University of HERTFORDSHIRE UH 25

Research Archive

Citation for published version:

Anastasios Papazafeiropoulos, 'Impact of General Channel Aging Conditions on the Downlink Performance of Massive MIMO', *IEEE Transactions on Vehicular Technology,* Vol. 66 (2): 1428-1442, February 2017.

DOI:

https://doi.org/10.1109/TVT.2016.2570742

Document Version:

This is the Accepted Manuscript version.
The version in the University of Hertfordshire Research Archive may differ from the final published version.

Copyright and Reuse:

Enquiries

If you believe this document infringes copyright, please contact the Research & Scholarly Communications Team at rsc@herts.ac.uk

Impact of General Channel Aging Conditions on the Downlink Performance of Massive MIMO

Anastasios K. Papazafeiropoulos, *Member, IEEE*

Abstract—Recent works have identified massive multiple-input-multiple-output (MIMO) as a key technology for achieving substantial gains in spectral and energy efficiency. Additionally, the turn to low-cost transceivers, being prone to hardware impairments is the most effective and attractive way for cost-efficient applications concerning massive MIMO systems. In this context, the impact of channel aging, which severely affects the performance, is investigated herein by considering a generalized model. Specifically, we show that both Doppler shift because of the users' relative movement as well as phase noise due to noisy local oscillators (LOs) contribute to channel aging. To this end, we first propose a joint model, encompassing both effects, in order to investigate the performance of a massive MIMO system based on the inevitable time-varying nature of realistic mobile communications. Then, we derive the deterministic equivalents (DEs) for the signal-to-noise-and-interference ratios (SINRs) with maximum ratio transmission (MRT) and regularized zero-forcing precoding (RZF). Our analysis not only demonstrates a performance comparison between MRT and RZF under these conditions, but most importantly, it reveals interesting properties regarding the effects of user mobility and phase noise. In particular, the large antenna limit behavior depends profoundly on both effects, but the burden due to user mobility is much more detrimental than phase noise even for moderate user velocities (≈ 30 km/h), while the negative impact of phase noise is noteworthy at lower mobility conditions. Moreover, massive MIMO systems are favorable even in general channel aging conditions. Nevertheless, we demonstrate that the transmit power of each user to maintain a certain quality of service can be scaled down at most by $1\sqrt{M}$ (M is the number of BS antennas), which indicates that the joint effects of phase noise and user mobility do not degrade the power scaling law, but only the achievable sum-rate.

Index Terms—Channel estimation, TDD systems, massive MIMO, Doppler shift, hardware impairments, phase noise, linear precoding.

I. Introduction

The increasing demand for wireless services and higher data volume per area has brought to the research forefront of both the scientific community (e.g., the METIS project [2]) and the standardization bodies a new technology known as massive multiple-input-multiple-output (MIMO) networks [3], [4]. The keystone of this concept, currently considered for the design of next generation wireless networks [5]–[7], is based on that the numbers of base station (BS) antennas and users are scaled

up by 1-2 orders of magnitude under the same time-frequency resources. This scientific area is motivated by the resulting gains stemming from large channel matrices, i.e., the arising advantages due to the asymptotics of random matrix theory [8]. Specifically, by allowing the number of BS antennas tend to infinity, small-scale fading and thermal noise average out, as well as the analysis is simplified.

Notably, the channel state information (CSI) is pivotal to the performance of such systems that take into account for multi-user MIMO (MU-MIMO) [9]. Thus, recognizing the need for a realistic study with imperfect CSI, we focus on an important factor-phenomenon, being present in time-varying channels, that deteriorates the quality of CSI available at the BS in time-division-duplex (TDD) systems. Our consideration includes TDD systems and not frequency-division-duplex (FDD) systems, since the application of the latter in massive MIMO systems is meaningful only when the channel matrices are sparse [10][1]. This phenomenon, known as *channel aging*, describes the mismatch appearing between the current and the estimated channel due to the relative movement between the users and the BS antennas in addition to any processing delays. In other words, it models the divergence occurring between the channel estimate used in the precoder/detector, and the channel over which the data transmission actually takes place.

Although the significance of channel aging due to user mobility, its effect on the performance of massive MIMO systems has not been addressed adequately as few studies on this area exist. In fact, only the user mobility has been considered as the only channel aging cause [11]–[18]. Basically, channel aging due to only user mobility was the focal point in [11], where an application of the deterministic equivalent (DE) analysis was presented by considering linear techniques in the uplink and downlink in terms of maximum ratio combining (MRC) detector and maximum ratio transmission (MRT) precoder, respectively. Further investigations have been performed in [14], where more sophisticated linear techniques are employed, i.e., minimum-mean-square-error (MMSE) receivers (uplink) and regularized zero-forcing precoders (downlink) are considered, and a comparison between the various linear transceiver techniques is performed. Moreover, in [15], the optimal linear receiver in the case of cellular massive MIMO systems has been derived by exploiting the correlation between the channel estimates and the interference from other cells, while in [16], the uplink analysis of a cellular network with zero-forcing (ZF) receivers that holds for any finite as well as infinite number

Anastasios K. Papazafeiropoulos is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, United Kingdom (e-mail:a.papazafeiropoulos@imperial.ac.uk).

[1]The hypothesis of sparsity is still an open question requiring further research and channel measurements.

of BS antennas has been provided.

In general, hardware undergoes several types of impairments such as high power amplifier nonlinearities and I/Q imbalance [19]–[37] [2]. These hardware impairments can be classified into two categories. In the first category, the variable, describing the hardware impairments, is multiplied with the channel vector, and it might cause channel attenuations and phase shifts. In the case of slow variation of these impairments, they can be characterized as sufficiently static, and thus, can be assimilated by the channel vector by an appropriate scaling of its covariance matrix or due to the property of circular symmetry of the channel distribution. Note that this factor cannot be incorporated by the channel vector by an appropriate scaling of its covariance matrix or due to the property of circular symmetry of the channel distribution, when it changes faster than the channel. Focusing on this scenario, we include in our analysis the phase noise in the case which results to its accumulation within the channel coherence period [22], [28]–[30]. Given that additive transceiver impairments are not time-dependent, we have not included them in our analysis [26], [29], [31].

Hence, deeper consideration reveals that user mobility is not the only source of channel aging. Interestingly, the unavoidable *phase noise*, being one of the most severe transceiver impairments [32], contributes to the further degradation of the system performance due to channel aging because it accumulates over time. Ideally, local oscillators (LOs), used during the conversion of the baseband signal to bandpass and vice-versa, should output a sinusoid with stable amplitude, frequency, and phase. Unfortunately, the presence of phase noise is inevitable due to the inherent imperfections in the circuitry of LOs that bring up to the sinusoid a random phase drift. A lot of research has been conducted regarding phase noise in conventional MIMO systems. In fact, a big part of the literature has dealt with multi-carrier MIMO systems that employ orthogonal frequency-division multiplexing (OFDM) [33]–[36]. For example, in [34], the signal-to-interference-and-noise-ratio (SINR) degradation in OFDM is studied, and a method to mitigate the effect of phase noise is proposed. In [22], a method to jointly estimate the channel coefficients and the phase noise in a point-to-point MIMO system is presented. Not to mention that the degradation appears to be more significant in coherent communications such as massive MIMO systems. Specifically, although massive MIMO technology is attractive (cost-efficient) for network deployment, only if the antenna elements consist of inexpensive hardware (more imperfections), most of the research contributions are based on the strong assumption of perfect hardware without phase noise, which is quite idealistic in practice. Reasonably, the introduction of a larger number of antennas will have a more

severe contribution to the degradation provoked by phase noise, which could be partially avoided in conventional MIMO systems, where more expensive hardware with less induced phase is affordable. Hence, the larger the number of antennas and the corresponding employed LOs, the larger the impact of phase noise. Thus, in the case of massive MIMO systems, it is a dire necessity to study the impact of channel aging induced by phase noise. Furthermore, motivated by the need of employing low-quality RF elements (e.g., oscillators) for cost-effective construction of large BS antenna arrays, it is expected that the phase noise will be more severe in a massive MIMO setting. In particular, the large number of antennas regime has been partially investigated [28]–[30], [37]. Specifically, the phase noise has been taken into account for single carrier uplink massive MIMO by considering ZF and time-reversal MRC in [37] and [28], respectively. Regarding the downlink, an effort to address the effect of phase noise, being of great importance in massive MIMO systems, has taken place only in [30]. However, the analysis therein does not account for fast time-varying conditions due to the mobility of the users.

In this paper, we provide a novel realistic CSI model that allows studying channel aging thoroughly by means of the simultaneous impact of user mobility and phase noise at the LOs. Taking into account that next generation wireless systems need to be cost-efficient, implying that real systems may include low-quality RF elements with more imperfections, makes the significance of the proposed model in describing the ineluctable time-varying nature of realistic scenarios remarkable. Under these conditions, we analyze the downlink performance of a massive MIMO system with imperfect CSI using linear precoders (MRT and RZF). To this end, we invoke random matrix theory (RMT) tools, and we derive the DEs (asymptotically tight approximations as system's dimensions increase). Actually, these approximations agree with simulation results as has been already shown in the literature [8], [11], [14]. As a result, these deterministic approximations make any lengthy simulations unnecessary. Furthermore, as far as phase noise is concerned, we address two system arrangements: a) all BS antennas are connected to a common LO (CLO), and b) each BS antenna includes its own separate LO (SLOs setup). The contributions of this paper can be summarized as follows:

- Given a single cell MU-MIMO realistic system model with path loss and shadowing, we present a joint channel-phase noise estimate describing the impact of channel aging by incorporating the time-varying effects of Doppler shift and phase noise at the BS and user elements (UEs).
- Using tools of large RMT, we obtain the DEs of the downlink SINRs, when the BS applies MRT and RZF precoders under imperfect CSI due to uplink training and time variation of the channel (generalized channel aging). Moreover, we account for special cases of the channel covariance, corresponding to various interesting effects described by the large-scale component, that simplify a lot the resulting expressions.
- We demonstrate that the square root power-law, generally holding in imperfect CSI scenarios, stands for the case of generalized channel aging as well. Hence, the transmit

power per user can be cut down by scaling the number of BS antennas to the inverse of their square root.

- We corroborate our analysis concerning deterministic approximations by performing simulations coinciding with the analytical expressions.
- We elaborate on insightful conclusions, drawn from our analysis. For example, Doppler shift has a far more severe impact than phase noise even for low vehicle velocities. In addition, massive MIMO systems outperform conventional MIMO even in generalized channel aging conditions with RZF precoder behaving better than MRC for various Doppler shifts and phase noise severities.

The remainder of this paper is organized as follows. Section II introduces the system and signal models accounting for both the effects of user mobility and phase noise, which allow a more complete characterization of channel aging. Next, Section III presents the downlink transmission, in terms of the received signal, the applied precoders, and the achievable user rate. In Section IV, we proceed with the main results that include the derivation of the DEs of the downlink SINRs, as well as the investigation of certain simpler cases such as the scenario where the large-scale fading component is absent. Section V reports a set of numerical results that confirms the validity of our analytical results, and sheds light on the behavior of channel aging for various mobility environments, phase noise settings, and a number of antennas in a massive MIMO system. We summarize our results and observations in Section VI. In Appendix A, some useful results from the literature are presented, while the following appendices provide the proofs for the main analytical results.

Notation: Vectors and matrices are represented by boldface lower and upper case symbols. $(\cdot)^\mathsf{T}$, $(\cdot)^\mathsf{H}$, and $\mathrm{tr}\cdot$ denote the transpose, Hermitian transpose, and trace operators, respectively. The expectation and variance operators, as well as the spectral norm of a matrix are denoted by $\mathbb{E}[\cdot]$ and $\mathrm{var}[\cdot]$, as well as $\|\cdot\|$, respectively, while $\mathrm{Re}\{\cdot\}$ is the real part of a complex number. The $\mathrm{diag}\{\cdot\}$ operator generates a diagonal matrix from a given vector, and the symbol $\triangleq$ declares definition. The notations $\mathcal{C}^M$ and $\mathcal{C}^{M\times N}$ refer to complex M-dimensional vectors and $M\times N$ matrices, respectively. Finally, $\mathbf{b}\sim\mathcal{CN}(\mathbf{0},\mathbf{\Sigma})$ or $\mathbf{b}\sim\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ denote a circularly symmetric complex Gaussian or a real Gaussian vector $\mathbf{b}$ with zero-mean and covariance matrix $\mathbf{\Sigma}$.

II. SYSTEM MODEL

This work considers a single-cell system, where a BS communicates with a number of single-antenna non-cooperative UEs belonging to the set $\mathcal{K}$ with cardinality $K=|\mathcal{K}|$ being the number of UEs. In particular, each UE is assigned an index k in $\mathcal{K}$, while the BS is deployed with an array of M antennas[3]. Since each UE has a single antenna, the terms user and user element, are used interchangeably in the following analysis. Our interest focuses on large-scale topologies, where both number of antennas M and users K grow infinitely large

while keeping a finite ratio β, i.e., $M,K\to\infty$ with $K/M=\beta$ fixed such that $\limsup_M K/M<\infty$[4].

During the transmission of the nth symbol, the channel gain vector between the BS and user k, denoted by $\mathbf{h}_{k,n}\in\mathbb{C}^M$ exhibits flat-fading, and it is assumed constant for the symbol period, while it may vary slowly from symbol to symbol. Basically, given the time-frequency resources of the system, the symbol duration is assumed smaller or equal to the coherence time of all the users. Moreover, in order to account for certain inevitable effects such as path loss and lognormal shadowing, which are dependent on the distance of the users from the BS, we express the channel between user k and the BS as

$$\mathbf{h}_k=\mathbf{R}_k^{1/2}\mathbf{w}_k, \tag{1}$$

where $\mathbf{R}_k=\mathbb{E}\left[\mathbf{h}_{k,n}\mathbf{h}_{k,n}^\mathsf{H}\right]\in\mathbb{C}^{M\times M}$ is a deterministic Hermitian-symmetric positive-definite matrix representing the aforementioned effects[5].

A. Phase Noise Model

In practice, both the transmitter and the receiver are impaired by phase noise induced during the up-conversion of the baseband signal to passband and vice-versa[6]. As a result, the received signal is distorted during the reception processing, while at the transmitter side, a mismatch appears between the signal that is intended to be transmitted and the generated signal.

In our analysis regarding the nth time slot, we take into account for non-synchronous operation at the BS, i.e., the BS antennas have independent phase noise processes $\phi_{m,n}, m=1,\ldots,M$ with $\phi_{m,n}$ being the phase noise process at the mth antenna. Note that the phase noise processes are considered as mutually independent, since each antenna has its own oscillator, i.e., separate LOs (SLOs) at each antenna [28], [30]. In case that all the antennas have a common LO (CLO), the phase noise processes $\phi_{m,n}$ are identical for all $m=1,\ldots,M$ (synchronous operation). Similarly, we denote $\varphi_{k,n}, k=1,\ldots,K$ the phase noise process at the single-antenna user k.

Phase noise during the nth symbol can be described by a discrete-time independent Wiener process, i.e., the phase noises at the LOs of the mth antenna of the BS and kth user are modeled as

$$\phi_{m,n}=\phi_{m,n-1}+\delta_n^{\phi_m} \tag{2}$$

$$\varphi_{k,n}=\varphi_{k,n-1}+\delta_n^{\varphi_k}, \tag{3}$$

---

[3]The antennas can be co-located at the BS or can be distributed as fully-coordinated, multiple, smaller BSs [29].

[4] In the following, in order to simplify notation, $M\to\infty$ will be used instead of $M,K\to\infty$, unless stated otherwise.

[5]The independence of $\mathbf{R}_k$ from the time index, n, originates from the assumption that the shadowing vary with time in a longer slower pace than the coherence time [11]–[13]. Interestingly, this model is quite versatile, since $\mathbf{R}_k$ can also describe the antenna correlation due to either insufficient antenna spacing or a lack of scattering.However, the following results focus only on diagonal $\mathbf{R}_k$ matrices describing only large-scale fading, and $\mathbf{w}_{k,n}\in\mathbb{C}^M$ is an uncorrelated fast-fading Gaussian channel vector drawn as a realization from a zero-mean circularly symmetric complex Gaussian distribution, i.e., $\mathbf{w}_{k,n}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_M)$.

[6]The conversion takes place by multiplying the signal with the LO's output.

where $\delta_n^{\phi_m} \sim \mathcal{N}(0, \sigma_{\phi_m}^2)$ and $\delta_n^{\varphi_k} \sim \mathcal{N}(0, \sigma_{\varphi_k}^2)$ [28], [30], [32]. Note that $\sigma_i^2 = 4\pi^2 f_c c_i T_s$, $i = \phi_m, \varphi_k$ describes the phase noise increment variance with T_s, c_i, and f_c being the symbol interval, a constant dependent on the oscillator, and the carrier frequency, respectively.

The phase noise, being the distortion in the phase due to the random phase drift in the signal coming from the LOs of the BS and user k, is expressed as a multiplicative factor to the channel vector as

$$\mathbf{g}_{k,n} = \mathbf{\Theta}_{k,n}\mathbf{h}_{k,n}, \tag{4}$$

where $\mathbf{\Theta}_{k,n} \triangleq \mathrm{diag}\left\{e^{j\theta_{k,n}^{(1)}}, \ldots, e^{j\theta_{k,n}^{(M)}}\right\} = e^{j\varphi_{k,n}}\mathbf{\Phi}_n \in \mathbb{C}^{M \times M}$ with $\mathbf{\Phi}_n \triangleq \mathrm{diag}\left\{e^{j\phi_{1,n}}, \ldots, e^{j\phi_{M,n}}\right\}$ being the phase noise sample matrix at time n because of the imperfections in the LOs of the BS, while $e^{j\varphi_{k,n}}$ corresponds to the phase noise contribution by the LO of user k [28], [30].

B. Channel Estimation

In realistic conditions, where no perfect CSI is available, the BS needs to estimate the uplink channel [14], [38], [39]. We take advantage of the TDD operation scheme, where an uplink training phase during each transmission block (coherence period) exists by means of transmit pilot symbols. Since in TDD the number of required pilots scales with the number of terminals K, but not the number of BS antennas M, not only the analysis of systems with large number of BS antennas is possible because of the finite duration of the coherence time[7], but also the knowledge of the downlink channel is considered known due to the property of reciprocity.

By assuming that the channel estimation takes place at time 0, we now derive the joint channel-phase noise linear minimum mean-square error detector (LMMSE) estimate of the effective channel $\mathbf{g}_{k,0} = \mathbf{\Theta}_{k,0}\mathbf{h}_{k,0}$ in the presence of phase noise, small-scale fading, and channel impairments such as path loss. During this phase, we neglect the channel aging effect by assuming that both the channel and phase noise remain constant [40]. This is a valid assumption since the duration of the training phase is small, and the consequent time-variation of the channel is unnoticeable. However, in the data transmission phase, taking place for $(T_c - \tau)$ symbols, the channel is supposed to vary from symbol to symbol, where τ and T_c are the duration of the training sequence during the training phase and the channel coherence time, respectively.

Proposition 1: The LMMSE estimator of $\mathbf{g}_{k,0}$, obtained during the training phase, is

$$\hat{\mathbf{g}}_{k,0} = \left(\mathbf{I}_M + \frac{\sigma_b^2}{p_p}\mathbf{R}_k^{-1}\right)^{-1}\tilde{\mathbf{y}}_{,k,0}^{p}, \tag{5}$$

where σ_b^2 is the variance of the post-processed noise at base station, $\tilde{\mathbf{y}}_{,k,0}^{p}$ is a noisy observation of the effective channel from user k to the BS, and $p_p = \tau p_u$ with p_u being the power per user in the uplink data transmission phase.

Proof: : The proof of Proposition 1 is given in Appendix B. ∎

C. Channel Aging

The relative movement between the users and the BS inflicts a phenomenon, known as channel aging in the literature [11]–[13]. Specifically, the Doppler shift arising from this movement causes the channel to vary between when it is learned by estimation, and when the estimate is applied for detection/precoding. Especially, channel aging is more severe in massive MIMO systems, where narrow transmit beams appear due to the high angular resolution. Thus, the precoders that are used for achieving this spatial focus, bring a critical performance loss, which worsens as the users move with increasing velocity [41]. An autoregressive model of order 1 enables the joint channel-phase noise process $\mathbf{g}_{k,n}$ between the BS and the kth user at time n (current time) to be approximated as[8]

$$\mathbf{g}_{k,n} = \mathbf{A}_n\mathbf{g}_{k,0} + \mathbf{e}_{k,n}, \tag{6}$$

where, initially, $\mathbf{A}_n = \mathrm{J}_0(2\pi f_D T_s n)\mathbf{I}_M$ is supposed to model 2-D isotropic scattering [43][9], $\mathbf{g}_{k,0}$ is the channel vector received during the training phase, and $\mathbf{e}_{k,n} \in \mathbb{C}^M$ is the uncorrelated channel error vector due to the channel variation modelled as a stationary Gaussian random process with independent and identically distributed (i.i.d.) entries and distribution $\mathcal{CN}(\mathbf{0}, \mathbf{R}_k - \mathbf{A}_n\mathbf{R}_k\mathbf{A}_n)$.

This model can be enriched by incorporating other interesting effects. By introducing the effect of channel estimation to the autoregressive model of (6), the BS takes into account for both estimated and delayed CSI, in order to design the detector $\mathbf{W}_n \in \mathbb{C}^{M \times M}$ or the precoder $\mathbf{F}_n \in \mathbb{C}^{M \times M}$ at time instance n for the uplink and downlink, respectively. Hence, the effective channel at time n is expressed by

$$\mathbf{g}_{k,n} = \mathbf{A}_n\mathbf{g}_{k,0} + \mathbf{e}_{k,n} \tag{7}$$
$$= \mathbf{A}_n\hat{\mathbf{g}}_{k,0} + \tilde{\mathbf{e}}_{k,n}, \tag{8}$$

where $\hat{\mathbf{g}}_{k,0} \sim \mathcal{CN}(\mathbf{0}, \mathbf{D}_k)$ with $\mathbf{D}_k = \left(\mathbf{I}_M + \frac{\sigma_b^2}{p_p}\mathbf{R}_k^{-1}\right)^{-1}\mathbf{R}_k$ and $\tilde{\mathbf{e}}_{k,n} \triangleq \mathbf{A}_n\tilde{\mathbf{g}}_{k,n} + \mathbf{e}_{k,n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k - \mathbf{A}_n\mathbf{D}_k\mathbf{A}_n)$ are mutually independent, while $\mathbf{A}_n$, is assumed to be known to the BS, now incorporates both the effects of 2-D isotropic scattering, user mobility, and phase noise, as shown below by Theorem 1. In other words, the combined error $\tilde{\mathbf{e}}_{k,n}$ depends on both the imperfect and the general delayed CSI effects, allowing the elicitation of interesting outcomes during the ensuing analysis in Section IV.

---

[7]TDD is the only viable solution in fast time-varying channel conditions with massive MIMO systems, where the limitations for the coherence time are more stringent, while FDD can be considered only when there is some kind of channel sparsity [10].

[8]Contrary to other works [11]–[13], [42], $\mathbf{A}_n$, being the factor of $\mathbf{g}_{k,0}$, is described by a matrix instead of a scalar to account for the setting of SLOs. Moreover, in this work, $\mathbf{A}_n$ is time-dependent by the delay n in terms of the number of symbols transmitted, in order to include the accumulative channel aging effect [18]. Therefore, the longer the delay, the smaller the $\mathbf{A}_n$, which in turn results in less accurate CSI. Hence, the channel estimated at time 0, is used at time 1, 2, and so on till n. It is intuitive that the CSI will be much less reliable at time n, which results to a more realistic model. We assume that it can be embodied in the channel vector $\mathbf{g}_{k,0}$ by employing the circular symmetry of the channel distribution, or mainly, by the inclusion of an appropriate scaling of the covariance matrix.

[9]Normally, the BS antennas are deployed in a fixed space instead of considering distantly distributed small BSs constituting one massive MIMO BS. In such case, $\mathbf{A}_n$ is a scalar, since it can be reasonably assumed that all the BS antennas have the same relative movement comparing to the user.

We turn our attention to the derivation of $\mathbf{A}_n$ by accounting for user mobility, phase noise, and the general channel effects modelled by $\mathbf{R}_k$.

Theorem 1: The effective joint channel-phase noise vector at time n is expressed as

$$\mathbf{g}_{k,n}^{\text{Joint}} = \mathbf{A}_n \mathbf{g}_{k,0} \tag{9}$$

with

$$\mathbf{A}_n = \mathrm{J}_0(2\pi f_{\mathrm{D}} T_s n) e^{-\frac{\sigma_{\varphi_k}^2}{2} n} \Delta\boldsymbol{\Phi}_n, \tag{10}$$

where $\Delta\boldsymbol{\Phi}_n = \mathrm{diag}\left\{ e^{-\frac{\sigma_{\phi_1}^2}{2} n}, \ldots, e^{-\frac{\sigma_{\phi_M}^2}{2} n} \right\}$.

Proof: The total mean square error (MSE), which is actually given by means of the error covariance matrix, is $\mathrm{MSE} = \mathrm{tr}\,\mathbf{E}_{k,n}$, where $\mathbf{E}_{k,n}$ is the error covariance written as

$$
\begin{aligned}
\mathbf{E}_{k,n} &= \mathbb{E}\left[\left(\mathbf{g}_{k,n} - \mathbf{A}_n \mathbf{g}_{k,0}\right) \left(\mathbf{g}_{k,n} - \mathbf{A}_n \mathbf{g}_{k,0}\right)^{\mathsf{H}} \right] \\
&= \mathbb{E}|\mathbf{g}_{k,n}|^2 + \mathbf{A}_n \mathbb{E}|\mathbf{g}_{k,0}|^2 \mathbf{A}_n^{\mathsf{H}} - 2\mathbb{E}\left[\mathrm{Re}\left\{\mathbf{A}_n \mathbf{g}_{k,0} \mathbf{g}_{k,n}^{\mathsf{H}}\right\}\right] \\
&= \mathbf{R}_k + \mathbf{A}_n \mathbf{R}_k \mathbf{A}_n^{\mathsf{H}} - 2\mathbb{E}\left[\mathrm{Re}\left\{\mathbf{A}_n \mathbf{g}_{k,0} \mathbf{g}_{k,n}^{\mathsf{H}}\right\}\right] \\
&= \mathbf{R}_k + \mathbf{A}_n \mathbf{R}_k \mathbf{A}_n^{\mathsf{H}} - 2\mathrm{J}_0(2\pi f_{\mathrm{D}} T_s n) e^{-\frac{\sigma_{\varphi_k}^2}{2} n} \mathbf{A}_n \mathbf{R}_k \Delta\boldsymbol{\Phi}_n. \tag{11}
\end{aligned}
$$

Note that the expectation is taken over all channel and phase noise realizations. Also, we have used that $\mathbf{g}_{k,n} = \boldsymbol{\Theta}_{k,n} \mathbf{h}_{k,n}$ with $\boldsymbol{\Theta}_{k,n}$ and $\mathbf{h}_{k,n}$ being uncorrelated. Also, we have denoted $\Delta\boldsymbol{\Phi}_n = \mathrm{diag}\left\{ e^{-\frac{\sigma_{\phi_1}^2}{2} n}, \ldots, e^{-\frac{\sigma_{\phi_M}^2}{2} n} \right\}$, and

$$\mathbb{E}[\mathbf{h}_{k,0} \mathbf{h}_{k,n}^{\mathsf{H}}] = \mathrm{J}_0(2\pi f_D T_s n) \mathbf{R}_k. \tag{12}$$

The multiplcative parameter matrix $\mathbf{A}_n$ is determined by minimizing the MSE in (11). Thus, if we differentiate (11) with respect to $\mathbf{A}_n$, and equate the resulting expression to zero, we obtain the Hermitian matrix $\mathbf{A}_n$ as in (10). ∎

Remark 1: During the data transmission phase, the available CSI at time n is obtained in terms of the estimated CSI at time 0. The parameter matrix $\mathbf{A}_n$ has a physical meaning, since it describes the effects of 2-D isotropic scattering, user mobility, and phase noise impairments at both the UE and the BS.

Remark 2: Notably, phase noise induces an extra loss due to the additional deviation between the actual channel at time n, and the channel estimate from time 0 because of the symbol-by-symbol random phase drift. In other words, the phase noise contributes to the channel aging phenomenon that imposes a further challenge to investigate the realistic potentials of massive MIMO systems.

Corollary 1: In the case where all the BS antennas are connected to the same oscillator or all the oscillators are identical, $\mathbf{A}_n$ degenerates to a scalar $\alpha_n = \mathrm{J}_0(2\pi f_{\mathrm{D}} T_s n) e^{-\frac{\sigma_{\varphi_k}^2 + \sigma_{\phi_k}^2}{2} n}$ or a scaled identity matrix $\alpha_n \mathbf{I}_M$, respectively.

III. DOWNLINK TRANSMISSION

The BS transmits data simultaneously to all users in its cell by employing space-division multiple access (SDMA). Applying reciprocity, the downlink channel is written as the Hermitian transpose of the uplink channel. The coherence time,

the received signal $y_{k,n} \in \mathbb{C}$ by the kth UE during the data transmission phase ($n = \tau + 1, \ldots, T_{\mathrm{c}}$) is

$$y_{k,n} = \sqrt{p_{\mathrm{d}}} \mathbf{h}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n} \mathbf{s}_n + z_{k,n} \tag{13}$$

where $\mathbf{s}_n = \sqrt{\lambda} \mathbf{F}_n \mathbf{x}_n$ denotes the signal vector transmitted by the BS with λ, $\mathbf{F}_n \in \mathbb{C}^{M \times K}$ and $\mathbf{x}_n = [x_{1,n},\ x_{2,n}, \cdots,\ x_{K,n}]^{\mathsf{T}} \in \mathbb{C}^K \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ being a normalization parameter, the linear precoding matrix, and the data symbol vector to its K served users, respectively. Moreover, we have assumed that the BS transmits data to all the users with the same power p_{d}, and $z_{k,n} \sim \mathcal{CN}(0, \sigma_k^2)$ is complex Gaussian noise at user k. The need for constraining the transmit power per user to p_{d}, i.e., $\mathbb{E}\left[\frac{p_{\mathrm{d}}}{K} \mathbf{s}_n^{\mathsf{H}} \mathbf{s}_n\right] = p_{\mathrm{d}}$, provides the normalization parameter as

$$\lambda = \frac{1}{\mathbb{E}\left[\frac{1}{K} \mathrm{tr}\, \mathbf{F}_n \mathbf{F}_n^{\mathsf{H}}\right]}. \tag{14}$$

Henceforth, by taking into account that $\mathbf{g}_{k,n} = \boldsymbol{\Theta}_{k,n} \mathbf{h}_{k,n}$, we proceed with the following presentation of the received signal, obtained by (35). Thus, we have

$$y_{k,n} = \sqrt{\lambda p_{\mathrm{d}}} \mathbf{g}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n}^2 \mathbf{f}_{k,n} x_{k,n} + \sum_{i \neq k} \sqrt{\lambda p_{\mathrm{d}}} \mathbf{g}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n}^2 \mathbf{f}_{i,n} x_{i,n} + z_{k,n} \tag{15}$$

$$
\begin{aligned}
&= \underbrace{\sqrt{\lambda p_{\mathrm{d}}} \mathbb{E}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n}^2 \mathbf{f}_{k,n}\right] x_{k,n}}_{\text{desired signal}} + \underbrace{z_{k,n}}_{\text{noise}} + \sum_{i \neq k} \sqrt{\lambda p_{\mathrm{d}}} \mathbf{g}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n}^2 \mathbf{f}_{i,n} x_{i,n} \\
&\quad + \sqrt{\lambda p_{\mathrm{d}}} \left(\mathbf{g}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n}^2 \mathbf{f}_{k,n} x_{k,n} - \mathbb{E}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \boldsymbol{\Theta}_{k,n}^2 \mathbf{f}_{k,n}\right] x_{k,n}\right). \tag{16}
\end{aligned}
$$

In (15), we have applied a similar technique to [44] because UEs do not have access to instantaneous CSI during the downlink phase, but we can assume that, in particular, UE k is aware of only $\mathbb{E}[\mathbf{g}_{k,n}^{\mathsf{H}} \mathbf{f}_{jm}]$.

Given that the input symbols are Gaussian, we provide the achievable user rate based on the analysis in [45, Theorem 1]. Specifically, the worst case uncorrelated additive noise is also zero-mean with variance equal to the variance of interference plus noise. In addition, since $\mathbf{g}_{k,n}$ is circularly symmetric, we let $\mathbf{g}_{k,n} = \boldsymbol{\Theta}_{k,0}^2 \mathbf{g}_{k,n}$. Thus, by denoting $\widetilde{\boldsymbol{\Theta}}_{k,n} = \Delta\boldsymbol{\Theta}_{k,n}$, where $\Delta\boldsymbol{\Theta}_{k,n} = (\boldsymbol{\Theta}_{k,n} \boldsymbol{\Theta}_{k,0}^{\mathsf{H}})^2 = \mathrm{diag}\left\{ e^{2j\left(\theta_{k,n}^{(1)} - \theta_{k,0}^{(1)}\right)}, \ldots, e^{2j\left(\theta_{k,n}^{(1)} - \theta_{k,0}^{(1)}\right)} \right\}$, we may consider a SISO model with the desired signal power and the interference plus noise power at UE k given by

$$S_{k,n} = \lambda \left| \mathbb{E}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \mathbf{f}_{k,n}\right] \right|^2, \tag{17}$$

and

$$I_{k,n} = \lambda \mathrm{var}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \mathbf{f}_{k,n}\right] + \frac{\sigma_k^2}{p_{\mathrm{d}}} + \sum_{i \neq k} \lambda \mathbb{E}\left[\left|\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \mathbf{f}_{i,n}\right|^2\right]. \tag{18}$$

Since the available CSI at the BS at time n is $\mathbf{A}_n \hat{\mathbf{g}}_{k,0}$, as can be seen from (8), the expression corresponding to the MRT precoder is

$$\mathbf{F}_n = \mathbf{A}_n \hat{\mathbf{G}}_0. \tag{19}$$

In a similar way, the BS designs its RZF precoder as [39]

$$\mathbf{F}_n = \left(\mathbf{A}_n\hat{\mathbf{G}}_0\hat{\mathbf{G}}_0^{\mathsf{H}}\mathbf{A}_n + \mathbf{Z} + Ma\mathbf{I}_M\right)^{-1}\mathbf{A}_n\hat{\mathbf{G}}_0$$

$$= \boldsymbol{\Sigma}\mathbf{A}_n\hat{\mathbf{G}}_0, \tag{20}$$

where we define $\boldsymbol{\Sigma} \triangleq \left(\mathbf{A}_n\hat{\mathbf{G}}_0\hat{\mathbf{G}}_0^{\mathsf{H}}\mathbf{A}_n + \mathbf{Z} + aM\mathbf{I}_M\right)^{-1}$ with $\mathbf{Z} \in \mathbb{C}^{M\times M}$ being an arbitrary Hermitian nonnegative definite matrix and a being a regularization scaled by M, in order to converge to a constant, as M, $K \to \infty$. Note that both $\mathbf{Z}$ and a could be optimized, but this is outside the scope of this paper and left to future work. One possible common choise for these parameters is provided in [39].

Following a similar approach to [28], [29], we compute the achievable rate for each time instance of the data transmission phase. Thus, the downlink ergodic achievable rate of user k, is given by

$$R_k = \frac{1}{T_c}\sum_{n=1}^{T_c-\tau} R_{k,n}$$

$$= \frac{1}{T_c}\sum_{n=1}^{T_c-\tau} \log_2\left(1 + \gamma_{k,n}\right), \tag{21}$$

where $\gamma_{k,n} = \frac{S_{k,n}}{I_{k,n}}$ is the instantaneous downlink SINR at time n. In other words, the mutual information between the received signal and the transmitted symbols is lower bounded by the achievable rate per user provided in (21).

IV. DETERMINISTIC EQUIVALENT ANALYSIS

In this section, we provide the derivations of the downlink SINRs after applying MRT and RZF precoders in the presence of imperfect and delayed CSI due to user mobility and phase noise. Use of RMT in terms of DEs enables the extraction of asymptotic expressions as $K, M \to \infty$, while keeping their ratio $K/M = \beta$ finite. Given that the theory of DEs provides tight approximations even for moderate system dimensions, the significance of our results is of great importance. Confirmation of this tightness is provided in Section V by simulations. Interestingly, the results mirror the dependence of the harmful effects described by channel aging.

The DE of the SINR $\gamma_{k,n}$ is such that $\gamma_{k,n} - \bar{\gamma}_{k,n} \xrightarrow[M\to\infty]{\text{a.s.}} 0$[10], while the deterministic rate of user k is obtained by the dominated convergence [46] and the continuous mapping theorem [47] by means of (21)

$$R_k - \frac{1}{T_c}\sum_{n=1}^{T_c-\tau} \log_2(1 + \bar{\gamma}_{k,n}) \xrightarrow[N\to\infty]{\text{a.s.}} 0. \tag{22}$$

The DE downlink achievable user rates corresponding to MRT and RZF can be obtained by means of the following theorems. Hereafter, for the sake of exposition and without loss of generality, we assume that in the case of SLOs, the phase noises obey to identical statistics. However, in Appendix C,

[10]Note that $\xrightarrow[M\to\infty]{\text{a.s.}}$ denotes almost sure convergence, and $a_n \asymp b_n$ expresses the equivalence relation $a_n - b_n \xrightarrow[M\to\infty]{\text{a.s.}} 0$ with a_n and b_n being two infinite sequences.

we scrutinize the general setting, where phase noises across different LOs are independent, but not identically distributed. First, we consider MRT precoding.

A. MRT

This section presents the general DE regarding the SINR with MRT precoding. Then, this asymptotic result particularizes to a number of cases of interest. In additions, it is investigated the dependence of the transmit power per user with the number of BS antennas (power scaling law), in order to support a specific desired rate.

Theorem 2: The downlink DE of the SINR of user k at time n with MRT precoding , accounting for imperfect CSI and delayed CSI due to phase noise and user mobility, is given by

$$\bar{\gamma}_{k,n} = \frac{e^{-2\left(\sigma_{\varphi_k}^2 + \sigma_\phi^2\right)n}\delta_k^2}{\frac{1}{M}\delta_k' + \frac{\sigma_k^2}{p_{\text{d}}\bar{\lambda}M} + \sum_{i\neq k}\frac{1}{M}\delta_i''}, \tag{23}$$

with

$$\bar{\lambda} = \left(\frac{1}{K}\sum_{i=1}^K \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_k\right)^{-1},$$

$\delta_k = \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_k$, $\delta_k' = \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_k\left(\mathbf{R}_k - \mathbf{A}_n^2\mathbf{D}_k\right)$, and $\delta_i'' = \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_i\mathbf{R}_k$.

Proof: : The proof of Theorem 2 is given in Appendix C. ∎

Corollary 2: In case that $\mathbf{R}_k = \mathbf{I}_M$ (no large-scale component), and $\mathbf{A}_n$ degenerates to a scalar α_n or a $\alpha_n\mathbf{I}_M$, i.e., the BS employs a CLO or a SLOs setting with identical LOs, respectively, we obtain

$$\bar{\gamma}_{k,n} = \frac{\alpha_n^2 d e^{-2\left(\sigma_{\varphi_k}^2 + \sigma_\phi^2\right)n}}{\frac{1-\alpha_n^2 d}{M} + \frac{\sigma_k^2}{p_{\text{d}}M} + \frac{K-1}{M}}, \tag{24}$$

where $d = \left(\frac{\tau p_{\text{u}}}{\tau p_{\text{u}} + \sigma_b^2}\right)^{-1}$ and $\alpha_n = \mathrm{J}_0(2\pi f_{\text{D}}T_s n)e^{-\frac{\sigma_{\varphi_k}^2 + \sigma_\phi^2}{2}n}$.

Proof: Using $\mathbf{R}_k = \mathbf{I}_M$ leads to $\mathbf{D}_k = d\mathbf{I}_M = \left(\frac{\tau p_{\text{u}}}{\tau p_{\text{u}} + \sigma_b^2}\right)^{-1}\mathbf{I}_M$. Moreover, if $\mathbf{A}_n$ is replaced by the scalar α_n or the scaled identity matrix $\alpha_n\mathbf{I}_M$, Theorem 2 allows to simplify enough the deterministic SINR. Hence, by substituting $\bar{\lambda} = \left(\alpha_n^2 d\right)^{-1}$, $\delta = \alpha_n^2 d$, $\delta' = \alpha_n^2 d\left(1 - \alpha_n^2 d\right)$, and $\delta'' = \alpha_n^2 d$, we obtain $\bar{\gamma}_{k,n}$ as in (24). ∎

Nevertheless, it is interesting to investigate the effect of the combined impairments of phase noise and user mobility on the asymptotic power scaling law. For the sake of exposition, we focus on MRT precoding, however similar results can be obtained in the case of RZF precoding as well.

Proposition 2: For the identical SLOs or CLO setting, the downlink achievable SINR of user k at time n with MRT precoding subject to imperfect CSI, delayed CSI due to phase noise and user mobility, and collocated BS antennas becomes

$$\gamma_{k,n} = \frac{\tau E_{\text{d}}E_{\text{u}}}{\sigma^4 M^{2q-1}}\mathrm{J}_0^2(2\pi f_{\text{D}}T_s n)e^{-3\left(\sigma_{\varphi_k}^2 + \sigma_\phi^2\right)n}[\mathbf{R}_k^2]_{mm}, \tag{25}$$

where the transmit uplink and downlink powers are scaled proportionally to $1/M^q$, i.e., $p_{\text{u}} = E_{\text{u}}/M^q$ and $p_{\text{d}} = E_{\text{d}}/M^q$ for fixed E_{u} and E_{d}, and $q > 0$.

Proof: Let us first substitute the MRT precoder, given by (19), in (16). Then, we divide both the desired and the interference parts by λ and $1/M^2$. The desired signal power is written as

$$S_{k,n}^{\mathrm{MRT}} = \frac{1}{M^2}\left|\mathbf{g}_{k,n}^{\mathsf{H}}\widetilde{\mathbf{\Theta}}_{k,n}\mathbf{A}_n\hat{\mathbf{g}}_{k,0}\right|^2$$

$$= \frac{1}{M^2}\left|\hat{\mathbf{g}}_{k,0}^{\mathsf{H}}\widetilde{\mathbf{\Theta}}_{k,n}\mathbf{A}_n^2\hat{\mathbf{g}}_{k,0}\right|^2 \qquad (26)$$

$$= \frac{1}{M^2}\mathrm{J}_0^4 2(2\pi f_{\mathrm{D}}T_{\mathrm{s}}n)e^{-4\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}[\mathbf{D}_k]_{mm}^2, \qquad (27)$$

where $[\mathbf{D}_k]_{mm}$ is the mth diagonal element of the matrix $\mathbf{D}_k$ expressing the variance of the mth element. Given that both $\widetilde{\mathbf{\Theta}}_{k,n}$ and $\mathbf{A}_n$ are diagonal matrices, we have taken into account in (27) that they can commute. Moreover, in the last step of (27), we have used the law of large numbers for large M in the case of collocated BS antennas, as well as that $\mathbf{A}_n$ is a scaled identity matrix or a scalar in the case of identical SLOs or CLO, respectively. In other words, $\hat{\mathbf{g}}_{k,0}$ has i.i.d. elements with variance $[\mathbf{D}_k]_{mm}$. As far as the interference is concerned, the first and third terms of (18) vanish to zero as $M \to \infty$ by means of the same law. Thus, we have

$$I_{k,n}^{\mathrm{MRT}} = \frac{\sigma_k^2}{M^2 p_{\mathrm{d}}\lambda}$$

$$= \frac{\sigma_k^2\mathrm{J}_0^2(2\pi f_{\mathrm{D}}T_{\mathrm{s}}n)e^{-\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}[\mathbf{D}_k]_{mm}}{M p_{\mathrm{d}}}, \qquad (28)$$

where we have applied the law of large numbers to obtain $\lambda = \left(\frac{1}{M}\mathrm{J}_0^2(2\pi f_{\mathrm{D}}T_{\mathrm{s}}n)e^{-\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}[\mathbf{D}_k]_{mm}\right)^{-1}$ from (14), and we have substituted $\mathbf{D}_k = \frac{\tau E_{\mathrm{u}}}{M^q\sigma_k^2}\mathbf{R}_k^2$, since p_{p} depends on p_{u}, the result in (25) is obtained. $\blacksquare$

Proposition 2 reveals that the selection of q affects heavily the achievable rate per user. In fact, proper selection allows maintaining the same performance, even by further scaling down the transmit power of each user. Specifically, if q is set less than $1/2$, $\gamma_{k,n}$ is unbounded, i.e., it tends to infinity. On the contrary, the SINR of user k diminishes to zero, if $q > 1/2$. This clearly indicates that the transmit powers of each user during the training and downlink phases have been reduced over the required value. Most importantly, in the special case that $q = 1/2$, the SINR approaches a non-zero limit given by the following corollary.

Corollary 3: In the presence of phase noise and user mobility, when the transmit uplink and downlink powers are scaled down by $p_{\mathrm{u}} = E_{\mathrm{u}}/\sqrt{M}$ and $p_{\mathrm{d}} = E_{\mathrm{d}}/\sqrt{M}$ for fixed E_{u} and E_{d}, the downlink SINR per user with MRT can be finite as

$$\gamma_{k,n} = \frac{\tau^2 E_{\mathrm{d}}E_{\mathrm{u}}}{\sigma_k^4}\mathrm{J}_0^2(2\pi f_{\mathrm{D}}T_{\mathrm{s}}n)e^{-3\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}[\mathbf{R}_k^2]_{mm}. \qquad (29)$$

Evidently, phase noise and user mobility reduce the SINR, but the power scaling law is not affected.

B. RZF

As far as RZF is concerned, the analysis is more complex, as shown below. First we present the DE of the SINR in general channel aging conditions by means of the following theorem.

Theorem 3: The downlink DE of the SINR user k at time n with RZF precoding, accounting for imperfect CSI and delayed CSI due to phase noise and user mobility, is given by

$$\bar{\gamma}_{k,n} = \frac{e^{-2\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}\delta_k^2}{\frac{1}{M}\delta_k' + \frac{\sigma_k^2(1+\delta_k)^2}{p_{\mathrm{d}}\bar{\lambda}} + \sum_{i\neq k}\frac{Q_{ik}(1+\delta_k)^2}{M(1+\delta_i)^2}}, \qquad (30)$$

with

$$\bar{\lambda} = \frac{K}{\left(\frac{1}{M}\operatorname{tr}\mathbf{T} - \frac{1}{M}\operatorname{tr}\left(\frac{\mathbf{Z}}{M}+a\mathbf{I}_M\right)\mathbf{C}\right)}$$

$$Q_{ik} \asymp \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_i\mathbf{C}''' + \frac{|\delta_k|^2\delta_k''}{(1+\delta_k)^2} - 2\mathrm{Re}\left\{\frac{\delta_k\delta_k''}{(1+\delta_k)}\right\},$$

$\delta_k = \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_k\mathbf{T}, \delta_k' = \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_k\mathbf{C}',$ and $\delta_k'' = \frac{1}{M}\operatorname{tr}\mathbf{A}_n^2\mathbf{D}_k\mathbf{C}'',$ where

* $\mathbf{T} = \mathbf{T}(a)$ and $\boldsymbol{\delta} = [\delta_1,\cdots,\delta_K]^{\mathsf{T}} = \boldsymbol{\delta}(a) = \boldsymbol{e}(a)$ are given by Theorem 4 for $\mathbf{L} = \mathbf{A}_n^2\mathbf{D}_k$, $\mathbf{S} = \mathbf{Z}/M$, $\mathbf{R}_k = \mathbf{A}_n^2\mathbf{D}_k \, \forall k \in \mathcal{K}$,
* $\mathbf{C} = \mathbf{T}'(a)$ is given by Theorem 5 for $\mathbf{L} = \mathbf{I}_M$, $\mathbf{S} = \mathbf{Z}/M$, $\mathbf{K} = \mathbf{I}_M$, $\mathbf{R}_k = \mathbf{A}_n^2\mathbf{D}_k \, \forall k \in \mathcal{K}$,
* $\mathbf{C}' = \mathbf{T}''(a)$ is given by Theorem 5 for $\mathbf{L} = \mathbf{A}_n^2\mathbf{D}_k$, $\mathbf{S} = \mathbf{Z}/M$, $\mathbf{K} = \mathbf{R}_k - \mathbf{A}_n^2\mathbf{D}_k$, $\mathbf{R}_k = \mathbf{A}_n^2\mathbf{D}_k \, \forall k \in \mathcal{K}$,
* $\mathbf{C}'' = \mathbf{T}'''(a)$ is given by Theorem 5 for $\mathbf{L} = \mathbf{A}_n^2\mathbf{D}_k$, $\mathbf{S} = \mathbf{Z}_j/M$, $\mathbf{K} = \mathbf{A}_n^2\mathbf{D}_i$, $\mathbf{R}_k = \mathbf{A}_n^2\mathbf{D}_k \, \forall k \in \mathcal{K}$.
* $\mathbf{C}''' = \mathbf{T}''''(a)$ is given by Theorem 5 for $\mathbf{L} = \mathbf{A}_n^2\mathbf{D}_i$, $\mathbf{S} = \mathbf{Z}_j/M$, $\mathbf{K} = \mathbf{R}_k$, $\mathbf{R}_k = \mathbf{A}_n^2\mathbf{D}_k \, \forall k \in \mathcal{K}$.

Proof: : The proof of Theorem 3 is given in Appendix D. $\blacksquare$

Corollary 4: Let $\mathbf{R}_k = \mathbf{R}$, i.e., the large-scale effects (path loss and shadowing) affect the same all users, then $\mathbf{D}_k = \mathbf{D}$. In such case, the deterministic SINR of Theorem 3 $\bar{\gamma}_{k,n}$ can be simplified to

$$\bar{\gamma}_{k,n} = \frac{e^{-2\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}\delta^2}{\frac{1}{M}\delta' + \frac{\sigma_k^2(1+\delta)^2}{p_{\mathrm{d}}\bar{\lambda}} + (K-1)\frac{Q}{M}}, \qquad (31)$$

with $\delta = \frac{1}{M}\mathbf{A}_n^2\mathbf{DT} \triangleq e, \delta' = \frac{1}{M}\mathbf{A}_n^2\mathbf{DC}', \delta'' = \frac{1}{M}\mathbf{A}_n^2\mathbf{DC}''$, $\mathbf{T} = \left(\mathbf{A}_n^2\mathbf{D}/\beta(1+\delta) + \mathbf{Z}/M + a\mathbf{I}_M\right)^{-1}$, $\bar{\lambda} = K/\left(\frac{1}{M}\operatorname{tr}\mathbf{T} - \frac{1}{M}\operatorname{tr}\left(\frac{\mathbf{Z}}{M}+a\mathbf{I}_M\right)\mathbf{C}\right)$, and

$$Q \asymp \frac{1}{M}\operatorname{tr}\mathbf{RC}'' + \frac{|\delta|^2\delta''}{(1+\delta)^2} - 2\mathrm{Re}\left\{\frac{\delta\delta''}{(1+\delta)}\right\},$$

while the general expression of $\mathbf{T}'$ is given by

$$\mathbf{T}' = \mathbf{TKT} + e_{\mathbf{K}}'/\beta(1+\delta)^2\mathbf{TA}_n^2\mathbf{DT},$$

where $e_{\mathbf{K}}' = \beta(1+\delta)^2 e_{111}^{\mathbf{K}}/\left(\beta - e_{201}^{\mathbf{K}}\right)$ with $e_{ijm}^{\mathbf{K}} = 1/(1+\delta)^{j+m}\frac{1}{M}\operatorname{tr}\left(\mathbf{A}_n^2\mathbf{D}\right)^i\mathbf{TK}^j\mathbf{T}^m$.

Corollary 5: If no large-scale component ($\mathbf{R}_k = \mathbf{I}_M$) are assumed, and the phase noise from all the oscillators is considered identical or the BS has only a CLO, i.e., $\mathbf{A}_n$ becomes $\alpha_n\mathbf{I}_M$ or a scalar α_n, we obtain

$$\bar{\gamma}_{k,n} = \frac{e^{-2\left(\sigma_{\varphi_k}^2+\sigma_\phi^2\right)n}\delta^2}{\frac{1}{M}\delta' + \frac{\sigma_k^2(1+\delta)^2}{p_{\mathrm{d}}\bar{\lambda}} + (K-1)\frac{Q}{M}}, \qquad (32)$$

where $d = \left(\frac{\tau p_{\mathrm{u}}}{\tau p_{\mathrm{u}}+\sigma_b^2}\right)^{-1}$ and $\alpha_n = \mathrm{J}_0(2\pi f_{\mathrm{D}}T_{\mathrm{s}}n)e^{-\frac{\sigma_{\varphi_k}^2+\sigma_\phi^2}{2}n}$.

Proof: When $\mathbf{R}_k = \mathbf{I}_M$, $\mathbf{D}_k$ becomes a scaled identity matrix, i.e., $\mathbf{D}_k = d\mathbf{I}_M = \left(\frac{p_\mathrm{p}\tau}{p_\mathrm{p}\tau+\sigma_b^2}\right)\mathbf{I}_M$. In addition, replacing $\mathbf{A}_n$ by the scalar $\alpha_n = \mathrm{J}_0(2\pi f_\mathrm{D}T_\mathrm{s}n)e^{-\frac{\sigma_{\varphi_k}^2+\sigma_{\phi_k}^2}{2}n}$, we obtain $\delta = \alpha_n^2 dt$, $\delta' = \alpha_n^2 dt_1'$, $\delta'' = \alpha_n^2 dt_2'$, $\bar{\lambda} = \frac{K}{\left(t-t_3'a-\frac{t_3'z}{M}\right)}$,

$$Q \asymp \frac{\beta d^4 \delta^2 \left(\delta^2 + \delta + 1\right)}{\beta(\delta+1)^2 - d^4\delta^2}.$$

By considering the unique positive root of the quadratic equation $\delta = \alpha_n^2 dt$, we obtain (33). Note that the general expression of t_i' for $i = 1, 2, 3$ is given by

$$t_i' = \frac{\beta d^2 \delta^2 (\delta+1)^2 k_i}{a^4 \left(\beta(\delta+1)^2 - d^4\delta^2\right)},$$

since $e_{k_i}' = e_\mathbf{K}' = \beta d\delta^2 k_i/(\alpha_n^2(\beta - (d^2\delta^2)/(1+\delta)^2))$. Here, $k_1 = (1 - \alpha_n^2 d)$, $k_2 = \alpha_n^2 d$, and $k_3 = 1$. ∎

V. Numerical Results

The purpose of this section is to present some representative numerical examples enabling the study of the time variation of the channel due to the effects of phase noise and user mobility on the performance of massive MIMO systems with MRT and RZF precoders. The metrics under study are the achievable sum-rates and the required transmit power achieving a certain user rate. Note that the achievable sum-rates are described by means of the SINRs, given by Theorems 2 and 3. In addition, the correctness of the proposed results is validated by Monte-Carlo simulations. Notably, despite that the analytical results are obtained by assuming that $M, K \to \infty$, they coincide with the simulations even for finite values of M and K. Actually, this is a known observation in the literature concerning the DEs [8], [11]–[13], [39].

A. Simulation Setup

Given that the interest of this work is based on the investigation of the impact of practical channel impairments, the setting of our scenario includes long-term evolution (LTE) system specifications [3]. Specifically, the simulation setup considers a single cell with a radius of $R = 1000$ meters, and a guard range of $r_0 = 100$ meters specifying the distance between the nearest user and the BS. The BS, comprised of M antennas, broadcasts to K users that are uniformly distributed within the cell. According to the system model, the channel vector between the BS and the UE k in the nth time slot describes the large-scale fading and path loss or spatial correlation by means of $\mathbf{R}_k$. Herein, $\mathbf{R}_k$ describes large-scale fading modelled as $\mathbf{R}_k = l_k\mathbf{I}_M$ with $l_k = q_k/(r_k/r_0)^\upsilon$. Especially, q_k is a log-normal random variable with standard deviation σ ($\sigma = 8$ dB) expressing the shadow-fading effect, r_k denotes the distance between UE k and the BS, and υ ($\upsilon = 3.8$) is the path loss exponent. The uplink and downlink powers are $p_\mathrm{u} = p_\mathrm{d} = 46$dBm, while the thermal noise density is -174 dBm/Hz. The length of the training duration is $\tau = K$ symbols, and the phase noises at the BS and user LOs are simulated as discrete Wiener processes given by (2) and (3), with increment variances in the interval $0°$-$2°$ [48]. The coherence

time is $T_\mathrm{c} = 1/4f_\mathrm{D} = 1$ ms, where $f_\mathrm{D} = 250$ Hz is the Doppler spread corresponding to a relative velocity of 135 km/h between the BS and the users, if the center frequency is assumed to be $f_\mathrm{c} = 2$ GHz. Moreover, given that the bandwidth for LTE-A is W $= 20$MHz, the symbol time is $T_s = 1/(2\mathrm{W}) = 0.025~\mu$s. In order to account for fast varying channels, where high mobility occurs, the coherence block is assumed to include $T = 196$ channel uses corresponding to the coherence bandwidth $B_\mathrm{c} = 196$ KHz.

One useful metric, providing the means for study the considered system, is

$$\mathcal{S} \triangleq \sum_{k=1}^K \bar{R}_k, \tag{34}$$

where $\mathcal{S}$ is the sum-rate and $\bar{R}_k$ is the DE rate of UE k given by (22), and Theorems (2) and (3) for the cases of MRT and RZF, respectively.

In the following figures, the red and black line patterns correspond to the rates with RZF and MRT precoding for various phase noise nominal values, respectively. Thus, the "solid", "dash", and "dot" lines designate the analytical results with no channel aging, $\phi_{m,n} = 0°$, $\varphi_{k,n} = 2°$, and $\phi_{m,n} = \varphi_{k,n} = 2°$, respectively. The bullets represent the simulation results. Finally, the green "solid" and "dot" lines, where applicable, depict a scenario with channel aging but not phase noise in the cases of MRT and RZF, respectively.

B. Rate Comparison between MRT and RZF

The rate comparison between MRT and RZF considers the impact of both Doppler shift and phase noise. Clearly, the RZF outperforms the MRT for the scenarios considered. Especially, Fig. 1 depicts the achievable rates by varying the number of BS antennas for different values of phase noise, when the users are assumed to be static ($v = 0$ Km/h), or equivalently, in the case of no Doppler shift ($f_\mathrm{D} = 0$ Hz). By increasing the hardware imperfection in terms of phase noise, the sum-rate decreases. As far as the RZF is concerned, when the phase noise at the user side has a variance of $\sigma_{\varphi_k}^2 = 2°$ the rate loss is 75% with the BS having $M = 30$ antennas, but when M reaches 300, the loss is smaller, i.e., 52%. Moreover, when the phase noise increases, the sum-rate saturates faster. Similar conclusions can be made in the case of MRT precoding. The perfect match between Monte Carlo simulations and analytical results validates the analytical results.

On the other hand, Fig. 5 provides the variation of $\mathcal{S}$ versus the normalized Doppler shift $f_\mathrm{D}T_\mathrm{s}$ when $M = 60$ for different values of phase noise starting from perfect LOs (no phase noise) to high phase noise ($\sigma_{\phi_k}^2 = \sigma_{\varphi_k}^2 = 2°$). It is revealed that the effect of the Doppler shift on the achievable rate is more detrimental than that of phase noise. In fact, for low velocities in the order of 30 km/h equivalent to $f_\mathrm{D}T_\mathrm{s} \approx 0.2$, the degradation due to phase noise starts to become insignificant, but then the achievable rate can become so low that it is inadequate for practical applications, i.e., it is not worthy to investigate the impact of phase noise in high mobility environments. Hence, the higher the mobility, the less important role the phase noise plays. In the same figure, the

$$\delta = \frac{\sqrt{4\alpha_n^2\beta^2 d(a+z) + (\alpha_n^2 d(1-\beta) + \beta(a+z))^2} + \alpha_n^2(\beta-1)d + \alpha_n\beta - \beta(a+z)}{2\beta(a+z)}. \tag{33}$$
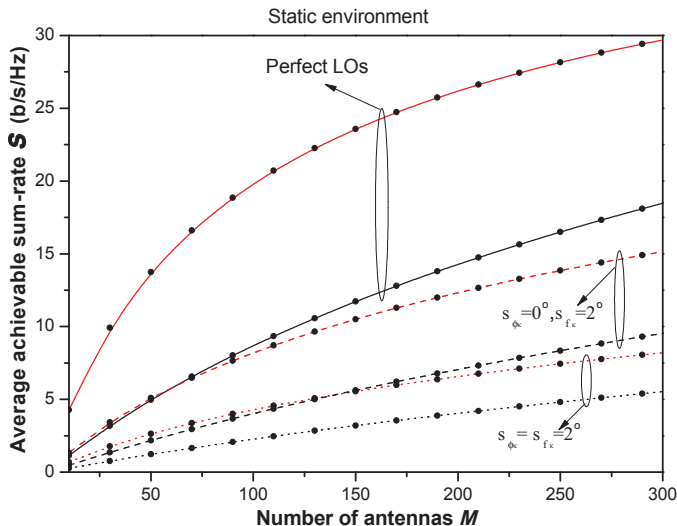

Fig. 1. Simulated and DE downlink sum-rates with MRT and RZF precoders in a static environment versus the number of BS antennas for various values of phase noise. Red and black lines correspond to the theoretical sum-rates with RZF and MRT precoding, respectively, while the black bullets refer to the simulation results.
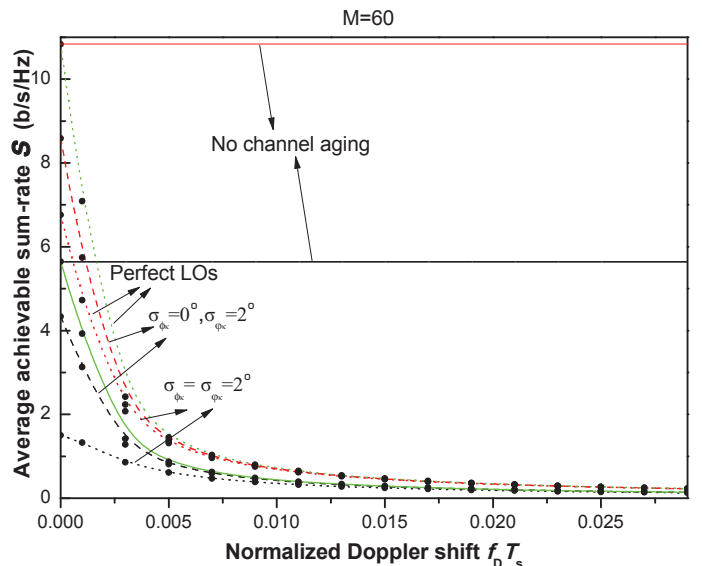

Fig. 2. Simulated and DE downlink sum-rates with MRT and RZF precoders when $M = 60$ as a function of the normalized Doppler shift for various values of phase noise. Red and black lines correspond to the theoretical sum-rates with RZF and MRT precoding, respectively, while the black bullets refer to the simulation results. The green "solid" and "dot" lines mirror a scenario with channel aging but not phase noise in the cases of MRT and RZF, respectively. Lines parallel to x-axis represent scenarios with no channel aging.

straight lines illustrate the sum-rate with imperfect CSI, but with no user mobility and with perfect LOs at both BS and user ends. Furthermore, it is evident that as phase noise increases, the performance worsens. However, the loss due to phase noise is prominent in static environments.

C. Required Transmit Power Comparison between MRT and RZF

Regarding the required transmit power achieving a specific rate per user equal to 1 bit/s/Hz, it is quite insightful to investigate how it changes by varying the number of BS antennas, the amount of Doppler shift, and the severity of phase noise.

Hence, Fig. 3 illustrates the variation of the transmit power p_d versus the number of BS antennas in a static environment ($f_D T_s = 0$). Specifically, p_d decreases considerably when we increase M. Especially, a closer observation shows a reduction in the transmit power by approximately 1.5dB after doubling the number of BS antennas, which agrees with previously known results e.g., [38]. Notably, the more severe is the phase noise at the user, the more the required transmit power should be.

Fig. 4 depicts p_d versus the varying normalized Doppler shift $f_D T_s$ and phase noise, when $M = 60$. In particular, the straight lines represent the required transmit power with no channel aging. As a result, these do not depend on $f_D T_s$ or phase noise. However, as the Doppler shift increases, p_d becomes higher, and very soon (low velocities), it saturates to

a constant. In other words, after a specific value of $f_D T_s$, any increase of p_d cannot achieve any benefit. Furthermore, the higher the phase noise, the faster the saturation ensues.

D. Extension to Multi-Cell Large MIMO Systems

Herein, we focus on the impact of channel aging in the practical case of a hexagonal cellular system. Especially, we consider the downlink of a cellular MIMO system with L cells operating under the same frequency band, i.e., the system is impaired by pilot contamination. Actully, each cell can be assumed as an instance of the single-cell setting provided in Section II. The received signal at the kth user, located at the ith cell, is given by

$$y_{ik,n} = \sqrt{p_d} \sum_{l=1}^{L} \mathbf{h}_{lik,n}^{\mathsf{H}} \mathbf{\Theta}_{lk,n} \mathbf{s}_{l,n} + z_{ik,n} \tag{35}$$

where $\mathbf{s}_{l,n}$ denotes the unit-norm signal vector transmitted by the BS. Moreover, we have assumed that the BS transmits data to all the users with the same power p_d, and $z_{ik,n} \sim \mathcal{CN}(0, \sigma_k^2)$ is complex Gaussian noise at user k, found in the ith cell.

The LMMSE estimator of $\mathbf{g}_{ik,0}$, obtained during the training phase, is

$$\hat{\mathbf{g}}_{ik,0} = \left(\mathbf{I}_M + \frac{\sigma_b^2}{p_p} \sum_l \mathbf{R}_{lik}^{-1} \right)^{-1} \tilde{\mathbf{y}}_{iik,0}^{\mathrm{p}}, \tag{36}$$
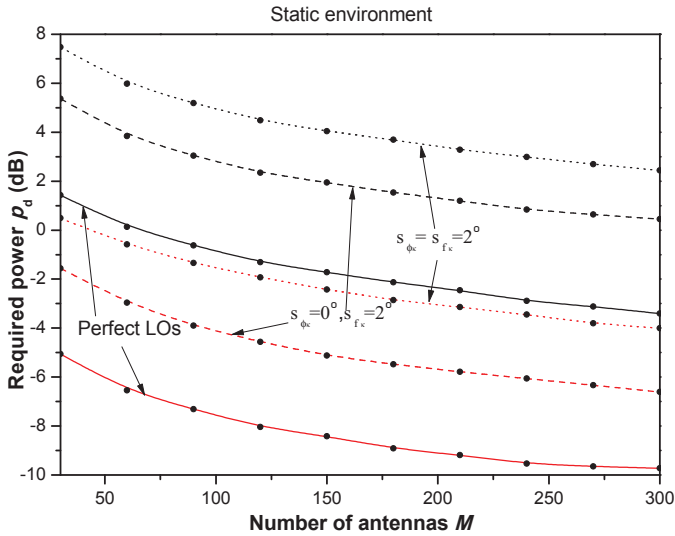
Fig. 3. Required transmit power to achieve 1 bit/s/Hz per user with MRT and RZF precoders in a static environment versus the number of BS antennas and various values of phase noise. Red and black lines correspond to the theoretical sum-rates with RZF and MRT precoding, respectively, while the black bullets refer to the simulation results.

Fig. 5. Simulated downlink sum-rates with MRT and RZF precoders in a cellular setting with $L = 7$ when $M = 60$ as a function of the normalized Doppler shift for various values of phase noise. Red and black lines correspond to the simulated sum-rates with RZF and MRT precoding, respectively. The green "solid" and "dot" lines mirror a scenario with channel aging but not phase noise in the cases of MRT and RZF, respectively. Lines parallel to x-axis represent scenarios with no channel aging.

particular, we consider the same design parameters with the setup of the single cell, but now we employ $L = 7$ cells. Moreover, the average sum-rate is going to be investigated for the central cell, while it has to be mentioned that in the multi-cell setting the large-scale fading depends on the distance of not only the distance of kth user from its associated BS, but also from the neighbor BSs. In such scenario, the equivalent SISO model for UE k in the central cell has desired signal power and interference plus noise power given by

$$S_{ik,n} = \lambda \left| \mathbb{E} \left[\mathbf{g}_{iik,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{ik,n} \mathbf{f}_{ik,n} \right] \right|^2 \qquad (37)$$

and

$$I_{ik,n} = \lambda \mathrm{var} \left[\mathbf{g}_{iik,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{ik,n} \mathbf{f}_{ik,n} \right] + \frac{\sigma_k^2}{p_{\mathrm{d}}}$$
$$+ \sum_{(l,m \neq i,k)} \lambda \mathbb{E} \left[\left| \mathbf{g}_{lim,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{lm,n} \mathbf{f}_{lm,n} \right|^2 \right], \qquad (38)$$

where $\widetilde{\boldsymbol{\Theta}}_{ik,n} = \Delta \boldsymbol{\Theta}_{ik,n}$ and $\mathbf{f}_{ik,n}$ represent the accumulated phase noise and the precoder applied by the ith BS.

Obviously, the difference between the single-cell and multi-cell analyses is limited to the term describing the inter-cell interference, or in other words, the pilot contamination. Fig. 5 illustrates the sum-rate of the ith (central) cell versus the normalized Doppler shift. As can be seen, the impact of channel aging on the inter-cell interference is similar to the intra-cell interference, and results only in an extra degradation of the system performance. In addition, the system is more sensitive to the normalized Doppler shift. In other words, the more severe the channel aging, and basically the user mobility, the higher the degradation of the achievable sum-rate because it affects the inter-cell interference term.
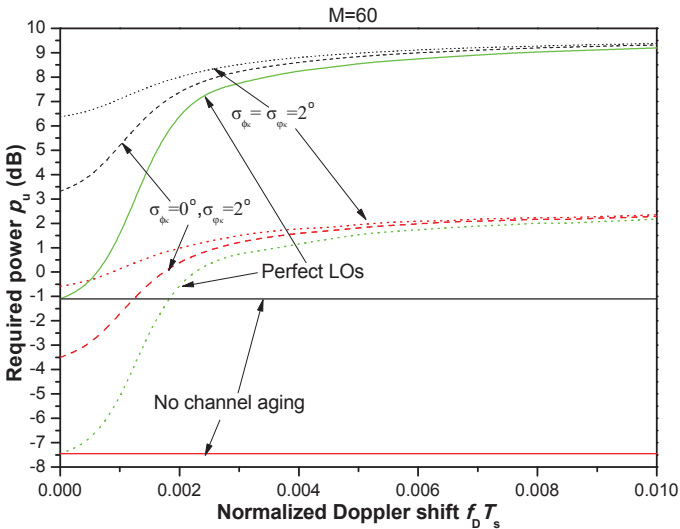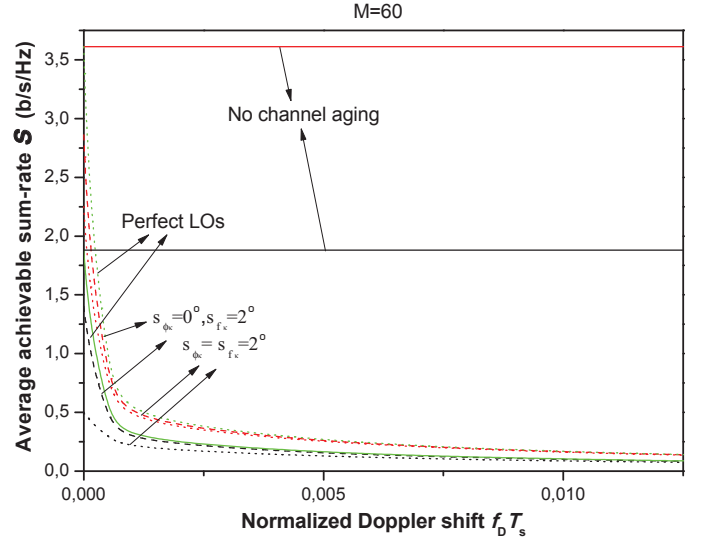

Fig. 4. Required transmit power to achieve 1 bit/s/Hz per user with MRT and RZF precoders when $M = 60$ as a function of the normalized Doppler shift for various values of phase noise. Red and black lines correspond to the theoretical sum-rates with RZF and MRT precoding, respectively. The green "solid" and "dot" lines mirror a scenario with channel aging but not phase noise in the cases of MRT and RZF, respectively. Lines parallel to x-axis represent scenarios with no channel aging.

where σ_{b}^2 is the variance of the post-processed noise at base station, $\tilde{\mathbf{y}}_{ik,0}^{\mathrm{p}}$ is a noisy observation of the effective channel from user k to the ith BS, and $p_{\mathrm{p}} = \tau p_{\mathrm{u}}$ with p_{u} being the power per user in the uplink data transmission phase.

Given that the focal point of this work is to conduct an investigation of a generalized channel aging model, and shed deeper light on channel aging as well as reveal new properties, we present only simulation results and not the corresponding deterministic equivalent analysis, which is straightforward and will distract the reader from the main objective of this part. In

VI. Conclusions

In this work, we modeled channel aging by incorporating both effects of Doppler shift coming from user mobility and phase noise due to circuitry imperfections of LOs, since both the effects contribute to the time variation of the effective channel multiplying/affecting the transmitted data. Given this novel integrated framework, we provided a joint channel-phase noise estimate. Next, the new CSI model was exploited to construct the MRT and the RZF precoders, and derived the DEs of the corresponding downlink SINRs. As attested by Monte Carlo simulations, these DEs are tight approximations for the rate performance of the studied system. In addition, the numerical results showed that the degradation due to Doppler shift dominates against phase noise. As a result, the detrimental effect of phase noise is meaningful only in low mobility conditions. Notably, the use of massive MIMO systems should be preferred even in general channel aging conditions, where, in addition, RZF behaves better than MRT as expected. Finally, we showed that in the case of MRT precoding, the required transmit power per user to achieve a certain rate can be scaled down at most by the inverse of the square root of the number of antennas, while a similar result should hold for RZF precoding as well.

Appendix A
Useful Lemmas

Lemma 1 (Matrix inversion lemma (I) [49, Eq. 2.2]): Let $\mathbf{B} \in \mathbb{C}^{M \times M}$ be Hermitian invertible. Then, for any vector $\mathbf{x} \in \mathbb{C}^M$, and any scalar $\tau \in \mathbb{C}^M$ such that $\mathbf{B} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}}$ is invertible,

$$\mathbf{x}^{\mathsf{H}}(\mathbf{B} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}})^{-1} = \frac{\mathbf{x}^{\mathsf{H}}\mathbf{B}^{-1}}{1 + \tau \mathbf{x}^{\mathsf{H}}\mathbf{B}^{-1}\mathbf{x}}.$$

Lemma 2 (Matrix inversion lemma (II) [39, Lemma 2]): Let $\mathbf{B} \in \mathbb{C}^{M \times M}$ be Hermitian invertible. Then, for any vector $\mathbf{x} \in \mathbb{C}^M$, and any scalar $\tau \in \mathbb{C}^M$ such that $\mathbf{B} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}}$ is invertible,

$$(\mathbf{B} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}})^{-1} = \mathbf{B} - \frac{\mathbf{B}^{-1}\tau \mathbf{x}\mathbf{x}^{\mathsf{H}}\mathbf{B}^{-1}}{1 + \tau \mathbf{x}^{\mathsf{H}}\mathbf{B}^{-1}\mathbf{x}}.$$

Lemma 3 (Rank-1 perturbation lemma [50, Lemma 2.1]): Let $z \in< 0$, $\mathbf{B} \in \mathbb{C}^{M \times M}$, $\mathbf{B} \in \mathbb{C}^{M \times M}$ with $\mathbf{B}$ Hermitian nonnegative-definite, and $\mathbf{x} \in \mathbb{C}^M$. Then,

$$|\operatorname{tr}\left((\mathbf{B} - z\mathbf{I}_M)^{-1} - (\mathbf{B} + \mathbf{x}\mathbf{x}^{\mathsf{H}} - z\mathbf{I}_M)^{-1}\mathbf{B}\right)| \leq \frac{\|\mathbf{B}\|}{|z|}.$$

Lemma 4 ([39, Thm. 3.7], [11, Lem. 1]): Let $\mathbf{B} \in \mathbb{C}^{M \times M}$ with uniformly bounded spectral norm (with respect to M). Consider $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}, \mathbf{y} \in \mathbb{C}^M$, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_x)$ and $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_y)$, are mutually independent and independent of $\mathbf{B}$. Then, we have

$$\frac{1}{M}\mathbf{x}^{\mathsf{H}}\mathbf{B}\mathbf{x} - \frac{1}{M}\operatorname{tr}\mathbf{B}\mathbf{\Phi}_x \xrightarrow[M \to \infty]{\text{a.s.}} 0 \qquad (39)$$

$$\frac{1}{M}\mathbf{x}^{\mathsf{H}}\mathbf{B}\mathbf{y} \xrightarrow[M \to \infty]{\text{a.s.}} 0 \qquad (40)$$

$$\mathbb{E}\left[\left|\left(\frac{1}{M}\mathbf{x}^{\mathsf{H}}\mathbf{B}\mathbf{x}\right)^2 - \left(\frac{1}{M}\operatorname{tr}\mathbf{B}\mathbf{\Phi}_x\right)^2\right|\right] \xrightarrow[M \to \infty]{\text{a.s.}} 0 \qquad (41)$$

$$\frac{1}{M^2}|\mathbf{x}^{\mathsf{H}}\mathbf{B}\mathbf{y}|^2 - \frac{1}{M^2}\operatorname{tr}\mathbf{B}\mathbf{\Phi}_x\mathbf{B}^{\mathsf{H}}\mathbf{\Phi}_y \xrightarrow[M \to \infty]{\text{a.s.}} 0. \qquad (42)$$

Lemma 5 ([51, p. 207]): Let $\mathbf{A}$, $\mathbf{B} \in \mathcal{C}^{M \times M}$ be freely independent random matrices with uniformly bounded spectral norm for all M. Further, let all the moments of the entries of $\mathbf{A}$, $\mathbf{B}$ be finite. Then,

$$\frac{1}{M}\operatorname{tr}\mathbf{A}\mathbf{B} - \frac{1}{M}\operatorname{tr}\mathbf{A}\frac{1}{M}\operatorname{tr}\mathbf{B} \xrightarrow[M \to \infty]{\text{a.s.}} 0. \qquad (43)$$

Theorem 4 ([52, Theorem 1]): Let $\mathbf{L} \in \mathbb{C}^{M \times M}$ and $\mathbf{S} \in \mathbb{C}^{M \times M}$ be Hermitian nonnegative definite matrices, and let $\mathbf{H} \in \mathbb{C}^{M \times K}$ be a random matrix with columns $\mathbf{v}_k \sim \mathcal{CN}\left(0, \frac{1}{M}\mathbf{R}_k\right)$. Assume that $\mathbf{L}$ and the matrices $\mathbf{R}_k$, $k = 1, \ldots, K$, have uniformly bounded spectral norms (with respect to M). Then, for any $\rho > 0$,

$$\frac{1}{M}\operatorname{tr}\mathbf{L}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{S} + \rho\mathbf{I}_M\right)^{-1} - \frac{1}{M}\operatorname{tr}\mathbf{L}\mathbf{T}(\rho) \xrightarrow[M_t \to \infty]{\text{a.s.}} 0,$$

where $\mathbf{T}(\rho) \in \mathbb{C}^{M \times M}$ is defined as

$$\mathbf{T}(\rho) = \left(\frac{1}{M}\sum_{k=1}^K \frac{\mathbf{R}_k}{1 + e_k(\rho)} + \mathbf{S} + \rho\mathbf{I}_M\right)^{-1},$$

and the elements of $\boldsymbol{e}(\rho) = [e_1(\rho) \cdots e_K(\rho)]^{\mathsf{T}}$ are defined as $e_k(\rho) = \lim_{t \to \infty} e_k^{(t)}(\rho)$, where for $t = 1, 2, \ldots$

$$e_k^{(t)}(\rho) = \frac{1}{M}\operatorname{tr}\mathbf{R}_k\left(\frac{1}{M}\sum_{j=1}^K \frac{\mathbf{R}_j}{1 + e_j^{(t-1)}(\rho)} + \mathbf{S} + \rho\mathbf{I}_M\right)^{-1} \quad (44)$$

with initial values $e_k^{(0)}(\rho) = \frac{1}{\rho}$ for all k.

Theorem 5 ([39, Theorem 2]): Let $\mathbf{\Theta} \in \mathbb{C}^{M \times M}$ be a Hermitian nonnegative definite matrix with uniformly bounded spectral norm (with respect to M). Under the same conditions as in Theorem 4, we have

$$\frac{1}{M}\operatorname{tr}\mathbf{L}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{S} + \rho\mathbf{I}_M\right)^{-1}\mathbf{K}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}} + \mathbf{S} + \rho\mathbf{I}_M\right)^{-1}$$
$$- \frac{1}{M}\operatorname{tr}\mathbf{L}\mathbf{T}'(\rho) \xrightarrow{\text{a.s.}} 0,$$

where $\mathbf{T}'(\rho) \in \mathbb{C}^{M \times M}$ is defined as

$$\mathbf{T}'(\rho) = \mathbf{T}(\rho)\mathbf{K}\mathbf{T}(\rho) + \mathbf{T}(\rho)\frac{1}{M}\sum_{k=1}^K \frac{\mathbf{R}_k e_k'(\rho)}{(1 + e_k(\rho))^2}\mathbf{T}(\rho)$$

with $\mathbf{T}(\rho)$ and $\boldsymbol{e}_k(\rho)$ as defined in Theorem 4 and $\boldsymbol{e}'(\rho) = \left[e_1'(\rho) \cdots e_K'(\rho)\right]^{\mathsf{T}}$ given by

$$\boldsymbol{e}'(\rho) = (\mathbf{I}_K - \mathbf{J}(\rho))^{-1}\mathbf{v}(\rho). \qquad (45)$$

The elements of $\mathbf{J}(\rho) \in \mathbb{C}^{K \times K}$ and $\mathbf{v}(\rho) \in \mathbb{C}^K$ are defined as

$$[\mathbf{J}(\rho)]_{kl} = \frac{\frac{1}{M} \operatorname{tr} \mathbf{R}_k \mathbf{T}(\rho) \mathbf{R}_l \mathbf{T}(\rho)}{M \left(1 + e_k(\rho)\right)^2}$$

and

$$[\mathbf{v}(\rho)]_k = \frac{1}{M} \operatorname{tr} \mathbf{R}_k \mathbf{T}(\rho) \mathbf{K} \mathbf{T}(\rho).$$

APPENDIX B
PROOF OF PROPOSITION 1

During the training phase, users transmit mutually orthogonal training sequences consisting of τ symbols, and it is assumed that the channel remains constant during this phase. In particular, the pilot sequences can be represented by $\mathbf{\Psi} = [\boldsymbol{\psi}_1; \cdots; \boldsymbol{\psi}_K] \in \mathbb{C}^{K \times \tau}$ with $\mathbf{\Psi}$ normalized, i.e., $\mathbf{\Psi}\mathbf{\Psi}^\mathsf{H} = \mathbf{I}_K$. Given that the channel estimation takes place at time 0, the received signal at the BS is written as

$$\mathbf{Y}_{\mathrm{p},0} = \sqrt{p_{\mathrm{p}}} \mathbf{\Theta}_{k,0} \mathbf{H}_0 \mathbf{\Psi} + \mathbf{Z}_{\mathrm{p},0}, \tag{46}$$

where p_{p} is the common average transmit power for all users, $\mathbf{\Theta}_{k,0} = \operatorname{diag}\left\{ e^{j\theta_{k,0}^{(1)}}, \dots, e^{j\theta_{k,0}^{(M)}} \right\}$ is the phase noise because of the BS and user k LOs at time 0, and $\mathbf{Z}_{\mathrm{p},0} \in \mathbb{C}^{M \times \tau}$ is the spatially white additive Gaussian noise matrix at the BS during the training phase. Note that $\theta_{k,0}^{(m)} = \phi_{m,0} + \varphi_{k,0}$, $m = 1, \dots, M$. By correlating the received signal with the training sequence $\frac{1}{\sqrt{p_{\mathrm{p}}}} \boldsymbol{\psi}_k^\mathsf{H}$ of user k, and by substituting $\mathbf{g}_{k,0} = \mathbf{\Theta}_{k,0} \mathbf{h}_{k,0}$ the BS obtains

$$\tilde{\mathbf{y}}_{k,0}^{\mathrm{p}} = \mathbf{g}_{k,0} + \frac{1}{\sqrt{p_{\mathrm{p}}}} \tilde{\mathbf{z}}_{\mathrm{p},0}, \tag{47}$$

where $\tilde{\mathbf{z}}_{\mathrm{p},0} \triangleq \mathbf{Z}_{\mathrm{p},0} \boldsymbol{\psi}_k^\mathsf{H} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{b}}^2 \mathbf{I}_M)$. After applying the MMSE estimation method [53], the effective channel estimated at the BS can be written as (36). Employing the orthogonality principle, the channel decomposes as

$$\mathbf{g}_{k,0} = \hat{\mathbf{g}}_{k,0} + \tilde{\mathbf{g}}_{k,0}, \tag{48}$$

where $\hat{\mathbf{g}}_{k,0}$ is distributed as $\mathcal{CN}\left(\mathbf{0}, \mathbf{D}_k\right)$ with $\mathbf{D}_k = \left(\mathbf{I}_M + \frac{\sigma_b^2}{p_{\mathrm{p}}} \mathbf{R}_k^{-1}\right)^{-1} \mathbf{R}_k$, and $\tilde{\mathbf{g}}_{k,0} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_k - \mathbf{D}_k)$ is the channel estimation error vector. Note that $\hat{\mathbf{g}}_k$ and $\tilde{\mathbf{g}}_k$ are statistically independent because they are uncorrelated and jointly Gaussian.

APPENDIX C
PROOF OF THEOREM 2

We consider the desired signal power

$$S_{k,n} = \frac{\lambda}{M} \left| \mathbb{E}\left[\mathbf{g}_{k,n}^\mathsf{H} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right] \right|^2. \tag{49}$$

First, we derive the DE of the normalization parameter λ. Hence, the normalization parameter can be written by means

of (14) and (19) as

$$\lambda = \frac{K}{\mathbb{E}\left[\frac{1}{M} \operatorname{tr} \mathbf{A}_n \hat{\mathbf{G}}_0 \hat{\mathbf{G}}_0^\mathsf{H} \mathbf{A}_n \right]}$$

$$= \frac{K}{\mathbb{E}\left[\sum_{i=1}^{K} \frac{1}{M} \mathbf{A}_n^2 \hat{\mathbf{g}}_{k,0} \hat{\mathbf{g}}_{k,0}^\mathsf{H} \right]}$$

$$\asymp \left(\frac{1}{K} \sum_{i=1}^{K} \frac{1}{M} \operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \right)^{-1}, \tag{50}$$

where we have applied Lemma 4. As far as the other term in (49) is concerned, we obtain

$$\frac{1}{M} \mathbb{E}\left[\mathbf{g}_{k,n}^\mathsf{H} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right] = \frac{1}{M} \mathbb{E}\left[(\hat{\mathbf{g}}_{k,0}^\mathsf{H} \mathbf{A}_n + \tilde{\mathbf{e}}_{k,n}^\mathsf{H}) \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right]$$

$$= \frac{1}{M} \mathbb{E}\left[\hat{\mathbf{g}}_{k,0}^\mathsf{H} \mathbf{A}_n \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right],$$

$$\asymp \frac{1}{M} \mathbb{E}\left[\operatorname{tr} \mathbf{A}_n \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \mathbf{D}_k \right] \tag{51}$$

$$\asymp \frac{e^{-\left(\sigma_{\varphi_k}^2 + \sigma_\phi^2\right)n}}{M} \operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k, \tag{52}$$

where in (51) and (52), we have applied Lemmas 4 and 5, respectively. Note that $\mathbf{A}_n$ and $\widetilde{\mathbf{\Theta}}_{k,n}$ commute because both are diagonal matrices. In particular, regarding the CLO setting, it holds that $\frac{1}{M} \operatorname{tr} \widetilde{\mathbf{\Theta}}_{k,n} = e^{2j(\varphi_{k,n} - \varphi_{k,0} + \phi_{k,n} - \phi_{k,0})}$, while in the case of the SLOs setup, we obtain $\frac{1}{M} \operatorname{tr} \widetilde{\mathbf{\Theta}}_{k,n} = e^{2j(\varphi_{k,n} - \varphi_{k,0}) - \sigma_\phi^2 n}$ or $\frac{1}{M} \operatorname{tr} \widetilde{\mathbf{\Theta}}_{k,n} = e^{2j(\varphi_{k,n} - \varphi_{k,0})} \frac{1}{M} \prod_{i=1}^{M} e^{-\sigma_{\phi_i}^2 n}$, when the LOs obey to identical or non-identical statistics, respectively. Herein, we focus on the former scenario regarding the BS antennas. Especially, in such case their increment variance is σ_ϕ^2. Application of the expectation operator to $\frac{1}{M} \operatorname{tr} \widetilde{\mathbf{\Theta}}_{k,n}$ gives $e^{-\left(\sigma_{\varphi_k}^2 + \sigma_\phi^2\right)n}$ for both the CLO and SLOs settings. Finally, after simple algebraic maipulations, we lead to (52). By denoting $\bar{S}_{k,n} = \lim_{M \to \infty} S_{k,n}$ the DE signal power, and by using (50) and (52), we have

$$\bar{S}_{k,n} \asymp \bar{\lambda} e^{-2\left(\sigma_{\varphi_k}^2 + \sigma_\phi^2\right)n} \delta_k^2, \tag{53}$$

where $\delta_k = \frac{1}{M} \operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k$. Now, we proceed with the derivation of each term in (18). The first term on the right hand side can be written as

$$\frac{1}{M} \operatorname{var}\left[\mathbf{g}_{k,n}^\mathsf{H} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right]$$

$$- \frac{1}{M^2} \mathbb{E}\left[\left| \tilde{\mathbf{e}}_{k,n}^\mathsf{H} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right|^2 \right] \xrightarrow[M \to \infty]{\text{a.s.}} 0, \tag{54}$$

where the property of the variance operator $\operatorname{var}[x] = \mathbb{E}[x^2] - \mathbb{E}^2[x]$ together with (8) have been used. Lemmas 4 and 5 enable us to derive this DE as

$$\frac{\lambda}{M} \operatorname{var}\left[\mathbf{g}_{k,n}^\mathsf{H} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right] \asymp \frac{\bar{\lambda}}{M} \delta_k^{'}, \tag{55}$$

where $\delta_k^{'} = \frac{1}{M} \operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \left(\mathbf{R}_k - \mathbf{A}_n^2 \mathbf{D}_k \right)$. Similarly, use of Lemmas 4 and 5 in the last term of (18) completes the proof. Thus, we have

$$\frac{\lambda}{M^2} \mathbb{E}\left[\left| \mathbf{g}_{k,n}^\mathsf{H} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n \hat{\mathbf{g}}_{i,0} \right|^2 \right] \asymp \frac{\bar{\lambda}}{M^2} \operatorname{tr} \widetilde{\mathbf{\Theta}}_{k,n} \mathbf{A}_n^2 \mathbf{D}_i \widetilde{\mathbf{\Theta}}_{k,n}^\mathsf{H} \mathbf{R}_k,$$

$$= \frac{\bar{\lambda}}{M} \delta_i^{''}, \tag{56}$$

since $\mathbf{g}_{k,n}$ and $\hat{\mathbf{g}}_{i,0}$ are mutually independent. Note here that $\delta_i'' = \frac{1}{M} \operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_i \mathbf{R}_k$.

APPENDIX D
PROOF OF THEOREM 3

First, we obtain the DE of the normalization parameter λ. Taking into account for (14) and (20), and making a simple algebraic manipulation, we lead to

$$\lambda = \frac{K}{\mathbb{E}\left[\operatorname{tr} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{G}}_0 \hat{\mathbf{G}}_0^{\mathsf{H}} \mathbf{A}_n \boldsymbol{\Sigma}\right]} \tag{57}$$

$$= \frac{K}{\mathbb{E}\left[\operatorname{tr} \boldsymbol{\Sigma} - \operatorname{tr}\left(\mathbf{Z} + aM\mathbf{I}_M\right)\boldsymbol{\Sigma}^2\right]}, \tag{58}$$

while its DE is

$$\bar{\lambda} = \frac{K}{\left(\frac{1}{M}\operatorname{tr} \mathbf{T} - \frac{1}{M}\operatorname{tr}\left(\frac{\mathbf{Z}}{M} + a\mathbf{I}_M\right)\mathbf{C}\right)}, \tag{59}$$

where we have applied Theorems 1 and 2 for $\mathbf{K} = \mathbf{I}_M$. Regarding the other term of the desired signal power, we have

$$\mathbb{E}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}\right] = \mathbb{E}\left[(\hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n + \tilde{\mathbf{e}}_{k,n}^{\mathsf{H}})\widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}\right]$$

$$= \mathbb{E}\left[\frac{\hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_k \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}{1 + \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_k \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}\right], \tag{60}$$

where Lemmas 1 and 4 have been applied in (60), while $\boldsymbol{\Sigma}_k$ is defined as

$$\boldsymbol{\Sigma}_k = \left(\widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{G}}_0 \hat{\mathbf{G}}_0^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} - \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,n} \hat{\mathbf{g}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n}\right.$$
$$\left. + \mathbf{Z} + Ma\mathbf{I}_M\right)^{-1}.$$

Exploiting Lemmas 4, 5, and Theorem 4 gives

$$\mathbb{E}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}\right] \asymp \mathbb{E}\left[\frac{\frac{1}{M}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \widetilde{\boldsymbol{\Theta}}_{k,n} \mathbf{T}}{1 + \frac{1}{M}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \mathbf{T}}\right]. \tag{61}$$

Hence, use of (59) and (61) provides the DE signal power $\bar{S}_{k,n}$ as

$$\bar{S}_{k,n} \asymp \bar{\lambda}\left(\frac{e^{-(\sigma_{\varphi_k}^2 + \sigma_\phi^2)n}\delta_k}{1 + \delta_k}\right)^2, \tag{62}$$

where $\delta_k = \frac{1}{M}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \mathbf{T}$. Note that the manipulation regarding $\widetilde{\boldsymbol{\Theta}}_{k,n}$ follows a similar analysis to (52). We continue with the derivation of each term of (18). Specifically, after applying Lemmas 1 and 4, as well as (8) to the first term in (18), we have

$$\operatorname{var}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}\right]$$
$$- \mathbb{E}\left[\left|\frac{\tilde{\mathbf{e}}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_k \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}{1 + \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_k \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}\right|^2\right] \xrightarrow[M\to\infty]{\text{a.s.}} 0, \tag{63}$$

which yields

$$\lambda\operatorname{var}\left[\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}\right] \asymp \bar{\lambda}\frac{\frac{1}{M}\delta_k'}{(1 + \delta_k)^2}, \tag{64}$$

where we have used Theorems 4 and 5 as well as Lemmas 4 and 5. Note that $\delta_k' = \frac{1}{M}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \mathbf{T}'$ and $\mathbf{K} = \mathbf{R}_k - \mathbf{A}_n^2 \mathbf{D}_k$. Next, we focus on the last term of (18), where we make use

of Theorems 4 and 5 as well as Lemmas 1 and 4 as before. In particular, if $i \neq k$, we have

$$\mathbb{E}\left[\left|\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{i,0}\right|^2\right] = \mathbb{E}\left[\left|\frac{\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_i \mathbf{A}_n \hat{\mathbf{g}}_{i,0}}{1 + \hat{\mathbf{g}}_{i,0}^{\mathsf{H}} \mathbf{A}_n \boldsymbol{\Sigma}_i \mathbf{A}_n \hat{\mathbf{g}}_{i,0}}\right|^2\right]$$

$$= \mathbb{E}\left[\frac{\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_i \mathbf{A}_n \hat{\mathbf{g}}_{i,0} \hat{\mathbf{g}}_{i,0}^{\mathsf{H}} \mathbf{A}_n \boldsymbol{\Sigma}_i \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{g}_{k,n}}{\left(1 + \hat{\mathbf{g}}_{i,0}^{\mathsf{H}} \mathbf{A}_n \boldsymbol{\Sigma}_i \mathbf{A}_n \hat{\mathbf{g}}_{i,0}\right)^2}\right]$$

$$\asymp \mathbb{E}\left[\frac{\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_i \mathbf{A}_n \mathbf{D}_i \mathbf{A}_n \boldsymbol{\Sigma}_i \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{g}_{k,n}}{(1 + \delta_i)^2}\right], \tag{65}$$

where we have taken into consideration that $\mathbf{g}_{k,n}$ and $\hat{\mathbf{g}}_{i,0}$ are mutually independent. Unfortunately, upon inspecting (65), we observe that $\boldsymbol{\Sigma}_i$ is not independent of $\hat{\mathbf{g}}_{k,0}$. For this reason, we use Lemma 2, which gives

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{ik} - \frac{\boldsymbol{\Sigma}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik}}{1 + \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}, \tag{66}$$

which introduces a new matrix $\boldsymbol{\Sigma}_{ik}$ to (65) defined as

$$\boldsymbol{\Sigma}_{ik} = \left(\widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{G}}_0 \hat{\mathbf{G}}_0^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} - \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{i,0} \hat{\mathbf{g}}_{i,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n}\right.$$
$$\left. - \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} + \mathbf{Z} + Ma\mathbf{I}_M\right)^{-1}. \tag{67}$$

By substituting (66) into (65), we obtain

$$\mathbb{E}\left[\left|\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma} \mathbf{A}_n \hat{\mathbf{g}}_{i,0}\right|^2\right] = \frac{Q_{ik}}{M(1 + \delta_i)^2}, \tag{68}$$

where Q_{ik} is given in (69) with $\widetilde{\boldsymbol{\Sigma}}_{ik} = \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik}$. The DE of each term in (69) is obtained as

$$\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik} \mathbf{A}_n^2 \mathbf{D}_i \boldsymbol{\Sigma}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{g}_{k,n} \asymp \frac{1}{M^2}\operatorname{tr} \mathbf{R}_k \mathbf{C}'' \tag{70}$$

$$\hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik} \mathbf{A}_n^2 \mathbf{D}_i \boldsymbol{\Sigma}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \asymp \frac{1}{M^2}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \mathbf{C}''$$
$$= \frac{\delta_k''}{M} \tag{71}$$

$$\mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n^2 \mathbf{D}_i \widetilde{\boldsymbol{\Theta}}_{k,n} \boldsymbol{\Sigma}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}$$
$$\asymp \frac{1}{M^2}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \mathbf{C}''$$
$$= \frac{\delta_k''}{M} \tag{72}$$

$$Q_{ik} \asymp \frac{1}{M^2}\operatorname{tr} \mathbf{R}_k \mathbf{C}'' + \frac{|\delta_k|^2 \delta_k''}{M(1 + \delta_k)^2} - 2\operatorname{Re}\left\{\frac{\delta_k \delta_k''}{M(1 + \delta_k)}\right\}, \tag{73}$$

where $\mathbf{K} = \mathbf{A}_n^2 \mathbf{D}_i$ with $\delta_k'' = \frac{1}{M}\operatorname{tr} \mathbf{A}_n^2 \mathbf{D}_k \mathbf{C}''$. This concludes the proof for the derivation of $\bar{\gamma}_{k,n}$.

$$Q_{ik} = \mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Sigma}}_{ik} \mathbf{A}_n^2 \mathbf{D}_i \widetilde{\boldsymbol{\Sigma}}_{ik}^{\mathsf{H}} \mathbf{g}_{k,n} + \frac{\left| \mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Sigma}}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \mathbf{g}_{k,0} \right|^2 \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Sigma}}_{ik} \mathbf{A}_n^2 \mathbf{D}_i \widetilde{\boldsymbol{\Sigma}}_{ik}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}{\left(1 + \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Sigma}}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0} \right)^2}$$
$$- 2\mathrm{Re} \left\{ \frac{\hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Sigma}}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{g}_{k,n} \mathbf{g}_{k,n}^{\mathsf{H}} \widetilde{\boldsymbol{\Sigma}}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n^2 \mathbf{D}_i \widetilde{\boldsymbol{\Sigma}}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}}{1 + \hat{\mathbf{g}}_{k,0}^{\mathsf{H}} \mathbf{A}_n \widetilde{\boldsymbol{\Sigma}}_{ik} \widetilde{\boldsymbol{\Theta}}_{k,n}^{\mathsf{H}} \mathbf{A}_n \hat{\mathbf{g}}_{k,0}} \right\}. \tag{69}$$

REFERENCES

[1] A. K. Papazafeiropoulos, "Downlink performance of massive MIMO under general channel aging conditions," in *Proc. IEEE Global Communications Conference (GLOBECOM 2015)*, San Diego, USA, December 2015. [Online]. Available: http://arxiv.org/abs/1509.04303

[2] "FP7 integrating project METIS (lCT 317669)." [Online]. Available: https://www.metis2020.com

[3] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, November 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 Mag.*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[5] "The 5G infrastructure public private partnership." [Online]. Available: http://5g-ppp.eu/

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

[7] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[8] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[9] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sept 2007.

[10] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one grand question," 2015. [Online]. Available: http://arxiv.org/abs/1503.06854

[11] K. Truong and R. Heath, "Effects of channel aging in massive MIMO systems," *IEEE/KICS Journal of Communications and Networks, Special Issue on massive MIMO*, vol. 15, no. 4, pp. 338–351, Aug 2013.

[12] A. Papazafeiropoulos and T. Ratnarajah, "Uplink performance of massive MIMO subject to delayed CSIT and anticipated channel prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, May 2014, pp. 3162–3165.

[13] ——, "Linear precoding for downlink massive MIMO with delayed CSIT and channel prediction," in *IEEE Wireless Communications and Networking Conference (WCNC), 2014*, April 2014, pp. 809–914.

[14] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, 2015.

[15] A. K. Papazafeiropoulos, "Impact of user mobility on optimal linear receivers in cellular networks," in *2015 IEEE International Conference on Communications, ICC 2015, London, United Kingdom, June 8-12, 2015*, 2015, pp. 2239–2244. [Online]. Available: http://dx.doi.org/10.1109/ICC.2015.7248658

[16] A. K. Papazafeiropoulos, H. Q. Ngo, M. Matthaiou, and T. Ratnarajah, "Uplink performance of conventional and massive MIMO cellular systems with delayed CSIT," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2014*, Washington, D.C., September 2014, pp. 574–579.

[17] C. Kong, C. Zhong, A. K. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Trans. on Commun.*, vol. 63, no. 12, pp. 4879–4893, 2015.

[18] ——, "Effect of channel aging on the sum-rate of uplink massive MIMO systems," in *IEEE International Symposium on Information Theory (ISIT) 2015*. IEEE, 2015, pp. 1222–1226.

[19] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.

[20] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *ITG/IEEE Work. Smart Ant. (WSA)*. IEEE, 2010, pp. 189–196.

[21] J. Qi and S. Aïssa, "On the power amplifier nonlinearity in MIMO transmit beamforming systems," *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 876–887, 2012.

[22] 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 Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sept 2012.

[23] B. Goransson, S. Grant, E. Larsson, and Z. Feng, "Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA," in *IEEE 9th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2008, pp. 496–500.

[24] E. Björnson, P. Zetterberg, and M. Bengtsson, "Optimal coordinated beamforming in the multicell downlink with transceiver impairments," in *IEEE Global Commun.Conf. (GLOBECOM), 2012*, Dec 2012, pp. 4775–4780.

[25] J. Qi and S. Aïssa, "Analysis and compensation of I/Q imbalance in MIMO transmit-receive diversity systems," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1546–1556, 2010.

[26] E. Björnson, P. Zetterberg, M. Bengtsson, and B. Ottersten, "Capacity limits and multiplexing gains of MIMO channels with transceiver impairments," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 91–94, 2013.

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

[28] A. Pitarokoilis, S. Mohammed, and E. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 711–723, Feb 2015.

[29] E. Bjornson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. on Wireless Commun.,*, vol. 14, no. 8, pp. 4353–4368, 2015.

[30] R. Krishnan, M. Khanzadi, N. Krishnan, Y. Wu, A. Amat, T. Eriksson, and R. Schober, "Linear massive MIMO precoders in the presence of phase noise–A large-scale analysis," *accepted in IEEE Trans. on Veh. Tech.*, 2015. [Online]. Available: http://arxiv.org/abs/1501.05461

[31] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inform. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov 2014.

[32] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterization," *IEEE Trans. Circuits Syst. I*, vol. 47, no. 5, pp. 655–674, May 2000.

[33] S. Bittner, E. Zimmermann, and G. Fettweis, "Iterative phase noise mitigation in MIMO-OFDM systems with pilot aided channel estimation," in *IEEE 66th Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007*, Sept 2007, pp. 1087–1091.

[34] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: Consequences and solutions," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1988–1996, Nov 2004.

[35] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1607–1616, Aug 2007.

[36] R. Krishnan, M. R. Khanzadi, N. Krishnan, A. Amat, T. Eriksson, N. Mazzali, and G. Colavolpe, "On the impact of oscillator phase noise on the uplink performance in a massive MIMO-OFDM system," *Under Review in IEEE Signal Processing Letters*, 2014. [Online]. Available: http://arxiv.org/abs/1405.0669

[37] A. Pitarokoilis, S. Mohammed, and E. Larsson, "Achievable rates of ZF receivers in massive MIMO with phase noise impairments," in *Asilomar*

Conference on Signals, Systems and Computers, 2013, Nov 2013, pp. 1004–1008.

[38] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[39] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, February 2013.

[40] K. T. Truong, A. Lozano, and R. W. Heath, "Optimal training in continuous flat-fading massive MIMO systems," in *Proceedings of 20th European Wireless Conference European Wireless 2014*, May 2014, pp. 1–6.

[41] D. Ying, F. Vook, T. Thomas, and D. Love, "Sub-sector-based codebook feedback for massive MIMO with 2D antenna arrays," in *IEEE Global Communications Conference (GLOBECOM), 2014*, Dec 2014, pp. 3702–3707.

[42] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1269–1283, September 2007.

[43] W. C. Jakes, *Microwave mobile communications*. New York: Wiley, 1974.

[44] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE T. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[45] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[46] P. Billingsley, *Probability and measure*, 3rd ed. John Wiley & Sons, Inc., 2008.

[47] V. d. Vaart and W. Aad, *Asymptotic statistics (Cambridge series in statistical and probabilistic mathematics)*. New York: Cambridge University Press, 2000.

[48] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Select. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sept 2005.

[49] J. W. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[50] Z. Bai, J. W. Silverstein *et al.*, "On the signal-to-interference ratio of CDMA systems in wireless communications," *The Annals of Applied Probability*, vol. 17, no. 1, pp. 81–101, 2007.

[51] T. Tao, *Topics in random matrix theory*. American Mathematical Soc., 2012, vol. 132.

[52] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[53] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River: Prentice Hall PTR, 1993.

Anastasios Papazafeiropoulos is currently a Research Fellow in IDCOM at the University of Edinburgh, U.K. He obtained the B.Sc in physics and the M.Sc. in electronics and computers science both with distinction from the University of Patras, Greece in 2003 and 2005, respectively. He then received the Ph.D. degree from the same university in 2010. From November 2011 through December 2012 he was with the Institute for Digital Communications (IDCOM) at the University of Edinburgh, U.K. working as a postdoctoral Research Fellow, while during 2012-2014 he was a Marie Curie Fellow at Imperial College London, U.K. His research interests span massive MIMO, 5G wireless networks, full-duplex radio, mmWave communications, random matrices theory, signal processing for wireless communications, hardware-constrained communications, and performance analysis of fading channels.