Uplink Performance of Time-Reversal MRC in Massive MIMO Systems subject to Phase Noise

Antonios Pitarokoilis, Saif Khan Mohammed and Erik G. Larsson

Linköping University Post Print

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

Uplink Performance of Time-Reversal MRC in Massive MIMO Systems Subject to Phase Noise

Antonios Pitarokoilis, Saif Khan Mohammed and Erik G. Larsson

Abstract—**Multi-user multiple-input multiple-output (MU-MIMO) cellular systems with an excess of base station (BS) antennas (Massive MIMO) offer unprecedented multiplexing gains and radiated energy efficiency. Oscillator phase noise is introduced in the transmitter and receiver radio frequency chains and severely degrades the performance of communication systems. We study the effect of oscillator phase noise in frequency-selective Massive MIMO systems with imperfect channel state information (CSI). In particular, we consider two distinct operation modes, namely when the phase noise processes at the M BS antennas are identical (synchronous operation) and when they are independent (non-synchronous operation). We analyze a linear and low-complexity time-reversal maximum-ratio combining (TR-MRC) reception strategy. For both operation modes we derive a lower bound on the sum-capacity and we compare their performance. Based on the derived achievable sum-rates, we show that with the proposed receive processing an $O(\sqrt{M})$ array gain is achievable. Due to the phase noise drift the estimated effective channel becomes progressively outdated. Therefore, phase noise effectively limits the length of the interval used for data transmission and the number of scheduled users. The derived achievable rates provide insights into the optimum choice of the data interval length and the number of scheduled users.**

Index Terms—**Receiver algorithns, MU-MIMO, phase noise.**

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology offers substantial performance gains in wireless links [1]. The spatial degrees of freedom enable many users to share the same time-frequency resources, paving the way for multi-user MIMO (MU-MIMO) systems [2]. MU-MIMO systems with an excess of BS antennas, termed as Massive MIMO or large-scale MIMO, have recently attracted significant interest [3]–[5]. They promise a significant increase in the total cell throughput by means of simple signal processing. At the same time, the radiated power can be scaled down with the number of BS antennas, M, while maintaining a desired sum-rate. More specifically, in [6] the authors show that in a MU-MIMO uplink with linear receivers and imperfect channel state information (CSI), by increasing the number of BS antennas from 1 to M, one can reduce the total transmit power by a factor $O(\sqrt{M})$ while maintaining a fixed per-user information rate.

In [7] the authors report an improved result for channels with arbitrary channel covariance matrices. The crucial assumption in Massive MIMO is that the squared Euclidean norm of the channel vector of each user grows as $O(M)$, whereas the inner products between channel vectors of different users grow at a lesser rate. This assumption can be justified in the MU-MIMO setting since the users are typically separated by many wavelengths, which implies that their channel vectors become asymptotically (in the number of BS antennas) orthogonal. Extensive measurements have confirmed the validity of this assumption [4], [5].

Phase noise is inevitable in communication systems due to imperfections in the circuitry of the local oscillators that are used for the conversion of the baseband signal to passband and vice versa. To be specific, phase noise is the instantaneous drift of the phase of the carrier wave and results in a widening of the power spectral density of the generated waveform. Phase noise causes a partial loss of coherency between the channel estimate and the true channel gain during data transmission. This can result in severe degradation of the system performance.

In MIMO an array power gain is obtained by coherently combining signals received by several antennas, using estimated channel responses. Since phase noise distorts the received data, it is crucial to examine its effect on the performance. Significant research work is available on phase noise. However, most of it is concerned with single-user single-antenna multi-carrier transmission, since multi-carrier transmission is more sensitive to phase noise compared to single-carrier transmission [8]. In [9] a method to calculate the bit-error-rate (BER) of a single-user orthogonal frequency division multiplexing (OFDM) system impaired with phase noise is provided. Reference [10] studies the signal-to-interference-and-noise-ratio (SINR) degradation in OFDM and proposes a method to mitigate the effect of phase noise. In [11] a method to characterize phase noise in OFDM systems is developed and an algorithm to compensate for the degradation is described. Finally, in [12] the authors propose a method to jointly estimate the channel coefficients and the phase noise in a single-user MIMO system and an associated phase noise mitigation algorithm.

From an information-theoretic point of view, the calculation of capacity of phase noise channels is challenging. To the best of our knowledge, the exact capacity of typical phase noise-impaired channels under realistic models is not known. The behavior of the capacity of such channels is only known asymptotically for some cases in the high signal-to-noise-ratio (SNR) regime [13]. In [14] the authors derive a non-asymptotic upper bound on the capacity of a single-user deterministic MIMO channel impaired with Wiener phase noise, which is

tight in the high-SNR regime. In [15], the authors consider the performance of Massive MIMO systems with hardware impairments. Their model is suitable for the residual hardware impairments after the application of appropriate compensation algorithms.

To the authors' knowledge, we present the first analysis of the effect of Wiener phase noise in a multi-user multi-antenna scenario with imperfect channel state information where single-carrier transmission is used. Specifically, we consider a single-cell frequency-selective MU-MIMO uplink, where a number of non-cooperative users transmit independent data streams to a base station having a large number of antennas. Since the channel is assumed to be unknown, CSI is acquired via uplink training. There are phase noise sources both at the transmitters and at the receiver. In this paper we extend the work presented in [16]. We consider and compare two distinct cases. In the first case, which is termed *synchronous operation* mode, the phase noise processes at the BS antennas are identical. In the second case, which is termed *non-synchronous operation* mode, the phase noise processes at the BS antennas are independent. These two operation modes correspond to the cases of a common phase reference versus independent phase references, respectively. A time-reversal maximum-ratio combining (TR-MRC) strategy is proposed and achievable sum-rates are derived for both operation modes.

Based on the derived expressions of the achievable sum-rates, we show that for a fixed desired per-user information rate, by doubling the number of BS antennas, the total transmit power can be reduced by a factor of $\sqrt{2}$. This is the same scaling law as without phase noise [6]. We observe that the use of independent phase noise sources can yield higher sum-rate performance and we support this interesting result by a simple toy example for which the exact capacity is calculated. Furthermore, the achievable rate expressions reveal a fundamental trade-off between the length of the time interval spent on data transmission and the sum-rate performance. The rate expressions also provide valuable insight into the optimum number of scheduled users.

II. System Model

We consider a frequency-selective MU-MIMO uplink channel with M BS antennas and K single-antenna users. The channel between the k-th user and the m-th BS antenna is modeled as a finite impulse response (FIR) filter with L symbol-spaced channel taps. The l-th channel tap is given by $g_{m,k,l} \triangleq \sqrt{d_{k,l}}h_{m,k,l}$, where $h_{m,k,l}$ and $d_{k,l}$ model the fast and slow time-varying components, respectively. We assume a block fading model where $h_{m,k,l}$ is fixed during the transmission of a block of $N_c \triangleq N_D + (K+3)L - 3$ symbols and varies independently from one block to another. N_D denotes the number of channel uses utilized for data transmission (see Fig. 1). We further assume that the channel fading process is ergodic. The parameters $d_{k,l} \geq 0$, $l = 0, \ldots, L-1$ model the power delay profile (PDP) of the frequency-selective channel for the k-th user. Since $\{d_{k,l}\}$ vary slowly with time and spatial location, we assume them to be fixed for the entire communication and independent of m. We further

assume $h_{m,k,l}$ to be independent and identically distributed (i.i.d.) zero-mean and unit-variance proper complex random variables. The i.i.d. assumption is justified in [4], [5], [17].[1] Further, the PDP for every user is normalized such that the average received power is independent of the length of the channel impulse response, L. Therefore, it holds that

$$\sum_{l=0}^{L-1} \mathbb{E}\left[|\sqrt{d_{k,l}}h_{m,k,l}|^2\right] = \sum_{l=0}^{L-1} d_{k,l} = \alpha_k, \qquad (1)$$

for $1 \leq k \leq K$. The positive constants, α_k, account for different propagation losses between users and are assumed to be fixed throughout the communication. The BS is assumed to have perfect knowledge of all the PDPs. Finally, we assume exact knowledge of the channel statistics at the BS, but not of the particular channel realizations.

A. Phase Noise Model

Phase noise is introduced at the transmitter during up-conversion, when the baseband signal is multiplied with the carrier generated by the local oscillator. The phase of the generated carrier drifts randomly, resulting in a phase distortion of the transmitted signal. A similar phenomenon also happens at the receiver side during down-conversion of the bandpass signal to baseband. In the following, θ_k, $k = 1, \ldots, K$ denotes the phase noise process at the k-th single-antenna user. Since the users have different local oscillators, the transmitter phase noise processes are assumed to be mutually independent. On the other hand, at the receiver side two distinct operation modes are considered. We term these operation modes as *synchronous* and *non-synchronous* operation depending on whether the phase noise processes at the BS antennas are identical or independent. For the synchronous case, all BS antennas are subject to the same phase noise process and ϕ denotes this common phase noise process at each BS antenna. This models the scenario of a centralized BS with a single oscillator feeding the down-conversion module in each receiver. For the case of non-synchronous operation, ϕ_m, $m = 1, \ldots, M$ denotes the phase noise process at the m-th BS antenna. This models a completely distributed scenario where each BS antenna uses a distinct oscillator for down-conversion. We further assume that the phase noise processes θ_k, $k = 1, \ldots, K$ and ϕ (or ϕ_m, $m = 1, \ldots, M$) for the case of synchronous (or non-synchronous) operation mode are mutually independent.

In this study each phase noise process is modeled as an independent Wiener process, which is a well-established model [11], [18]. Therefore, the discrete-time phase noise process at the k-th user at time i is given by[2]

$$\theta_k[i] = \theta_k[i-1] + w_k^t[i], \qquad (2)$$

---

[1]We note that with the i.i.d. assumption on the channel gains, the captured energy increases linearly with the number of BS antennas, M. This is not reasonable if M grows unbounded. However, this deficiency of the model takes effect only for exorbitantly large values of M which do not lie in the regime of our interest [5], [4], [7].

[2]The discrete-time phase noise model is used since we will be working with the discrete-time complex baseband representation of the transmit and receive signals.

where $w_k^t[i] \sim \mathcal{N}(0, \sigma_\theta^2)$ are independent identically distributed zero-mean Gaussian increments with variance $\sigma_\theta^2 \triangleq 4\pi^2 f_c^2 c_\theta T_s$, f_c is the carrier frequency, T_s is the symbol interval and c_θ is a constant that depends on the oscillator. Depending on the operation mode, the phase noise processes $\phi[i]$ and $\phi_m[i]$ at the M BS antennas are defined in a manner similar to (2), where the increments have variance $\sigma_\phi^2 \triangleq 4\pi^2 f_c^2 c_\phi T_s$.

B. Received Signal

Let $x_k[i]$ be the symbol transmitted from the k-th user at time i. The received sample at the m-th BS antenna element at time i is then given by, for the non-synchronous operation

$$y_m[i] = \sqrt{P} \sum_{k=1}^{K} \sum_{l=0}^{L-1} e^{-j\phi_m[i]} g_{m,k,l} e^{j\theta_k[i-l]} x_k[i-l] + n_m[i],$$
(3)

where $n_m[i] \sim \mathcal{CN}(0, \sigma^2)$ represents noise at the m-th receiver at time i, which is distributed as circularly symmetric complex Gaussian.[3] Each user transmits a stream of i.i.d. $\mathcal{CN}(0,1)$ information symbols (i.e., $x_k[i] \sim \mathcal{CN}(0,1)$), that are independent of the information symbols of the other users. P denotes the average uplink transmitted power from each user.

III. TRANSMISSION SCHEME AND RECEIVE PROCESSING

We consider a block-based uplink transmission scheme. A transmission block of N_c channel uses consists of KL channel uses dedicated to uplink channel training followed by a preamble of $L-1$ channel uses, where i.i.d. $\mathcal{CN}(0,1)$ non-information symbols are sent. The data interval of N_D channel uses comes after that and a postamble of $L-1$ channel uses is appended at the end of the coherence interval, where i.i.d. $\mathcal{CN}(0,1)$ non-information symbols are sent. The inclusion of the preamble and postamble accounts for the edge effects introduced due to the intersymbol interference. This way the subsequent analysis is valid for all the N_D channel uses during data transmission and no separate analysis for the edges of the data interval is required. At the beginning of each coherence interval an all-zero block of $L-1$ channel uses is prepended to eliminate inter-block interference (IBI) (see Fig. 1).

A. Channel Estimation

For coherent demodulation, the BS needs to estimate the uplink channel. This is facilitated through the transmission of uplink pilot symbols during the training phase of each transmission block.[4] The users transmit uplink training signals

---

[3]In the following we will present only the expressions of the non-synchronous mode. The expressions for the synchronous operation are obtained easily by substituting $\phi_1[i] \equiv \ldots \equiv \phi_M[i] \equiv \phi[i]$. In Sections IV–VI, when the expressions of the two distinct modes differ in a non-obvious way, both expressions will be given explicitly.

[4]In this paper we deal only with uplink transmission. In Massive MIMO Time Division Duplex (TDD) operation pilots are transmitted on the uplink. The number of required pilots scales with the number of terminals, K, but not the number of BS antennas, M, making Massive MIMO scalable with respect to M [3], [4].
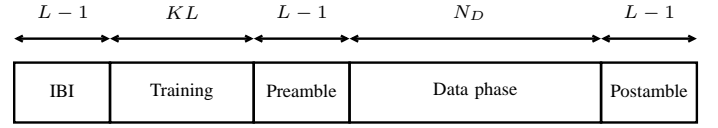

Fig. 1: The transmission block is assumed to span a coherence interval, $N_c \triangleq N_D + (K+3)L - 3$. In each block, the first KL channel uses (cu) are utilized for pilot based channel estimation and N_D cu are utilized for data transmission. An all-zero block, a preamble and a postamble of $L-1$ cu each are added due to the edge effects of the channel.

sequentially in time, i.e., at any given time only one user is transmitting uplink training signals and all other users are silent. To be precise, the k-th user sends an impulse of amplitude $\sqrt{P_p KL}$ at the $(k-1)L$-th channel use and is idle for the remaining portion of the training phase. Here, P_p is the average power transmitted by a user during the training phase. We choose the proposed training sequence since it allows for a very simple channel estimation scheme at the BS and since it facilitates our derivation of achievable rates. However, many of our results, such as partial loss of coherency due to Wiener phase noise and monotonic decrease in performance with increased variance of the phase noise increments, are expected to be qualitatively valid also for other (but not necessarily all possible) training schemes. Therefore, using (3), the signal received at the m-th BS receiver at time $(k-1)L+l$, $l = 0, \ldots, L-1$, $k = 1, \ldots, K$ is given by, for non-synchronous operation

$$y_m[(k-1)L+l] = \sqrt{P_p KL} g_{m,k,l} e^{j(\theta_k[(k-1)L] - \phi_m[(k-1)L+l])} + n_m[(k-1)L+l].$$
(4)

Based on (4), we derive the maximum likelihood (ML) estimate of the effective channel $g_{m,k,l} e^{j(\theta_k[(k-1)L] - \phi_m[(k-1)L+l])}$. The corresponding channel estimates are then given by, for non-synchronous operation

$$\hat{g}_{m,k,l} = \frac{1}{\sqrt{P_p KL}} y_m[(k-1)L+l]$$
$$= g_{m,k,l} e^{-j\phi_m[(k-1)L+l]} e^{j\theta_k[(k-1)L]}$$
$$+ \frac{1}{\sqrt{P_p KL}} n_m[(k-1)L+l].$$
(5)

We observe that the channel estimate is distorted by the AWGN and by the phase noise of the local oscillators at the user and at the BS.

B. Time-Reversal Maximum Ratio Combining (TR-MRC)

Using (3), the received signal during the data phase is given by, for non-synchronous operation

$$y_m[i] = \sqrt{P_D} \sum_{k=1}^{K} \sum_{l=0}^{L-1} e^{-j\phi_m[i]} g_{m,k,l} e^{j\theta_k[i-l]} x_k[i-l] + n_m[i],$$
(6)

where $i \in \mathcal{I}_d$, $\mathcal{I}_d \triangleq \{(K+1)L-1, \ldots, (K+1)L+N_D-2\}$ and P_D is the per-user average transmit power constraint during the data phase. Motivated by the need for low-complexity

detection, we consider the TR-MRC receiver at the BS. The TR-MRC receiver convolves the received symbols, $y_m[i]$, with the complex conjugate of the time-reversed estimated channel impulse response. The detected symbol, $\hat{x}_k[i]$, is given by

$$\hat{x}_k[i] = \sum_{l=0}^{L-1}\sum_{m=1}^{M} \hat{g}_{m,k,l}^* y_m[i+l], \qquad (7)$$

where $(\cdot)^*$ denotes the complex conjugation operation.

IV. ACHIEVABLE SUM-RATE

We use the information sum-rate as the performance metric for quantifying the effects of phase noise. To this end, using (5) and (6) for the non-synchronous operation, (7) is written as

$$\hat{x}_k[i] = A_k[i]x_k[i] + \mathtt{ISI}_k[i] + \mathtt{MUI}_k[i] + \mathtt{AN}_k[i], \quad (8)$$

where it holds for the non-synchronous operation that

$$A_k[i] \triangleq \sqrt{P_D}\sum_{m=1}^{M}\sum_{l=0}^{L-1}|g_{m,k,l}|^2 \vartheta\binom{m,k,k}{i,l,l} \qquad (9)$$

$$\mathtt{ISI}_k[i] \triangleq \sqrt{P_D}\sum_{m=1}^{M}\sum_{l=0}^{L-1}\sum_{\substack{p=0\\p\neq l}}^{L-1} g_{m,k,l}^* g_{m,k,p}\vartheta\binom{m,k,k}{i,l,p}x_k[i+l-p]$$
$$\qquad (10)$$

$$\mathtt{MUI}_k[i] \triangleq \sqrt{P_D}\sum_{m=1}^{M}\sum_{\substack{q=1\\q\neq k}}^{K}\sum_{l=0}^{L-1}\sum_{p=0}^{L-1} g_{m,k,l}^* g_{m,q,p}\times$$
$$\vartheta\binom{m,k,q}{i,l,p}x_q[i+l-p] \qquad (11)$$

$$\mathtt{AN}_k[i] \triangleq \sqrt{\frac{P_D}{P_p KL}}\sum_{m=1}^{M}\sum_{q=1}^{K}\sum_{l=0}^{L-1}\sum_{p=0}^{L-1} g_{m,q,p}\times$$
$$e^{-j(\phi_m[i+l]-\theta_q[i+l-p])}n_m[(k-1)L+l]x_q[i+l-p]$$
$$+\sum_{m=1}^{M}\sum_{l=0}^{L-1}\hat{g}_{m,k,l}^* n_m[i+l], \qquad (12)$$

where $\vartheta\binom{m,k,q}{i,l,p}\triangleq e^{j(\theta_q[i+l-p]-\theta_k[(k-1)L]-\phi_m[i+l]+\phi_m[(k-1)L+l])}$. In (8), $A_k[i]x_k[i]$ is the desired signal term for the k-th user, $\mathtt{ISI}_k[i]$ stands for the intersymbol interference for user k at time i, caused by the information symbols of the k-th user transmitted at other time instances, $\mathtt{MUI}_k[i]$ denotes the multi-user interference due to the information symbols of the other users and finally $\mathtt{AN}_k[i]$ is an aggregate noise term that incorporates the effects of the channel estimation error and the receiver AWGN noise, $n_m[i]$. The expressions for the terms in (8) for the synchronous operation are obtained from (9)-(12) by substituting $\phi_1[i]\equiv\ldots\equiv\phi_M[i]\equiv\phi[i]$.

In the following, we derive an achievable information rate for the k-th user. Similar capacity bounding techniques have been used earlier in e.g. [19], [20]. In (8), we add and subtract the term $\mathbb{E}\left[A_k[i]\right]x_k[i]$, where the expectation is taken over the channel gains, $g_{m,k,l}$, and the phase noise processes,

θ_k, ϕ for the synchronous operation and θ_k, ϕ_m for the non-synchronous operation. We relegate the variation around this term, i.e., $\mathtt{IF}_k[i] \triangleq (A_k[i] - \mathbb{E}\left[A_k[i]\right])x_k[i]$, to an effective noise term. This results in the following equivalent expression

$$\hat{x}_k[i] = \mathbb{E}\left[A_k[i]\right]x_k[i] + \mathtt{EN}_k[i], \qquad (13)$$

where

$$\mathtt{EN}_k[i] \triangleq \mathtt{IF}_k[i] + \mathtt{ISI}_k[i] + \mathtt{MUI}_k[i] + \mathtt{AN}_k[i], \qquad (14)$$

is the effective additive noise term. In (13) the detected symbol, $\hat{x}_k[i]$, is a sum of two *uncorrelated* terms (i.e., $\mathbb{E}\left[(\mathbb{E}[A_k[i]]x_k[i])(\mathtt{EN}_k[i])^*\right] = 0$). The importance of the equivalent representation in (13) is that the scaling factor $\mathbb{E}[A_k[i]]x_k[i]$ of the desired information symbol is a constant, which is known at the BS since the BS has knowledge of the channel statistics. The exact probability distribution of $\mathtt{EN}_k[i]$ is difficult to compute. However, its variance can be easily calculated given that the channel statistics is known at the BS. Therefore, (13) describes an effective single-user single-input single-output (SISO) additive noise channel, where the noise is zero mean, has known variance and is uncorrelated with the desired signal term. From the expressions for $A_k[i]$ and $\mathtt{EN}_k[i]$ in (9) and (14), the mean value of $A_k[i]$ and the variance of $\mathtt{EN}_k[i]$ is given by two propositions that follow.

Proposition 1. *The mean value of $A_k[i]$ in both operation modes is given by*

$$\mathbb{E}[A_k[i]] = \sqrt{P_D}M\alpha_k e^{-\frac{\sigma_\phi^2+\sigma_\theta^2}{2}(i-(k-1)L)}. \qquad (15)$$

Proof: We prove the statement for the non-synchronous operation. The proof for the synchronous operation is nearly identical. From (9), we have

$$\mathbb{E}[A_k[i]] = \mathbb{E}\left[\sqrt{P_D}\sum_{m=1}^{M}\sum_{l=0}^{L-1}|g_{m,k,l}|^2 \vartheta\binom{m,k,k}{i,l,l}\right]$$

$$\stackrel{(a)}{=}\sqrt{P_D}\mathbb{E}\left[e^{-j(\theta_k[(k-1)L]-\theta_k[i])}\right]\sum_{m=1}^{M}\sum_{l=0}^{L-1}\mathbb{E}\left[|g_{m,k,l}|^2\right]$$

$$\cdot\mathbb{E}\left[e^{-j(\phi_m[i+l]-\phi_m[(k-1)L+l])}\right]$$

$$\stackrel{(b)}{=}\sqrt{P_D}e^{-\frac{\sigma_\theta^2}{2}(i-(k-1)L)}\sum_{m=1}^{M}\sum_{l=0}^{L-1}d_{k,l}e^{-\frac{\sigma_\phi^2}{2}(i-(k-1)L)}$$

$$\stackrel{(c)}{=}\sqrt{P_D}M\alpha_k e^{-\frac{\sigma_\phi^2+\sigma_\theta^2}{2}(i-(k-1)L)}.$$

In (a) we have used the fact that the channel realizations, $g_{m,k,l}$, the phase noise at the BS, ϕ_m, and the phase noise at the k-th user, θ_k, are mutually independent random processes. The equality (b) is a consequence of the Wiener phase noise model. That is, after a time interval, $\Delta t = i - (k-1)L$, the phase drift of an oscillator is a zero mean Gaussian random variable with variance that is proportional to Δt,

$$U_{\phi_m} \triangleq \phi_m[i+l]-\phi_m[(k-1)L+l]\sim\mathcal{N}(0,\sigma_\phi^2(i-(k-1)L)),$$
$$U_{\theta_k} \triangleq \theta_k[i]-\theta_k[(k-1)L]\sim\mathcal{N}(0,\sigma_\theta^2(i-(k-1)L)).$$

Henceforth $\mathbb{E}\left[e^{-jU_{\phi_m}}\right] = \varphi_{\phi_m}(-1) = e^{-\frac{\sigma_\phi^2}{2}(i-(k-1)L)}$ and

$\mathbb{E}\left[e^{jU_{\theta_k}}\right] = \varphi_{\theta_k}(1) = e^{-\frac{\sigma_\theta^2}{2}(i-(k-1)L)}$, where $\varphi_{\phi_m}(\cdot)$ and $\varphi_{\theta_k}(\cdot)$ are the characteristic functions of U_{ϕ_m} and U_{θ_k}, respectively. The equality (c) follows from (1). ∎

In (15), the factor M signifies the combining gain in a coherent receiver (i.e., when $\sigma_\phi = \sigma_\theta = 0$). The factor $e^{-\frac{\sigma_\phi^2+\sigma_\theta^2}{2}(i-(k-1)L)}$ signifies the loss in effective amplitude gain due to the non-coherency between the received data samples and the estimated channel gains. Note that this non-coherency arises due to the fact that the channel gains for the k-th user are estimated at $t = (k-1)L + l$, $l = 0, \ldots, L-1$ and the samples for detecting $x_k[i]$ are received at $t = i + l$, $l = 0, \ldots, L-1$, that is, $i - (k-1)L$ samples later. The oscillator phase drift in this time period results in a partial non-coherency. It is clear that the larger this time difference is the smaller the effective amplitude gain is (the effective amplitude $M\alpha_k e^{-\frac{\sigma_\phi^2+\sigma_\theta^2}{2}(i-(k-1)L)}$ decreases exponentially with increasing time difference $i - (k-1)L$).

Proposition 2. *The variance* $\mathtt{Var}(\mathtt{EN}_k[i]) \triangleq \mathbb{E}\left[|\mathtt{EN}_k[i] - \mathbb{E}[\mathtt{EN}_k[i]]|^2\right]$ *satisfies, for synchronous operation*

$$\varsigma_k^s[i] \triangleq \mathtt{Var}(\mathtt{EN}_k^s[i]) = P_D M^2 \kappa_k[i] + C_k, \qquad (16)$$

and for non-synchronous operation

$$\varsigma_k^{ns}[i] \triangleq \mathtt{Var}(\mathtt{EN}_k^{ns}[i]) = P_D M^2 \alpha_k^2 \varpi_k[i] + P_D M \xi_k[i] + C_k, \qquad (17)$$

where $\kappa_k[i] \triangleq \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|} - \alpha_k^2 e^{-(\sigma_\phi^2+\sigma_\theta^2)(i-(k-1)L)}$,
$\xi_k[i] \triangleq \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|} - \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}$,
$\varpi_k[i] \triangleq e^{-\sigma_\phi^2(i-(k-1)L)}\left(1 - e^{-\sigma_\theta^2(i-(k-1)L)}\right)$,
$C_k \triangleq P_D M \alpha_k \sum_{q=1}^K \alpha_q + \sigma^2 M \left(\frac{P_D}{P_p K}\sum_{q=1}^K \alpha_q + \alpha_k + \frac{\sigma^2}{KP_p}\right)$.

Proof: See the Appendix. ∎

The second term of the constant C_k in Proposition 2 is the contribution of the additive noise term $\mathtt{AN}_k[i]$. This contribution has variance $\mathbb{E}\left[|\mathtt{AN}_k[i]|^2\right] = \sigma^2 M\left(\frac{P_D}{P_p K}\sum_{q=1}^K \alpha_q + \alpha_k + \frac{\sigma^2}{KP_p}\right)$. The term $\sigma^2 M\frac{P_D}{P_p K}\sum_{q=1}^K \alpha_q$ corresponds to the cross-correlation between the channel estimation error in (5) and the received symbols in (6). The term $\sigma^2 M \alpha_k$ corresponds to the filtered noise (7). Finally, the last term $\sigma^2 M \frac{\sigma^2}{KP_p}$ corresponds to the variance of the channel estimation error.

In the following we provide a coding strategy that justifies the achievable rates we are interested in deriving. From Propositions 1 and 2, it is obvious that $\mathbb{E}[A_k[i]]$ and $\mathtt{Var}(\mathtt{EN}_k[i])$ depend on i and are different for different $i \in \mathcal{I}_d$. Further, for a given i, across multiple transmission blocks, the terms $\mathbb{E}[A_k[i]]$ and $\mathtt{Var}(\mathtt{EN}_k[i])$ are the same and the realizations of $\mathtt{EN}_k[i]$ are i.i.d. Hence, for each i, we have an additive noise SISO channel. This motivates us to consider N_D channel codes for each user, one for each $i \in \mathcal{I}_d$. At the k-th transmitter (user), the symbols of the i-th channel code ($x_k[i]$) are transmitted only during the i-th channel use of each transmission block. Similarly, at the BS, for the k-th user, the i-th received and processed symbols (i.e., $\hat{x}_k[i]$) across

different transmission blocks are jointly decoded. Essentially, this implies that, at the BS we have N_D parallel channel decoders for each user. We propose the above scheme of N_D parallel channel codes for each user only to derive a lower bound on the achievable information rate. In practice, due to reasons of complexity, channel coding/decoding would not only be performed across different transmission blocks, but also across consecutive channel uses within each transmission block.

Given the previously described coding strategy, we are now interested in computing a lower bound on the reliable rate of communication for each of the N_D channel codes. Since the data symbols $x_k[i]$ are Gaussian, for each $i \in \mathcal{I}_d$ a lower bound on the information rate for the effective channel in (13) can be computed by considering the worst case (in terms of mutual information) uncorrelated additive noise. With Gaussian information symbols, it is known that the worst case uncorrelated noise is Gaussian with the same variance as that of $\mathtt{EN}_k[i]$ [19]. Consequently, a lower bound on $I(\hat{x}_k[i]; x_k[i])$ (i.e., the mutual information rate for the i-th channel code for user k) is given by Proposition 3.

Proposition 3. *The achievable rate for the i-th channel code for the k-th user is given by*

$$I(\hat{x}_k[i]; x_k[i]) \geq R_k^\times[i]$$
$$\triangleq \log_2\left(1 + \frac{P_D M^2 \alpha_k^2 e^{-(\sigma_\phi^2+\sigma_\theta^2)(i-(k-1)L)}}{\varsigma_k^\times[i]}\right), \qquad (18)$$

where $\times = s$ for synchronous operation and $\times = ns$ for non-synchronous operation and $\varsigma_k^\times$ are given in Proposition 2.

Corollary 1. *Based on the lower bounds (18), the proposed TR-MRC receiver exhibits better performance in the case of non-synchronous operation.*

Proof of Corollary 1:
$$\varsigma_k^s[i] - \varsigma_k^{ns}[i] = P_D M(M-1)\xi_k[i]$$
$$\overset{(a)}{\geq} P_D M(M-1)\left(\alpha_k^2 e^{-\sigma_\phi^2(L-1)} - \alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}\right)$$
$$\overset{(b)}{\geq} P_D M(M-1)\left(\alpha_k^2 e^{-\sigma_\phi^2(L-1)} - \alpha_k^2 e^{-\sigma_\phi^2(2L-1)}\right) \geq 0.$$

The inequality (a) follows from the fact that $|l-l'| \leq L-1$ and (1). The inequality (b) follows since $i \geq KL + L - 1 \Rightarrow i - (k-1)L \geq (K-k)L + 2L - 1 \geq 2L - 1$ and $k \leq K$. ∎

Note that Corollary 1 compares two lower bounds. However, there are good reasons to expect that these lower bounds are actually quite good predictions of the performance that could be achieved in reality. This is so because substantially we make a Gaussianity assumption on the effective noise. This is also very likely the type of approximation that would be used when deriving a soft decoding (LLR) metric for insertion into for example, a turbo decoder. Hence, using this Gaussian approximation would predict quite well the performance achievable with good channel codes and standard decoding metrics assuming Gaussian noise. Also note that comparing lower bounds that are reasonably tight is a standard
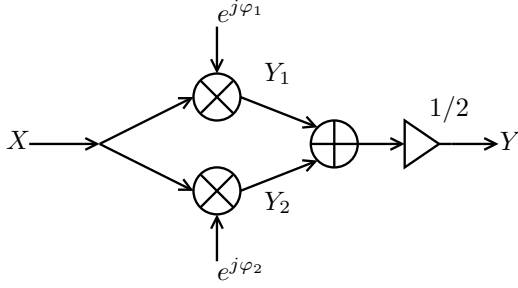
Fig. 2: System model for the example.

practice in the communication theory literature.

Corollary 1 conveys an interesting result that the performance is better when the phase noise processes at the different BS antennas are uncorrelated. However, this is not the first time that such a result is reported. In [21, Section III.A] the authors study the effect of phase noise in single-user beamforming. The performance measure they use is the error vector magnitude (EVM) and they show that EVM is smallest in the desired direction when uncorrelated phase noise sources are used. In [15, Section VI.D] the authors consider the impact of phase noise distortion in a flat fading channel with maximum ratio combining, using a small phase noise approximation. They also observe that by using separate oscillators the distortion scales as $O(t)$, where t is the time elapsed from channel estimation to data detection. On the other hand, when a common oscillator is used the distortion scales as $O(tM)$. (Note that in contrast to our analysis, [15] used a much simpler model that did not include the effects of intersymbol interference, nor of multiuser interference.) From Corollary 1, it can be argued that the use of independent oscillators at the BS can be beneficial when TR-MRC is used. Also, for a desired sum-rate performance one can choose between a high quality single oscillator or many oscillators of lower quality.

1) Achievable Sum-Rate: Since no data transmission happens during the training phase, the overall effective information rate achievable by the k-th user is given by,

$$R_k^\times \triangleq \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} R_k^\times[i]. \tag{19}$$

The achievable sum-rate is therefore given by

$$R^\times = \sum_{k=1}^{K} R_k^\times = \frac{1}{N_c} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_d} R_k^\times[i]. \tag{20}$$

It is clear that phase noise degrades the sum-rate performance both with synchronous and non-synchronous operation. To see this formally, note that the sum-rate for the no-phase-noise case can be derived from (18), (19) and (20) by setting $\sigma_\phi^2 = \sigma_\theta^2 = 0$ and is given by

$$\mathcal{R} = \frac{N_D}{N_c} \sum_{k=1}^{K} \log_2 \left(1 + \frac{P_D M^2 \alpha_k^2}{C_k} \right). \tag{21}$$

Since, $\frac{P_D}{\sigma^2} M^2 \alpha_k^2 \geq \frac{P_D}{\sigma^2} M^2 \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}$ and $\varsigma_k^s[i] \geq \varsigma_k^{ns}[i] \geq C_k$ we have that $\mathcal{R} \geq R^\times$.

A. Exact Analysis of Synchronous versus Non-Synchronous Operation for a Toy Channel Model

In the following, we provide a simple example to illustrate that the conclusion drawn from Corollary 1 is the result of a fundamental phenomenon and not an artifact of the techniques used to derive the lower bounds on the information rate. We consider a very simple channel with only phase noise and no AWGN, see Fig. 2. Here $X \in \{\pm 1\}$, $\Pr\{X = +1\} = p$, $\Pr\{X = -1\} = 1 - p$ is the input to the channel. The input X is rotated by φ_1 and φ_2 to form Y_1 and Y_2, respectively. Let the random variables φ_1, φ_2 model the phase noise, with the following probability mass functions (p.m.f.): $\varphi_i \in \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$, $\Pr\{\varphi_i = -\frac{\pi}{2}\} = \Pr\{\varphi_i = 0\} = \Pr\{\varphi_i = \frac{\pi}{2}\} = \frac{1}{3}$, $i = 1, 2$. The output of this discrete memoryless channel (DMC) is given by

$$Y = \frac{1}{2} \left(e^{j\varphi_1} + e^{j\varphi_2} \right) X. \tag{22}$$

We now consider two cases, firstly when the two phase noise processes are synchronous (i.e., $\varphi_1 \equiv \varphi_2$) and secondly when they are non-synchronous and mutually independent. In the synchronous case, $\varphi_1 \equiv \varphi_2$ so $Y = e^{j\varphi_1} X$. Then Y takes values in $\mathcal{Y}_s = \{+1, +j, -1, -j\}$. The output symbols have the p.m.f.: $\Pr\{Y = +1\} = p/3$, $\Pr\{Y = -1\} = (1-p)/3$, $\Pr\{Y = \pm j\} = 1/3$. The capacity of this channel can be calculated as follows

$$C_s = \max_p I(X;Y) = \max_p H(Y) - H(Y|X)$$
$$= \max_p \frac{1}{3} H_2(p) = 1/3 \text{ bits},$$

where $H_2(p)$ is the binary entropy function.

In the non-synchronous case, where φ_1 and φ_2 are independent of each other, the output variable takes values in $\mathcal{Y}_{ns} = \{+1, \frac{1}{2}(1+j), \frac{1}{2}(1-j), j, 0, -j, -\frac{1}{2}(1-j), -\frac{1}{2}(1+j), -1\}$. The p.m.f. of the output is $\Pr\{Y = +1\} = p/9$, $\Pr\{Y = (1 \pm j)/2\} = 2p/9$, $\Pr\{Y = \pm j\} = 1/9$, $\Pr\{Y = 0\} = 2/9$, $\Pr\{Y = -(1 \pm j)/2\} = 2(1-p)/9$, and $\Pr\{Y = -1\} = (1-p)/9$. We find that $H(Y) = \frac{5}{9} H_2(p) + \log_2 9 - 6/9$ and $H(Y|X = \pm 1) = \log_2 9 - 6/9$. Then, the capacity is given by

$$C_{ns} = \max_p I(X;Y) = \max_p H(Y) - H(Y|X)$$
$$= \max_p \frac{5}{9} H_2(p) = 5/9 \text{ bits}.$$

Since $C_s < C_{ns}$, it is concluded that the capacity of the channel in Fig. 2 is strictly larger in the non-synchronous case than in the synchronous case.

Note that the example does not show that the capacity *always* increases if we use independent phase noise sources. However, it shows that *there are* cases where the use of independent phase noise sources can be beneficial.

V. ASYMPTOTIC RESULTS

The achievable rates presented in Proposition 3 hold for any M. In this section we present some asymptotic (in M)

results based on these achievable rates in order to investigate the Massive MIMO effect in the system under study. In the following $\beta \triangleq \frac{P_p}{P_D} > 0$ denotes the ratio between the per-user average transmit power during the training phase and during the transmission phase.

We first note that in the low SNR regime, the performance loss due to phase noise is not significant. To see this quantitatively, consider the sum-rate when phase noise is present, given by (20). From (18) it is clear that in the low-SNR regime, i.e., when $P_D/\sigma^2 \ll 1$, the dominating factor in the denominator of the argument of the $\log_2$ function is, in both operation modes, the term $\frac{\sigma^4 M}{K\beta P_D}$. From (21) (after the substitution $P_p = \beta P_D$) it is clear that the term $\frac{\sigma^4 M}{K\beta P_D}$ is also the dominating term in the denominator of the achievable rate expression in the no-phase-noise case. Therefore, the performance loss of both operation modes compared to the no-phase-noise scenario is small. The result is of particular importance since this work focuses mainly on the low SNR (per degree of freedom). This is also often the foreseen operating point of Massive MIMO [5], [22].

We proceed with a result on the sum-rate performance in the high-SNR regime.

Proposition 4. Saturation in the high-SNR regime. *In the presence of phase noise the effective information rate of the k-th user saturates for $\frac{P_D}{\sigma^2} \to \infty$ to the values, for synchronous operation*

$$R_k^s \to \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{M\alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{M\kappa_k[i] + \alpha_k \sum_{q=1}^K \alpha_q} \right),$$
(23)

and for non-synchronous operation

$$R_k^{ns} \to \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{M\alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{M\alpha_k^2 \varpi_k[i] + \xi_k[i] + \alpha_k \sum_{q=1}^K \alpha_q} \right).$$
(24)

Proof: The result follows from (18) and the definitions of R_k^s and R_k^{ns} in (19). ∎

In the high-SNR regime, MRC is known to be suboptimal since intersymbol interference and multi-user interference dominate the effective noise term. Therefore saturation in the high-SNR regime is observed also in the no-phase-noise case due to the MRC reception strategy.

A particularly desirable property of massive MIMO systems is the array power gain that they offer. The following proposition shows that the phase-noise-impaired single-carrier massive MIMO uplink with TR-MRC receive processing and estimated CSI offers an array gain of $O(\sqrt{M})$—the same scaling law as for flat fading channels without phase noise, derived in [6].

Proposition 5. *Under the assumptions made in Section III, an $O(\sqrt{M})$ array gain is achievable.*

Proof: We start by proving the proposition for the synchronous case. Let $P_D = \frac{E_u}{M^\eta}$, where E_u is fixed. Based on the derived achievable rates in Proposition 3, we compute the maximum possible exponent, $\eta > 0$, such that a fixed,

non-zero rate for the i-th code of user k can be achieved, while the transmit power of each user is scaled as $1/M^\eta$ with increasing M. From argument of the log expression in (18), i.e. the effective SINR, we have

$$\text{SINR}_k[i] = \frac{\frac{E_u M \alpha_k^2}{\sigma^2 M^\eta} e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u M \kappa_k[i]}{\sigma^2 M^\eta} + \frac{E_u \alpha_k \sum_q \alpha_q}{\sigma^2 M^\eta} + \alpha_k + \frac{\sum_q \alpha_q}{\beta K} + \frac{M^\eta \sigma^2}{K\beta E_u}}$$

$$= \frac{\frac{E_u \alpha_k^2}{\sigma^2} e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u \kappa_k[i]}{\sigma^2} + \frac{E_u \alpha_k \sum_q \alpha_q}{M\sigma^2} + M^{\eta-1}\left(\alpha_k + \frac{\sum_q \alpha_q}{\beta K}\right) + \frac{M^{2\eta-1}\sigma^2}{K\beta E_u}}.$$

As $M \to \infty$ we have $\lim_{M\to\infty} R_k^s[i] > 0$ if $\eta - 1 \le 0$ and $2\eta - 1 \le 0 \Rightarrow \eta \le 1/2$. For $\eta = 1/2$ the rate R_k^s converges to the value (as $M \to \infty$)

$$R_k^s \to \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u}{\sigma^2} \kappa_k[i] + \frac{\sigma^2}{K\beta E_u}} \right).$$
(25)

Similarly, it can be proved that the array gain for the non-synchronous operation is $O(\sqrt{M})$ and the rate approaches (as $M \to \infty$) the value

$$R_k^{ns} \to \frac{1}{N_c} \sum_{i \in \mathcal{I}_d} \log_2 \left(1 + \frac{\frac{E_u}{\sigma^2} \alpha_k^2 e^{-(\sigma_\phi^2 + \sigma_\theta^2)(i-(k-1)L)}}{\frac{E_u}{\sigma^2} \alpha_k^2 \varpi_k[i] + \frac{\sigma^2}{K\beta E_u}} \right).$$
(26)

It is clear that for $\eta > 1/2$ the achievable rates approach 0 as $M \to \infty$. ∎

VI. IMPACT OF PHASE NOISE SEPARATELY AT THE BS AND AT THE USER TERMINALS

Based on the preceding analysis, we examine two special cases of particular interest. Namely, we study the impact on sum-rate performance, when there is phase noise only at the user terminals (UTs) and not at the BS (i.e. $\sigma_\phi^2 = 0$ and $\sigma_\theta^2 \ne 0$) and vice versa (i.e. $\sigma_\phi^2 \ne 0$ and $\sigma_\theta^2 = 0$).

A. Special Case 1: Phase Noise Only at the UTs, $\sigma_\phi^2 = 0$

If the oscillators at the BS are ideal, there is no distinction between synchronous and non-synchronous operation. From (18) it follows immediately that the lower bound in this case is given by

$$R_k[i] = \log_2 \left(1 + \frac{\frac{P_D M \alpha_k^2}{\sigma^2} e^{-\sigma_\theta^2(i-(k-1)L)}}{\frac{P_D M}{\sigma^2} \alpha_k^2 \left(1 - e^{-\sigma_\theta^2(i-(k-1)L)}\right) + \frac{C_k}{\sigma^2 M}} \right).$$
(27)

In the high SNR limit the rate saturates at the value

$$R_k[i] \to \log_2 \left(1 + \frac{M\alpha_k e^{-\sigma_\theta^2(i-(k-1)L)}}{M\alpha_k \left(1 - e^{-\sigma_\theta^2(i-(k-1)L)}\right) + \sum_{q=1}^K \alpha_q} \right).$$
(28)

Further, by scaling the transmit power as $P_D = E_u/\sqrt{M}$ we have the limiting expression as $M \to \infty$

$$R_k[i] \to \log_2\left(1 + \frac{\frac{E_u}{\sigma^2}\alpha_k^2 e^{-\sigma_\theta^2(i-(k-1)L)}}{\frac{E_u}{\sigma^2}\alpha_k^2\left(1 - e^{-\sigma_\theta^2(i-(k-1)L)}\right) + \frac{\sigma^2}{K\beta E_u}}\right). \tag{29}$$

The expressions in (27), (28) and (29) are qualitatively similar to the case of synchronous operation at the BS. In the following we provide an intuitive explanation of this similarity. Consider the link between user k and the BS. Irrespectively of whether there is phase noise at the BS or not, the distortion in the received signal at each BS antenna due to the phase noise at the user adds up after TR-MRC processing, giving an additional interference term (see $\mathtt{IF}_k[i]$ in (14)) with a standard deviation that scales as $O(M)$.

B. Special Case 2: Phase Noise Only at the BS, ($\sigma_\phi^2 \neq 0$ and $\sigma_\theta^2 = 0$)

In this case the achievable rate for the synchronous case is given by

$$R_k^s[i] = \log_2\left(1 + \frac{\frac{P_D M \alpha_k^2}{\sigma^2} e^{-\sigma_\phi^2(i-(k-1)L)}}{\frac{P_D M}{\sigma^2}\xi_k[i] + \frac{C_k}{\sigma^2 M}}\right), \tag{30}$$

and for the non-synchronous case

$$R_k^{ns}[i] = \log_2\left(1 + \frac{\frac{P_D M \alpha_k^2}{\sigma^2} e^{-\sigma_\phi^2(i-(k-1)L)}}{\frac{P_D}{\sigma^2}\xi_k[i] + \frac{C_k}{\sigma^2 M}}\right). \tag{31}$$

In the high SNR regime the above rates saturate at the following values

$$R_k^s[i] \to \log_2\left(1 + \frac{M\alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{M\xi_k[i] + \alpha_k \sum_{q=1}^K \alpha_q}\right) \tag{32}$$

$$R_k^{ns}[i] \to \log_2\left(1 + \frac{M\alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{\xi_k[i] + \alpha_k \sum_{q=1}^K \alpha_q}\right). \tag{33}$$

Further, by scaling the transmit power as $P_D = E_u/\sqrt{M}$ we have the limiting expressions as $M \to \infty$ for the synchronous operation

$$R_k^s[i] \to \log_2\left(1 + \frac{\frac{E_u}{\sigma^2}\alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}}{\frac{E_u}{\sigma^2}\xi_k[i] + \frac{\sigma^2}{K\beta E_u}}\right), \tag{34}$$

and for the non-synchronous operation

$$R_k^{ns}[i] \to \log_2\left(1 + \left(\frac{E_u}{\sigma^2}\right)^2 K\beta\alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}\right). \tag{35}$$

The expressions in (30), (32) and (34) are qualitatively similar to the case of phase noise only at the user terminals and to the general case with synchronous operation at the BS. In fact, it is the symmetric case as in Section VI-A. This behavior can be explained by arguments similar to the ones used there.

However, in the expressions for the non-synchronous operation (31), (33) and (35) we observe a fundamentally different behavior. Firstly, in (33) we note that by increasing the number of BS antennas, we can increase the high-SNR saturation

value of the achievable rate arbitrarily. In addition, from (35) it is clear that in the large array regime we can arbitrarily increase the limiting expression by appropriately selecting the value E_u. These observations lead to the conclusion that the distortions introduced by independent oscillators at the BS asymptotically vanish, when TR-MRC reception is used. We remark that similar behavior was also noted in [15], where the authors demonstrate that the dominating impairment is the one at the hardware of the user equipment, while impairments at the BS from independent sources asymptotically vanish as $M \to \infty$.

VII. NUMERICAL EXAMPLES

In this section, we present numerical examples of the main results presented in Sections IV–VI. Throughout the section we selected $T_s = 0.1\mu s$ and $f_c = 2$ GHz, which correspond to typical values of wideband wireless communication systems, such as the WLAN IEEE 802.11. The reference value of the oscillator parameter c_ϕ (and c_θ) is set to $c_\phi = 4.7 \times 10^{-18}(\text{rad Hz})^{-1}$, which also corresponds to a typical oscillator in WLAN IEEE 802.11 equipment [23, Table 1]. However, we will refer to the standard deviation of the phase noise innovations, i.e. σ_ϕ and σ_θ, since this is a more intuitive measure of the oscillator quality. For the parameters selected above and the relations in Section II-A, $\sigma_\phi = 0.49^o$. In typical cellular systems the delay spread is of the order of microseconds. We select $L = 20$, which corresponds to $2\mu s$ of delay spread for the selected symbol rate. We selected the large scale fading factors as $\alpha_k = 1, \ \forall k \in \{1, ..., K\}$, since the main purpose of this work is to understand the effect of phase noise and not of large scale fading. However, the same relations can be used with other choices of α_k's, when the study of particular propagation conditions is of interest. Further, we have selected a common power delay profile of every user as $d_{k,l} = e^{-0.35l}/\sum_{p=0}^{L-1} e^{-0.35p}, \ l = \{0, ..., L-1\}$. We note that the power delay profile enters the rate expressions through the terms $\kappa_k[i]$ and $\xi_k[i]$ (see Proposition 2). For most reasonable choices of σ_ϕ the choice of a particular PDP has a negligible effect on the achievable sum-rate. This choice of PDP and large scale fading is the same for all the figures that follow.

In Fig. 3 the sum-rate performance of the system, as given by (20), is plotted as a function of $\frac{P_D}{\sigma^2}$ for $N_D = 1000$ with $M = 200$, $K = 10$. The sum-rate achieved without phase noise (21) is plotted for the sake of comparison. We observe that at low SNR, the loss in sum-rate performance is insignificant. This observation supports our argument on the low SNR performance at the beginning of Section V. We plot the sum-rate as a function of $\frac{P_D}{\sigma^2}$ for various choices of σ_ϕ and σ_θ. It is clear from Fig. 3b that when the phase noise at the user terminals is dominant both operation modes have similar performance. On the other hand, when the phase noise at the BS is dominant, as in Fig. 3c, the sum-rate of the non-synchronous operation is significantly higher than the synchronous operation mode. This is in agreement with the discussion in Section VI.

A significant desirable property of massive MIMO systems is the array power gain that they offer, facilitating the design
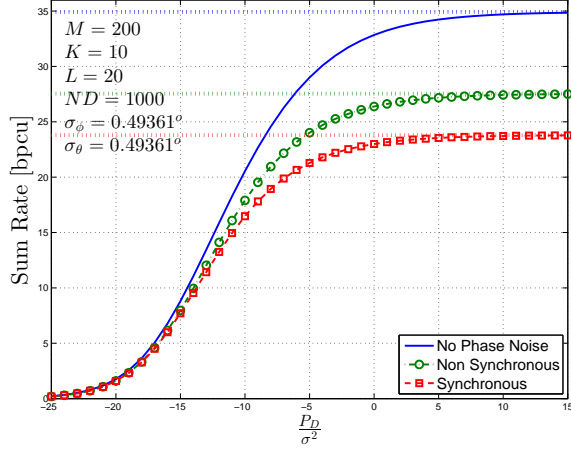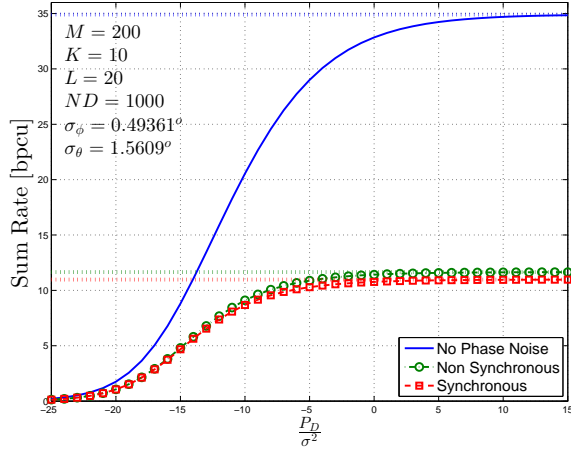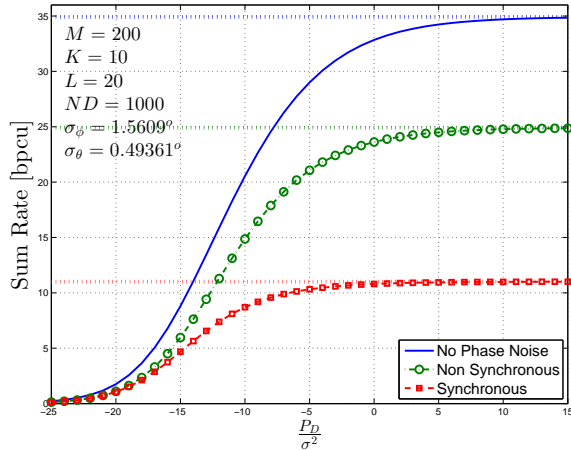
(a) $\sigma_\phi = \sigma_\theta = 0.49^o$

(b) $\sigma_\phi = 0.49^o$, $\sigma_\theta = 1.56^o$

(c) $\sigma_\phi = 1.56^o$, $\sigma_\theta = 0.49^o$

Fig. 3: Sum-rate as a function of $\frac{P_D}{\sigma^2}$ [dB] for $M = 200$, $K = 10$, $L = 20$ and $N_D = 1000$. The dotted vertical lines denote the high SNR asymptotic values of the achievable sum-rates.
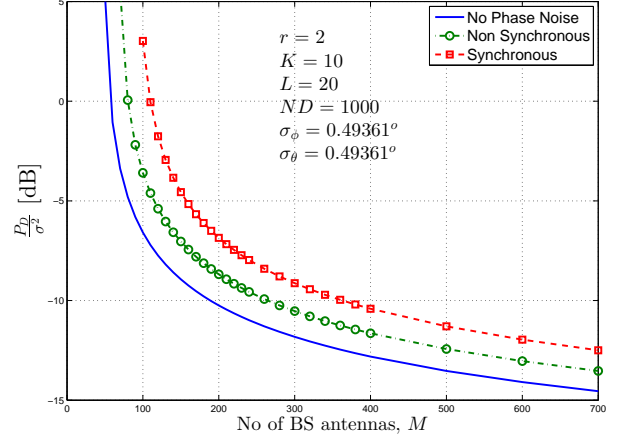

Fig. 4: Minimum required $\frac{P_D}{\sigma^2}$ to achieve a fixed per-user information rate of $r = 2$ bpcu as a function of M for fixed $K = 10$ users, $\sigma_\phi = \sigma_\theta = 0.49^o$ and $N_D = 1000$.

of highly power-efficient communication systems [3], [6], [24]. Proposition 5 extends this result to the case of single-carrier frequency-selective Massive MU-MIMO systems impaired with phase noise. The above observation is further supported through Fig. 4, where the minimum per-user $\frac{P_D}{\sigma^2}$ required to achieve a fixed per-user information rate of $r = 2$ bpcu is plotted as a function of the number of BS antennas for $N_D = 1000$ and $K = 10$ for $\sigma_\phi = \sigma_\theta = 0.49^o$. The plot for the phase-noise-free case is also given for the sake of comparison. We observe that by doubling the number of BS antennas we can reduce the per-user required $\frac{P_D}{\sigma^2}$ by 1.5dB, for sufficiently large M. This illustrates the validity of Proposition 5.

From Fig. 4 we are motivated to study the gap in required $\frac{P_D}{\sigma^2}$ between the phase-noise-impaired cases and the no-phase-noise operation. In Table I we present numerical results on this gap. Each row corresponds to a different oscillator constant $c_\phi = c_\theta$, namely, 9.4×10^{-19}, 4.7×10^{-18} and $2.35 \times 10^{-17}(\text{rad Hz})^{-1}$, which correspond to standard deviation of phase noise innovations of 0.22^o, 0.49^o and 1.1^o, respectively. In order to give a more intuitive measure of the disturbance introduced by phase noise, we list the vertical $\frac{P_D}{\sigma^2}$ gap as a function of the standard deviation of the accumulated phase noise drift at a time difference of $N_D + L - 1$ channel uses (i.e., the time difference between the end of the training phase and the end of the data phase). This result is shown in Table I. As expected, the performance gap is minimal for small phase noise drift and increases as the standard deviation of the phase noise drift increases.

It is also interesting to study the gap in required $\frac{P_D}{\sigma^2}$ as a function of the desired per-user information rate. For this purpose we provide Table II. There, we tabulate the gap in required $\frac{P_D}{\sigma^2}$ in dB for various values of the per-user desired information rate for the synchronous and non-synchronous mode, for $N_D = 1000$ channel uses, $\sigma_\phi = \sigma_\theta = 0.49^o$, $K = 10$ users and $M = 500$ BS antennas. In the low spectral efficiency regime this gap is minimal. However, as the desired

TABLE I: Gap in required $\frac{P_D}{\sigma^2}$ due to phase noise for $N_D = 1000$ and a fixed per-user information rate $r = 1$ bpcu. The number of users is fixed to $K = 10$.

| $\sigma_\phi \sqrt{N_D}$ | Gap in required $\frac{P_D}{\sigma^2}$ [dB] | | | |
| (degrees) | Synchronous | | Non-Synchronous | |
	M=500	M=2500	M=500	M=2500
7.05°	0.1174	0.1055	0.0828	0.0744
15.76°	0.6145	0.5492	0.4192	0.3753
35.23°	4.7459	3.9629	2.3071	2.0116

per-user information rate increases the gap increases at a faster rate. When the desired per-user information rate increases from 2 bpcu to 2.5 bpcu, which corresponds to 25% increase, the gap in dB in the case of non-synchronous operation doubles, whereas in the synchronous operation mode the vertical gap increases more than two times. This happens because the desired per-user rate is close to the high-SNR saturation rate for the case of synchronous receivers[5]. As a result, a large increase in the transmit power is required in order to achieve the desired information rate.

TABLE II: Gap in required $\frac{P_D}{\sigma^2}$ due to phase noise for $N_D = 1000$, $\sigma_\phi = \sigma_\theta = 0.49^o$, $K = 10$ users and $M = 500$ BS antennas for various values of the desired per-user information rate in bits per channel use [bpcu].

| | Gap in required $\frac{P_D}{\sigma^2}$ [dB] | |
Per-user rate	Synchronous	Non-Synchronous
0.25	0.2768	0.2481
0.5	0.3625	0.2941
1	0.6145	0.4192
2	2.2356	1.0987
2.5	6.8694	2.1749

For fixed M, K and L there is a fundamental trade-off between the length of the data interval, N_D, and the achievable sum-rate performance. A fraction $\frac{KL}{N_c}$ of each coherence interval is spent on training. Since a fixed time interval of KL channel uses is required for channel estimation, a small data interval, N_D, leads to underutilization of the available resources, yielding a low sum-rate performance. As N_D increases, more resources are utilized for the data transmission, increasing the sum-rate performance. However, as it can be seen from (18), $R_k^s[i] < R_k^s[i-1]$ and $R_k^{ns}[i] < R_k^{ns}[i-1]$, which implies that the gain of increasing the data interval diminishes with increasing N_D. In fact, the individual rates $R_k^s[i]$ and $R_k^{ns}[i]$ approach 0 as $i \to \infty$. This phenomenon occurs because with large N_D, the phase noise drift in the oscillators is so large such that there is a total loss of coherency between the received symbols during the data phase and the estimated channel at the beginning of the transmission block.

[5]With the selected parameters, the high-SNR saturation value for the synchronous operation is 2.66 bpcu per user.
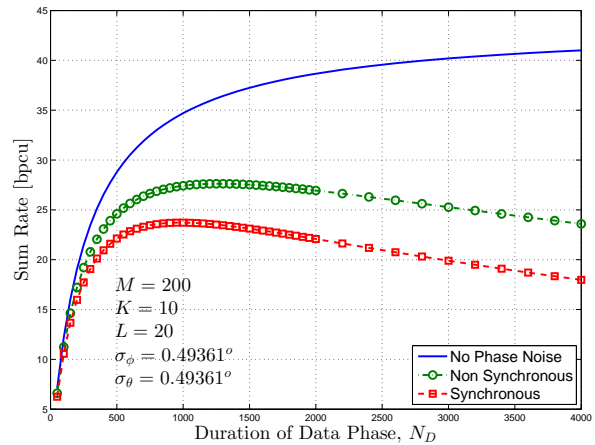

Fig. 5: Sum-rate performance as a function of N_D, with fixed, $\sigma_\phi = \sigma_\theta = 0.49^o$, $\frac{P_D}{\sigma^2} = 10$ dB, $M = 200$ BS antennas, $K = 10$ users and $L = 20$ taps.

In Fig. 5 the sum-rate performance is plotted as a function of N_D for $\sigma_\phi = \sigma_\theta = 0.49^o$. In the no-phase-noise case the optimal value of N_D is infinity. However, there is a clear trade-off between the sum-rate and the length of the data interval in the phase-noise-impaired operation modes.

Further insight can be obtained by considering the optimum number of scheduled users. In practice, the coherence interval is finite and therefore the training overhead upper-bounds the optimum number of scheduled users. Now, consider the case where the coherence interval is arbitrarily long. Then for the no-phase noise case, the optimal N_D is unbounded. In that case one can increase the number of users, thereby achieving an increase in the sum-rate performance due to the spatial multiplexing of more users in the same time-frequency resource. In the presence of phase noise increasing the number of scheduled users, K, not only increases the length of the training overhead, but it also increases the phase drift between the estimated channel coefficients and the actual realizations of the effective channel impulse responses during the data interval. That is, by increasing the number of users, K, the partial loss of coherency between the estimated channel coefficients and the actual effective channels during data transmission is also increased. As a result, with increasing K the increase in the achievable sum-rate during the data interval may eventually become insignificant to compensate for the reduction in sum-rate due to this partial loss of coherency. In Fig. 6, for every K the maximum achievable sum-rate performance is found by maximizing with respect to N_D and, subsequently, this maximum sum-rate performance is plotted as a function of K for $\frac{P_D}{\sigma^2} = 10$ dB, $M = 200$ BS antennas and $L = 20$ taps for the no phase noise case, the synchronous operation mode and the non-synchronous operation mode. It is clear that the sum-rate performance is not monotonically increasing in the phase-noise-impaired cases as it is in the no phase noise case. However, it has a unimodal shape. This implies that in practice the optimum number of scheduled users is not only upper-bounded by the length of the coherence interval, but it
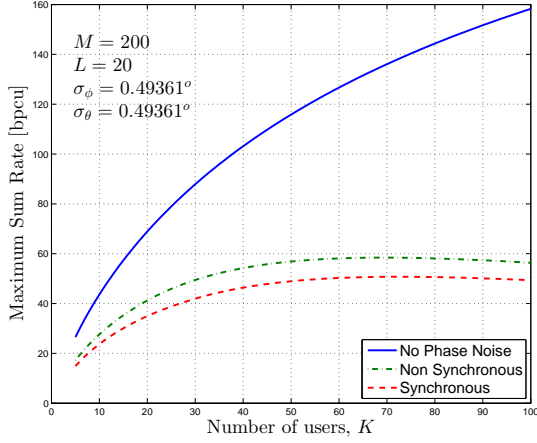
Fig. 6: Maximum sum-rate performance as a function of K, with fixed $\frac{P_D}{\sigma^2} = 10$ dB, $\sigma_\phi = \sigma_\theta = 0.49^o$, $M = 200$ BS antennas and $L = 20$ taps. For each K, N_D is optimally chosen.

is also upper-bounded as a consequence of the phase noise.

VIII. CONCLUSIONS

Phase noise is an inevitable hardware impairment in communication systems. We studied the effect of phase noise on the sum-rate performance of single-carrier transmission in a MU-MIMO uplink with an excess of BS antennas. Two distinct operation modes in terms of the phase noise processes at the BS antennas are considered, namely, synchronous and non-synchronous operation. Since the knowledge of the exact channel realizations is not available, CSI is acquired via uplink training. The BS uses TR-MRC receive processing to detect the information symbols. An analytical expression for the achievable sum-rate is rigorously derived for both operation modes. Based on the derived achievable sum-rates, we observe that it can be beneficial to use independent instead of fully synchronous phase noise sources. It is also shown that at low SNR, phase noise has little impact on the sum-rate performance. Further, the proposed receive processing achieves an $O(\sqrt{M})$ array power gain, extending earlier results where phase noise was not considered. Finally, due to the progressive phase noise drift in the oscillators, there is a fundamental trade-off between the length of the time interval used for data transmission and the sum-rate performance.

APPENDIX

In this appendix we state the proof of Proposition 2. For both operation modes, we have

$$\text{Var}\left(\text{EN}_k[i]\right) \overset{\Delta}{=} \mathbb{E}\left[|\text{EN}_k[i] - \mathbb{E}\left[\text{EN}_k[i]\right]|^2\right] = \text{Var}\left(\text{IF}_k[i]\right)$$
$$+ \text{Var}\left(\text{ISI}_k[i]\right) + \text{Var}\left(\text{MUI}_k[i]\right) + \text{Var}\left(\text{AN}_k[i]\right)$$

since the terms in $\text{EN}_k[i]$ are mutually uncorrelated. We start by computing the terms $\text{Var}\left(\text{ISI}_k[i]\right)$, $\text{Var}\left(\text{MUI}_k[i]\right)$, $\text{Var}\left(\text{AN}_k[i]\right)$ for the non-synchronous case, which are the same for both operation modes and conclude with the term

$\text{Var}\left(\text{IF}_k[i]\right)$, the calculation of which is different depending on the operation mode. First we compute the variance of the ISI term.

$$\mathbb{E}[|\text{ISI}_k[i]|^2] = \mathbb{E}[|\sqrt{P_D} \sum_{m=1}^{M} \sum_{l=0}^{L-1} \sum_{\substack{q=0 \\ q \neq l}}^{L-1} g_{m,k,l}^* g_{m,k,q} \vartheta\binom{m,k,k}{i,l,p}$$

$$\cdot x_k[i+l-q]|^2] = P_D \sum_{m=1}^{M} \sum_{m'=1}^{M} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \sum_{\substack{p=0 \\ p \neq l}}^{L-1} \sum_{\substack{p'=0 \\ p' \neq l'}}^{L-1}$$

$$\cdot \mathbb{E}\left[g_{m,k,l}^* g_{m,k,p} g_{m',k,p'}^* g_{m',k,l'}\right]$$

$$\cdot \mathbb{E}\left[e^{-j(\phi_m[i+l]-\phi_{m'}[i+l']-\phi_m[(k-1)L+l]+\phi_{m'}[(k-1)L+l'])}\right]$$

$$\cdot \mathbb{E}\left[e^{j(\theta_k[i+l-p]-\theta_k[(k-1)L]-\theta_k[i+l'-p']+\theta_k[(k-1)L])}\right]$$

$$\cdot \mathbb{E}\left[x_k[i+l-p]x_k^*[i+l'-p']\right] = P_D \sum_{m=1}^{M} \sum_{l=0}^{L-1} \sum_{\substack{q=0 \\ q \neq l}}^{L-1} d_{k,l} d_{k,q}$$

$$= P_D M \left(\alpha_k^2 - \sum_{l=0}^{L-1} d_{k,l}^2\right),$$

where we have used the fact that the channel coefficients, the phase noise processes and the data symbols are mutually independent. The last step follows from the normalization of the PDP (see (1)). We will make use of these facts in all the following derivations as well. We proceed with the calculation of the multi-user interference.

$$\mathbb{E}[|\text{MUI}_k[i]|^2] = \mathbb{E}[|\sqrt{P_D} \sum_{m=1}^{M} \sum_{\substack{q=1 \\ q \neq k}}^{K} \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} g_{m,k,l}^* g_{m,q,p} \vartheta\binom{m,k,q}{i,l,p}$$

$$\cdot x_q[i+l-p]|^2] = P_D \sum_{m=1}^{M} \sum_{m'=1}^{M} \sum_{\substack{q=1 \\ q \neq k}}^{K} \sum_{\substack{q'=1 \\ q' \neq k}}^{K} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \sum_{\substack{p=0 \\ p \neq l}}^{L-1} \sum_{\substack{p'=0 \\ p' \neq l}}^{L-1}$$

$$\cdot \mathbb{E}\left[g_{m,k,l}^* g_{m,q,p} g_{m',q',p'}^* g_{m',k,l'}\right]$$

$$\cdot \mathbb{E}\left[e^{-j(\phi_m[i+l]-\phi_{m'}[i+l']-\phi_m[(k-1)L+l]+\phi_{m'}[(k-1)L+l'])}\right]$$

$$\cdot \mathbb{E}\left[e^{j(\theta_q[i+l-p]-\theta_k[(k-1)L]-\theta_{q'}[i+l'-p']+\theta_k[(k-1)L])}\right]$$

$$\cdot \mathbb{E}\left[x_q[i+l-p]x_{q'}^*[i+l'-p']\right]$$

$$= P_D \sum_{m=1}^{M} \sum_{\substack{q=1 \\ q \neq k}}^{K} \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} d_{k,l} d_{q,p} = P_D M \alpha_k \sum_{\substack{q=1 \\ q \neq k}}^{K} \alpha_q$$

We conclude the first part of the proof with the calculation of the variance of the additive noise term.

$$\mathbb{E}[|\text{AN}_k[i]|^2] = \mathbb{E}[|\sqrt{\frac{P_D}{P_p K L}} \sum_{m=1}^{M} \sum_{q=1}^{K} \sum_{l=0}^{L-1} \sum_{p=0}^{L-1} g_{m,q,p}$$

$$\cdot e^{-j(\phi_m[i+l]-\theta_q[i+l-p])} n_m[(k-1)L+l] x_q[i+l-p]|^2$$

$$+ \mathbb{E}[|\sum_{m=1}^{M} \sum_{l=0}^{L-1} \hat{g}_{m,k,l}^* n_m[i+l]|^2]$$

$$= \frac{P_D}{P_p K L} \sum_{m=1}^{M} \sum_{m'=1}^{M} \sum_{q=1}^{K} \sum_{q'=1}^{K} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \sum_{p=0}^{L-1} \sum_{p'=0}^{L-1}$$

$$\mathbb{E}[(g_{m,q,p}e^{-j(\phi_m[i+l]-\theta_q[i+l-p])}n_m[(k-1)L+l]x_q[i+l-p])$$

$$\cdot (g_{m',q',p'}e^{-j(\phi_{m'}[i+l']-\theta_{q'}[i+l'-p'])}n_{m'}[(k-1)L+l']$$

$$\cdot x_{q'}[i+l'-p'])^*] + \sigma^2 \sum_{m=1}^{M}\sum_{l=0}^{L-1}\mathbb{E}[|\hat{g}_{m,k,l}|^2]$$

$$= \frac{P_D\sigma^2}{P_pKL}\sum_{m=1}^{M}\sum_{q=1}^{K}\sum_{l=0}^{L-1}\sum_{\substack{a=1-L\\0\le l-a\le L-1}}^{L-1}d_{q,l-a}$$

$$+ \sigma^2\sum_{m=1}^{M}\sum_{l=0}^{L-1}\left(\frac{\sigma^2}{P_pKL}+\mathbb{E}[|g_{m,k,l}|^2]\right)$$

$$= \sigma^2 M\left(\frac{P_D}{P_pK}\sum_{q=1}^{K}\alpha_q+\frac{\sigma^2}{P_pK}+\alpha_k\right)$$

We proceed by calculating the variance of the term $\mathtt{IF}_k[i]$. It holds

$$\mathtt{Var}(\mathtt{IF}_k[i]) = \mathbb{E}\left[|(A_k[i]-\mathbb{E}[A_k[i]])x_k[i]|^2\right]$$
$$= \mathbb{E}\left[|A_k[i]|^2\right]-|\mathbb{E}[A_k[i]]|^2.$$

Based on the result of Proposition 1 it is sufficient to calculate $\mathbb{E}\left[|A_k[i]|^2\right]$ for each operation mode. We start with the synchronous operation.

$$\mathbb{E}\left[|A_k[i]|^2\right] = P_D\sum_{m=1}^{M}\sum_{l=0}^{L-1}\mathbb{E}[|g_{m,k,l}|^4]$$

$$+ P_D\sum_{m=1}^{M}\sum_{l=0}^{L-1}\sum_{\substack{l'=0\\l'\ne l}}^{L-1}\mathbb{E}[|g_{m,k,l}|^2]\mathbb{E}[|g_{m,k,l'}|^2]$$

$$\cdot \mathbb{E}[e^{-j(\phi[i+l]-\phi[i+l']-\phi[(k-1)L+l]+\phi[(k-1)L+l'])}]$$

$$+ P_D\sum_{m=1}^{M}\sum_{\substack{m'=1\\m'\ne m}}^{M}\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}\mathbb{E}[|g_{m,k,l}|^2]\mathbb{E}[|g_{m',k,l'}|^2]$$

$$\cdot \mathbb{E}[e^{-j(\phi[i+l]-\phi[i+l']-\phi[(k-1)L+l]+\phi[(k-1)L+l'])}]$$

$$= P_DM\sum_{l=0}^{L-1}2d_{k,l}^2 + P_DM\sum_{l=0}^{L-1}\sum_{\substack{l'=0\\l'\ne l}}^{L-1}d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|}$$

$$+ P_DM(M-1)\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|}$$

$$= P_DM\sum_{l=0}^{L-1}d_{k,l}^2 + P_DM^2\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|}$$

Finally, for the synchronous operation, the effective noise variance, is given by

$$\varsigma_k^s[i] \triangleq \mathtt{Var}(\mathtt{EN}_k^s[i]) = P_DM^2\kappa_k[i]+C_k.$$

We conclude with the calculation of the term $\mathbb{E}\left[|A_k[i]|^2\right]$ for

the non-synchronous mode.

$$\mathbb{E}\left[|A_k[i]|^2\right] = P_D\sum_{m=1}^{M}\sum_{l=0}^{L-1}\mathbb{E}[|g_{m,k,l}|^4]$$

$$+ P_D\sum_{m=1}^{M}\sum_{l=0}^{L-1}\sum_{\substack{l'=0\\l'\ne l}}^{L-1}\mathbb{E}[|g_{m,k,l}|^2]\mathbb{E}[|g_{m,k,l'}|^2]$$

$$\cdot \mathbb{E}[e^{-j(\phi_m[i+l]-\phi_m[i+l']-\phi_m[(k-1)L+l]+\phi_m[(k-1)L+l'])}]$$

$$+ P_D\sum_{m=1}^{M}\sum_{\substack{m'=1\\m'\ne m}}^{M}\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}\mathbb{E}[|g_{m,k,l}|^2]\mathbb{E}[|g_{m',k,l'}|^2]$$

$$\cdot \mathbb{E}[e^{-j(\phi_m[i+l]-\phi_{m'}[i+l']-\phi_m[(k-1)L+l]+\phi_{m'}[(k-1)L+l'])}]$$

$$= P_DM\sum_{l=0}^{L-1}2d_{k,l}^2 + P_DM\sum_{l=0}^{L-1}\sum_{\substack{l'=0\\l'\ne l}}^{L-1}d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|}$$

$$+ P_DM(M-1)\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}d_{k,l}d_{k,l'}e^{-\sigma_\phi^2(i-(k-1)L)}$$

$$= P_DM\sum_{l=0}^{L-1}d_{k,l}^2 + P_DM\sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}d_{k,l}d_{k,l'}e^{-\sigma_\phi^2|l-l'|}$$

$$+ P_DM(M-1)\alpha_k^2 e^{-\sigma_\phi^2(i-(k-1)L)}.$$

The variance for the non-synchronous operation is

$$\varsigma_k^{ns}[i] \triangleq \mathtt{Var}(\mathtt{EN}_k^{ns}[i]) = P_DM\xi_k[i]+P_DM^2\varpi_k[i]+C_k.$$

REFERENCES

[1] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, Mar. 1998.

[2] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO Paradigm," *IEEE Signal Processing Magazine*, vol. 24, pp. 36 –46, September 2007.

[3] 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.

[4] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.

[5] 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, February 2014.

[6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[8] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and wiener phase noise," *IEEE Transactions on Communications*, vol. 43, pp. 191 –193, Feb/Mar/Apr 1995.

[9] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Transactions on Communications*, vol. 46, pp. 580 –583, May 1998.

[10] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: consequences and solutions," *IEEE Transactions on Communications*, vol. 52, pp. 1988 – 1996, Nov. 2004.

[11] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Transactions on Communications*, vol. 55, pp. 1607 –1616, Aug. 2007.

[12] H. Mehrpouyan, A. Nasir, S. Blostein, T. Eriksson, G. Karagiannidis, and T. Svensson, "Joint Estimation of Channel and Oscillator Phase Noise in MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 60, pp. 4790–4807, Sep. 2012.

[13] A. Lapidoth, "On phase noise channels at high SNR," in *Proceedings of the 2002 IEEE Information Theory Workshop*, pp. 1–4, Oct. 2002.

[14] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Transactions on Communications*, vol. 62, pp. 920–929, Mar. 2014.

[15] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits," *IEEE Transactions on Information Theory*, vol. PP, no. 99, pp. 1–1, 2014.

[16] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on the uplink performance of large MU-MIMO systems," in *50th Allerton Conference on Communication Control and Computing*, Oct. 2012.

[17] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear Pre-Coding Performance in Measured Very-Large MIMO Channels," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pp. 1–5, Sept 2011.

[18] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, pp. 655–674, May 2000.

[19] 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.

[20] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Fortieth Asilomar Conference on Signals, Systems and Computers, 2006. ACSSC '06.*, pp. 359 –363, November 2006.

[21] T. Höhne and V. Ranki, "Phase noise in beamforming," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3682 – 3689, Dec. 2010.

[22] H. Yang and T. L. Marzetta, "Total energy efficiency of cellular large-scale antenna system multiple access mobile networks," in *Proc. IEEE Online Conference on Green Communications*, Oct. 2013.

[23] D. Petrovic, W. Rave, and G. Fettweis, "Common phase error due to phase noise in OFDM-estimation and suppression," in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2004*, vol. 3, pp. 1901 – 1905 Vol.3, Sept. 2004.

[24] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Communications Letters*, vol. 1, pp. 276 –279, August 2012.