beta}\mathbf{h}x[i] + \sqrt{\rho\beta_\epsilon}\mathbf{h}_\epsilon\mathbf{r}_{d,i}^T\mathbf{x}_\epsilon + \mathbf{n}[i], \quad (15)$$

where $\mathbf{x}_\epsilon = [x_\epsilon[0], \dots, x_\epsilon[N_u - 1]]^T$ and $\mathbf{r}_{d,i}^T$ is the cyclic shift of the N_u -dimensional row vector \mathbf{r}_d^T i positions to the right, where $\mathbf{r}_d^T = [r_{\epsilon,0}, r_{\epsilon,-1}, \dots, r_{\epsilon,-L_1}, 0, \dots, 0, r_{\epsilon,L_2}, \dots, r_{\epsilon,1}]$. For example, $\mathbf{r}_{d,1}^T = [r_{\epsilon,1}, r_{\epsilon,0}, \dots, r_{\epsilon,-L_1}, 0, \dots, 0, r_{\epsilon,L_2}, \dots, r_{\epsilon,2}]$.

V. ACHIEVABLE RATES

In this section, achievable rates for the SU are presented for different receiver strategies at the BS.

A. Maximum Ratio Combining

Given the channel knowledge, $\hat{\mathbf{h}}_p$, acquired via uplink training (see Section IV-B), the BS processes the received signal, $\mathbf{y}[i]$, in (15) using maximum ratio combining (MRC):

$$\begin{aligned}\hat{x}[i] &= \hat{\mathbf{h}}^H \mathbf{y}[i] \\ &= \sqrt{\rho\beta} \hat{\mathbf{h}}^H \mathbf{h} x[i] + \sqrt{\rho\beta\epsilon} \hat{\mathbf{h}}^H \mathbf{h}_\epsilon \mathbf{r}_{d,i}^T \mathbf{x}_\epsilon + \hat{\mathbf{h}}^H \mathbf{n}[i].\end{aligned}\quad (16)$$

The detected symbol $\hat{x}[i]$ can be further written as

$$\hat{x}[i] = \sqrt{\rho\beta} \left\| \hat{\mathbf{h}} \right\|^2 x[i] + \tilde{n}[i], \quad (17)$$

where

$$\begin{aligned}\tilde{n}[i] &= \sqrt{\rho\beta\epsilon} \hat{\mathbf{h}}^H \hat{\mathbf{h}}_\epsilon \mathbf{r}_{d,i}^T \mathbf{x}_\epsilon + \sqrt{\rho\beta} \hat{\mathbf{h}}^H \tilde{\mathbf{h}} x[i] \\ &\quad + \sqrt{\rho\beta\epsilon} \hat{\mathbf{h}}^H \tilde{\mathbf{h}}_\epsilon \mathbf{r}_{d,i}^T \mathbf{x}_\epsilon + \hat{\mathbf{h}}^H \mathbf{n}[i].\end{aligned}\quad (18)$$

With $x[i] \sim \mathcal{N}_C(0, 1)$ and $\mathbb{E} \left\{ \left\| \hat{\mathbf{h}} \right\|^2 x[i] \tilde{n}^*[i] \right\} = 0$, an achievable rate [12] for the SU is given by

$$R_{\text{MRC}} = \mathbb{E} \left\{ \log_2 \left(1 + \frac{\rho\beta \left\| \hat{\mathbf{h}} \right\|^2}{\sigma_\epsilon^2} \right) \right\}, \quad (19)$$

where the expectation is over $\hat{\mathbf{h}}_p$ and

$$\begin{aligned}\sigma_\epsilon^2 &= \rho\beta\epsilon \left\| \mathbf{r}_d \right\|^2 \frac{\left\| \hat{\mathbf{h}}_\epsilon^H \hat{\mathbf{h}} \right\|^2}{\left\| \hat{\mathbf{h}} \right\|^2} + \rho\beta(1 - \hat{c}) \\ &\quad + \rho\beta\epsilon \left\| \mathbf{r}_d \right\|^2 (1 - \hat{c}_\epsilon) + 1.\end{aligned}\quad (20)$$

When the rate of the PCU is low enough²[13], the BS can first decode the information from the PCU and then subtract it by the detected symbol, $\hat{x}[i]$. An achievable rate with PCU interference cancellation is given by

$$R_{\text{SIC}} = \mathbb{E} \left\{ \log_2 \left(1 + \frac{\rho\beta \left\| \hat{\mathbf{h}} \right\|^2}{\sigma_{\text{SIC}}^2} \right) \right\}, \quad (21)$$

where

$$\sigma_{\text{SIC}}^2 = \rho\beta(1 - \hat{c}) + \rho\beta\epsilon \left\| \mathbf{r}_d \right\|^2 (1 - \hat{c}_\epsilon) + 1. \quad (22)$$

It is noted that the rate R_{SIC} is mainly provided as a measure of the performance when most of PC is efficiently suppressed. Indeed, as $M \rightarrow \infty$ and in spite of the fact that the estimate $\hat{\mathbf{h}}$ is contaminated by the channel \mathbf{h}_ϵ of PCU, the rate R_{SIC} becomes asymptotically $R_{\text{SIC}} \rightarrow \log_2 \left(1 + \frac{\rho\beta \hat{c} M}{\sigma_{\text{SIC}}^2} \right)$, i.e., it becomes unbounded with M , since the coherent interference is subtracted away. Hence, the PC only degrades the estimation quality in the asymptotic regime.

²The rate of the PCU must be lower than the rate that the BS can decode the information signal of the PCU by treating the desired signal as noise.

B. PC-Aware Reception (PCAR)

Define $\mathbf{u}_\epsilon \triangleq \frac{\hat{\mathbf{h}}_\epsilon}{\left\| \hat{\mathbf{h}}_\epsilon \right\|}$ and $\mathbf{P}_u \triangleq \mathbf{I}_M - \mathbf{u}_\epsilon \mathbf{u}_\epsilon^H$. The matrix \mathbf{P}_u is a projection matrix onto the orthogonal complement of the unit-norm vector \mathbf{u}_ϵ , i.e., of the normalized channel estimate of the PCU. The detector first projects the received vector $\mathbf{y}[i]$ onto the orthogonal complement of the estimated channel of the PCU and then performs MRC based on $\hat{\mathbf{h}}$, i.e.,

$$\begin{aligned}\hat{x}[i] &= \hat{\mathbf{h}}^H \mathbf{P}_u \mathbf{y}[i] = \sqrt{\rho\beta} \hat{\mathbf{h}}^H \mathbf{P}_u \hat{\mathbf{h}} x[i] + \sqrt{\rho\beta} \hat{\mathbf{h}}^H \mathbf{P}_u \tilde{\mathbf{h}} x[i] \\ &\quad + \sqrt{\rho\beta\epsilon} \hat{\mathbf{h}}^H \mathbf{P}_u \tilde{\mathbf{h}}_\epsilon \mathbf{r}_{d,i}^T \mathbf{x}_\epsilon + \hat{\mathbf{h}}^H \mathbf{P}_u \mathbf{n}[i].\end{aligned}\quad (23)$$

This strategy is coined as *pilot-contamination-aware reception* (PCAR). PCAR can be considered essentially as a multi-cell zero-forcing (ZF) scheme, as considered in [14]. An achievable rate for the SU when PCAR is used is given by

$$R_{\text{PCAR}} = \mathbb{E} \left\{ \log_2 \left(1 + \frac{\rho\beta \left| \hat{\mathbf{h}}^H \mathbf{P}_u \hat{\mathbf{h}} \right|^2}{\sigma_{\text{PCAR}}^2} \right) \right\}, \quad (24)$$

where

$$\begin{aligned}\sigma_{\text{PCAR}}^2 &= \rho\beta(1 - \hat{c}) \hat{\mathbf{h}}^H \mathbf{P}_u \hat{\mathbf{h}} \\ &\quad + \rho\beta\epsilon \left\| \mathbf{r}_d \right\|^2 (1 - \hat{c}_\epsilon) \hat{\mathbf{h}}^H \mathbf{P}_u \hat{\mathbf{h}} + \hat{\mathbf{h}}^H \mathbf{P}_u \hat{\mathbf{h}}.\end{aligned}\quad (25)$$

In order to include the loss in spectral efficiency due to training overhead and the pre-/postfixes, the average achievable rate over $N_c = N_u + L + 2(L_1 + L_2)$ channel uses is used as the figure of merit of interest, i.e.,

$$\bar{R}_\times \triangleq \frac{N_u}{N_c} R_\times = \left(1 - \frac{L + 2(L_1 + L_2)}{N_c} \right) R_\times, \quad (26)$$

where $\times \in \{\text{MRC}, \text{SIC}, \text{PCAR}\}$.

C. On the Choice of Training Sequence

It is clear from (20), (22), and (25) that the additive noise variances depend on the choice of the training sequences, since this choice affects \hat{c} , \hat{c}_ϵ and \mathbf{P}_u . In addition, this choice affects the alignment of the channel estimates $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}_\epsilon$. Consider, for instance, the all-ones training sequence³ being used for both users. Due to the fact that \mathbf{R}_p is a circulant matrix, \mathbf{A} becomes rank deficient since the second column is a scalar multiple of the first. As a result, the estimates $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}_\epsilon$ are co-linear and PC cannot be eliminated. An intuitive choice for the selection of the training sequence is the eigenvector of \mathbf{R}_p that corresponds to the eigenvalue with the minimum absolute value. Since \mathbf{R}_p is circulant [15], the training sequence is the column of the L -dimensional DFT matrix that corresponds to the eigenvalue with the minimum absolute value appropriately scaled to meet the energy constraint for the training sequences. It is noted that the matrix \mathbf{R}_p depends only on ϵ and $R_p(t)$ in (8). Hence, it can be computed by the BS as soon as ϵ has been estimated. In the numerical examples presented in Section VI both users, SU and PCU, use the same sequence based on the above criterion.

³The same holds for any scaled version of the all-ones sequence.

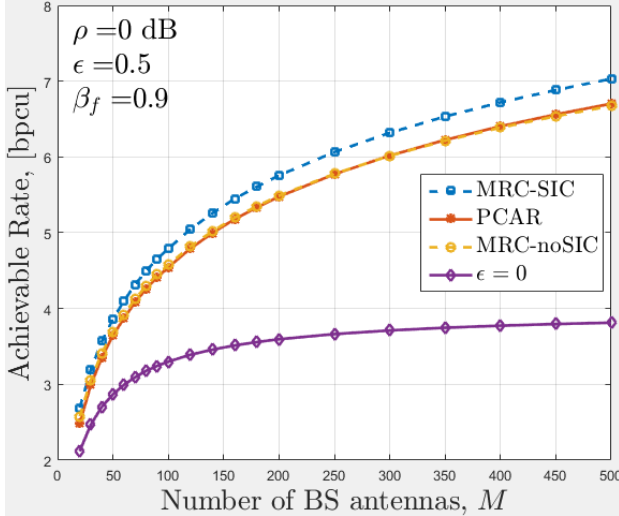


Fig. 3: Average achievable rate [bpsu] as a function of the number of BS antennas, M , for $\epsilon = 0.5$ and $\beta_f = 0.9$.

VI. NUMERICAL EXAMPLES

In the following, the achievable rate expressions in (19), (21), (24) and (26) are numerically evaluated under various choices of the parameters i) number of BS antennas, M , ii) timing mismatch, ϵ , and iii) Nyquist filter roll-off factor, β_f . Throughout the numerical examples, the large scale fading coefficients are kept fixed to the values $\beta = 0.5$ and $\beta_\epsilon = \beta/4$ the SNR, $\rho = 0$ dB, and the IR of the cascade of the transmit and receive pulse shaping filters is a raised cosine pulse with

$$R_{RC}(t) = \begin{cases} \frac{\pi}{4} \operatorname{sinc}\left(\frac{1}{2\beta_f}\right), & t = \pm \frac{T_s}{2\beta_f} \\ \operatorname{sinc}\left(\frac{t}{T_s}\right) \frac{\cos\left(\frac{\pi\beta_f t}{T_s}\right)}{1 - \left(\frac{2\beta_f t}{T_s}\right)^2}, & \text{otherwise.} \end{cases} \quad (27)$$

The coefficients r_{-L_1}, \dots, r_{L_2} are derived from (27) by evaluating $R_{RC}(t)$ at $t = lT_s + \epsilon T_s$, $l \in \mathbb{Z}$. The length, L , of the truncated IR is such that at least $\varepsilon = 99.9\%$ of the energy of the sampled IRs is contained in the truncated IR. The length L is variable and depends on ϵ and β_f . Both users are assigned the same training sequence as discussed in Section V-C. For $\epsilon = 0$, this choice of training sequences corresponds to the typical scenario with PC.

In Fig. 3 the average achievable rate in (26) is plotted as a function of the number of BS antennas, M , for $\epsilon = 0.5$ and $\beta_f = 0.9$. The “ $\epsilon = 0$ ” curve corresponds to the case when MRC without SIC is used (see (19)) and PCU is in perfect TS with the BS. In this case L channel uses are spent for training, $\mathbf{R}_p = \mathbf{I}_L$ and there is no ISI from the PCU. Hence, the pre/postfixes are redundant and the corresponding channel uses are spent for data transmission. Consequently, the penalty on the average achievable rate is $\left(1 - \frac{L}{N_c}\right)$ in contrast to (26). In Fig. 3 it is clear that timing mismatch is beneficial for SU and the gain in achievable rate is substantial as M increases. With the particular choice of ϵ and β_f , regular MRC and PCAR are very close to the MRC-SIC case, which is the benchmark for

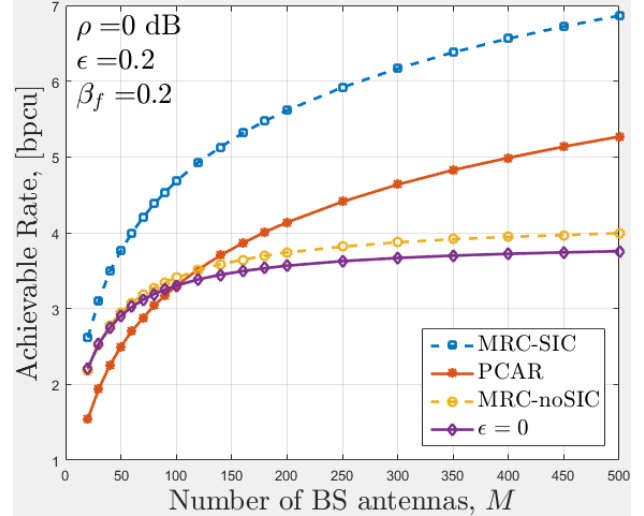


Fig. 4: Average achievable rate [bpsu] as a function of the number of BS antennas, M , for $\epsilon = 0.2$ and $\beta_f = 0.2$.

the pilot-contamination-free case, in the sense described at the end of Section V-A.

In Fig. 4 a similar setup is presented for $\epsilon = 0.2$ and $\beta_f = 0.2$. In this case the regular MRC (see (19)) is only slightly better than the “ $\epsilon = 0$ ” case. The average achievable rate with PCAR is the lowest for small antenna arrays but it is clearly better than MRC and “ $\epsilon = 0$ ” when M increases. An approximately constant gap of 1 bit/channel use from the MRC-SIC average achievable rate is observed. Hence, PCAR can efficiently suppress PC. In Figs. 3 and 4 it is shown that timing mismatch can be leveraged in order to mitigate PC. Intuitively, this happens because timing mismatch causes ISI (spreading) to the signal from the PCU. This spreading gives more degrees of freedom along the temporal dimension for the BS. That is, the BS can use timing mismatch to get estimates for the SU and the PCU channels that are not co-linear. Hence, the signals from the SU and the PCU become distinguishable as M increases. The level of PC suppression depends on the parameters ϵ and β_f and on the decay of the pulse in (8).

In Fig. 5 the average achievable rate is plotted as a function of the timing mismatch, ϵ , for $M = 100$ and $\beta = 0.2$. Increasing ϵ is beneficial for all curves. In particular, it is observed that PCAR has the worst achievable rate performance for small ϵ , however, it approaches the achievable rate of MRC-SIC as ϵ increases. For fixed β_f , the length of the truncated IR increases as ϵ increases. Hence, in general, the BS can discriminate more accurately the signal originating from SU from the signal of the PCU. In particular, for small ϵ the length of the truncated IR is 1, hence, the BS does not have any additional degrees of freedom to get non-co-linear channel estimates and distinguish between the two signals. Hence, the performance of PCAR, which projects the received signal on the orthogonal complement of the estimated PCU channel, drops to 0. As ϵ increases the BS has more degrees of freedom to mitigate PC and the achievable rate increases.

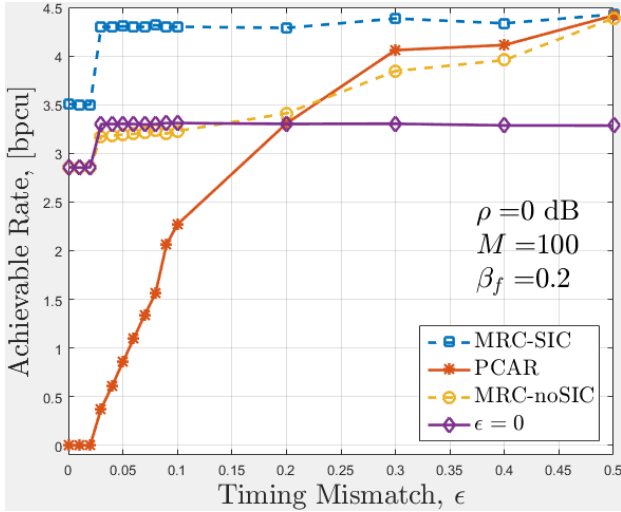


Fig. 5: Average achievable rate [bpsu] as a function of the timing mismatch, ϵ , for $M = 100$ and $\beta = 0.2$.

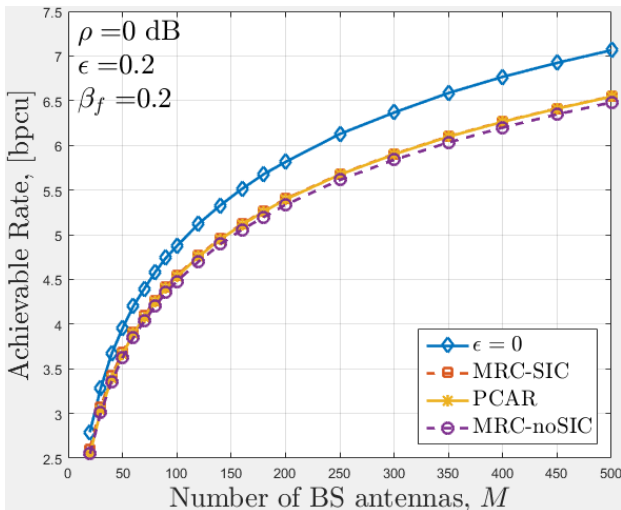


Fig. 6: Average achievable rate [bpsu] as a function of the number of BS antennas, M , for $\epsilon = 0.2$ and $\beta = 0.2$, when orthogonal training is used.

So far it has been observed and documented that imperfect TS causes co-linear training sequences to appear non-co-linear at the receiver, facilitating the mitigation of PC. Similarly, imperfect TS causes orthogonal training sequences to appear non-orthogonal at the receiver. This is harmful since interference caused by users with orthogonal sequences cannot be eliminated. Therefore, there is a clear trade-off between the level of PC suppression and the increase in the interference from users with orthogonal training sequences. In Fig. 6 the average achievable rate is plotted as a function of M for $\epsilon = 0.2$ and $\beta = 0.2$. The SU and the PCU use mutually orthogonal training sequences. It is clear that “ $\epsilon = 0$ ” has the best achievable rate performance since there is no pilot contamination in this scenario. However, it is observed that there is a moderate decrease in the achievable rates of the other

receivers. This is an indication that the increase in interference from users with orthogonal training sequences will be minor.

VII. SUMMARY

A two-user Massive MIMO uplink is considered, where one user is in perfect TS with the BS and the other is not. The discrete-time system model used in this work is derived based on the continuous-time system model in the presence of imperfect TS and it is shown that imperfect TS causes ISI. Both users communicate simultaneously and use non-orthogonal training sequences. It is shown that the ISI introduced by imperfect TS can be used to mitigate PC, which is a major limiting phenomenon in multi-cell Massive MIMO systems. On the other hand, imperfect TS breaks the orthogonality at the receiver side between orthogonal training sequences. This introduces interference to the channel estimation and to the data detection phases. However, the results presented in this work motivate the further study of imperfect TS as a means to mitigate PC.

REFERENCES

- [1] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.
- [2] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, Nov. 2010.
- [3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of Massive MIMO: Benefits and challenges,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 742–758, Oct. 2014.
- [4] 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, pp. 2640–2651, Aug. 2011.
- [5] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, “A coordinated approach to channel estimation in large-scale multiple-antenna systems,” *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 264–273, February 2013.
- [6] 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)*, pp. 3249–3252, March 2012.
- [7] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, “Blind pilot decontamination,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 773–786, Oct. 2014.
- [8] A. Ashikhmin and T. Marzetta, “Pilot contamination precoding in multi-cell large scale antenna systems,” in *2012 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 1137–1141, July 2012.
- [9] H. Zhang, N. B. Mehta, A. F. Molisch, J. Zhang, and S. H. Dai, “Asynchronous interference mitigation in cooperative base station systems,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 155–165, Jan 2008.
- [10] J. G. Proakis, *Digital Communications*. McGraw-Hill, 1995.
- [11] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.
- [12] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?,” *IEEE Transactions on Information Theory*, vol. 49, pp. 951 – 963, Apr. 2003.
- [13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2004.
- [14] E. Björnson, E. G. Larsson, and M. Debbah, “Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?,” *IEEE Transactions on Wireless Communications*, vol. 15, pp. 1293–1308, Feb 2016.
- [15] R. M. Gray, “Toeplitz and circulant matrices: A review,” *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.