Massive MIMO with Non-Ideal Arbitrary Arrays: Hardware Scaling Laws and Circuit-Aware Design

Published in:
IEEE Transactions on Wireless Communications

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Massive MIMO with Non-Ideal Arbitrary Arrays: Hardware Scaling Laws and Circuit-Aware Design

Emil Björnson, *Member, IEEE,* Michail Matthaiou, *Senior Member, IEEE,* and Mérouane Debbah, *Fellow, IEEE*

Abstract—Massive multiple-input multiple-output (MIMO) systems are cellular networks where the base stations (BSs) are equipped with unconventionally many antennas, deployed on co-located or distributed arrays. Huge spatial degrees-of-freedom are achieved by coherent processing over these massive arrays, which provide strong signal gains, resilience to imperfect channel knowledge, and low interference. This comes at the price of more infrastructure; the hardware cost and circuit power consumption scale linearly/affinely with the number of BS antennas N. Hence, the key to cost-efficient deployment of large arrays is low-cost antenna branches with low circuit power, in contrast to today's conventional expensive and power-hungry BS antenna branches. Such low-cost transceivers are prone to hardware imperfections, but it has been conjectured that the huge degrees-of-freedom would bring robustness to such imperfections. We prove this claim for a generalized uplink system with multiplicative phase-drifts, additive distortion noise, and noise amplification. Specifically, we derive closed-form expressions for the user rates and a scaling law that shows how fast the hardware imperfections can increase with N while maintaining high rates. The connection between this scaling law and the power consumption of different transceiver circuits is rigorously exemplified. This reveals that one can make the circuit power increase as $\sqrt{N}$, instead of linearly, by careful circuit-aware system design.

Index Terms—Achievable user rates, channel estimation, massive MIMO, scaling laws, transceiver hardware imperfections.

I. INTRODUCTION

Interference coordination is the major limiting factor in cellular networks, but modern multi-antenna base stations (BSs) can control the interference in the spatial domain by coordinated multipoint (CoMP) techniques [1]–[3]. The cellular networks are continuously evolving to keep up with the rapidly increasing demand for wireless connectivity [4].

Massive densification, in terms of more service antennas per unit area, has been identified as a key to higher area throughput in future wireless networks [5]–[7]. The downside of densification is that even stricter requirements on the interference co-ordination need to be imposed. Densification can be achieved by adding more antennas to the macro BSs and/or distributing the antennas by ultra-dense operator-deployment of small BSs. These two approaches are non-conflicting and represent the two extremes of the *massive MIMO* paradigm [7]: a large co-located antenna array or a geographically distributed array (e.g., using a cloud RAN approach [8]). The massive MIMO topology originates from [9] and has been given many alternative names; for example, large-scale antenna systems (LSAS), very large MIMO, and large-scale multi-user MIMO. The main characteristics of massive MIMO are that each cell performs coherent processing on an array of hundreds (or even thousands) of active antennas, while simultaneously serving tens (or even hundreds) of users in the uplink and downlink. In other words, the number of antennas, N, and number of users per BS, K, are unconventionally large, but differ by a factor two, four, or even an order of magnitude. For this reason, massive MIMO brings unprecedented spatial degrees-of-freedom, which enable strong signal gains from coherent reception/transmit beamforming, give nearly orthogonal user channels, and resilience to imperfect channel knowledge [10].

Apart from achieving high area throughput, recent works have investigated additional ways to capitalize on the huge degrees-of-freedom offered by massive MIMO. Towards this end, [5] showed that massive MIMO enables fully distributed coordination between systems that operate in the same band. Moreover, it was shown in [11] and [12] that the transmit uplink/downlink powers can be reduced as $\frac{1}{\sqrt{N}}$ with only a minor loss in throughput. This allows for major reductions in the emitted power, but is actually bad from an overall energy efficiency (EE) perspective—the EE is maximized by increasing the emitted power with N to compensate for the increasing circuit power consumption [13].

This paper explores whether the huge degrees-of-freedom offered by massive MIMO provide robustness to transceiver hardware imperfections/impairments; for example, phase noise, non-linearities, quantization errors, noise amplification, and inter-carrier interference. Robustness to hardware imperfections has been conjectured in overview articles, such as [7]. Such a characteristic is notably important since the deployment cost and circuit power consumption of massive MIMO scales linearly with N, unless the hardware accuracy constraints can be relaxed such that low-power, low-cost hardware is deployed which is more prone to imperfections. Constant envelope

precoding was analyzed in [14] to facilitate the use of power-efficient amplifiers in the downlink, while the impact of phase-drifts was analyzed and simulated for single-carrier systems in [15] and for orthogonal frequency-division multiplexing (OFDM) in [16]. A preliminary proof of the conjecture was given in [17], but the authors therein considered only additive distortions and, thus, ignored other important characteristics of hardware imperfections. That paper showed that one can tolerate distortion variances that increase as $\sqrt{N}$ with only minor throughput losses, but did not investigate what this implies for the design of different transceiver circuits.

In this paper, we consider a generalized uplink massive MIMO system with arbitrary array configurations (e.g., co-located or distributed antennas). Based on the extensive literature on modeling of transceiver hardware imperfections (see [3], [4], [15], [18]–[24] and references therein), we propose a tractable system model that jointly describes the impact of multiplicative phase-drifts, additive distortion noise, noise amplification, and inter-carrier interference. This stands in contrast to the previous works [15]–[17], which each investigating only one of these effects. The following are the main contributions of this paper:

- We derive a new linear minimum mean square error (LMMSE) channel estimator that accounts for hardware imperfections and allows the prediction of the detrimental impact of phase-drifts.
- We present a simple and general expression for the achievable uplink user rates and compute it in closed-form, when the receiver applies maximum ratio combining (MRC) filters. We prove that the additive distortion noise and noise amplification vanish asymptotically as $N \to \infty$, while the phase-drifts remain but are not exacerbated.
- We obtain an intuitive scaling law that shows how fast we can tolerate the levels of hardware imperfections to increase with N, while maintaining high user rates. This is an analytic proof of the conjecture that massive MIMO systems can be deployed with inexpensive low-power hardware without sacrificing the expected major performance gains. The scaling law provides sufficient conditions that hold for any judicious receive filters.
- The practical implications of the scaling law are exemplified for the main circuits at the receiver, namely, the analog-to-digital converter (ADC), low noise amplifier (LNA), and local oscillator (LO). The main components of a typical receiver are illustrated in Fig. 1. The scaling law reveals the tradeoff between hardware cost, level of imperfections, and circuit power consumption. In particular, it shows how a circuit-aware design can make the circuit power consumption increase as $\sqrt{N}$ instead of N.
- The analytic results are validated numerically in a realistic simulation setup, where we consider different antenna deployment scenarios, common and separate LOs, different pilot sequence designs, and two types of receive filters. A key observation is that separate LOs can provide better performance than a common LO, since the phase-drifts average out and the interference is reduced. This is
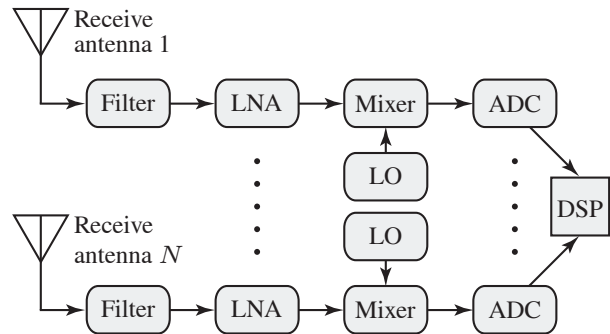
Fig. 1. Block diagram of a typical N-antenna receiver. The main circuits are shown, but these can be complemented with additional intermediate filters and amplifiers depending on the implementation. Most of the circuits affect only one antenna, whilst the LO can be either common for all antennas or different.

also rigorously supported by the analytic scaling law.

This paper extends substantially our conference papers [25] and [26], by generalizing the propagation model, generalizing the analysis according to the new model, and providing more comprehensive simulations. The paper is organized as follows: In Section II, the massive MIMO system model under consideration is presented. In Section III, a detailed performance analysis of the achievable uplink user rates is pursued and the impact of hardware imperfections is characterized, while in Section IV we provide guidelines for circuit-aware design in order to minimize the power dissipation of receiver circuits. Our theoretical analysis is corroborated with simulations in Section V, while Section VI concludes the paper.

Notation: The following notation is used throughout the paper: Boldface (lower case) is used for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. Let $\mathbf{X}^{\mathrm{T}}$, $\mathbf{X}^*$, and $\mathbf{X}^{\mathrm{H}}$ denote the transpose, conjugate, and conjugate transpose of $\mathbf{X}$, respectively. A diagonal matrix with $a_1, \ldots, a_N$ on the main diagonal is denoted as $\mathrm{diag}(a_1, \ldots, a_N)$, while $\mathbf{I}_N$ is an $N \times N$ identity matrix. The set of complex-valued $N \times K$ matrices is denoted by $\mathbb{C}^{N \times K}$. The expectation operator is denoted $\mathbb{E}\{\cdot\}$ and $\triangleq$ denotes definitions. The matrix trace function is $\mathrm{tr}(\cdot)$ and $\otimes$ is the Kronecker product. A Gaussian random variable x is denoted $x \sim \mathcal{N}(\bar{x}, q)$, where $\bar{x}$ is the mean and q is the variance. A circularly symmetric complex Gaussian random vector $\mathbf{x}$ is denoted $\mathbf{x} \sim \mathcal{CN}(\bar{\mathbf{x}}, \mathbf{Q})$, where $\bar{\mathbf{x}}$ is the mean and $\mathbf{Q}$ is the covariance matrix. The big $\mathcal{O}$ notation $f(x) = \mathcal{O}(g(x))$ means that $\left| \frac{f(x)}{g(x)} \right|$ is bounded as $x \to \infty$.

II. SYSTEM MODEL WITH HARDWARE IMPERFECTIONS

We consider the uplink of a cellular network with $L \geq 1$ cells. Each cell consists of K single-antenna user equipments (UEs) that communicate simultaneously with an array of N antennas, which can be either co-located at a macro BS or distributed over multiple fully coordinated small BSs. The analysis of our paper holds for any N and K, but we are primarily interested in massive MIMO topologies, where $N \gg K \gg 1$. The frequency-flat channel from UE k in cell l to BS j is denoted as $\mathbf{h}_{jlk} \triangleq \left[h_{jlk}^{(1)} \ldots h_{jlk}^{(N)} \right]^{\mathrm{T}} \in \mathbb{C}^{N \times 1}$

and is modeled as Rayleigh block fading. This means that it has a static realization for a coherence block of T channel uses and independent realizations between blocks.[1] The UEs' channels are independent. Each realization is complex Gaussian distributed with zero mean and covariance matrix $\mathbf{\Lambda}_{jlk} \in \mathbb{C}^{N \times N}$:

$$\mathbf{h}_{jlk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda}_{jlk}). \tag{1}$$

The covariance matrix $\mathbf{\Lambda}_{jlk} \triangleq \operatorname{diag}\left(\lambda_{jlk}^{(1)}, \ldots, \lambda_{jlk}^{(N)}\right)$ is assumed to be diagonal, which holds if the inter-antenna distances are sufficiently large and the multi-path scattering environment is rich [27].[2] The average channel attenuation $\lambda_{jlk}^{(n)}$ is different for each combination of cells, UE index, and receive antenna index n. It depends, for example, on the array geometry and the UE location. Even for co-located antennas one might have different values of $\lambda_{jlk}^{(n)}$ over the array, because of the large aperture that may create variations in the shadow fading.

The received signal $\mathbf{y}_j(t) \in \mathbb{C}^{N \times 1}$ in cell j at a given channel use $t \in \{1, \ldots, T\}$ in the coherence block is conventionally modeled as [9]–[12]

$$\mathbf{y}_j(t) = \sum_{l=1}^{L} \mathbf{H}_{jl} \mathbf{x}_l(t) + \mathbf{n}_j(t) \tag{2}$$

where the transmit signal in cell l is $\mathbf{x}_l(t) = [x_{l1}(t) \ldots x_{lK}(t)]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ and we use the notation $\mathbf{H}_{jl} = [\mathbf{h}_{jl1} \ldots \mathbf{h}_{jlK}] \in \mathbb{C}^{N \times K}$ for brevity. The scalar signal $x_{lk}(t)$ sent by UE k in cell l at channel use t is either a deterministic pilot symbol (used for channel estimation) or an information symbol from a Gaussian codebook; in any case, we assume that the expectation of the transmit energy per symbol is bounded as $\mathbb{E}\{|x_{lk}(t)|^2\} \leq p_{lk}$. The thermal noise vector $\mathbf{n}_j(t) \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ is spatially and temporally independent and has variance σ^2.

The conventional model in (2) is well-accepted for small-scale MIMO systems, but has an important drawback when applied to massive MIMO topologies: it assumes that the large antenna array consists of N high-quality antenna branches which are all perfectly synchronized. Consequently, the deployment cost and total power consumption of the circuits attached to each antenna would *at least* grow linearly with N, thereby making the deployment of massive MIMO rather questionable, if not prohibitive, from an overall cost and efficiency perspective.

In this paper, we analyze the far more realistic scenario of having inexpensive hardware-constrained massive MIMO arrays. More precisely, each receive array experiences hardware imperfections that distort the communication. The exact distortion characteristics depend generally on which modulation scheme is used; for example, OFDM [18], filter bank multicarrier (FBMC) [28], or single-carrier transmission [15].

Nevertheless, the distortions can be classified into three distinct categories: 1) received signals are shifted in phase; 2) distortion noise is added with a power proportional to the total received signal power; and 3) thermal noise is amplified and channel-independent interference is added. To draw general conclusions on how these distortion categories affect massive MIMO systems, we consider a generic system model with hardware imperfections. The received signal in cell j at a given channel use $t \in \{1, \ldots, T\}$ is modeled as

$$\mathbf{y}_j(t) = \mathbf{D}_{\boldsymbol{\phi}_j(t)} \sum_{l=1}^{L} \mathbf{H}_{jl} \mathbf{x}_l(t) + \boldsymbol{v}_j(t) + \boldsymbol{\eta}_j(t) \tag{3}$$

where the channel matrices $\mathbf{H}_{jl}$ and transmitted signals $\mathbf{x}_l(t)$ are exactly as in (2). The hardware imperfections are defined as follows:

1) The matrix $\mathbf{D}_{\boldsymbol{\phi}_j(t)} \triangleq \operatorname{diag}\left(e^{\imath \phi_{j1}(t)}, \ldots, e^{\imath \phi_{jN}(t)}\right)$ describes multiplicative phase-drifts, where $\imath$ is the imaginary unit. The variable $\phi_{jn}(t)$ is the phase-drift at the nth receive antenna in cell j at time t. Motivated by the standard phase-noise models in LOs [21], $\phi_{jn}(t)$ follows a Wiener process

$$\phi_{jn}(t) \sim \mathcal{N}(\phi_{jn}(t-1), \delta) \tag{4}$$

which equals the previous realization $\phi_{jn}(t-1)$ plus an independent Gaussian innovation of variance δ. The phase-drifts can be either independent or correlated between the antennas; for example, co-located arrays might have a common LO (CLO) for all antennas which makes the phase-drifts $\phi_{jn}(t)$ identical for all $n = 1, \ldots, N$. In contrast, distributed arrays might have separate LOs (SLOs) at each antenna, which make the drifts independent, though we let the variance δ be equal for simplicity. Both cases are considered herein.

2) The distortion noise $\boldsymbol{v}_j(t) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon}_j(t))$, where

$$\boldsymbol{\Upsilon}_j(t) \triangleq \kappa^2 \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbb{E}\{|x_{lk}(t)|^2\} \operatorname{diag}\left(|h_{jlk}^{(1)}|^2, \ldots, |h_{jlk}^{(N)}|^2\right) \tag{5}$$

for given channel realizations, where the double-sum gives the received power at each antenna. Thus, the distortion noise is independent between antennas and channel uses, and the variance at a given antenna is proportional to the current received signal power at this antenna. This model can describe the quantization noise in ADCs with gain control [19], approximate generic non-linearities [4, Chapter 14], and approximate the leakage between subcarriers due to calibration errors. The parameter $\kappa \geq 0$ describes how much weaker the distortion noise magnitude is compared to the signal magnitude.

3) The receiver noise $\boldsymbol{\eta}_j(t) \sim \mathcal{CN}(\mathbf{0}, \xi \mathbf{I}_N)$ is independent of the UE channels, in contrast to the distortion noise. This term includes thermal noise, which typically is amplified by LNAs and mixers in the receiver hardware, and interference leakage from other frequency bands and/or other networks. The receiver noise variance must

---

[1]The size of the time/frequency block where the channels are static depends on UE mobility and propagation environment: T is the product of the coherence time $\tilde{\tau}_c$ and coherence bandwidth $\tilde{W}_c$, thus $\tilde{\tau}_c = 5$ ms and $\tilde{W}_c = 100$ kHz gives $T = 500$.

[2]The analysis and main results of this paper can be easily extended to arbitrary non-diagonal covariance matrices as in [11] and [17], but at the cost of complicating the notation and expressions.
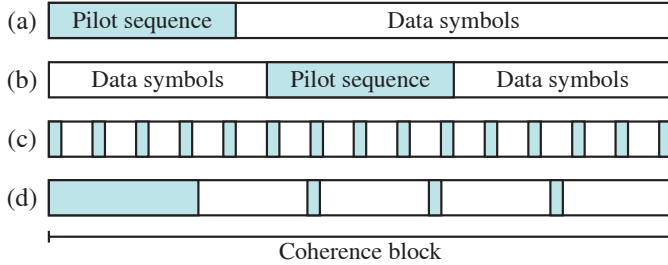
Fig. 2. Examples of different ways to distribute the B pilot symbols over the coherence block of length T: (a) beginning of block; (b) middle of block; (c) uniform pilot distribution; (d) preamble and a few distributed pilot symbols.

satisfy $\xi \geq \sigma^2$. If there is no interference leakage, $F = \frac{\xi}{\sigma^2}$ is called the noise amplification factor.

This tractable generic model of hardware imperfections at the BSs is inspired by a plethora of prior works [3], [4], [15], [18]–[24] and characterizes the joint behavior of all hardware imperfections at the BSs—these can be uncalibrated imperfections or residual errors after calibration. The model in (3) is characterized by three parameters: δ, κ, and ξ. The model is compatible with the conventional model in (2), which is obtained by setting $\xi = \sigma^2$ and $\delta = \kappa = 0$. The analysis in this paper holds for arbitrary parameter values. Section IV exemplifies the connection between imperfections in the main transceiver circuits of the BSs and the three parameters. These connections allow for circuit-aware design of massive MIMO systems.

In the next section, we derive a channel estimator and achievable UE rates for the system model in (3). By analyzing the performance as $N \to \infty$, we bring new insights into the fundamental impact of hardware imperfections (in particular, in terms of δ, κ, and ξ).

III. PERFORMANCE ANALYSIS

In this section, we derive achievable UE rates for the uplink multi-cell system in (3) and analyze how these depend on the number of antennas and hardware imperfections. We first need to specify the transmission protocol.[3] The T channel uses of each coherence block are split between transmission of uplink pilot symbols and uplink data symbols. It is necessary to dedicate $B \geq K$ channel uses for pilot transmission if the receiving array should be able to spatially separate the different UEs in the cell. The remaining $T - B$ channel uses are allocated for data transmission. The pilot symbols can be distributed in different ways: for example, placed in the beginning of the block [17], in the middle of the block [29], uniformly distributed as in the LTE standard [30], or a combination of these approaches [22]. These different cases are illustrated in Fig. 2. The time indices used for pilot transmission are denoted by $\tau_1, \ldots, \tau_B \in \{1, \ldots, T\}$, while $\mathcal{D} \triangleq \{1, \ldots, T\} \setminus \{\tau_1, \ldots, \tau_B\}$ are the time indices for data transmission.

[3]We assume that the same protocol is used in all cells, for analytic simplicity. It was shown in [12, Remark 5] that nothing substantially different will happen if this assumption is relaxed.

A. Channel Estimation under Hardware Imperfections

Based on the transmission protocol, the pilot sequence of UE k in cell j is $\tilde{\mathbf{x}}_{jk} \triangleq [x_{jk}(\tau_1) \ldots x_{jk}(\tau_B)]^T \in \mathbb{C}^{B \times 1}$. The pilot sequences are predefined and can be selected arbitrarily under the power constraints. Our analysis supports any choice, but it is reasonable to make $\tilde{\mathbf{x}}_{j1}, \ldots, \tilde{\mathbf{x}}_{jK}$ in cell j mutually orthogonal to avoid intra-cell interference (this is the reason to have $B \geq K$).

Example 1: Let $\widetilde{\mathbf{X}}_j \triangleq [\tilde{\mathbf{x}}_{j1} \ldots \tilde{\mathbf{x}}_{jK}]$ denote the pilot sequences in cell j. The simplest example of linearly independent pilot sequences (with $B = K$) is

$$\widetilde{\mathbf{X}}_j^{\text{temporal}} \triangleq \text{diag}(\sqrt{p_{j1}}, \ldots, \sqrt{p_{jK}}) \tag{6}$$

where the different sequences are temporally orthogonal since only UE k transmits at time τ_k. Alternatively, the pilot sequences can be made spatially orthogonal so that all UEs transmit at every pilot transmission time, which effectively increases the total pilot energy by a factor K. The canonical example is to use a scaled discrete Fourier transform (DFT) matrix [31]:

$$\widetilde{\mathbf{X}}_j^{\text{spatial}} \triangleq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_K & \cdots & W_K^{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & W_K^{B-1} & \cdots & W_K^{(B-1)(K-1)} \end{bmatrix} \widetilde{\mathbf{X}}_j^{\text{temporal}} \tag{7}$$

where $W_K \triangleq e^{-\iota 2\pi/K}$.

The pilot sequences can also be jointly designed across cells, to reduce inter-cell interference during pilot transmission. Since network-wide pilot orthogonality requires $B \geq LK$, which typically is much larger than the coherence block length T, practical networks need to balance between pilot orthogonality and inter-cell interference. A key design goal is to allocate non-orthogonal pilot sequences to UEs that have nearly orthogonal channel covariance matrices; for example, by making $\text{tr}(\mathbf{\Lambda}_{jjk}\mathbf{\Lambda}_{jlm})$ small for any combination of a UE k in cell j and a UE m in cell l, as suggested in [32].

For any given set of pilot sequences, we now derive estimators of the effective channels

$$\mathbf{h}_{jlk}(t) \triangleq \mathbf{D}_{\boldsymbol{\phi}_j(t)} \mathbf{h}_{jlk} \tag{8}$$

at any channel use $t \in \{1, \ldots, T\}$ and for all j, l, k. The conventional multi-antenna channel estimators from [33]–[35] cannot be applied in this paper since the generalized system model in (3) has two non-standard properties: the pilot transmission is corrupted by random phase-drifts and the distortion noise is statistically dependent on the channels. Therefore, we derive a new LMMSE estimator for the system model at hand.

Theorem 1: Let $\boldsymbol{\psi}_j \triangleq [\mathbf{y}_j^T(\tau_1) \ldots \mathbf{y}_j^T(\tau_B)]^T \in \mathbb{C}^{BN}$ denote the combined received signal in cell j from the pilot transmission. The LMMSE estimate of $\mathbf{h}_{jlk}(t)$ at any channel use $t \in \{1, \ldots, T\}$ for any l and k is

$$\hat{\mathbf{h}}_{jlk}(t) = \left(\tilde{\mathbf{x}}_{lk}^H \mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \mathbf{\Lambda}_{jlk}\right) \mathbf{\Psi}_j^{-1} \boldsymbol{\psi}_j \tag{9}$$

$$\text{SINR}_{jk}(t) = \frac{p_{jk}|\mathbb{E}\{\mathbf{v}_{jk}^{\text{H}}(t)\mathbf{h}_{jjk}(t)\}|^2}{\sum_{l=1}^{L}\sum_{m=1}^{K} p_{lm}\mathbb{E}\{|\mathbf{v}_{jk}^{\text{H}}(t)\mathbf{h}_{jlm}(t)|^2\} - p_{jk}|\mathbb{E}\{\mathbf{v}_{jk}^{\text{H}}(t)\mathbf{h}_{jjk}(t)\}|^2 + \mathbb{E}\{|\mathbf{v}_{jk}^{\text{H}}(t)\boldsymbol{v}_j(t)|^2\} + \xi\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\}} \quad (20)$$

where

$$\mathbf{D}_{\boldsymbol{\delta}(t)} \triangleq \text{diag}\left(e^{-\frac{\delta}{2}|t-\tau_1|}, \ldots, e^{-\frac{\delta}{2}|t-\tau_B|}\right), \quad (10)$$

$$\boldsymbol{\Psi}_j \triangleq \sum_{\ell=1}^{L}\sum_{m=1}^{K} \mathbf{X}_{\ell m} \otimes \boldsymbol{\Lambda}_{j\ell m} + \xi\mathbf{I}_{BN}, \quad (11)$$

$$\mathbf{X}_{\ell m} \triangleq \bar{\mathbf{X}}_{\ell m} + \kappa^2\mathbf{D}_{|\tilde{\mathbf{x}}_{\ell m}|^2}, \quad (12)$$

$$\mathbf{D}_{|\tilde{\mathbf{x}}_{\ell m}|^2} \triangleq \text{diag}\left(|x_{\ell m}(\tau_1)|^2, \ldots, |x_{\ell m}(\tau_B)|^2\right), \quad (13)$$

while the element (b_1, b_2) of $\bar{\mathbf{X}}_{\ell m} \in \mathbb{C}^{B\times B}$ is

$$[\bar{\mathbf{X}}_{\ell m}]_{b_1,b_2} = \begin{cases} |x_{\ell m}(\tau_{b_1})|^2, & b_1 = b_2, \\ x_{\ell m}(\tau_{b_1})x_{\ell m}^*(\tau_{b_2})e^{-\frac{\delta}{2}|\tau_{b_1}-\tau_{b_2}|}, & b_1 \neq b_2. \end{cases} \quad (14)$$

The corresponding error covariance matrix is

$$\mathbf{C}_{jlk}(t) = \mathbb{E}\left\{\left(\mathbf{h}_{jlk}(t) - \hat{\mathbf{h}}_{jlk}(t)\right)\left(\mathbf{h}_{jlk}(t) - \hat{\mathbf{h}}_{jlk}(t)\right)^{\text{H}}\right\}$$
$$= \boldsymbol{\Lambda}_{jlk} - \left(\tilde{\mathbf{x}}_{lk}^{\text{H}}\mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \boldsymbol{\Lambda}_{jlk}\right)\boldsymbol{\Psi}_j^{-1}(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\text{H}}\tilde{\mathbf{x}}_{lk} \otimes \boldsymbol{\Lambda}_{jlk}) \quad (15)$$

and the mean-squared error (MSE) is $\text{MSE}_{jlk}(t) = \text{tr}(\mathbf{C}_{jlk}(t))$.

Proof: The proof is given in Appendix B. ∎

It is important to note that although the channels are block fading, the phase-drifts caused by hardware imperfections make the effective channels $\mathbf{h}_{jlk}(t)$ change between every channel use. The new LMMSE estimator in Theorem 1 provides different estimates for each time index $t \in \mathcal{D}$ used for data transmission—this is a prediction, interpolation, or retrospection depending on how the pilot symbols are distributed in the coherence block (recall Fig. 2). The LMMSE estimator is the same for systems with independent and correlated phase-drifts which brings robustness to modeling errors, but also means that there exist better non-linear estimators that can exploit phase-drift correlations, though we do not pursue this issue further in this paper.

The estimator expression is simplified in the special case of co-located arrays, as shown by the following corollary.

Corollary 1: If $\boldsymbol{\Lambda}_{jlk} = \lambda_{jlk}\mathbf{I}_N$ for all j, l, and k, the LMMSE estimate in (9) simplifies to

$$\hat{\mathbf{h}}_{jlk}(t) = \left(\lambda_{jlk}\tilde{\mathbf{x}}_{lk}^{\text{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1} \otimes \mathbf{I}_N\right)\boldsymbol{\psi}_j \quad (16)$$

and the error covariance matrix in (15) becomes

$$\mathbf{C}_{jlk}(t) = \lambda_{jlk}\left(1 - \lambda_{jlk}\tilde{\mathbf{x}}_{lk}^{\text{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1}\mathbf{D}_{\boldsymbol{\delta}(t)}^{\text{H}}\tilde{\mathbf{x}}_{lk}\right)\mathbf{I}_N \quad (17)$$

where $\boldsymbol{\Omega}_j$ is the Hermitian matrix

$$\boldsymbol{\Omega}_j \triangleq \sum_{\ell=1}^{L}\sum_{m=1}^{K} \lambda_{j\ell m}\mathbf{X}_{\ell m} + \xi\mathbf{I}_B. \quad (18)$$

Next, we use these channel estimates to design receive filters and derive achievable UE rates.

B. Achievable UE Rates under Hardware Imperfections

It is difficult to compute the maximum achievable UE rates when the receiver has imperfect channel knowledge [36], and hardware imperfections are not simplifying this task. Upper bounds on the achievable rates were obtained in [17] and [37]. In this paper, we want to guarantee certain performance and thus seek simple achievable (but suboptimal) rates. The following lemma provides such rate expressions and builds upon well-known techniques from [9], [15], [36], [38], [39] for computing lower bounds on the mutual information.

Lemma 1: Suppose the receiver in cell j has complete statistical channel knowledge and applies the linear receive filters $\mathbf{v}_{jk}^{\text{H}}(t) \in \mathbb{C}^{1\times N}$, for $t \in \mathcal{D}$, to detect the signal from its kth UE. An ergodic achievable rate for this UE is

$$R_{jk} = \frac{1}{T}\sum_{t\in\mathcal{D}} \log_2\left(1 + \text{SINR}_{jk}(t)\right) \quad \text{[bit/channel use]} \quad (19)$$

where $\text{SINR}_{jk}(t)$ is given in (20) at the top of this page and all UEs use full power (i.e., $\mathbb{E}\{|x_{lk}(t)|^2\} = p_{lk}$ for all l, k).

Proof: The proof is given in Appendix C. ∎

The achievable UE rates in Lemma 1 can be computed for any choice of receive filters, using numerical methods; the MMSE filter is simulated in Section V. Note that the sum in (19) has $|\mathcal{D}| = T - B$ terms, while the pre-log factor $\frac{1}{T}$ also accounts for the B channel uses of pilot transmissions. The next theorem gives new closed-form expressions for all the expectations in (20) when using MRC receive filters.

Theorem 2: The expectations in the SINR expression (20) are given in closed form by (21)–(24), at the top of the next page, when the MRC receive filter $\mathbf{v}_{jk}^{\text{MRC}}(t) = \hat{\mathbf{h}}_{jjk}(t)$ is used in cell j. The nth column of $\mathbf{I}_N$ is denoted by $\mathbf{e}_n \in \mathbb{C}^{N\times 1}$ in this paper.

Proof: The proof is given in Appendix D. ∎

By substituting the expressions from Theorem 2 into (20), we obtain closed-form UE rates that are achievable using MRC filters. Although the expressions in (21)–(24) are easy to compute, their interpretation is non-trivial. The size of each term depends on the setup and scales differently with N; note that each trace-expression and/or sum over the antennas give a scaling factor of N. This property is easily observed in the special case of co-located antennas:

Corollary 2: If $\boldsymbol{\Lambda}_{jlk} = \lambda_{jlk}\mathbf{I}_N$ for all j, l, and k, the MRC receive filter yields:

$$\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\} = \mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\boldsymbol{\Lambda}_{jjk}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\boldsymbol{\Lambda}_{jjk}\right)\right) \tag{21}$$

$$\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jjk}(t)\} = \mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\} \tag{22}$$

$$\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\} = \mathrm{tr}\left(\boldsymbol{\Lambda}_{jlm}\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\boldsymbol{\Lambda}_{jjk}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\boldsymbol{\Lambda}_{jjk}\right)\right) \tag{23}$$

$$+ \begin{cases} \sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\lambda_{jjk}^{(n_1)}\lambda_{jlm}^{(n_1)}\lambda_{jjk}^{(n_2)}\lambda_{jlm}^{(n_2)}\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\mathbf{e}_{n_1}^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}\left(\bar{\mathbf{X}}_{lm}\otimes\mathbf{e}_{n_1}\mathbf{e}_{n_2}^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\mathbf{e}_{n_2}\right) & \text{if a CLO} \\ \left(\mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\boldsymbol{\Lambda}_{jjk}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}\otimes\boldsymbol{\Lambda}_{jlm}\right)\right)\right)^2 & \text{if SLOs} \end{cases}$$

$$+ \begin{cases} \sum_{n=1}^{N}\left(\lambda_{jjk}^{(n)}\lambda_{jlm}^{(n)}\right)^2\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\mathbf{e}_n^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}\left(\kappa^2\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\mathbf{e}_n\right) & \text{if a CLO} \\ \sum_{n=1}^{N}\left(\lambda_{jjk}^{(n)}\lambda_{jlm}^{(n)}\right)^2\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\mathbf{e}_n^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}\left(\left(\mathbf{X}_{lm}-\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}\tilde{\mathbf{x}}_{lm}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\right)\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\mathbf{e}_n\right) & \text{if SLOs} \end{cases}$$

$$\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\boldsymbol{v}_j(t)|^2\} = \kappa^2\sum_{l=1}^{L}\sum_{m=1}^{K}p_{lm}\mathrm{tr}\left(\boldsymbol{\Lambda}_{jlm}\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\boldsymbol{\Lambda}_{jjk}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\boldsymbol{\Lambda}_{jjk}\right)\right) \tag{24}$$

$$+ \kappa^2\sum_{l=1}^{L}\sum_{m=1}^{K}\sum_{n=1}^{N}p_{lm}\left(\lambda_{jjk}^{(n)}\lambda_{jlm}^{(n)}\right)^2\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\mathbf{e}_n^{\mathrm{H}}\right)\boldsymbol{\Psi}_j^{-1}(\mathbf{X}_{lm}\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}})\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}\otimes\mathbf{e}_n\right)$$

$$\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\} = N\lambda_{jjk}^2\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1}\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}$$

$$\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jjk}(t)\} = \mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\}$$

$$\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\} = \lambda_{jlm}\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\}$$

$$+ N\lambda_{jjk}^2\lambda_{jlm}^2\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1}\mathbf{X}_{lm}\boldsymbol{\Omega}_j^{-1}\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk} + N(N-1)$$

$$\times \begin{cases} \lambda_{jjk}^2\lambda_{jlm}^2\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1}\bar{\mathbf{X}}_{lm}\boldsymbol{\Omega}_j^{-1}\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk} & \text{if a CLO} \\ \lambda_{jjk}^2\lambda_{jlm}^2|\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1}\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}|^2 & \text{if SLOs} \end{cases}$$

$$\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\boldsymbol{v}_j(t)|^2\} = \kappa^2\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\}\sum_{l=1}^{L}\sum_{m=1}^{K}p_{lm}\lambda_{jlm}$$

$$+ \kappa^2\sum_{l=1}^{L}\sum_{m=1}^{K}p_{lm}N\lambda_{jjk}^2\lambda_{jlm}^2\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\boldsymbol{\Omega}_j^{-1}\mathbf{X}_{lm}\boldsymbol{\Omega}_j^{-1}\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk}.$$

As seen from this corollary, most terms scale linearly with N but there are a few terms that scale as N^2. The latter terms dominate in the asymptotic analysis below.

The difference between having a CLO and SLOs only manifests itself in the second-order moments $\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\}$. Hence, the desired signal quality is the same in both cases, while the interference terms are different; the case with the smallest interference variance $\sum_{l=1}^{L}\sum_{m=1}^{K}p_{lm}\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\}$ gives the largest rate for UE k in cell j. These second-order moments depend on the pilot sequences, channel covariance matrices, and phase-drifts. By looking at (23) in Theorem 2 (or the corresponding expression in Corollary 2), we see that the only difference is that two occurrences of $\bar{\mathbf{X}}_{\ell m}$ in the case of a CLO are replaced by $\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{\ell m}\tilde{\mathbf{x}}_{\ell m}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}$ in the case of SLOs. These terms are equal when there are no phase-drifts (i.e., $\delta = 0$), while the difference grows larger with δ. In particular, the term $\bar{\mathbf{X}}_{\ell m}$ is unaffected by the time index t, while the corresponding terms for SLOs decay as $e^{-\delta t}$ (from $\mathbf{D}_{\boldsymbol{\delta}(t)}$). The following example provides the intuition behind this result.

Example 2: The interference power in (20) consists of multiple terms of the form $\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{D}_{\boldsymbol{\phi}_j(t)}\mathbf{h}_{jlm}|^2\}$. Suppose that the receive filter is set to some constant $\mathbf{v}_{jk}(t) = \mathbf{v}_{jk}$. If a CLO is used, we have $\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\phi}_j(t)}\mathbf{h}_{jlm}|^2\} = \mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jlm}|^2\}$, which is independent of the phase-drifts since all elements of $\mathbf{v}_{jk}$ are rotated in the same way. In contrast, each component of $\mathbf{v}_{jk}$ is rotated in an independent random manner with SLOs, which reduces the average interference power since the components of the inner product $\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\phi}_j(t)}\mathbf{h}_{jlm}$ add up incoherently. Consequently, the received interference power is reduced by SLOs while it remains the same with a CLO.

To summarize, we expect SLOs to provide larger UE rates than a CLO, because the interference reduces with t when the phase-drifts are independent, at the expense of increasing the deployment cost by having N LOs. This observation is validated by simulations in Section V.

C. Asymptotic Analysis and Hardware Scaling Laws

The closed-form expressions in Theorem 2 and Corollary 2 can be applied to cellular networks of arbitrary (finite) dimensions. In massive MIMO, the asymptotic behavior of large antenna arrays is of particular interest. In this section, we assume that the N receive antennas in each cell are distributed over $A \geq 1$ spatially separated subarrays, where each subarray contains $\frac{N}{A}$ antennas. This assumption is made for analytic tractability, but also makes sense in many practical scenarios. Each subarray is assumed to have an inter-antenna distance much smaller than the propagation distances to the UEs, such that $\tilde{\lambda}_{jlk}^{(a)}$ is the average channel attenuation to all antennas in subarray a in cell j from UE k in cell l. Hence, the channel covariance matrix $\boldsymbol{\Lambda}_{jlk} \in \mathbb{C}^{N \times N}$ can be factorized as

$$\boldsymbol{\Lambda}_{jlk} = \underbrace{\mathrm{diag}\left(\tilde{\lambda}_{jlk}^{(1)}, \ldots, \tilde{\lambda}_{jlk}^{(A)}\right)}_{\triangleq \tilde{\boldsymbol{\Lambda}}_{jlk}^{(A)} \in \mathbb{C}^{A \times A}} \otimes \mathbf{I}_{\frac{N}{A}}. \tag{25}$$

By letting the number of antennas in each subarray grow large, we obtain the following property.

Corollary 3: If the MRC receive filter is used and the channel covariance matrices can be factorized as in (25), then

$$\mathrm{SINR}_{jk}(t) = \frac{p_{jk}\mathrm{Sig}_{jk}}{\sum\limits_{l=1}^{L}\sum\limits_{m=1}^{K} p_{lm}\mathrm{Int}_{jklm} - p_{jk}\mathrm{Sig}_{jk} + \mathcal{O}\left(\frac{1}{N}\right)} \quad (26)$$

where the signal part is

$$\mathrm{Sig}_{jk} = \left(\mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathsf{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\tilde{\boldsymbol{\Lambda}}_{jjk}^{(A)}\right)\widetilde{\boldsymbol{\Psi}}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}\tilde{\mathbf{x}}_{jk}\otimes\tilde{\boldsymbol{\Lambda}}_{jjk}^{(A)}\right)\right)\right)^2 \quad (27)$$

the interference terms with a CLO are

$$\mathrm{Int}_{jklm}^{\mathrm{CLO}} = \sum_{a_1=1}^{A}\sum_{a_2=1}^{A}\tilde{\lambda}_{jjk}^{(a_1)}\tilde{\lambda}_{jlm}^{(a_1)}\tilde{\lambda}_{jjk}^{(a_2)}\tilde{\lambda}_{jlm}^{(a_2)}\left(\tilde{\mathbf{x}}_{jk}^{\mathsf{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\mathbf{e}_{a_1}^{\mathsf{H}}\right)$$
$$\times \widetilde{\boldsymbol{\Psi}}_j^{-1}\left(\bar{\mathbf{X}}_{lm}\otimes\mathbf{e}_{a_1}\mathbf{e}_{a_2}^{\mathsf{H}}\right)\widetilde{\boldsymbol{\Psi}}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathsf{H}}\tilde{\mathbf{x}}_{jk}\otimes\mathbf{e}_{a_2}\right) \quad (28)$$

and the interference terms with SLOs are

$$\mathrm{Int}_{jklm}^{\mathrm{SLOs}} = \left(\mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathsf{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\tilde{\boldsymbol{\Lambda}}_{jjk}^{(A)}\right)\widetilde{\boldsymbol{\Psi}}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathsf{H}}\tilde{\mathbf{x}}_{lm}\otimes\tilde{\boldsymbol{\Lambda}}_{jlm}^{(A)}\right)\right)\right)^2 \quad (29)$$

where we have $\widetilde{\boldsymbol{\Psi}}_j \triangleq \sum_{\ell=1}^{L}\sum_{m=1}^{K}\mathbf{X}_{\ell m}\otimes\tilde{\boldsymbol{\Lambda}}_{j\ell m}^{(A)} + \xi\mathbf{I}_{AB}$, $\mathbf{e}_a$ is the ath column of $\mathbf{I}_A$, and the big $\mathcal{O}$ notation $\mathcal{O}(\frac{1}{N})$ denotes terms that go to zero as $\frac{1}{N}$ or faster when $N\to\infty$.

Proof: The proof is given in Appendix E. ∎

This corollary shows that the distortion noise and receiver noise vanish as $N\to\infty$. The phase-drifts remain, but have no dramatic impact since these affect the numerator and denominator of the asymptotic SINR in (26) in similar ways. The simulations in Section V show that the phase-drift degradations are not exacerbated in massive MIMO systems with SLOs, while the performance with a CLO improves with N but at a slower pace due to the phase-drifts.

The asymptotic SINRs are finite because both the signal power and parts of the inter-cell and intra-cell interference grow quadratically with N. This interference scaling behavior is due to so-called *pilot contamination* (PC) [9], [40], which represents the fact that a BS cannot fully separate signals from UEs that interfered to each other during pilot transmission.[4] Intra-cell PC is, conventionally, avoided by making the pilot sequences orthogonal in space; for example, by using the DFT pilot matrix $\widetilde{\mathbf{X}}_j^{\mathrm{spatial}}$ in Example 1. Unfortunately, the phase-drifts break any spatial pilot orthogonality. Hence, it is reasonable to remove intra-cell PC by assigning temporally orthogonal sequences, such as $\widetilde{\mathbf{X}}_j^{\mathrm{temporal}}$ in Example 1. Note that with temporal orthogonality the total pilot energy per UE, $\|\tilde{\mathbf{x}}_{jk}\|^2$, is reduced by $\frac{1}{K}$ since the energy per pilot symbol is constrained. Consequently, the simulations in Section V reveal that temporally orthogonal pilot sequences are only beneficial for extremely large arrays. Inter-cell PC cannot generally be removed, because there are only $B\le T$ orthogonal sequences in the whole network, but it can be mitigated by allocating the same pilot to UEs that are well separated (e.g., in terms

[4]Pilot contamination can be mitigated through semi-blind channel estimation as proposed in [41], but the UE rates will still be limited by hardware imperfections [17].

of second-order channel statistics such as different path-losses and spatial correlation [32]).

Apparently, the detrimental impact of hardware imperfections vanishes almost completely as N grows large. This result holds for any fixed values of the parameters δ, κ, and ξ. In fact, the hardware imperfections may even vanish when the hardware quality is gradually decreased with N. The next corollary formulates analytically such an important hardware scaling law.

Corollary 4: Suppose the hardware imperfection parameters are replaced as $\kappa^2\mapsto\kappa_0^2 N^{z_1}$, $\xi\mapsto\xi_0 N^{z_2}$, and $\delta\mapsto\delta_0(1+\log_e(N^{z_3}))$, for some given scaling exponents $z_1, z_2, z_3\ge 0$ and some initial values $\kappa_0, \xi_0, \delta_0\ge 0$. Moreover, let all pilot symbols be non-zero: $x_{jk}(\tau_b) > 0$ for all j, k, and b. Then, all the SINRs, $\mathrm{SINR}_{jk}(t)$, under MRC receive filtering converge to non-zero limits as $N\to\infty$ if

$$\begin{cases} \max(z_1,z_2)\le\frac{1}{2} \ \text{ and } \ z_3 = 0 & \text{for a CLO} \\ \max(z_1,z_2) + z_3\min\limits_{\tau\in\{\tau_1,\dots,\tau_B\}}\frac{\delta_0|t-\tau|}{2}\le\frac{1}{2} & \text{for SLOs.} \end{cases} \quad (30)$$

Proof: The proof is given in Appendix F. ∎

This corollary proves that we can tolerate stronger hardware imperfections as the number of antennas increases. This is a very important result for practical deployments, because we can relax the design constraints on the hardware quality as N increases. In particular, we can achieve better energy efficiency in the circuits and/or lower hardware costs by accepting larger distortions than conventionally. This property has been conjectured in overview articles, such as [7], and was proved in [17] using a simplified system model with only additive distortion noise. Corollary 4 shows explicitly that the conjecture is also true for multiplicative phase-drifts, receiver noise, and inter-carrier interference. Going a step further, Section IV exemplifies how the scaling law may impact the circuit design in practical deployments.

Since Corollary 4 is derived for MRC filtering, (30) provides a *sufficient* scaling condition for any receive filter that performs better than MRC. The scaling law for SLOs consists of two terms: $\max(z_1,z_2)$ and $z_3\min\limits_{\tau\in\{\tau_1,\dots,\tau_B\}}\frac{\delta_0|t-\tau|}{2}$. The first term $\max(z_1,z_2)$ shows that the additive distortion noise and receiver noise can be increased simultaneously and independently (as fast as $\sqrt{N}$), while the sum of the two terms manifests a tradeoff between allowing hardware imperfections that cause additive and multiplicative distortions. The scaling law for a CLO allows only for increasing the additive distortion noise and receiver noise, while the phase-drift variance should not be increased because only the signal gain (and not the interference) is reduced by phase-drifts in this case; see Example 2. Clearly, the system is particularly vulnerable to phase-drifts due to their accumulation and since they affect the signal itself; even in the case of SLOs, the second term of (30) increases with T and the variance δ can scale only logarithmically with N. Note that we can accept larger phase-drift variances if the coherence block T is small and the pilot symbols are distributed over the coherence block, which is in line with the results in [15].

IV. UTILIZING THE SCALING LAW:

CIRCUIT-AWARE DESIGN

The generic system model with hardware imperfections in (3) describes a flat-fading multi-cell channel. This channel can describe either single-carrier transmission over the full available flat-fading bandwidth as in [22] or one of the subcarriers in a system based on multi-carrier modulation; for example, OFDM or FBMC as in [18], [28]. To some extent, it can also describe single-carrier transmission over frequency-selective channels as in [15]. The mapping between the imperfections in a certain circuit in the receiving array to the three categories of distortions (defined in Section II) depends on the modulation scheme. For example, the multiplicative distortions caused by phase-noise leads also to inter-carrier interference in OFDM which is an additive noise-like distortion.

In this section, we exemplify what the scaling law in Corollary 4 means for the circuits depicted in Fig. 1. In particular, we show that the scaling law can be utilized for circuit-aware system design, where the cost and power dissipation per circuit will be gradually decreased to achieve a sub-linear cost/power scaling with the number of antennas. For clarity of presentation, we concentrate on single-carrier transmission over flat-fading channels, but mention briefly if the interpretation might change for multi-carrier modulation.

A. Analog-to-Digital Converter (ADC)

The ADC quantizes the received signal to a b bit resolution. Suppose the received signal power is P_{signal} and that automatic gain control is used to achieve maximum quantization accuracy irrespective of the received signal power. In terms of the originally received signal power P_{signal}, the quantization in single-carrier transmission can be modeled as reducing the signal power to $(1 - 2^{-2b})P_{\text{signal}}$ and adding uncorrelated quantization noise with power $2^{-2b}P_{\text{signal}}$ [19, Eq. (17)]. This model is particularly accurate for high ADC resolutions. We can include the quantization noise in the channel model (3) by normalizing the useful signal. The quantization noise is included in the additive distortion noise $v_j(t)$ and contributes to κ^2 with $\frac{2^{-2b}}{1-2^{-2b}}$, while the receiver noise variance ξ is scaled by a factor $\frac{1}{1-2^{-2b}}$ due to the normalization. The scaling law in Corollary 4 allows us to increase the variance κ^2 as N^{z_1} for $z_1 \leq \frac{1}{2}$. This corresponds to reducing the ADC resolution by around $\frac{z_1}{2}\log_2(N)$ bits, which reduces cost and complexity. For example, we can reduce the ADC resolution per antenna by 2 bits if we deploy 256 antennas instead of one. For very large arrays, it is even sufficient to use 1-bit ADCs (cf. [42]).

The power dissipation of an ADC, P_{ADC}, is proportional to 2^{2b} [19, Eq. (14)] and can, thus, be decreased approximately as $1/N^{z_1}$. If each antenna has a separate ADC, the total power NP_{ADC} increases with N but proportionally to N^{1-z_1}, for $z_1 \leq \frac{1}{2}$, instead of N, due to the gradually lower ADC resolution. The scaling can thus be made as small as $\sqrt{N}$.

B. Low Noise Amplifier (LNA)

The LNA is an analog circuit that amplifies the received signal. It is shown in [43] that the behavior of an LNA is characterized by the figure-of-merit (FoM) expression

$$\text{FoM}_{\text{LNA}} = \frac{G}{(F-1)P_{\text{LNA}}} \quad (31)$$

where $F \geq 1$ is the noise amplification factor, G is the amplifier gain, and P_{LNA} is the power dissipation in the LNA. Using this notation, the LNA contributes to the receiver noise variance ξ with $F\sigma^2$. For optimized LNAs, FoM_{LNA} is a constant determined by the circuit architecture [43]; thus, FoM_{LNA} basically scales with the hardware cost. The scaling law in Corollary 4 allows us to increase ξ as N^{z_2} for $z_2 \leq \frac{1}{2}$. The noise figure, defined as $10\log_{10}(F)$, can thus be increased by $z_2 10\log_{10}(N)$ dB. For example, at $z_2 = \frac{1}{2}$ we can allow an increase by 10 dB if we deploy 100 antennas instead of one.

For a given circuit architecture, the invariance of the FoM_{LNA} in (31) implies that we can decrease the power dissipation (roughly) proportional to $1/N^{z_2}$. Hence, we can make the total power dissipation of the N LNAs, NP_{LNA}, increase as N^{1-z_2} instead of N by tolerating higher noise amplification. The scaling can thus be made as small as $\sqrt{N}$.

C. Local Oscillator (LO)

Phase noise in the LOs is the main source of multiplicative phase-drifts and changes the phases gradually at each channel use. The average amount of phase-drifts that occurs under a coherence block is δT and depends on the phase-drift variance δ and the block length T. If the LOs are free-running, the phase noise is commonly modeled by the Wiener process (random walk) defined in Section II [15], [21]–[23], [44] and the phase noise variance is given by

$$\delta = 4\pi^2 f_c^2 T_s \zeta \quad (32)$$

where f_c is the carrier frequency, T_s is the symbol time, and ζ is a constant that characterizes the quality of the LO [21]. If δ and/or T are small, such that $\delta T \approx 0$, the channel variations dominate over the phase noise. However, phase noise can play an important role when modeling channels with large coherence time (e.g., fixed indoor users, line-of-sight, etc.) and as the carrier frequency increases (since $\delta = \mathcal{O}(f_c^2)$ while the Doppler spread reduces T as $\mathcal{O}(f_c^{-1})$ [22]). Relevant examples are mobile broadband access to homes and WiFi at millimeter frequencies.

The power dissipation P_{LO} of the LO is coupled to ζ, such that $P_{\text{LO}}\zeta \approx \text{FoM}_{\text{LO}}$ where the FoM value FoM_{LO} depends on the circuit architecture [21], [45] and naturally on the hardware cost. For a given architecture, we can allow larger δ and, thereby, decrease the power P_{LO}. The scaling law in Corollary 4 allows us to increase δ as $(1 + \log_e(N^{z_3}))$ when using SLOs. The power dissipation per LO can then be reduced as $\frac{1}{1+z_3\log_e(N)}$. This reduction is only logarithmic in N, which stands in contrast to the $1/\sqrt{N}$ scalings for ADCs and LNAs (achieved by $z_1 = z_2 = \frac{1}{2}$). Since linear increase is much faster than logarithmic decay, the total power NP_{LO} with SLOs increases almost linearly with N; thus, the benefit is mostly cost and design related. In contrast, the phase noise variance cannot be scaled when having a CLO, because

massive MIMO only relaxes the design of circuits that are placed independently at each antenna branch.

Imperfections in the LOs also cause inter-carrier interference in OFDM systems, since the subcarrier orthogonality is broken [18]. When inter-carrier interference is created at the receiver side it depends on the channels of other subcarriers. It is thus uncorrelated with the useful channel in (3) and can be included in the receiver noise term. Irrespective of the type of LOs, the severity of inter-carrier interference is suppressed by $z_2 10 \log_{10}(N)$ dB according to Corollary 4. Hence, massive MIMO is less vulnerable to in-band distortions than conventional systems.

The phase-noise variance formula in (32) gives other possibilities than decreasing the circuit power. In particular, one can increase the carrier frequency f_c with N by using Corollary 4. This is an interesting observation since massive MIMO has been identified as a key enabler for millimeter-wave communications [6], in which the phase noise is more severe since the variance in (32) increases quadratically with the carrier frequency f_c. Fortunately, massive MIMO with SLOs has an inherent resilience to phase noise.

D. Non-Linearities

Although the physical propagation channel is linear, practical systems can exhibit non-linear behavior due to a variety of reasons; for examples, non-linearities in filters, converters, mixers, and amplifiers [18] as well as passive intermodulation caused by various electro-thermal phenomena [46]. Such non-linearities are often modeled by power series or Volterra series [46], but since we consider a system with Gaussian transmit signals the Bussgang theorem can be applied to simplify the characterization [4], [24]. For a Gaussian variable X and any non-linear function $g(\cdot)$, the Bussgang theorem implies that $g(X) = cX + V$, where c is a scaling factor and V is a distortion uncorrelated with X; see [24, Eq. (15)]. If we let $g(X)$ describe a nonlinear component and let X be the useful signal, the impact of non-linearities can be modeled by a scaling of the useful signal and an additional distortion term. Depending on the nature of each non-linearity, the corresponding distortion is either included in the distortion noise or the receiver noise.[5] The scaling factor c of the useful signal is removed by scaling κ^2 and ξ by $\frac{1}{|c|^2}$.

V. NUMERICAL ILLUSTRATIONS

Our analytic results are corroborated in this section by studying the uplink in a cell surrounded by 24 interfering cells, as shown in Fig. 3. Each cell is a square of $250\,\text{m} \times 250\,\text{m}$ and we compare two topologies: (a) co-located deployment of N antennas in the middle of the cell; and (b) distributed deployment of 4 subarrays of $\frac{N}{4}$ antennas at distances of 62.5 m from the cell center. To mimic a simple user scheduling algorithm, each cell is divided into 8 virtual sectors and one UE is picked with a uniform distribution in each sector (with a

---

[5]The distortion from non-linearities are generally non-Gaussian, but this has no impact on our analysis because the achievable rates in Lemma 1 were obtained by making the worst-case assumption of all additive distortions being Gaussian distributed.

minimum distance of $25\,\text{m}$ from any array location). We thus have $K = 8$ and use $B = 8$ as pilot length in this section. Each sector is allocated an orthogonal pilot sequence, while the same pilot is reused in the same sector of all other cells. The channel attenuations are modeled as [47]

$$\lambda_{jlk}^{(n)} = \frac{10^{s_{jlk}^{(n)} - 1.53}}{(d_{jlk}^{(n)})^{3.76}} \tag{33}$$

where $d_{jlk}^{(n)}$ is the distance in meters between receive antenna n in cell j and UE k in cell l and $s_{jlk}^{(n)} \sim \mathcal{N}(0, 3.16)$ is shadow-fading (it is the same for co-located antennas but independent between the 4 distributed arrays). We consider statistical power control with $p_{jk} = \frac{\rho}{\frac{1}{N}\sum_{n=1}^{N}\lambda_{jjk}^{(n)}}$ to achieve an average received signal power of ρ over the receive antennas. The thermal noise variance is $\sigma^2 = -174$ dBm/Hz. We consider average SNRs, ρ/σ^2, of 5 and 15 dB, leading to reasonable transmit powers (below 200 mW over a 10 MHz bandwidth) for UEs at cell edges.

A. Comparison of Deployment Scenarios

We first compare the co-located and distributed deployments in Fig. 3. We consider the MRC filter, set the coherence block to $T = 500$ channel uses (e.g., 5 ms coherence time and 100 kHz coherence bandwidth), use the DFT-based pilot sequences of length $B = 8$, and send these in the beginning of the coherence block. The results are averaged over different UE locations.

The average achievable rates per UE are shown in Fig. 4 for $\rho/\sigma^2 = 5$ dB, using either ideal hardware or imperfect hardware with $\kappa = 0.0156$, $\xi = 1.58\sigma^2$, and $\delta = 1.58 \cdot 10^{-4}$. These parameter values were not chosen arbitrarily, but based on the circuit examples in Section IV. More specifically, we obtained $\kappa = 2^{-b}/\sqrt{1 - 2^{-2b}} = 0.0156$ by using $b = 6$ bit ADCs and $\xi = \frac{F\sigma^2}{1 - 2^{-2b}} = 1.58\sigma^2$ for a noise amplification factor of $F = 2$ dB. The phase noise variance $\delta = 1.58 \cdot 10^{-4}$ was obtained from (32) by setting $f_c = 2\,\text{GHz}$, $T_s = 10^{-7}\,\text{s}$, and $\zeta = 10^{-17}$. Note that the curves in Fig. 4 are based on the analytic results in Theorem 2, while the marker symbols correspond to Monte Carlo simulations of the expectations in (20). The perfect match validates the analytic results.

Looking at Fig. 4, we see that the tractable ergodic rate from Lemma 1 approaches well the slightly higher achievable rate from [12, Eq. (39)]. Moreover, we see that the hardware imperfections cause small rate losses when the number of antennas, N, is small. However, the large-N behavior depends strongly on the oscillators: the rate loss is small for SLOs at any N, while it can be very large if a CLO is used when N is large (e.g., 25% rate loss at $N = 400$). This important property was explained in Example 2 and the simple explanation is that the effect of phase noise averages out with SLOs, but at the cost of adding more hardware.

Fig. 4 also shows that the distributed massive MIMO deployment achieves roughly twice the rates of co-located massive MIMO. This is because distributed arrays can exploit both the proximity gains (normally achieved by small cells) and the array gains and spatial resolution of coherent processing over many antennas.
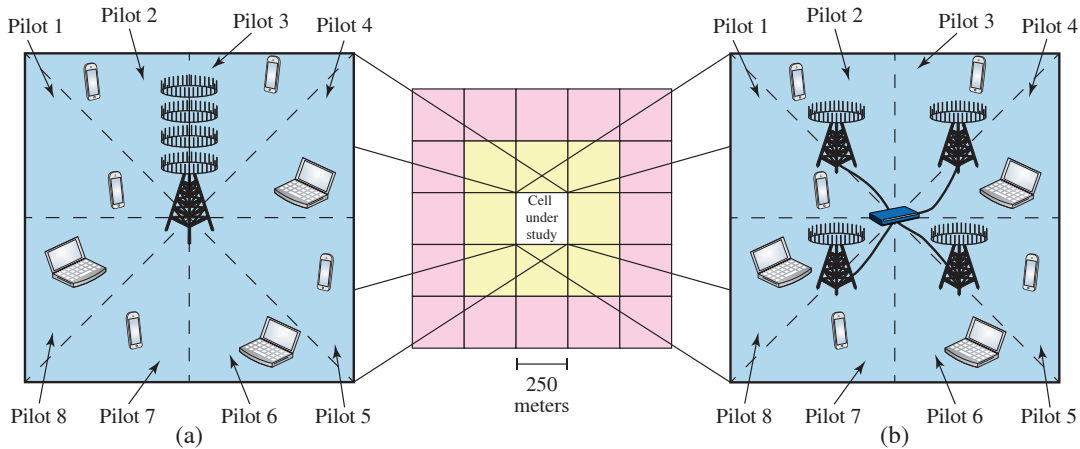
Fig. 3. The simulations consider the uplink of a cell surrounded by two tiers of interfering cells. Each cell contains $K = 8$ UEs that are uniformly distributed in different parts of the cell. Two site deployments are considered: (a) N co-located antennas in the middle of the cell; and (b) $N/4$ antennas at 4 distributed arrays.
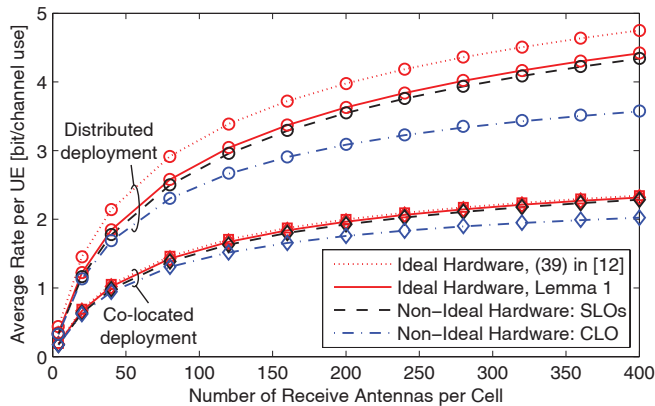

Fig. 4. Achievable rates with MRC filter and either ideal hardware or imperfections given by $(\kappa, \xi, \delta) = (0.0156, 1.58\sigma^2, 1.58 \cdot 10^{-4})$. Co-located and distributed antenna deployments are compared, as well as, a CLO and SLOs.
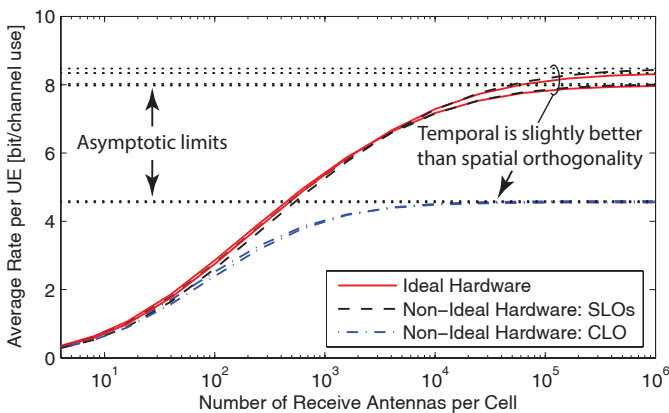

Fig. 5. Average UE rate with MRC filter for different numbers of antennas, different hardware imperfections, and spatially or temporally orthogonal pilots. Note the logarithmic horizontal scale which is used to demonstrate the asymptotic behavior.

B. Validation of Asymptotic Behavior

Next, we illustrate the asymptotic behavior of the UE rates (with MRC filter) as $N \to \infty$. For the sake of space, we only

consider the distributed deployment in Fig. 3, while a similar figure for the co-located deployment is available in [25]. Fig. 5 shows the UE rates as a function of the number of antennas, for ideal hardware and the same hardware imperfections as in the previous figure. The simulation validates the convergence to the limits derived in Corollary 3, but also shows that the convergence is very slow—we used logarithmic scale on the horizontal axis because $N = 10^6$ antennas are required for convergence for ideal hardware and for hardware imperfections with SLOs, while $N = 10^4$ antennas are required for hardware imperfections with a CLO. The performance loss for hardware imperfections with SLOs is almost negligible, while the loss when having a CLO grows with N and approaches 50 %.

Two types of pilot sequences are also compared in Fig. 5: the temporally orthogonal pilots in (6) and the spatially orthogonal DFT-based pilots in (7). As discussed in relation to Corollary 3, temporal orthogonality provides slightly higher rates in the asymptotic regime (since the phase noise cannot break the temporal pilot orthogonality). However, this gain is barely visible in Fig. 5 and only kicks in at impractically large N. Since temporally orthogonal pilots use K times less pilot energy, they are the best choice in this simulation. However, if the average SNR is decreased then spatially orthogonal pilots can be used to improve the estimation accuracy.

C. Impact of Coherence Block Length

Next, we illustrate how the length of the coherence block, T, affects the UE rates with the MRC filter. We consider a practical number of antennas, $N = 240$, while having $\rho/\sigma^2 = 5$ dB and imperfections with $\{\kappa, \xi, \delta\} = \{0.0156, 1.58\sigma^2, 1.58 \cdot 10^{-4}\}$, as before. The UE rates are shown in Fig. 6 as a function of T. We compare two ways of distributing the pilot sequences over the coherence block: in the beginning or in the middle (see (a) and (b) in Fig. 2).

With ideal hardware, the pilot distribution has no impact. We observe in Fig. 6 that the average UE rates are slightly increasing with T. This is because the pre-log penalty of using only $T - B$ out of T channel uses for data transmission is
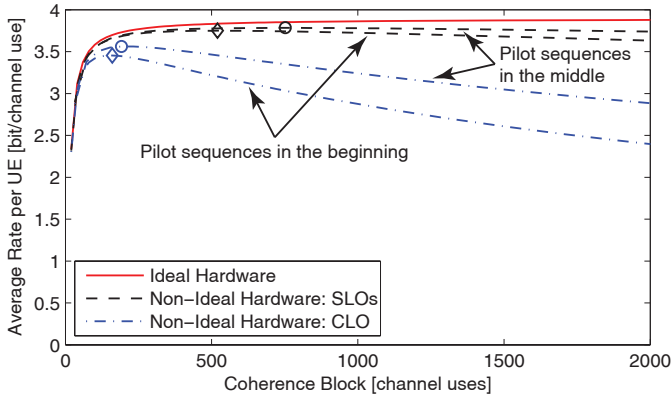
Fig. 6. Average UE rate with MRC filter as a function of the coherence block length, for different pilot sequence distributions. The maximum is marked at each curve and is the preferable operating point for the transmission protocol.

Fig. 7. Average UE rate with MRC filter for different numbers of antennas, N, and with hardware imperfections that are either fixed or increase with N as in Corollary 4.

smaller when T is large (and B is fixed). In the case with hardware imperfections, Fig. 6 shows that the rates increase with T for small T (for the same reason as above) and then decrease with T since phase-drifts accumulate over time.

Interestingly, slightly higher rates are achieved and larger coherence blocks can be handled if the pilot sequences are sent in the middle of the coherence block (instead of the beginning) since the phase drifts only accumulate half as much. From an implementation perspective it is, however, better to put pilot sequences in the beginning, since then there is no need to buffer the incoming signals while waiting for the pilots that enable computation of receive filters.

Fig. 6 shows, once again, that systems with SLOs have higher robustness to phase-drifts than systems with a CLO. To make a fair comparison, we need to consider that the coherence block is a modeling concept—we can always *choose* a transmission protocol with a smaller T than prescribed by the coherence block length, at the cost of increasing the pilot overhead B/T. Hence, it is the maximum at each curve, indicated by markers, which is the operating point to compare. The difference between SLOs and a CLO is much smaller when comparing the maxima, but these are achieved at very different T-values; the transmission protocol should send pilots much more often when having a CLO. The true optimum is achieved by maximizing over T and B, and probably by spreading the pilots to reduce the accumulation of phase drifts.

D. Hardware Scaling Laws with Different Receive Filters

Next, we illustrate the scaling laws for hardware imperfections established in Corollary 4 and set $\rho/\sigma^2 = 15$ dB to emphasize this effect. We focus on the distributed scenario for $T = 500$, since the co-located scenario behaves similarly and can be found in [26]. Using the notation from the scaling law, we set the baseline hardware imperfections to $(\kappa_0, \xi_0, \delta_0) = (0.05, 3\sigma^2, 7 \cdot 10^{-5})$ and increase these with N using different values on the exponents z_1, z_2, and z_3.

The UE rates with MRC filters are given in Fig. 7 for ideal hardware, fixed hardware imperfections, and imperfections that either increase according to the scaling law or faster than the law (observe that we always have $z_3 = 0$ for a CLO).
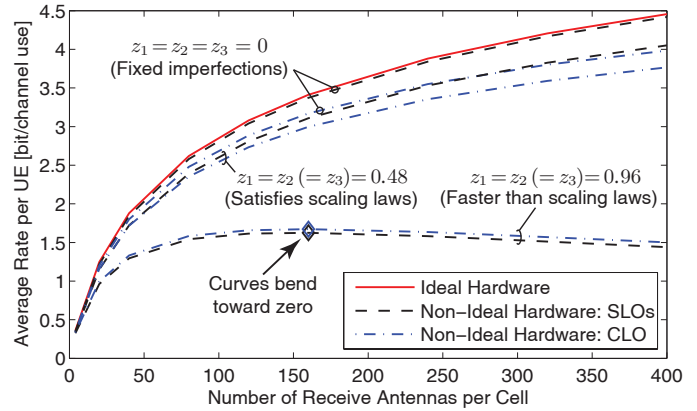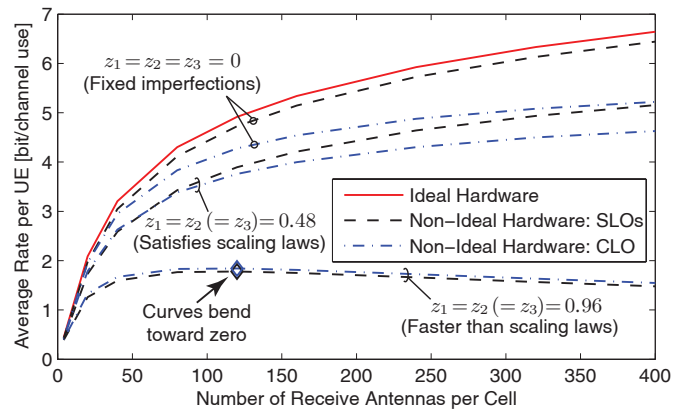
Fig. 8. Average UE rate with MMSE filter (in (34)) for different numbers of antennas, N, and with hardware imperfections that are either fixed or increase with N as in Corollary 4.

As expected, the z-combinations that satisfy the scaling laws give small performance losses, while the bottom curves go asymptotically to zero since the law is not fulfilled (the points where the curves bend downwards are marked).

We have considered the MRC filter since its low computational complexity is attractive for massive MIMO topologies. MRC provides a performance baseline for other receive filters which typically have higher complexity. In Fig. 8 we consider the (approximate) MMSE receive filter

$$\mathbf{v}_{jk}^{\mathrm{MMSE}}(t)$$
$$\triangleq \left(\sum_{l=1}^{L} \sum_{m=1}^{K} p_{lm} \big(\mathbf{G}_{jlm}(t) + \kappa^2 \mathbf{D}_{\mathbf{G}_{jlm}(t)} \big) + \xi \mathbf{I}_M \right)^{-1} \hat{\mathbf{h}}_{jjk}(t) \tag{34}$$

where $\mathbf{G}_{jlm}(t) \triangleq \hat{\mathbf{h}}_{jlm}(t)\hat{\mathbf{h}}_{jlm}^{\mathrm{H}}(t) + \mathbf{C}_{jlm}(t)$ and $\mathbf{D}_{\mathbf{G}_{jlm}(t)}$ is a diagonal matrix where the diagonal elements are the same as in $\mathbf{G}_{jlm}(t)$. By comparing Figs. 7 and 8 (notice the different scales), we observe that the MMSE filter provides higher rates than the MRC filter. Interestingly, the losses due to hardware imperfections behave similarly but are larger for MMSE. This is because the MMSE filter exploits spatial interference suppression which is sensitive to imperfections.

VI. Conclusion

Massive MIMO technology can theoretically improve the spectral and energy efficiencies by orders of magnitude, but to make it a commercially viable solution it is important that the N antenna branches can be manufactured using low-cost and low-power components. As exemplified in Section IV, such components are prone to hardware imperfections that distort the communication and limit the achievable performance.

In this paper, we have analyzed the impact of such hardware imperfections at the BSs by studying an uplink communication model with multiplicative phase-drifts, additive distortion noise, noise amplifications, and inter-carrier interference. The system model can be applied to both co-located and distributed antenna arrays. We derived a new LMMSE channel estimator/predictor and the corresponding achievable UE rates with MRC. Based on these closed-form results, we prove that only the phase-drifts limit the achievable rates as $N \to \infty$. This showcases that massive MIMO systems are robust to hardware imperfections, which is a property that has been conjectured in prior works (but only proved for simple models with one type of imperfection). This phenomenon can be attributed to the fact that distortions are uncorrelated with the useful signals and, thus, add non-coherently during the receive processing.

Particularly, we established a scaling law showing that the variance of the distortion noise and receiver noise can increase simultaneously as $\sqrt{N}$. If the phase-drifts are independent between the antennas, we can also tolerate an increase of the phase-drift variance with N, but only logarithmically. If the phase-drifts are the same over the antennas (e.g., if a CLO is used), then the phase-drift variance cannot increase. The numerical results show that there are substantial performance benefits of using separate oscillators at each antenna branch instead of a common oscillator. The difference in performance might be smaller if the LMMSE estimator is replaced by a Kalman filter that exploits the exact distribution of the phase-drifts [16], [17]. Interestingly, the benefit of having SLOs remains also under idealized uplink conditions (e.g., perfect CSI, no interference, and high SNR [48]). In any case, the transmission protocol must be adapted to how fast the phase-drifts deteriorate performance. The scaling law was derived for MRC but provides a sufficient condition for other judicious receive filters, like the MMSE filter. We also exemplified what the scaling law means for different circuits in the receiver (e.g., ADCs, LNAs, and LOs). This quantifies how fast the requirements on the number of quantization bits and the noise amplification can be relaxed with N. It also shows that a circuit-aware design can make the total circuit power consumption of the N ADCs and LNAs increase as $\sqrt{N}$, instead of N which it would conventionally be the case.

A natural extension to this paper would be to consider also the downlink with hardware imperfections at the BSs. If maximum ratio transmission (MRT) is used for precoding, then more-or-less the same expectations as in the uplink SINRs in (20) will show up in the downlink SINRs but at different places [49]. Hence, we believe that similar closed-form rate expressions and scaling laws for the levels of hardware imperfections can be derived. The analytic details and interpretations are, however, outside the scope of this paper.

Appendix A: A Useful Lemma

Lemma 2: Let $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ and consider some deterministic matrix $\mathbf{M}$. It holds that

$$\mathbb{E}\left\{|\mathbf{u}^{\mathrm{H}}\mathbf{M}\mathbf{u}|^2\right\} = |\mathrm{tr}(\mathbf{\Lambda}\mathbf{M})|^2 + \mathrm{tr}(\mathbf{\Lambda}\mathbf{M}\mathbf{\Lambda}\mathbf{M}^{\mathrm{H}}). \qquad (35)$$

Proof: This lemma follows from straightforward computation, by exploiting that $\mathbf{u}^{\mathrm{H}}\mathbf{M}\mathbf{u} = \tilde{\mathbf{u}}^{\mathrm{H}}\mathbf{\Lambda}^{1/2}\mathbf{M}\mathbf{\Lambda}^{1/2}\tilde{\mathbf{u}} = \sum_{i,j} \tilde{u}_i^*[\mathbf{\Lambda}^{1/2}\mathbf{M}\mathbf{\Lambda}^{1/2}]_{i,j}\tilde{u}_j$ where $\tilde{\mathbf{u}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. ∎

Appendix B: Proof of Theorem 1

We exploit the fact that $\hat{\mathbf{h}}_{jlk}(t) = \mathbb{E}\{\mathbf{h}_{jlk}(t)\boldsymbol{\psi}_j^{\mathrm{H}}\}\left(\mathbb{E}\{\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}}\}\right)^{-1}\boldsymbol{\psi}_j$ is the general expression of an LMMSE estimator [33, Ch. 12]. Since the additive distortion and receiver noises are uncorrelated with $\mathbf{h}_{jlk}(t)$ and the UEs' channels are independent, we have that

$$\mathbb{E}\{\mathbf{h}_{jlk}(t)\boldsymbol{\psi}_j^{\mathrm{H}}\}$$
$$= \mathbb{E}\{\mathbf{D}_{\boldsymbol{\phi}_j(t)}\mathbf{h}_{jlk}\mathbf{h}_{jlk}^{\mathrm{H}}[\mathbf{D}_{\boldsymbol{\phi}_j(\tau_1)}^{\mathrm{H}}x_{lk}^*(\tau_1) \ldots \mathbf{D}_{\boldsymbol{\phi}_j(\tau_B)}^{\mathrm{H}}x_{lk}^*(\tau_B)]\}$$
$$= \mathbf{\Lambda}_{jlk}\mathbb{E}\{[\mathbf{D}_{\boldsymbol{\phi}_j(t)}\mathbf{D}_{\boldsymbol{\phi}_j(\tau_1)}^{\mathrm{H}}x_{lk}^*(\tau_1) \ldots \mathbf{D}_{\boldsymbol{\phi}_j(t)}\mathbf{D}_{\boldsymbol{\phi}_j(\tau_B)}^{\mathrm{H}}x_{lk}^*(\tau_B)]\}$$
$$= \mathbf{\Lambda}_{jlk}[x_{lk}^*(\tau_1)e^{-\frac{\delta}{2}|t-\tau_1|}\mathbf{I}_N \ldots x_{lk}^*(\tau_B)e^{-\frac{\delta}{2}|t-\tau_B|}\mathbf{I}_N]$$
$$= \tilde{\mathbf{x}}_{lk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \mathbf{\Lambda}_{jlk} \qquad (36)$$

since $\mathbb{E}\{e^{\iota\phi_{n,t_1}}e^{-\iota\phi_{n,t_2}}\} = e^{-\frac{\delta}{2}|t_1-t_2|}$ and by exploiting the fact that diagonal matrices commute. Furthermore, we have that

$$\mathbb{E}\{\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}}\}$$
$$= \sum_{\ell=1}^{L}\sum_{m=1}^{K}\mathbb{E}\left\{[\mathbf{D}_{\boldsymbol{\phi}_j(\tau_1)}^{\mathrm{H}}x_{\ell m}^*(\tau_1) \ldots \mathbf{D}_{\boldsymbol{\phi}_j(\tau_B)}^{\mathrm{H}}x_{\ell m}^*(\tau_B)]^{\mathrm{H}}\mathbf{h}_{j\ell m}\right.$$
$$\times \left. \mathbf{h}_{j\ell m}^{\mathrm{H}}[\mathbf{D}_{\boldsymbol{\phi}_j(\tau_1)}^{\mathrm{H}}x_{\ell m}^*(\tau_1) \ldots \mathbf{D}_{\boldsymbol{\phi}_j(\tau_B)}^{\mathrm{H}}x_{\ell m}^*(\tau_B)]\right\}$$
$$+ \mathbb{E}\{[\boldsymbol{v}_j^{\mathrm{H}}(\tau_1) \ldots \boldsymbol{v}_j^{\mathrm{H}}(\tau_B)]^{\mathrm{H}}[\boldsymbol{v}_j^{\mathrm{H}}(\tau_1) \ldots \boldsymbol{v}_j^{\mathrm{H}}(\tau_B)]\}$$
$$+ \mathbb{E}\{[\boldsymbol{\eta}_j^{\mathrm{H}}(\tau_1) \ldots \boldsymbol{\eta}_j^{\mathrm{H}}(\tau_B)]^{\mathrm{H}}[\boldsymbol{\eta}_j^{\mathrm{H}}(\tau_1) \ldots \boldsymbol{\eta}_j^{\mathrm{H}}(\tau_B)]\}$$
$$= \underbrace{\sum_{\ell=1}^{L}\sum_{m=1}^{K}\left(\mathbf{X}_{\ell m} \otimes \mathbf{\Lambda}_{j\ell m} + \kappa^2\mathbf{D}_{|\tilde{\mathbf{x}}_{\ell m}|^2} \otimes \mathbf{\Lambda}_{j\ell m}\right) + \xi\mathbf{I}_{BN}}_{\triangleq \mathbf{\Psi}_j}.$$
$$(37)$$

The LMMSE estimator in (9) now follows from (36) and (37). The error covariance matrix in (15) is computed as $\mathbf{\Lambda}_{jlk} - \mathbb{E}\{\mathbf{h}_{jlk}(t)\boldsymbol{\psi}_j^{\mathrm{H}}\}\left(\mathbb{E}\{\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}}\}\right)^{-1}(\mathbb{E}\{\mathbf{h}_{jlk}(t)\boldsymbol{\psi}_j^{\mathrm{H}}\})^{\mathrm{H}}$ [33, Ch. 12].

Appendix C: Proof of Lemma 1

Since the effective channels vary with t, we follow the approach in [15] and compute one ergodic achievable rate for each $t \in \mathcal{D}$. We obtain (19) by taking the average of these rates. The SINR in (20) is obtained by treating the uncorrelated inter-user interference and distortion noise as independent Gaussian noise, which is a worst-case assumption when computing the mutual information [39]. In addition, we

follow an approach from [38] and only exploit the knowledge of the average effective channel $\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jjk}(t)\}$ in the detection, while the deviation from the average effective channel is treated as worst-case Gaussian noise with variance $\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jjk}(t)|^2\} - |\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jjk}(t)\}|^2$.

APPENDIX D: PROOF OF THEOREM 2

The expressions in Theorem 2 are derived one at the time. For brevity, we use the following notations in the derivations:

$$\mathbf{A}_{jlk}(t) = \left(\tilde{\mathbf{x}}_{lk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \boldsymbol{\Lambda}_{jlk}\right)\boldsymbol{\Psi}_j^{-1} \tag{38}$$

$$\mathbf{D}_{|\mathbf{h}_{jlk}|^2} = \mathrm{diag}\left(|h_{jlk}^{(1)}|^2, \ldots, |h_{jlk}^{(N)}|^2\right) \tag{39}$$

$$\mathbf{B}_{jklm}(t) = \boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t) \tag{40}$$

$$\mathbf{M}_{jklm}(t) = \mathbf{D}_{\boldsymbol{\phi}_j(t)}^{\mathrm{H}}\mathbf{A}_{jjk}(t)$$
$$\times [\mathbf{D}_{\boldsymbol{\phi}_j(\tau_1)}^{\mathrm{T}}x_{lm}(\tau_1) \ldots \mathbf{D}_{\boldsymbol{\phi}_j(\tau_B)}^{\mathrm{T}}x_{lm}(\tau_B)]^{\mathrm{T}}. \tag{41}$$

We begin with (21) and exploit that $\mathbf{v}_{jk}(t) = \hat{\mathbf{h}}_{jlk}(t)$ is an LMMSE estimate to see that

$$\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\} = \mathrm{tr}(\boldsymbol{\Lambda}_{jjk} - \mathbf{C}_{jjk}(t))$$
$$= \mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \boldsymbol{\Lambda}_{jjk}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk} \otimes \boldsymbol{\Lambda}_{jjk}\right)\right) \tag{42}$$

which proves (21). Next, we exploit that $\hat{\mathbf{h}}_{jjk}(t) = \mathbf{A}_{jjk}(t)\boldsymbol{\psi}_j$ and note that

$$\mathbb{E}\{\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jjk}(t)\} = \mathrm{tr}\left(\mathbb{E}\{\mathbf{h}_{jjk}(t)\boldsymbol{\psi}_j^{\mathrm{H}}\}\mathbf{A}_{jjk}^{\mathrm{H}}(t)\right)$$
$$= \mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \boldsymbol{\Lambda}_{jjk}\right)\mathbf{A}_{jjk}^{\mathrm{H}}(t)\right)$$
$$= \mathrm{tr}\left(\left(\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)} \otimes \boldsymbol{\Lambda}_{jjk}\right)\boldsymbol{\Psi}_j^{-1}\left(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{jk} \otimes \boldsymbol{\Lambda}_{jjk}\right)\right) \tag{43}$$

where the second equality follows from (36) and the third equality follows from the full expression of $\mathbf{A}_{jjk}(t)$ in (38). Observe that the expression (43) is the same as for $\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\}$ in (42).

Next, the second-order moment in (23) can be expanded as

$$\mathbb{E}\left\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\right\}$$
$$= \mathbb{E}\{\mathrm{tr}(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)\mathbf{h}_{jlm}^{\mathrm{H}}(t)\mathbf{A}_{jjk}(t)\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}})\}$$
$$= \mathrm{tr}\left(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)\left(\boldsymbol{\Psi}_j - \mathbf{X}_{lm} \otimes \boldsymbol{\Lambda}_{jlm}\right)\right)$$
$$+ \mathbb{E}\left\{\mathrm{tr}(\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}}\mathbf{M}_{jklm}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}})\right\}$$
$$+ \kappa^2\mathbb{E}\left\{\mathrm{tr}(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2} \otimes \mathbf{D}_{|\mathbf{h}_{jlm}|^2}))\right\} \tag{44}$$

where the first term follows from computing separate expectations for the parts of $\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}}$ that are independent of $\mathbf{h}_{jlm}(t)\mathbf{h}_{jlm}^{\mathrm{H}}(t)$. The remaining two terms take care of the statistically dependent terms. The middle term is simplified as

$$\mathbb{E}\left\{\mathrm{tr}(\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}}\mathbf{M}_{jklm}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}})\right\}$$
$$= \mathbb{E}\left\{|\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}(t))|^2\right\}$$
$$+ \mathbb{E}\left\{\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}^{\mathrm{H}}(t))\right\} \tag{45}$$

by computing the expectation with respect to $\mathbf{h}_{jlm}$ using Lemma 2 in Appendix A. The first expectation in (45) is now

computed by expanding the expression as

$$\mathbb{E}\{|\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}(t))|^2\} = \mathbb{E}\Big\{\Big|\mathrm{tr}\Big(\mathbf{B}_{jklm}(t)$$
$$\times [\mathbf{D}_{\boldsymbol{\phi}_j(\tau_1)}^{\mathrm{T}}x_{lm}(\tau_1) \ldots \mathbf{D}_{\boldsymbol{\phi}_j(\tau_B)}^{\mathrm{T}}x_{lm}(\tau_B)]^{\mathrm{T}}\mathbf{D}_{\boldsymbol{\phi}_j(t)}^{\mathrm{H}}\Big)\Big|^2\Big\}$$
$$= \mathbb{E}\Big\{\sum_{n_1=1}^N\sum_{b_1=1}^B[\mathbf{B}_{jklm}(t)\mathbf{E}_{b_1}]_{n_1n_1}x_{lm}(\tau_{b_1})e^{\imath\phi_{n_1,\tau_{b_1}}}e^{-\imath\phi_{n_1,t}}$$
$$\times \sum_{n_2=1}^N\sum_{b_2=1}^B[\mathbf{E}_{b_2}^{\mathrm{H}}\mathbf{B}_{jklm}^{\mathrm{H}}(t)]_{n_2n_2}x_{lm}^*(\tau_{b_2})e^{-\imath\phi_{n_2,\tau_{b_2}}}e^{\imath\phi_{n_2,t}}\Big\}$$
$$= \sum_{n_1,n_2,b_1,b_2}[\mathbf{B}_{jklm}(t)\mathbf{E}_{b_1}]_{n_1n_1}[\mathbf{E}_{b_2}^{\mathrm{H}}\mathbf{B}_{jklm}^{\mathrm{H}}(t)]_{n_2n_2}$$
$$\times x_{lm}(\tau_{b_1})x_{lm}^*(\tau_{b_2})\mathbb{E}\left\{e^{\imath(\phi_{n_1,\tau_{b_1}}-\phi_{n_1,t}-\phi_{n_2,\tau_{b_2}}+\phi_{n_2,t})}\right\} \tag{46}$$

where $\mathbf{E}_b = \mathbf{e}_b \otimes \mathbf{I}_N$ and $\mathbf{e}_b \in \mathbb{C}^{B\times 1}$ is the bth column of $\mathbf{I}_B$. The phase-drift expectation depends on the use of a CLO or SLOs:

$$\mathbb{E}\left\{e^{\imath(\phi_{n_1,\tau_{b_1}}-\phi_{n_1,t}-\phi_{n_2,\tau_{b_2}}+\phi_{n_2,t})}\right\}$$
$$= \begin{cases} e^{-\frac{\delta}{2}|\tau_{b_1}-\tau_{b_2}|}, & \text{if a CLO,} \\ e^{-\frac{\delta}{2}|\tau_{b_1}-\tau_{b_2}|}, & \text{if SLOs and } n_1 = n_2, \\ e^{-\frac{\delta}{2}|t-\tau_{b_1}|}e^{-\frac{\delta}{2}|t-\tau_{b_2}|}, & \text{if SLOs and } n_1 \neq n_2. \end{cases} \tag{47}$$

Since $x_{lm}(\tau_{b_1})x_{lm}^*(\tau_{b_2})e^{-\frac{\delta}{2}|\tau_{b_1}-\tau_{b_2}|} = [\bar{\mathbf{X}}_{lm}]_{b_1,b_2} = \mathbf{e}_{b_1}^{\mathrm{H}}\bar{\mathbf{X}}_{lm}\mathbf{e}_{b_2}$ in the case of a CLO, (46) becomes

$$\sum_{n_1,n_2,b_1,b_2}[\mathbf{B}_{jklm}(t)\mathbf{E}_{b_1}]_{n_1n_1}[\mathbf{E}_{b_2}^{\mathrm{H}}\mathbf{B}_{jklm}^{\mathrm{H}}(t)]_{n_2n_2}\mathbf{e}_{b_1}^{\mathrm{H}}\bar{\mathbf{X}}_{lm}\mathbf{e}_{b_2}$$
$$= \sum_{n_1,n_2}\mathbf{e}_{n_1}^{\mathrm{H}}\mathbf{B}_{jklm}(t)(\bar{\mathbf{X}}_{lm} \otimes \mathbf{e}_{n_1}\mathbf{e}_{n_2}^{\mathrm{H}})\mathbf{B}_{jklm}^{\mathrm{H}}(t)\mathbf{e}_{n_2} \tag{48}$$

where $\mathbf{e}_n \in \mathbb{C}^{N\times 1}$ is the nth column of of $\mathbf{I}_N$ (recall also the definitions of $\mathbf{E}_b$ and $\mathbf{e}_b$ above).

Next, we note that $e^{-\frac{\delta}{2}|t-\tau_b|}x_{lm}(\tau_b) = [\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}]_b$. In the case of SLOs, (46) then becomes

$$\sum_{\substack{n_1,n_2,b_1,b_2 \\ n_1\neq n_2}}[\mathbf{B}_{jklm}(t)\mathbf{E}_{b_1}]_{n_1n_1}[\mathbf{E}_{b_2}^{\mathrm{H}}\mathbf{B}_{jklm}^{\mathrm{H}}(t)]_{n_2n_2}$$
$$\times [\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}]_{b_1}[\tilde{\mathbf{x}}_{lm}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}]_{b_2}$$
$$+ \sum_{n,b_1,b_2}[\mathbf{B}_{jklm}(t)\mathbf{E}_{b_1}]_{nn}[\mathbf{E}_{b_2}^{\mathrm{H}}\mathbf{B}_{jklm}^{\mathrm{H}}(t)]_{nn}[\bar{\mathbf{X}}_{lm}]_{b_1,b_2}$$
$$= \Big|\sum_{n,b}[\mathbf{B}_{jklm}(t)\mathbf{E}_b]_{nn}[\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}]_b\Big|^2$$
$$+ \sum_{n,b_1,b_2}[\mathbf{B}_{jklm}(t)\mathbf{E}_{b_1}]_{nn}[\mathbf{E}_{b_2}^{\mathrm{H}}\mathbf{B}_{jklm}^{\mathrm{H}}(t)]_{nn}$$
$$\times [\bar{\mathbf{X}}_{lm} - \mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}\tilde{\mathbf{x}}_{lm}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}]_{b_1,b_2}$$
$$= \left|\mathrm{tr}(\mathbf{A}_{jjk}(t)(\mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm} \otimes \boldsymbol{\Lambda}_{jlm}))\right|^2 + \sum_n\mathbf{e}_n^{\mathrm{H}}\mathbf{B}_{jklm}(t)$$
$$\times \left((\bar{\mathbf{X}}_{lm} - \mathbf{D}_{\boldsymbol{\delta}(t)}^{\mathrm{H}}\tilde{\mathbf{x}}_{lm}\tilde{\mathbf{x}}_{lm}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}) \otimes \mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}\right)\mathbf{B}_{jklm}^{\mathrm{H}}(t)\mathbf{e}_n. \tag{49}$$

The second expectation in (45) is computed along the same lines as in (37) and becomes

$$\begin{aligned}
&\mathbb{E}\{\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}^{\mathrm{H}}(t))\} \\
&= \mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\bar{\mathbf{X}}_{lm}\otimes\boldsymbol{\Lambda}_{jlm})\mathbf{A}_{jjk}^{\mathrm{H}}(t)).
\end{aligned} \tag{50}$$

It remains to compute the last term in (44). We exploit the following expansion of diagonal matrices: $\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2} = \sum_{b=1}^{B}|x_{lm}(\tau_b)|^2\mathbf{e}_b\mathbf{e}_b^{\mathrm{H}}$ and $\mathbf{D}_{|\mathbf{h}_{jlm}|^2} = \sum_{n=1}^{N}|\mathbf{e}_n^{\mathrm{H}}\mathbf{h}_{jlm}|^2\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}$, where $\mathbf{e}_b$ is the bth column of $\mathbf{I}_B$ and $\mathbf{e}_n$ is the nth column of $\mathbf{I}_N$. Plugging this into the last term in (44) yields

$$\begin{aligned}
&\sum_{b,n}|x_{lm}(\tau_b)|^2\mathbb{E}\left\{|\mathbf{h}_{jlm}^{\mathrm{H}}\mathbf{A}_{jjk}(t)(\mathbf{e}_b\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}})\mathbf{h}_{jlm}|^2\right\} \\
&= \sum_{b,n}|x_{lm}(\tau_b)|^2|\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\mathbf{e}_b\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}))|^2 \\
&\quad + \sum_{b,n}|x_{lm}(\tau_b)|^2\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\mathbf{e}_b\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}) \\
&\qquad \times \boldsymbol{\Lambda}_{jlm}(\mathbf{e}_b^{\mathrm{H}}\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}})\mathbf{A}_{jjk}^{\mathrm{H}}(t)) \\
&= \sum_{n}\mathbf{e}_n^{\mathrm{H}}\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}})\mathbf{A}_{jjk}^{\mathrm{H}}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{e}_n \\
&\quad + \mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\boldsymbol{\Lambda}_{jlm})\mathbf{A}_{jjk}^{\mathrm{H}}(t))
\end{aligned} \tag{51}$$

where the first equality follows from Lemma 2 and the second equality from reverting the matrix expansions wherever possible. Plugging (45)–(51) into (44) and utilizing $\bar{\mathbf{X}}_{lm} + \kappa^2\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2} = \mathbf{X}_{lm}$, we obtain (23) by removing the special notation that was introduced in the beginning of this appendix.

Finally, we compute the expectation in (24) by noting that

$$\begin{aligned}
&\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\boldsymbol{v}_j(t)|^2\} = \mathbb{E}\{\mathrm{tr}(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\boldsymbol{\Upsilon}_j(t)\mathbf{A}_{jjk}(t)\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}})\} \\
&= \kappa^2\sum_{l=1}^{L}\sum_{m=1}^{K}p_{lm} \\
&\quad \times \Bigg(\mathrm{tr}\Big(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)\big(\boldsymbol{\Psi}_j - \mathbf{X}_{lm}\otimes\boldsymbol{\Lambda}_{jlm}\big)\Big) \\
&\quad + \mathbb{E}\{\mathrm{tr}(\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{D}_{|\mathbf{h}_{jlm}|^2}\mathbf{M}_{jklm}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}})\} \\
&\quad + \kappa^2\mathbb{E}\{\mathrm{tr}\big(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\mathbf{D}_{|\mathbf{h}_{jlm}|^2}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\mathbf{D}_{|\mathbf{h}_{jlm}|^2})\big)\}\Bigg).
\end{aligned} \tag{52}$$

The first equality follows by taking the expectation with respect to $\boldsymbol{v}_j(t)$ for fixed channel realizations. The second equality follows by taking separate expectations with respect to the terms of $\boldsymbol{\Upsilon}_j = \kappa^2\sum_{l=1}^{L}\sum_{m=1}^{K}p_{lm}\mathbf{D}_{|\mathbf{h}_{jlm}|^2}$ and $\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{H}}$ that are independent. These give the first term in (52) while the last two terms take care of the statistically dependent terms.

The expectation in the middle term of (52) is computed as

$$\begin{aligned}
&\mathbb{E}\{\mathrm{tr}(\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{D}_{|\mathbf{h}_{jlm}|^2}\mathbf{M}_{jklm}(t)\mathbf{h}_{jlm}\mathbf{h}_{jlm}^{\mathrm{H}})\} \\
&= \sum_{n}\mathbb{E}\left\{|\mathbf{h}_{jlm}^{\mathrm{H}}\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}\mathbf{h}_{jlm}|^2\right\} \\
&= \sum_{n}\mathbb{E}\left\{|\mathbf{e}_n^{\mathrm{H}}\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{e}_n|^2\right\} \\
&\quad + \sum_{n}\mathbb{E}\left\{\mathbf{e}_n^{\mathrm{H}}\boldsymbol{\Lambda}_{jlm}\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}}\mathbf{M}_{jklm}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{M}_{jklm}^{\mathrm{H}}(t)\mathbf{e}_n\right\} \\
&= \mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\bar{\mathbf{X}}_{lm}\otimes\boldsymbol{\Lambda}_{jlm})\mathbf{A}_{jjk}^{\mathrm{H}}(t)) \\
&\quad + \sum_{n}\mathbf{e}_n^{\mathrm{H}}\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\bar{\mathbf{X}}_{lm}\otimes\mathbf{e}_n\mathbf{e}_n^{\mathrm{H}})\mathbf{A}_{jjk}^{\mathrm{H}}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{e}_n
\end{aligned} \tag{53}$$

where the first equality follows from the same diagonal matrix expansion as in (51), the second equality is due to Lemma 2 (and that diagonal matrices commute), and the third equality follows from computing the expectation with respect to phase-drifts as in (37) and then reverting the matrix expansions wherever possible.

Similarly, we have

$$\begin{aligned}
&\mathbb{E}\{\mathrm{tr}\big(\mathbf{A}_{jjk}^{\mathrm{H}}(t)\mathbf{D}_{|\mathbf{h}_{jlm}|^2}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\mathbf{D}_{|\mathbf{h}_{jlm}|^2})\big)\} \\
&= \sum_{n_1,n_2,b}|x_{lm}(\tau_b)|^2\mathbb{E}\left\{|\mathbf{h}_{jlm}^{\mathrm{H}}\mathbf{e}_{n_1}\mathbf{e}_{n_1}^{\mathrm{H}}\mathbf{A}_{jjk}(t)(\mathbf{e}_b\otimes\mathbf{e}_{n_2}\mathbf{e}_{n_2}^{\mathrm{H}})\mathbf{h}_{jlm}|^2\right\} \\
&= \sum_{n_1,n_2,b}|x_{lm}(\tau_b)|^2\Big(|\mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{e}_{n_1}\mathbf{e}_{n_1}^{\mathrm{H}}\mathbf{A}_{jjk}(t)(\mathbf{e}_b\otimes\mathbf{e}_{n_2}\mathbf{e}_{n_2}^{\mathrm{H}}))|^2 \\
&\quad + \mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{e}_{n_1}\mathbf{e}_{n_1}^{\mathrm{H}}\mathbf{A}_{jjk}(t)(\mathbf{e}_b\mathbf{e}_b^{\mathrm{H}}\otimes\mathbf{e}_{n_2}\mathbf{e}_{n_2}^{\mathrm{H}}\boldsymbol{\Lambda}_{jlm})\mathbf{A}_{jjk}^{\mathrm{H}}(t))\Big) \\
&= \sum_{n_1}\mathbf{e}_{n_1}^{\mathrm{H}}\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\mathbf{e}_{n_1}\mathbf{e}_{n_1}^{\mathrm{H}})\mathbf{A}_{jjk}^{\mathrm{H}}(t)\boldsymbol{\Lambda}_{jlm}\mathbf{e}_{n_1} \\
&\quad + \mathrm{tr}(\boldsymbol{\Lambda}_{jlm}\mathbf{A}_{jjk}(t)(\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2}\otimes\boldsymbol{\Lambda}_{jlm})\mathbf{A}_{jjk}^{\mathrm{H}}(t))
\end{aligned} \tag{54}$$

where the first equality follows from the same diagonal matrix expansions as above, the second equality follows from Lemma 2 (and that diagonal matrices commute), and the third equality from reverting the matrix expansions wherever possible.

By plugging (53) and (54) into (52) and utilizing $\bar{\mathbf{X}}_{lm} + \kappa^2\mathbf{D}_{|\tilde{\mathbf{x}}_{lm}|^2} = \mathbf{X}_{lm}$, we finally obtain (24).

APPENDIX E: PROOF OF COROLLARY 3

This corollary is obtained by dividing all the terms in $\mathrm{SINR}_{jk}(t)$ by $\frac{N^2}{A^2}$ and inspecting the scaling behavior as $N \to \infty$. Using the expressions in Theorem 2 and utilizing that $\boldsymbol{\Psi}_j^{-1} = \widetilde{\boldsymbol{\Psi}}_j^{-1}\otimes\mathbf{I}_{\frac{N}{A}}$, we observe that $\frac{\xi A^2}{N^2}\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\} = \frac{\xi A^2}{N^2}\mathrm{tr}(\mathbf{F}\otimes\mathbf{I}_{\frac{N}{A}}) = \frac{\xi A}{N}\mathrm{tr}(\mathbf{F}) = \mathcal{O}(\frac{1}{N})$, where $\mathbf{F} = (\tilde{\mathbf{x}}_{jk}^{\mathrm{H}}\mathbf{D}_{\boldsymbol{\delta}(t)}\otimes\tilde{\boldsymbol{\Lambda}}_{jjk}^{(A)})\widetilde{\boldsymbol{\Psi}}_j^{-1}(\mathbf{D}_{\boldsymbol{\delta}(t)}\tilde{\mathbf{x}}_{jk}\otimes\tilde{\boldsymbol{\Lambda}}_{jjk}^{(A)})$. Similarly, it is straightforward but lengthy to prove that $\frac{A^2}{N^2}\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\boldsymbol{v}_j(t)|^2\} = \mathcal{O}(\frac{1}{N})$. The only terms in the SINR that remain as $N \to \infty$ are $\frac{A^2}{N^2}(\mathbb{E}\{\|\mathbf{v}_{jk}(t)\|^2\})^2 = \mathrm{Sig}_{jk}$ and $\frac{A^2}{N^2}\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\} = \mathrm{Int}_{jklm} + \mathcal{O}(\frac{1}{N})$.

APPENDIX F: PROOF OF COROLLARY 4

The first step of the proof is to substitute the new parameters into the SINR expression in (20) and scale all terms by $1/N^{1+z_3\delta_0\min_\tau|t-\tau|}$. Since the distortion noise

and receiver noise terms normally behave as $\mathcal{O}(N)$, it is straightforward (but lengthy) to verify that the (scaled) distortion noise and receiver noise terms go to zero when $N \to \infty$. Similarly, the signal term in the numerator which normally behave as $\mathcal{O}(N^2)$, will after the scaling behave as $\mathcal{O}(N^{1-2\max(z_1,z_2)-z_3\delta_0\min_\tau |t-\tau|})$. In the case of SLOs, the second-order interference moments $\mathbb{E}\{|\mathbf{v}_{jk}^{\mathrm{H}}(t)\mathbf{h}_{jlm}(t)|^2\}$ in the denominator exhibit the same scaling as the signal term. The scaling law in (30) then follows from that we want the signal and interference terms to be non-vanishing in the asymptotic limit; that is, $1 - 2\max(z_1,z_2) - z_3\delta_0\min_\tau |t - \tau| > 1$. In the case of a CLO, the second-order interference moments behave as $\mathcal{O}(N^{1-2\max(z_1,z_2)})$ and do not depend on z_3. To make the signal and interference terms have the same scaling and be non-vanishing, we thus need to set $z_3 = 0$ and $\max(z_1,z_2) \leq \frac{1}{2}$.

REFERENCES

[1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[2] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[3] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[4] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, Wiley, 2nd edition, 2011.

[5] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5–21, Sep. 2013.

[6] R. Baldemair, E. Dahlman, G. Fodor, G. Mildh, S. Parkvall, Y. Selen, H. Tullberg, and K. Balachandran, "Evolving wireless communications: Addressing the challenges and expectations of the future," *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 24–30, Mar. 2013.

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

[8] China Mobile Research Institute, "C-RAN: The road towards green RAN," Tech. Rep., White Paper, Oct. 2011.

[9] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, 3590–3600, Nov. 2010.

[10] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[11] 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. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

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

[13] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," in *Proc. IEEE WCNC*, Apr. 2014, pp. 242–247.

[14] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[15] A. Pitarokoilis, S. K. Mohammed, and E. G. 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.

[16] R. Krishnan, M. R. Khanzadi, N. Krishnan, A. Graell i 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," *IEEE Signal Process. Lett.*, 2014, Submitted.

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

[18] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation*, Springer, 2008.

[19] A. Mezghani, N. Damak, and J. A. Nossek, "Circuit aware design of power-efficient short range communication systems," in *Proc. IEEE ISWCS*, Sep. 2010, pp. 869–873.

[20] M. Wenk, *MIMO-OFDM Testbed: Challenges, Implementations, and Measurement Results*, Series in microelectronics. Hartung-Gorre, 2010.

[21] 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.

[22] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. 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, Sep. 2012.

[23] G. Durisi, A. Tarable, and T. Koch, "On the multiplexing gain of MIMO microwave backhaul links affected by phase noise," in *Proc. IEEE ICC*, June 2013, pp. 3209–3214.

[24] W. Zhang, "A general framework for transmission with transceiver distortion and some applications," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 384–399, Feb. 2012.

[25] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO systems with hardware-constrained base stations," in *Proc. IEEE ICASSP*, May 2014, pp. 3142–3146.

[26] E. Björnson, M. Matthaiou, and M. Debbah, "Circuit-aware design of energy-efficient massive MIMO systems," in *Proc. IEEE ISCCSP*, May 2014, pp. 101–104.

[27] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO in real propagation environments," *IEEE Trans. Wireless Commun.*, 2014, Submitted, Available: http://arxiv.org/abs/1403.3376.

[28] A. Farhang, N. Marchetti, L. E. Doyle, and B. Farhang-Boroujeny, "Filter bank multicarrier for massive MIMO," 2014, Submitted, Available: http://arxiv.org/abs/1402.5881.

[29] H. Yang and T. L. Marzetta, "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," in *Proc. IEEE OnlineGreenComm*, Oct. 2013, pp. 27–32.

[30] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, Academic Press, 2nd edition, 2008.

[31] M. Biguesh and A. B. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 9, pp. 1330–1339, 2004.

[32] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[34] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[35] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[36] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[37] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 920–929, Mar. 2014.

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

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

[40] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[41] R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[42] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *IEEE Trans. Commun.*, 2014, Submitted, Available: http://arxiv.org/abs/1404.7736.

[43] I. Song, M. Koo, H. Jung, H.-S. Jhon, and H. Shinz, "Optimization of cascode configuration in CMOS low-noise amplifier," *Micr. Opt. Techn. Lett.*, vol. 50, no. 3, pp. 646–649, Mar. 2008.

[44] D. Petrovic, W. Rave, and G. Fettweis, "Common phase error due to phase noise in OFDM-estimation and suppression," in *Proc. IEEE PIMRC*, Sept. 2004, pp. 1901–1905.

[45] K.-G. Park, C.-Y. Jeong, J.-W. Park, J.-W. Lee, J.-G. Jo, and C. Yoo, "Current reusing VCO and divide-by-two frequency divider for quadrature LO generation," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 6, pp. 413–415, June 2008.

[46] J. R. Wilkerson, *Passive Intermodulation Distortion in Radio Frequency Communication Systems*, Ph.D. thesis, North Carolina State University, 2010.

[47] *Further advancements for E-UTRA physical layer aspects (Release 9)*, 3GPP TS 36.814, Mar. 2010.

[48] M. R. Khanzadi, G. Durisi, and T. Eriksson, "Capacity of multiple-antenna phase-noise channels with common/separate oscillators," *IEEE Trans. Commun.*, 2014, In press, Available: http://arxiv.org/abs/1409.0561.

[49] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, July 2014.

Michail Matthaiou (S'05–M'08–SM'13) was born in Thessaloniki, Greece in 1981. He obtained the Diploma degree (5 years) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece in 2004. He then received the M.Sc. (with distinction) in Communication Systems and Signal Processing from the University of Bristol, U.K. and Ph.D. degrees from the University of Edinburgh, U.K. in 2005 and 2008, respectively. From September 2008 through May 2010, he was with the Institute for Circuit Theory and Signal Processing, Munich University of Technology (TUM), Germany working as a Postdoctoral Research Associate. He is currently a Senior Lecturer at Queen's University Belfast, U.K. and also holds an adjunct Assistant Professor position at Chalmers University of Technology, Sweden. His research interests span signal processing for wireless communications, massive MIMO, hardware-constrained communications, and performance analysis of fading channels.

Dr. Matthaiou was the recipient of the 2011 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East and Africa Region and a co-recipient of the 2006 IEEE Communications Chapter Project Prize for the best M.Sc. dissertation in the area of communications. He was co-recipient of the Best Paper Award at the 2014 IEEE International Conference on Communications (ICC) and was an Exemplary Reviewer for IEEE COMMUNICATIONS LETTERS for 2010. In 2014, he received the Research Fund for International Young Scientists from the National Natural Science Foundation of China. He has been a member of Technical Program Committees for several IEEE conferences such as ICC, GLOBECOM, VTC etc. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS and was the Lead Guest Editor of the special issue on "Large-scale multiple antenna wireless systems" of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is an associate member of the IEEE Signal Processing Society SPCOM and SAM technical committees.

Mérouane Debbah (SM'08, F'15) entered the Ecole Normale Suprieure de Cachan (France) in 1996 where he received his M.Sc and Ph.D. degrees respectively. He worked for Motorola Labs (Saclay, France) from 1999-2002 and the Vienna Research Center for Telecommunications (Vienna, Austria) until 2003. From 2003 to 2007, he joined the Mobile Communications department of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. Since 2007, he is a Full Professor at Supelec (Gif-sur-Yvette, France). From 2007 to 2014, he was director of the Alcatel-Lucent Chair on Flexible Radio. Since 2014, he is Vice-President of the Huawei France R&D center and director of the Mathematical and Algorithmic Sciences Lab.

His research interests are in information theory, signal processing and wireless communications. He is an Associate Editor in Chief of the journal Random Matrix: Theory and Applications and was an associate and senior area editor for IEEE Transactions on Signal Processing respectively in 2011-2013 and 2013-2014. Mrouane Debbah is a recipient of the ERC grant MORE (Advanced Mathematical Tools for Complex Network Engineering). He is a IEEE Fellow, a WWRF Fellow and a member of the academic senate of Paris-Saclay. He is the recipient of the Mario Boella award in 2005, the 2007 IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award, the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC best paper award as well as the Valuetools 2007, Valuetools 2008, CrownCom2009 , Valuetools 2012 and SAM 2014 best student paper awards. In 2011, he received the IEEE Glavieux Prize Award and in 2012, the Qualcomm Innovation Prize Award.

Emil Björnson (S'07, M'12) received the M.S. degree in Engineering Mathematics from Lund University, Sweden, in 2007. He received the Ph.D. degree in Telecommunications from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2011. From 2012 to July 2014, he was a joint postdoc at Suplec, Gif-sur-Yvette, France, and at KTH Royal Institute of Technology. He is currently an Assistant Professor and Docent at the Department of Electrical Engineering (ISY) at Linköping University, Sweden.

His research interests include multi-antenna cellular communications, radio resource allocation, energy efficiency, massive MIMO, and network topology design. He is the first author of the textbook "Optimal Resource Allocation in Coordinated Multi-Cell System" published in Foundations and Trends in Communications and Information Theory, 2013. He is also dedicated to reproducible research and has made a large amount of simulation code publicly available. Dr. Björnson received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA and has received 4 best paper awards for novel research on multi-cell multi-antenna communications: WCNC 2014, SAM 2014, CAMSAP 2011, and WCSP 2009.