

Massive MIMO Systems With Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits

Emil Bjornson, Jakob Hoydis, Marios Kountouris, Merouane Debbah

▶ **To cite this version:**

Emil Bjornson, Jakob Hoydis, Marios Kountouris, Merouane Debbah. Massive MIMO Systems With Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits. IEEE Transactions on Information Theory, 2014, 60 (11), pp.7112-7139. 10.1109/TIT.2014.2354403 . hal-01108979

HAL Id: hal-01108979

https://centralesupelec.hal.science/hal-01108979

Submitted on 23 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits

Emil Björnson, *Member, IEEE,* Jakob Hoydis, *Member, IEEE,* Marios Kountouris, *Member, IEEE,* and Mérouane Debbah, *Senior Member, IEEE*

arXiv:1307.2584v3 [cs.IT] 2 Sep 2014

***Abstract*—The use of large-scale antenna arrays can bring substantial improvements in energy and/or spectral efficiency to wireless systems due to the greatly improved spatial resolution and array gain. Recent works in the field of massive multiple-input multiple-output (MIMO) show that the user channels decorrelate when the number of antennas at the base stations (BSs) increases, thus strong signal gains are achievable with little inter-user interference. Since these results rely on asymptotics, it is important to investigate whether the conventional system models are reasonable in this asymptotic regime. This paper considers a new system model that incorporates general transceiver hardware impairments at both the BSs (equipped with large antenna arrays) and the single-antenna user equipments (UEs). As opposed to the conventional case of ideal hardware, we show that hardware impairments create finite ceilings on the channel estimation accuracy and on the downlink/uplink capacity of each UE. Surprisingly, the capacity is mainly limited by the hardware at the UE, while the impact of impairments in the large-scale arrays vanishes asymptotically and inter-user interference (in particular, pilot contamination) becomes negligible. Furthermore, we prove that the huge degrees of freedom offered by massive MIMO can be used to reduce the transmit power and/or to tolerate larger hardware impairments, which allows for the use of inexpensive and energy-efficient antenna elements.**

***Index Terms*—Capacity bounds, channel estimation, energy efficiency, massive MIMO, pilot contamination, time-division duplex, transceiver hardware impairments.**

I. Introduction

The spectral efficiency of a wireless link is limited by the information-theoretic capacity [2], which depends not only on the signal-to-noise ratio (SNR) but also on spatial correlation in the propagation environment [3], [4], channel estimation accuracy [5], transceiver hardware impairments [6], [7], and signal processing resources [8], [9]. It is of profound importance to increase the spectral efficiency of future networks, to keep up with the increasing demand for wireless services. However, this is a challenging task and usually comes at the price of having stricter hardware and overhead requirements.

A new network architecture has recently been proposed with the remarkable potential of both increasing the spectral efficiency and relaxing the aforementioned implementation issues. It is known as massive MIMO, or large-scale MIMO, and is based on having a very large number of antennas at each BS and exploiting channel reciprocity in time-division duplex (TDD) mode [9]–[13]. Some key features are: 1) propagation losses are mitigated by a large array gain due to coherent beamforming/combining; 2) interference-leakage due to channel estimation errors vanish asymptotically in the large-dimensional vector space; 3) low-complexity signal processing algorithms are asymptotically optimal; and 4) inter-user interference is easily mitigated by the high beamforming resolution.

The amount of research on massive MIMO increases rapidly, but the impact of transceiver hardware impairments on these systems has received little attention so far—although large arrays might only be attractive for network deployment if each antenna element consists of inexpensive hardware. Cheap hardware components are particularly prone to the impairments that exist in any transceiver (e.g., amplifier non-linearities, I/Q-imbalance, phase noise, and quantization errors [14]–[23]). The influence of hardware impairments is usually mitigated by compensation algorithms [14], which can be implemented by analog and digital signal processing. These techniques cannot remove the impairments completely, but there remain residual impairments since the time-varying hardware characteristics cannot be fully parameterized and estimated, and because there is randomness induced by different types of noise. Transceiver impairments are known to fundamentally limit the capacity in the high-power regime [6], [24], while there are only a few publications that analyze the behavior in the large number of antenna regime. Lower bounds on the achievable uplink sum rate in massive single-cell systems with phase noise from free-running oscillators were derived in [25]. The impact of amplifier non-linearities in a transmitter can be reduced by having a low peak-to-average power ratio (PAPR). The excess degrees of freedom offered by massive MIMO were used in [26] to optimize the downlink precoding for low

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The Matlab code that reproduces all simulation results is available online, see https://github.com/emilbjornson/massive-MIMO-hardware-impairments/

E. Björnson was with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France, and with the Department of Signal Processing, KTH Royal Institute of Technology, Stockholm, Sweden. He is currently with the Department of Electrical Engineering (ISY), Linköping University, Sweden (email: emil.bjornson@liu.se).

J. Hoydis was with Bell Laboratories, Alcatel-Lucent, Germany. He is now with Spraed SAS, Orsay, France (email: hoydis@ieee.org).

M. Kountouris and M. Debbah are SUPELEC, Gif-sur-Yvette, France (e-mail: marios.kountouris@supelec.fr, merouane.debbah@supelec.fr).

This paper was presented in part at the International Conference on Digital Signal Processing (DSP), Santorini, Greece, July 2013 [1].

The work of E. Björnson was funded by the International Postdoc Grant 2012-228 from The Swedish Research Council. This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering). Parts of this work have been performed in the framework of the FP7 project ICT-317669 METIS. This work was supported by the Future and Emerging Technologies (FET) project HIATUS within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number 265578.

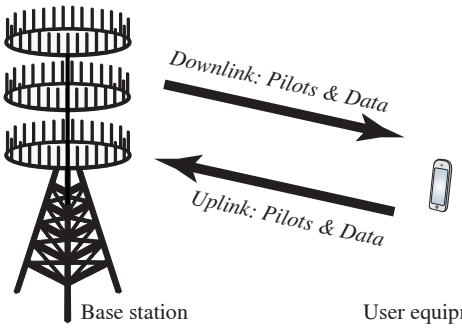


Fig. 1: Illustration of the reciprocal channel between a BS equipped with a large antenna array and a single-antenna UE.

PAPR, while [27] considered a constant-envelope precoding scheme designed for very low PAPR.

This paper analyzes the *aggregate* impact of different hardware impairments on systems with large antenna arrays, in contrast to the ideal hardware considered in [10]–[13] and the single type of impairments considered in [25]–[27]. We assume that appropriate compensation algorithms have been applied and focus on the residual hardware impairments. Motivated by the analytic analysis and experimental results in [14]–[18], the residual hardware impairments at the transmitter and receiver are modeled as additive distortion noises with certain important properties. The system model with hardware impairments is defined and motivated in Section II. Section III derives a new pilot-based channel estimator and shows that the estimation accuracy is limited by the levels of impairments. The focus of Section IV is on a single link in the system where we derive lower and upper bounds on the downlink and uplink capacities. Our results reveal the existence of finite capacity ceilings due to hardware impairments. Despite these discouraging results, Section V shows that a high energy efficiency and resilience towards hardware impairments at the BS can be achieved. Section VI puts these results in a multi-cell context and shows that inter-user interference (including pilot contamination) basically drowns in the distortion noise from hardware impairments. Section VII describes the impact of various refinements of the system model, while Section VIII summarizes the contributions and insights of the paper.

To encourage reproducibility and extensions to this paper, all the simulation results can be generated by the Matlab code that is available at https://github.com/emilbjornson/massive-MIMO-hardware-impairments/

Notation: Boldface (lower case) is used for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. Let $\mathbf{X}^T$, $\mathbf{X}^*$, and $\mathbf{X}^H$ denote the transpose, conjugate, and conjugate transpose of $\mathbf{X}$, respectively. $\mathbf{X}_1 \succeq \mathbf{X}_2$ means that $\mathbf{X}_1 - \mathbf{X}_2$ is positive semi-definite. A diagonal matrix with $a_1, \ldots, a_N$ on the main diagonal is denoted $\mathrm{diag}(a_1, \ldots, a_N)$ and $\mathbf{I}$ denotes an identity matrix (of appropriate dimensions). The Frobenius and spectral norms of a matrix $\mathbf{X}$ are denoted by $\|\mathbf{X}\|_F$ and $\|\mathbf{X}\|_2$, respectively, while $\|\mathbf{x}\|_k$ denotes the L_k norm of a vector $\mathbf{x}$. A stochastic variable $\mathbf{x}$ and its realization is denoted in the same way, for brevity. The expectation operator with respect to a stochastic variable $\mathbf{x}$ is denoted $\mathbb{E}\{\mathbf{x}\}$, while $\mathbb{E}\{\mathbf{x}|\mathbf{y}\}$ is the conditional expectation when $\mathbf{y}$ is given. A Gaussian stochastic 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 stochastic 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 empty set is denoted by $\emptyset$. 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. Channel and System Model

For analytical clarity, the major part of this paper analyzes the fundamental spectral and energy efficiency limits of a single link, which operates under arbitrary interference conditions. The link is established between an N-antenna BS and a single-antenna UE. A main characteristic in the analysis is that the number of antennas N can be very large. We consider a TDD protocol that toggles between uplink (UL) and downlink (DL) transmission on the same flat-fading subcarrier. This enables efficient channel estimation even when N is large, because the estimation accuracy and overhead in the UL is independent of N [9]. The acquired instantaneous channel state information (CSI) is utilized for UL data detection as well as DL data transmission, by exploiting channel reciprocity;[1] see Fig. 1. In Section VI, we put our results in a multi-cell context with many users, inter-cell interference, and pilot contamination.

We assume a block fading structure where each channel is static for a *coherence period* of T_{coher} channel uses. The channel realizations are generated randomly and are independent between blocks. For simplicity, T_{coher} is the same for the useful channel and any interfering channels, and the coherence periods are synchronized. We consider the conventional TDD protocol in Fig. 2, which can be found in many previous works; see for example [28] and [29]. Each block begins with UL pilot/control signaling for $T_{\mathrm{pilot}}^{\mathrm{UL}}$ channel uses, followed by UL data transmission for $T_{\mathrm{data}}^{\mathrm{UL}}$ channel uses. Next, the system toggles to the DL. This part begins with $T_{\mathrm{pilot}}^{\mathrm{DL}}$ channel uses of DL pilot/control signaling. These pilots are typically used by the UEs to estimate their effective channel (with precoding) and the current interference conditions, which enables coherent DL reception. Note that these quantities are scalars irrespective of N, thus the DL pilot signaling need not scale with N. The coherence period ends with DL data transmission for $T_{\mathrm{data}}^{\mathrm{DL}}$ channel uses. The four parameters satisfy $T_{\mathrm{pilot}}^{\mathrm{UL}} + T_{\mathrm{data}}^{\mathrm{UL}} + T_{\mathrm{pilot}}^{\mathrm{DL}} + T_{\mathrm{data}}^{\mathrm{DL}} = T_{\mathrm{coher}}$. The analysis of this paper is valid for arbitrary fixed values of those parameters, but we note that these can also be optimized dynamically based on T_{coher}, user load, user conditions, ratio of UL/DL traffic, etc.

The stochastic block-fading channel between the BS and the UE is denoted as $\mathbf{h} \in \mathbb{C}^{N \times 1}$. It is modeled as an ergodic process with a fixed independent realization $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ in each coherence period. This is known as Rayleigh block fading and $\mathbf{R} = \mathbb{E}\{\mathbf{h}\mathbf{h}^H\} \in \mathbb{C}^{N \times N}$ is the positive semi-definite covariance matrix. The statistical distribution is assumed to

[1]The physical channels are always reciprocal, but different transceiver chains are typically used in the UL and DL. Careful calibration is therefore necessary to utilize the reciprocity for transmission; see Section VII-E.

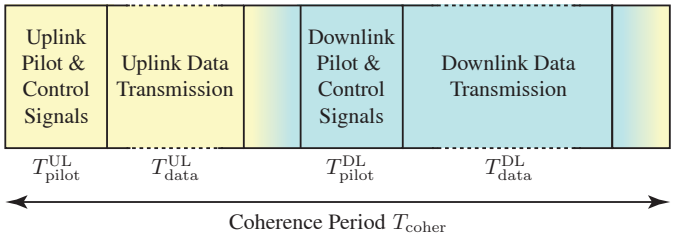


Fig. 2: Cyclic operation of a block-fading TDD system, where the coherence period T_{coher} is divided into phases for UL/DL pilot and data transmission.

be known at the BS. In the asymptotic analysis, we make the following technical assumptions:

- The spectral norm of $\mathbf{R}$ is uniformly bounded, irrespective of the number of antennas N (i.e., $\|\mathbf{R}\|_2 = \mathcal{O}(1)$);
- The trace of $\mathbf{R}$ scales linearly with N (i.e., $0 < \liminf_N \frac{1}{N}\text{tr}(\mathbf{R}) \leq \limsup_N \frac{1}{N}\text{tr}(\mathbf{R}) < \infty$) and $\mathbf{R}$ has strictly positive diagonal elements.

The first assumption is a necessary physical property that originates from the law of energy conservation. It is also a common enabler for asymptotic analysis (cf. [12]). The second assumption is a typical consequence of increasing the array size with N and thereby improving the spatial resolution and aperture [9].[2] These assumptions imply $0 < \liminf_N \frac{1}{N}\text{rank}(\mathbf{R}) \leq 1$, which means that $\mathbf{R}$ can be rank deficient but the rank increases with N such that $cN \leq \text{rank}(\mathbf{R}) \leq N$ for some $c > 0$. We stress that $\mathbf{R}$ is generally *not* a scaled identity matrix, but describes the spatial propagation environment and array geometry. It might be rank-deficient (e.g., have a large conditional number) for large arrays due to insufficient richness of the scattering [3], [4].

A. Transceiver Hardware Impairments

The majority of papers on massive MIMO systems considers channels with ideal transceiver hardware. However, practical transceivers suffer from hardware impairments that 1) create a mismatch between the intended transmit signal and what is actually generated and emitted; and 2) distort the received signal in the reception processing. In this paper, we analyze how these impairments impact the performance and key asymptotic properties of massive MIMO systems.

Physical transceiver implementations consist of many different hardware components (e.g., amplifiers, converters, mixers, filters, and oscillators [30]) and each one distorts the signals in its own way. The hardware imperfections are unavoidable, but the severity of the impairments depends on engineering decisions—larger distortions can be deliberately introduced to decrease the hardware cost and/or the power consumption [7]. The non-ideal behavior of each component can be modeled in detail for the purpose of designing compensation algorithms, but even after compensation there remain residual transceiver impairments [15], [17]; for example, due to insufficient modeling accuracy, imperfect estimation of model parameters, and time varying characteristics induced by noise.

From a system performance perspective, it is the aggregate effect of all the residual transceiver impairments that is important, not the individual behavior of each hardware component. Recently, a new system model has been proposed in [14]–[19] where the aggregate residual hardware impairments are modeled by independent additive distortion noises at the BS as well as at the UE. We adopt this model herein due its analytical tractability and the experimental verifications in [15]–[17]. The details of the DL and UL system models are given in the next subsections, and these are then used in Sections III–VI to analyze different aspects of massive MIMO systems. Possible model refinements are then provided in Section VII, along with discussions on how these might impact the main results of this paper.

B. Downlink System Model

The downlink channel is used for data transmission and pilot-based channel estimation; see Fig. 1. The received DL signal $y \in \mathbb{C}$ in a flat-fading multiple-input single-output (MISO) channel is conventionally modeled as

$$y = \mathbf{h}^T\mathbf{s} + n \tag{1}$$

where $\mathbf{s} \in \mathbb{C}^{N\times 1}$ is either a deterministic pilot signal (during channel estimation) or a stochastic zero-mean data signal; in any case, the covariance matrix is denoted $\mathbf{W} = \mathbb{E}\{\mathbf{s}\mathbf{s}^H\}$ and the average power is $p^{\text{BS}} = \text{tr}(\mathbf{W})$. $\mathbf{W}$ is a design parameter that might be a function of the channel realization $\mathbf{h}$ and the realizations of any other channel in the system (e.g., due to precoding); we let $\mathcal{H}$ denote the set of channel realizations for all useful and interfering channels (i.e., $\mathbf{h} \in \mathcal{H}$). Hence, $\mathbf{W}$ is constant within each coherence period but changes between coherence periods since $\mathcal{H}$ changes. The additive term $n = n_{\text{noise}} + n_{\text{interf}}$ is an ergodic stochastic process that consists of independent receiver noise $n_{\text{noise}} \sim \mathcal{CN}(0, \sigma_{\text{UE}}^2)$ and interference n_{interf} from simultaneous transmissions (e.g., to other UEs). The interference has zero mean and is independent of the data signal, but might depend on any channel in the system (e.g., such that carry interference). Hence, the conditional interference variance is $\mathbb{E}\{|n_{\text{interf}}|^2|\mathcal{H}\} = I_{\mathcal{H}}^{\text{UE}} \geq 0$ in the coherence period where the channel realizations are $\mathcal{H}$. The long-term interference variance is denoted $\mathbb{E}\{I_{\mathcal{H}}^{\text{UE}}\}$. It is only for brevity that we use a common notation n for interference and receiver noise—it does not mean that the interference must be treated as noise at the UE. A detailed interference model is provided in Section VI.

To model systems with non-ideal hardware more accurately, we consider the new system model from [14]–[19] where the received signal at the UE is

$$y = \mathbf{h}^T(\mathbf{s} + \boldsymbol{\eta}_t^{\text{BS}}) + \eta_r^{\text{UE}} + n. \tag{2}$$

The difference from the conventional model in (1) is the additive distortion noise terms $\boldsymbol{\eta}_t^{\text{BS}} \in \mathbb{C}^{N\times 1}$ and $\eta_r^{\text{UE}} \in \mathbb{C}$, which are ergodic stochastic processes that describe the residual transceiver impairments of the transmitter hardware at the BS and the receiver hardware at UE, respectively. We assume

[2]Although these assumptions make sense for practically large N [4], we cannot physically let $N \to \infty$ since the propagation environment is enclosed by a finite volume [9]. Nevertheless, our simulations reveal that the asymptotic analysis enabled by the technical assumptions is accurate at quite small N.

that these are independent of the signal $\mathbf{s}$, but depend on the channel $\mathbf{h}$ and thus are stationary only within each coherence period.[3] In particular, we consider the conditional distributions $\boldsymbol{\eta}_t^{\mathrm{BS}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon}_t^{\mathrm{BS}})$ and $\eta_r^{\mathrm{UE}} \sim \mathcal{CN}(0, \upsilon_r^{\mathrm{UE}})$ for given channel realizations $\mathcal{H}$. The Gaussian distributions of $\boldsymbol{\eta}_t^{\mathrm{BS}}$ and η_r^{UE} have been verified experimentally (see e.g., [17, Fig. 4.13]) and can be motivated analytically by the central limit theorem—the distortion noises describe the aggregate effect of many residual hardware impairments. A key property is that the distortion noise caused at an antenna is proportional to the signal power at this antenna (see [15]–[17] for experimental verifications), thus we have

$$\boldsymbol{\Upsilon}_t^{\mathrm{BS}} = \kappa_t^{\mathrm{BS}} \operatorname{diag}(W_{11}, \ldots, W_{NN}) \tag{3}$$

$$\upsilon_r^{\mathrm{UE}} = \kappa_r^{\mathrm{UE}} \mathbf{h}^T \mathbf{W} \mathbf{h}^* \tag{4}$$

where W_{ii} is the ith diagonal element of $\mathbf{W}$ and $\kappa_t^{\mathrm{BS}}, \kappa_r^{\mathrm{UE}} \geq 0$ are the proportionality coefficients. The intuition is that a fixed portion of the signal is turned into distortion; for example, due to quantization errors in automatic-gain-controlled analog-to-digital conversion (ADC), inter-carrier interference induced by phase noise, leakage from the mirror subcarrier under I/Q imbalance, and amplitude-amplitude nonlinearities in the power amplifier [14], [21], [31]. The proportionality coefficients are treated as constants in the analysis, but can generally increase with the signal power; see Section VII-B for details.

Remark 1 (Distortion Noise and EVM)**.** *Distortion noise is an alteration of the useful signal, while the classical receiver noise models random fluctuations in the electronic circuits at the receiver. A main difference is thus that the distortion noise power is non-stationary since it is proportional to the signal power p^{BS} and the current channel gain $\|\mathbf{h}\|_2^2$. The proportionality coefficients κ_t^{BS} and κ_r^{UE} characterize the levels of impairments and are related to the error vector magnitude (EVM) [15]; for example, the EVM at the BS is defined as*

$$\mathrm{EVM}_t^{\mathrm{BS}} = \sqrt{\frac{\mathbb{E}\{\|\boldsymbol{\eta}_t^{\mathrm{BS}}\|_2^2 | \mathcal{H}\}}{\mathbb{E}\{\|\mathbf{s}\|_2^2 | \mathcal{H}\}}} = \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Upsilon}_t^{\mathrm{BS}})}{\operatorname{tr}(\mathbf{W})}} = \sqrt{\kappa_t^{\mathrm{BS}}}. \tag{5}$$

The EVM is a common quality measure of transceivers and the 3GPP LTE standard specifies total EVM requirements in the range $[0.08, 0.175]$, where higher spectral efficiencies (modulations) are supported if the EVM is smaller [31, Sec. 14.3.4]. LTE transceivers typically support all the standardized modulations, thus the EVM is below 0.08. Larger EVMs are, however, of interest in massive MIMO systems since such relaxed hardware constraints enable the use of low-cost equipment. Therefore, the simulations in this paper consider κ-parameters in the range $[0, 0.15^2]$, where small values represent accurate and expensive transceiver hardware.

The system model in (2) captures the main characteristics of non-ideal hardware, in the sense that it allows us to identify some fundamental differences in the behavior of massive MIMO systems as compared to the case of ideal hardware. However, it cannot capture all practical characteristics of residual transceiver hardware impairments. Possible refinements, and their respective implications on our analytical results and observations, are outlined in Section VII.

C. Uplink System Model

The reciprocal UL channel is used for pilot-based channel estimation and data transmission; see Fig. 1 and Sections III–IV. Similar to (2), we consider a system model with the received signal $\mathbf{z} \in \mathbb{C}^N$ at the BS being

$$\mathbf{z} = \mathbf{h}(d + \eta_t^{\mathrm{UE}}) + \boldsymbol{\eta}_r^{\mathrm{BS}} + \boldsymbol{\nu} \tag{6}$$

where $d \in \mathbb{C}$ is either a deterministic pilot signal (used for channel estimation) or a stochastic data signal; in any case, the average power is $p^{\mathrm{UE}} = \mathbb{E}\{|d|^2\}$. The additive term $\boldsymbol{\nu} = \boldsymbol{\nu}_{\mathrm{noise}} + \boldsymbol{\nu}_{\mathrm{interf}} \in \mathbb{C}^{N \times 1}$ is an ergodic process that consists of independent receiver noise $\boldsymbol{\nu}_{\mathrm{noise}} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathrm{BS}}^2 \mathbf{I})$ as well as potential interference $\boldsymbol{\nu}_{\mathrm{interf}}$ from other simultaneous transmissions. The interference is independent of d but might depend on the channel realizations in $\mathcal{H}$. Moreover, the interference statistics can be different in the pilot and data transmission phases; for example, it is common to assume that each cell uses time-division multiple access (TDMA) for pilot transmission, since this can provide sufficient CSI accuracy to enable spatial-division multiple access (SDMA) for data transmission [9]–[13]. Therefore, we assume that $\boldsymbol{\nu}_{\mathrm{interf}}$ has zero mean and $\mathbf{S} = \mathbb{E}\{\boldsymbol{\nu}_{\mathrm{interf}} \boldsymbol{\nu}_{\mathrm{interf}}^H\}$ is that the covariance matrix during pilot transmission. We assume that $\mathbf{S}$ has a uniformly bounded spectral norm, $\|\mathbf{S}\|_2 = \mathcal{O}(1)$, for the same physical reasons as for $\mathbf{R}$. For data transmission, we define the conditional covariance matrix $\mathbf{Q}_{\mathcal{H}} = \mathbb{E}\{\boldsymbol{\nu}_{\mathrm{interf}} \boldsymbol{\nu}_{\mathrm{interf}}^H | \mathcal{H}\}$, in a coherence period with channel realizations $\mathcal{H}$, and the corresponding long-term covariance matrix $\mathbb{E}\{\mathbf{Q}_{\mathcal{H}}\}$. The covariance matrices $\mathbf{S}, \mathbf{Q}_{\mathcal{H}} \in \mathbb{C}^{N \times N}$ are positive semi-definite. The spectral norm of $\mathbf{Q}_{\mathcal{H}}$ might grow unboundedly with N due to pilot contamination in multi-cell scenarios [9]–[13]; see Section VI for further details.

Similar to the DL, the aggregate residual transceiver impairments in the hardware used for UL transmission are modeled by the independent distortion noises $\eta_t^{\mathrm{UE}} \in \mathbb{C}$ and $\boldsymbol{\eta}_r^{\mathrm{BS}} \in \mathbb{C}^{N \times 1}$ at the transmitter and receiver, respectively. These ergodic stochastic processes are independent of d, but depend on the channel realizations $\mathcal{H}$. The conditional distribution for a given $\mathcal{H}$ are $\eta_t^{\mathrm{UE}} \sim \mathcal{CN}(0, \upsilon_t^{\mathrm{UE}})$ and $\boldsymbol{\eta}_r^{\mathrm{BS}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon}_r^{\mathrm{BS}})$. Similar to (3) and (4), the conditional covariance matrices are modeled as

$$\upsilon_t^{\mathrm{UE}} = \kappa_t^{\mathrm{UE}} p^{\mathrm{UE}} \tag{7}$$

$$\boldsymbol{\Upsilon}_r^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}} p^{\mathrm{UE}} \operatorname{diag}(|h_1|^2, \ldots, |h_N|^2). \tag{8}$$

Note that the hardware quality is characterized by $\kappa_t^{\mathrm{BS}}, \kappa_r^{\mathrm{BS}}$ at the BS and by $\kappa_t^{\mathrm{UE}}, \kappa_r^{\mathrm{UE}}$ at the UE. We can have $\kappa_t^{\mathrm{BS}} \neq \kappa_r^{\mathrm{BS}}$ and $\kappa_t^{\mathrm{UE}} \neq \kappa_r^{\mathrm{UE}}$ since different transceiver chains are used for transmission and reception at a device.

Generally speaking, we would like to achieve high performance using cheap hardware. This is particularly evident in massive MIMO systems since the deployment cost of large

[3] These are model assumptions that originate from the experimental works of [15]–[17]. An analytic motivation of the assumptions (which should not be misinterpret as a proof) can be obtained from the Bussgang theorem; see Section VII.

antenna arrays might scale linearly with N unless we can accept higher levels of impairments, $\kappa_t^{\mathrm{BS}}, \kappa_r^{\mathrm{BS}}$, at the BSs than in conventional systems. This aspect is analyzed in Section V.

III. Uplink Channel Estimation

This section considers estimation of the current channel realization $\mathbf{h}$ by comparing the received UL signal $\mathbf{z}$ in (6) with the predefined UL pilot signal d (recall: $p^{\mathrm{UE}} = |d|^2$). The classic results on pilot-based channel estimation consider Rayleigh fading channels that are observed in independent complex Gaussian noise with known statistics [32]–[35]. However, this is not the case herein because the distortion noises η_t^{UE} and $\boldsymbol{\eta}_r^{\mathrm{BS}}$ effectively depend on the unknown stochastic channel $\mathbf{h}$. The dependence is either through the multiplication $\mathbf{h}\eta_t^{\mathrm{UE}}$ or the conditional variance of $\boldsymbol{\eta}_r^{\mathrm{BS}}$ in (8), which is essentially the same type of relation. Although the distortion noises are Gaussian when conditioned on a channel realization, the effective distortion is the product of Gaussian variables and, thus, has a *complex double Gaussian* distribution [36].[4] Consequently, an optimal channel estimator cannot be deduced from the standard results provided in [32]–[35].

We now derive the *linear* minimum mean square error (LMMSE) estimator of $\mathbf{h}$ under hardware impairments.

Theorem 1. *The LMMSE estimator of* $\mathbf{h}$ *from the observation of* $\mathbf{z}$ *in* (6) *is*

$$\hat{\mathbf{h}} = \underbrace{d^* \mathbf{R} \bar{\mathbf{Z}}^{-1}}_{\triangleq \mathbf{A}} \mathbf{z} \tag{9}$$

where $\mathbf{R}_{\mathrm{diag}} = \mathrm{diag}(r_{11}, \ldots, r_{NN})$ *consists of the diagonal elements of* $\mathbf{R}$ *and the covariance matrix of* $\mathbf{z}$ *is denoted as*

$$\bar{\mathbf{Z}} = \mathbb{E}\{\mathbf{z}\mathbf{z}^H\} = p^{\mathrm{UE}}(1+\kappa_t^{\mathrm{UE}})\mathbf{R} + p^{\mathrm{UE}}\kappa_r^{\mathrm{BS}}\mathbf{R}_{\mathrm{diag}} + \mathbf{S} + \sigma_{\mathrm{BS}}^2\mathbf{I}. \tag{10}$$

The total MSE is $\mathrm{MSE} = \mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_2^2\} = \mathrm{tr}(\mathbf{C})$*, where the error covariance matrix is*

$$\mathbf{C} = \mathbb{E}\{(\hat{\mathbf{h}} - \mathbf{h})(\hat{\mathbf{h}} - \mathbf{h})^H\} = \mathbf{R} - p^{\mathrm{UE}}\mathbf{R}\bar{\mathbf{Z}}^{-1}\mathbf{R}. \tag{11}$$

Proof: The LMMSE estimator has the form $\hat{\mathbf{h}} = \mathbf{A}\mathbf{z}$ where $\mathbf{A}$ minimizes the MSE. The MSE definition gives

$$\mathrm{MSE} = \mathrm{tr}\left(\mathbf{R} - d\mathbf{A}\mathbf{R} - d^*\mathbf{R}\mathbf{A}^H + \mathbf{A}\bar{\mathbf{Z}}\mathbf{A}^H \right) \tag{12}$$

where the expectations that involve $\eta_t^{\mathrm{UE}}, \boldsymbol{\eta}_r^{\mathrm{BS}}$ in $\mathrm{MSE} = \mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_2^2\}$ are computed by first having a fixed value of $\mathbf{h}$ and then average over $\mathbf{h}$. The LMMSE estimator in (9) is achieved by differentiation of (12) with respect to $\mathbf{A}$ and equating to zero. This vector minimizes the MSE since the Hessian is always positive definite. The error covariance matrix and the MSE are obtained by plugging (9) into the respective definitions. ∎

Based on Theorem 1, the channel can be decomposed as $\mathbf{h} = \hat{\mathbf{h}} + \boldsymbol{\epsilon}$ where $\hat{\mathbf{h}}$ is the LMMSE estimate in (9) and $\boldsymbol{\epsilon} \in \mathbb{C}^{N \times 1}$ denotes the unknown estimation error. Contrary to conventional estimation with independent Gaussian noise (cf. [32, Chapter 15.8]), $\hat{\mathbf{h}}$ and $\boldsymbol{\epsilon}$ are neither independent nor jointly complex Gaussian, but only uncorrelated and have zero mean. The covariance matrices are $\mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = \mathbf{R} - \mathbf{C}$ and $\mathbb{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H\} = \mathbf{C}$ where $\mathbf{C}$ was given in (11).

We remark that there might exist non-linear estimators that achieve smaller MSEs than the LMMSE estimator in Theorem 1. This stands in contrast to conventional channel estimation with independent Gaussian noise, where the LMMSE estimator is also the MMSE estimator [34]. However, the difference in MSE performance should be small, since the dependent distortion noises are relatively weak.

Corollary 1. *Consider the special case of* $\mathbf{R} = \lambda\mathbf{I}$ *and* $\mathbf{S} = \mathbf{0}$*. The error covariance matrix in* (11) *becomes*

$$\mathbf{C} = \lambda \left(1 - \frac{p^{\mathrm{UE}}\lambda}{p^{\mathrm{UE}}\lambda(1 + \kappa_t^{\mathrm{UE}} + \kappa_r^{\mathrm{BS}}) + \sigma_{\mathrm{BS}}^2} \right) \mathbf{I}. \tag{13}$$

In the high UL power regime, we have

$$\lim_{p^{\mathrm{UE}} \to \infty} \mathbf{C} = \lambda \left(1 - \frac{1}{1 + \kappa_t^{\mathrm{UE}} + \kappa_r^{\mathrm{BS}}} \right) \mathbf{I}. \tag{14}$$

This corollary brings important insights on the average estimation error per element in $\mathbf{h}$. The error variance is given by the factor in front of the identity matrix in (13). It is independent of the number of antennas N, thus letting N grow large neither increases nor decreases the estimation error *per element*.[5] The estimation error is clearly a decreasing function of the pilot power $p^{\mathrm{UE}} = |d|^2$, but contrary to the ideal hardware case the error variance is *not* converging to zero as $p^{\mathrm{UE}} \to \infty$. As seen in (14), there is a strictly positive error floor of $\lambda(1 - \frac{1}{1+\kappa_t^{\mathrm{UE}}+\kappa_r^{\mathrm{BS}}})$ due to the transceiver hardware impairments. Thus, perfect estimation accuracy cannot be achieved in practice, not even asymptotically. The error floor is characterized by the sum of the levels of impairments κ_t^{UE} and κ_r^{BS} in the transmitter and receiver hardware, respectively. In terms of estimation accuracy, it is thus equally important to have high-quality hardware at the BS and at the UE.

Non-ideal hardware exhibits an error floor also when $\mathbf{R}$ is non-diagonal and when there is interference such that $\mathbf{S} \neq \mathbf{0}$; the general high-power limit is easily computed from (11). In fact, the results hold for any zero-mean channel and interference distributions with covariance matrices $\mathbf{R}$ and $\mathbf{S}$, because the LMMSE estimator and its MSE are computed using only the first two moments of the statistical distributions [32], [34].

A. Impact of the Pilot Length

The LMMSE estimator in Theorem 1 considers a scalar pilot signal d, which is sufficient to excite all N channel dimensions in the UL and is used in Section IV-B to derive lower bounds on the UL and DL capacities. With ideal hardware and a total pilot energy constraint, a scalar pilot signal is also sufficient to minimize the MSE [34]. In contrast, we have non-ideal

[4] For example, the ith element of $\boldsymbol{\eta}_r^{\mathrm{BS}}$ can be expressed as $x_i = h_i\xi_i$, which is the product of the ith channel coefficient $h_i \sim \mathcal{CN}(0, r_{ii})$ and an independent variable $\xi_i \sim \mathcal{CN}(0, \kappa_r^{\mathrm{BS}}p^{\mathrm{UE}})$. The joint product distribution is complex double Gaussian with the PDF $f(x_i) = \frac{2}{\pi\mu_i} K_0\left(2\frac{|x_i|}{\sqrt{\mu_i}}\right)$, where $\mu_i = r_{ii}\kappa_r^{\mathrm{BS}}p^{\mathrm{UE}}$ is the variance and $K_0(\cdot)$ denotes the zeroth-order modified Bessel function of the second kind [36].

[5] The MSE per element is finite, i.e. $\frac{1}{N}\mathrm{tr}(\mathbf{C}) < \infty$, but the sum MSE behaves as $\mathrm{tr}(\mathbf{C}) \to \infty$ when $N \to \infty$ since the number of elements grows.

hardware and per-symbol energy constraints in this paper. In this case we can improve the MSE by increasing the pilot length.

Suppose we use a pilot signal $\mathbf{d} \in \mathbb{C}^{1\times B}$ that spans $1 \leq B \leq T_{\mathrm{pilot}}^{\mathrm{UL}}$ channel uses and where each element of $\mathbf{d}$ has squared norm p^{UE}. A simple estimation approach would be to compute B separate LMMSE estimates, $\hat{\mathbf{h}}_i = \mathbf{h} - \boldsymbol{\epsilon}_i$ for $i = 1, \ldots, B$, using Theorem 1. By averaging, we obtain

$$\widehat{\hat{\mathbf{h}}} = \frac{1}{B}\sum_{i=1}^{B} \hat{\mathbf{h}}_i = \mathbf{h} - \frac{1}{B}\sum_{i=1}^{B} \boldsymbol{\epsilon}_i. \tag{15}$$

If the distortion noises are temporally uncorrelated and identically distributed, the MSE of the estimate $\widehat{\hat{\mathbf{h}}}$ is

$$\mathbb{E}\left\{ \left(\frac{1}{B}\sum_{i=1}^{B} \boldsymbol{\epsilon}_i \right)^H \left(\frac{1}{B}\sum_{j=1}^{B} \boldsymbol{\epsilon}_j \right) \right\} = \frac{\mathrm{tr}(\mathbf{C})}{B}. \tag{16}$$

Hence, the MSE goes to zero as $1/B$ when we increase the pilot length B, although the MSE per pilot channel use is limited by the non-zero error floor demonstrated in Corollary 1. This is interesting because one pilot signal with energy Bp^{UE} exhibits a noise floor, while B pilot signals with energy p^{UE} per signal does not.[6] This stands in contrast to the case of ideal hardware, where the MSE is exactly the same in both cases [34]. The reason is that we can average out the distortion noise (similar to the law of large numbers) when we have B independent realizations.

Despite the averaging effect, we stress that $B \leq T_{\mathrm{coher}}$ and thus there is always an estimation error floor for non-ideal hardware—we can, at most, reduce the floor by a factor $1/T_{\mathrm{coher}}$ by increasing the pilot length. Moreover, the derivation above is based on having temporally uncorrelated distortions, but the distortions might be temporally correlated in practice (especially if the same pilot signal d is transmitted multiple times through the same hardware). In these cases, the benefit of increasing B is smaller and $\widehat{\hat{\mathbf{h}}}$ should be replaced by an estimator that exploits the temporal correlation by estimating $\mathbf{h}$ jointly from all the B observations. Finally, we note that it is of great interest to find the B that maximizes some measure of system-wide performance, but this is outside the scope of our current paper. We refer to [34], [35], [37], [38] for some relevant works in the case of ideal hardware.

B. Numerical Illustrations

This section exemplifies the impact of transceiver hardware impairments on the channel estimation accuracy.

In Fig. 3, we consider $N = 50$ antennas at the BS and no interference (i.e., $\mathbf{S} = \mathbf{0}$). The channel covariance matrix $\mathbf{R}$ is generated by the exponential correlation model from [39], which means that the (i,j)th element of $\mathbf{R}$ is

$$[\mathbf{R}]_{i,j} = \begin{cases} \delta\, r^{j-i}, & i \leq j, \\ \delta\, (r^{i-j})^*, & i > j, \end{cases} \tag{17}$$

where δ is an arbitrary scaling factor. This model basically describes a uniform linear array (ULA) where the correlation factor between adjacent antennas is given by $|r|$ (for $0 \leq |r| \leq 1$) and the phase $\angle r$ describes the angle of arrival/departure as seen from the array. The correlation factor $|r|$ determines the eigenvalue spread in $\mathbf{R}$, while $\angle r$ determines the corresponding eigenvectors. Since we simulate channel estimation without interference, the angle has no impact on the MSE and we can let r be real-valued without loss of generality. We consider a correlation coefficient of $r = 0.7$, which is a modest correlation in the sense of behaving similarly to an array with half-wavelength antenna spacings and a large angular spread of 45 degrees (cf. [40, Fig. 1] which shows how practical angular spreads map non-linearly to $|r|$).

Fig. 3 shows the relative estimation error per channel element, $\mathrm{MSE}_{\mathrm{rel}} = \frac{\mathrm{MSE}}{\mathrm{tr}(\mathbf{R})}$, as a function of the average SNR in the UL, defined as

$$\mathtt{SNR}^{\mathrm{UL}} = p^{\mathrm{UE}} \frac{\mathrm{tr}(\mathbf{R})}{N\sigma_{\mathrm{BS}}^2}. \tag{18}$$

Based on the typical EVM ranges described in Remark 1, we consider four hardware setups with different levels of impairments: $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} \in \{0, 0.05^2, 0.1^2, 0.15^2\}$. We compare the LMMSE estimator in Theorem 1 with the conventional impairment-ignoring MMSE estimator from [32]–[34].[7]

Fig. 3 confirms that there are non-zero error floors at high SNRs, as proved by Corollary 1 and the subsequent discussion. The error floor increases with the levels of impairments. The estimation error is very close to the floor when the uplink SNR reaches 20–30 dB, thus further increase in SNR only brings minor improvement. This tells us that we need an uplink SNR of at least 20 dB to fully utilize massive MIMO, because coherent transmission/reception requires accurate CSI. Lower SNRs can be compensated by adding extra antennas (see Fig. 6 in Section IV), but the practical performance not as large. Moreover, Fig. 3 shows that the conventional impairment-ignoring estimator is only slightly worse than the proposed LMMSE estimator. This indicates that although hardware impairments greatly affect the estimation performance, it only brings minor changes to the structure of the optimal estimator.

The influence of the estimation error floors depend on the anticipated spectral efficiency, the uplink SNR, and the number of antennas. To gain some insight, suppose we have ideal hardware and that the fraction of channel uses allocated for UL data transmission is $T_{\mathrm{data}}^{\mathrm{UL}}/T_{\mathrm{coher}} = 0.45$. The uplink spectral efficiency can then be approximated as

$$0.45 \log_2 \left(1 + \frac{1 - \mathrm{MSE}_{\mathrm{rel}}}{\mathrm{MSE}_{\mathrm{rel}} + \frac{1}{N\mathtt{SNR}^{\mathrm{UL}}}} \right) \tag{19}$$

by using [41, Lemma 1]. When the number of antennas is large, such that $N\,\mathtt{SNR}^{\mathrm{UL}} \to \infty$, this approximation gives a spectral efficiency of 1.5 [bit/channel use] for $\mathrm{MSE}_{\mathrm{rel}} = 10^{-1}$ and 4.5 [bit/channel use] for $\mathrm{MSE}_{\mathrm{rel}} = 10^{-3}$. The impact of the estimation errors on systems with non-ideal hardware is

[6] Since we have per-symbol energy constraints, what we really compare is one system that has an average symbol energy of Bp^{UE} and one with p^{UE}.

[7] Note that the MSE of any linear estimator $\tilde{\mathbf{A}}\mathbf{z}$ can be computed by plugging the matrix $\tilde{\mathbf{A}}$ into the general MSE expression in (12). The difference in MSE is easily quantified by comparing with $\mathrm{tr}(\mathbf{C})$ using (11).

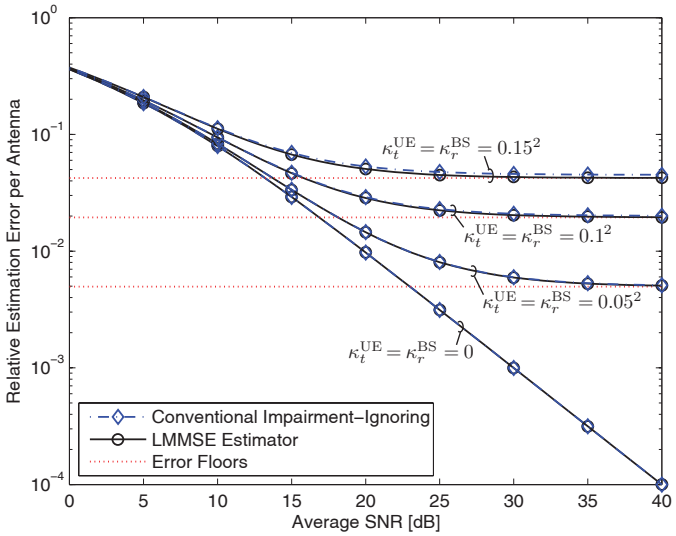


Fig. 3: Estimation error per antenna element for the LMMSE estimator in Theorem 1 and the conventional impairment-ignoring MMSE estimator. Transceiver hardware impairments create non-zero error floors.

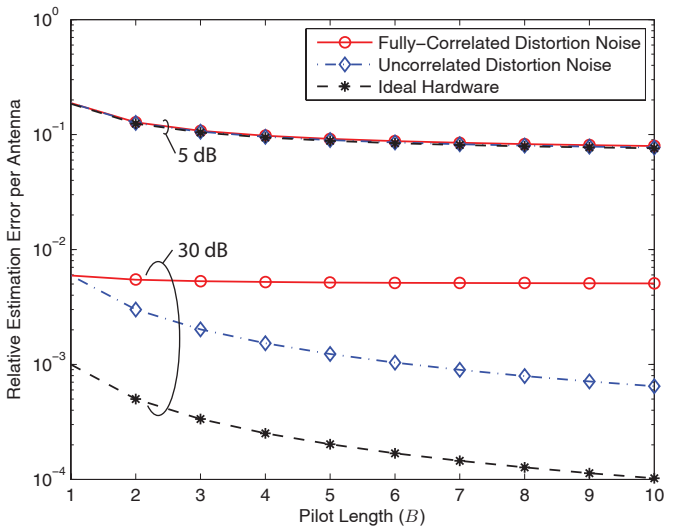


Fig. 4: Estimation error per antenna element for the LMMSE estimator in Theorem 1 as a function of the pilot length B. The levels of impairments are $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} = 0.05^2$ and different temporal correlations are considered.

considered in Section IV, where we derive lower and upper capacity bounds and analyze these for different SNRs and number of antennas.

Next, we illustrate the possible improvement in estimation accuracy by increasing the pilot length to comprise $B \geq 1$ channel uses. As discussed in Section III-A, it is not clear whether the distortion noise is temporally uncorrelated or correlated in practice. Therefore, we fix the levels of impairments at $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} = 0.05^2$ and consider the two extremes: temporally uncorrelated and fully correlated distortion noises. The latter means that the distortion noise realizations are the same for all B channel uses, since the same pilot signal is always distorted in the same way. The channel and interference statistics are as in the previous figure (i.e., $N = 50$, $\mathbf{S} = \mathbf{0}$, and $\mathbf{R}$ is given by the exponential correlation model with $r = 0.7$).

The relative estimation error per antenna element is shown in Fig. 4 as a function of the pilot length. We also show the performance with ideal hardware as a reference. At a low SNR of 5 dB, hardware impairments have little impact and there is a small but clear gain from increasing the pilot length because the total pilot energy increases as Bp^{UE}. At a high SNR of 30 dB, the temporal correlation has a large impact. Only small improvements are possible in the fully correlated case, since only the receiver noise can be mitigated by increasing B. In the uncorrelated case the distortion noise can be also mitigated by increasing B. This gives a logarithmic slope similar to the case of ideal hardware. We stress that the actual performance lies somewhere in between the two extremes.

Next, we consider different channel covariance models:

1) Uncorrelated antennas $\mathbf{R} = \mathbf{I}$. (Equivalent to the exponential correlation model in (17) with $r = 0$.)
2) Exponential correlation model with $r = 0.7$.
3) One-ring model with 20 degrees angular spread [42].
4) One-ring model with 10 degrees angular spread [42].

The exponential correlation model was defined in (17). The classic one-ring model assumes a ring of scatterers around the UE, while there is no scattering close to the BS [42]. From the BS perspective, the multipath components arrive from a main angle of arrival (here: 30 degrees) and a small angular spread around it (here: 10 or 20 degrees). The BS is assumed to have a ULA with half-wavelength antenna spacings. An important property of this model is that $\mathbf{R}$ might not have full rank as N grows large [43], [44], due to insufficient scattering.

The relative estimation error per channel element is shown in Fig. 5 for these four channel covariance models. We consider two SNRs (5 and 30 dB), hardware impairments with $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} = 0.05^2$, and show the estimation errors as a function of the number of BS antennas. The main observation from Fig. 5 is that the choice of covariance model has a large impact on the estimation accuracy. It was proved in [34] that spatially correlated channels are easier to estimate and this is consistent with our results; increasing the coefficient r in the exponential correlation model and decreasing the angular spread in the one-ring model lead to higher spatial correlation and smaller errors in Fig. 5. However, the error floors due to hardware impairments make the difference between the models reduce with the SNR. Moreover, the estimation error *per* antenna is virtually independent of N in the exponential correlation model, while increasing N improves the error *per* antenna in the one-ring model. This is explained by the limited richness of the propagation environment in the one-ring model, which is a physical property that we can expect in practice.

Remark 2 (Acquiring Large Covariance Matrices)**.** *The proposed channel estimator requires knowledge of the $N \times N$ covariance matrices $\mathbf{R}$ and $\mathbf{S}$. It becomes increasingly difficult to acquire consistent estimates of covariance matrices as their dimensions grow large [45]. Fortunately, the channel statistics have a much larger coherence time and coherence bandwidth than the channel realization itself; thus, one can obtain many more observations in the covariance estimation than in channel vector estimation. Robust covariance estimators for*

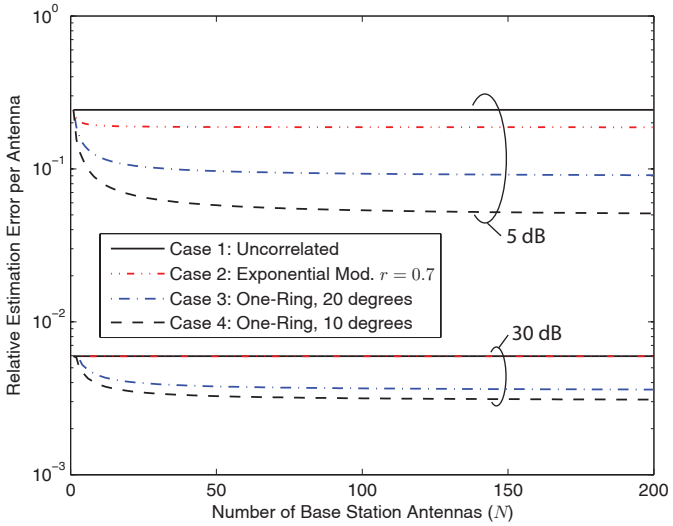


Fig. 5: Estimation error per antenna element for the LMMSE estimator in Theorem 1 as a function of the number of BS antennas. Four different channel covariance models are considered and $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} = 0.05^2$.

large matrices were recently considered in [46]. The impact of imperfect covariance information on the channel estimation accuracy was analyzed in [47]. The authors observe that the usual improvement in MSE from having spatial correlation vanishes if the covariance information cannot be trusted, but the MSE degradation is otherwise small (if the estimated covariance matrices are robustified). Another problem is that the large-dimensional matrix inversion in (9) *is very computationally expensive, but [47] proposed low-complexity approximations based on polynomial expansions.*

Instead of acquiring the covariance matrix of a user directly, the coverage area can be divided into "location bins" with approximately the same channel statistics within each bin [48]. By acquiring and storing the covariance matrices for each bin in advance, it is sufficient to estimate the location of a user and then associate the user with the corresponding bin.

IV. Downlink and Uplink Data Transmission

This section analyzes the ergodic channel capacities of the downlink in (2) and the uplink in (6), under the fixed TDD protocol depicted in Fig. 2. More precisely, we derive upper and lower capacity bounds that reveal the fundamental impact of non-ideal hardware. These bounds are based on having perfect CSI (i.e., exact knowledge of $\mathbf{h}$) and imperfect pilot-based CSI estimation (using the LMMSE estimator in Theorem 1), respectively. Since these are two extremes, the capacity bounds hold when using the channel estimation technique proposed in Section III and for any better CSI acquisition technique that can be derived in the future. We now define the DL and UL capacities for arbitrary CSI quality at the BS and UE.

We consider the ergodic capacity (in bit/channel use) of the memoryless DL system in (2). In each coherence period, the BS has some arbitrary imperfect knowledge $\mathcal{H}^{\mathrm{BS}}$ of the current channel states $\mathcal{H}$ and uses it to select the conditional distribution $f(\mathbf{s}|\mathcal{H}^{\mathrm{BS}})$ of the data signal $\mathbf{s}$. The UE has a separate arbitrary imperfect knowledge $\mathcal{H}^{\mathrm{UE}}$ of the current channel states $\mathcal{H}$ and uses it to decode the data. Based on the well-known capacity expressions in [49], the ergodic DL capacity is

$$\mathsf{C}^{\mathrm{DL}} = \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \mathbb{E}\left\{ \max_{f(\mathbf{s}|\mathcal{H}^{\mathrm{BS}}) : \mathbb{E}\{\|\mathbf{s}\|_2^2\} \leq p^{\mathrm{BS}}} \mathcal{I}(\mathbf{s}; y | \mathcal{H}, \mathcal{H}^{\mathrm{BS}}, \mathcal{H}^{\mathrm{UE}}) \right\} \tag{20}$$

where $\mathcal{I}(\mathbf{s}; y|\mathcal{H}, \mathcal{H}^{\mathrm{BS}}, \mathcal{H}^{\mathrm{UE}})$ denotes the conditional mutual information between the received signal y and data signal $\mathbf{s}$ for a given channel realization $\mathcal{H}$ and given channel knowledge $\mathcal{H}^{\mathrm{BS}}$ and $\mathcal{H}^{\mathrm{UE}}$. The expectation in (20) is taken over the joint distribution of $\mathcal{H}$, $\mathcal{H}^{\mathrm{BS}}$, and $\mathcal{H}^{\mathrm{UE}}$. Note that the factor $T_{\mathrm{data}}^{\mathrm{DL}}/T_{\mathrm{coher}}$ is the fixed fraction of channel uses allocated for DL data transmission.

In addition, the ergodic capacity (in bit/channel use) of the memoryless fading UL system in (6) is

$$\mathsf{C}^{\mathrm{UL}} = \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \mathbb{E}\left\{ \max_{f(d|\mathcal{H}^{\mathrm{UE}}) : \mathbb{E}\{|d|^2\} \leq p^{\mathrm{UE}}} \mathcal{I}(d; \mathbf{z} | \mathcal{H}, \mathcal{H}^{\mathrm{BS}}, \mathcal{H}^{\mathrm{UE}}) \right\} \tag{21}$$

where $\mathcal{I}(d; \mathbf{z}|\mathcal{H}, \mathcal{H}^{\mathrm{BS}}, \mathcal{H}^{\mathrm{UE}})$ denotes the conditional mutual information between the received signal $\mathbf{z}$ and data signal d for a given channel realization $\mathcal{H}$ and given channel knowledge $\mathcal{H}^{\mathrm{BS}}$ and $\mathcal{H}^{\mathrm{UE}}$. The conditional probability distribution of the data signal is denoted $f(d|\mathcal{H}^{\mathrm{UE}})$ and the expectation in (21) is taken over the joint distribution of $\mathcal{H}, \mathcal{H}^{\mathrm{BS}}, \mathcal{H}^{\mathrm{UE}}$. The fraction of channel uses allocated for UL data transmission is $T_{\mathrm{data}}^{\mathrm{UL}}/T_{\mathrm{coher}}$.

There are a few implicit properties in the capacity definitions. Firstly, the interference variance $I_{\mathcal{H}}^{\mathrm{UE}}$ and covariance matrix $\mathbf{Q}_{\mathcal{H}}$ depend on the channel realizations $\mathcal{H}$ and change between coherence periods. We are not limiting the analysis to any specific interference models but take care of it in the capacity bounds; the lower bounds treat the interference as Gaussian noise, while the upper bounds assume perfect interference suppression. Section VI describes the interference in multi-cell scenarios in detail. Secondly, we assume that the distortion noises are temporally independent, which is a good model when the data signals are also temporally independent.

The next subsections study the capacity behavior in the limit of infinitely many BS antennas ($N \to \infty$), which bring insights on how hardware impairments affect channels with large antenna arrays. The DL and UL are analyzed side-by-side since the results follow from similar derivations.

A. Upper Bounds on Channel Capacities

Upper bounds on the capacities in (20) and (21) can be obtained by adding extra channel knowledge and removing all interference (i.e., $I_{\mathcal{H}}^{\mathrm{UE}} = 0$ and $\mathbf{Q}_{\mathcal{H}} = \mathbf{0}$). We assume that the UL/DL pilot signals provide the BS and UE with perfect channel knowledge in each coherence period: $\mathcal{H}^{\mathrm{BS}} = \mathcal{H}^{\mathrm{UE}} = \mathcal{H}$. Since the receiver noise and distortion noises in (2) and (6) are circularly symmetric complex Gaussian distributed and independent of the useful signals under perfect CSI, we deduce that Gaussian signaling is optimal in the DL and UL [2] and that single-stream transmission with $\mathrm{rank}(\mathbf{W}) = 1$ is sufficient to achieve optimality [6]; that is, we can set $\mathbf{s} = \mathbf{w}s$

for $s \sim \mathcal{CN}(0, p^{\mathrm{BS}})$ and some unit-norm beamforming vector $\mathbf{w}$ in the DL and $d \sim \mathcal{CN}(0, p^{\mathrm{UE}})$ in the UL. This gives us the following initial upper bounds.

Lemma 1. *The downlink and uplink capacities in* (20) *and* (21)*, respectively, are bounded as*

$$\mathtt{C}^{\mathrm{DL}} \leq \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \times \mathbb{E}\left\{ \log_2(1 + \mathbf{h}^H \left(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \kappa_r^{\mathrm{UE}} \mathbf{h}\mathbf{h}^H + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I} \right)^{-1} \mathbf{h} \right\} \tag{22}$$

$$\mathtt{C}^{\mathrm{UL}} \leq \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \times \mathbb{E}\left\{ \log_2 \left(1 + \mathbf{h}^H \left(\kappa_t^{\mathrm{UE}} \mathbf{h}\mathbf{h}^H + \kappa_r^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{BS}}^2}{p^{\mathrm{UE}}} \mathbf{I} \right)^{-1} \mathbf{h} \right) \right\} \tag{23}$$

where $\mathbf{D}_{|\mathbf{h}|^2} = \mathrm{diag}(|h_1|^2, \ldots, |h_N|^2)$ *with* $\mathbf{h} = [h_1 \ \ldots \ h_N]^T$*. These upper bounds are achieved with equality under perfect CSI, using the beamforming vector*

$$\mathbf{w}_{\mathrm{upper}}^{\mathrm{DL}} = \frac{(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}^*}{\left\| \left(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I} \right)^{-1} \mathbf{h}^* \right\|_2} \tag{24}$$

in the downlink and by applying a receive combining vector

$$\mathbf{w}_{\mathrm{upper}}^{\mathrm{UL}} = \frac{(\kappa_r^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{BS}}^2}{p^{\mathrm{UE}}} \mathbf{I})^{-1} \mathbf{h}}{\left\| (\kappa_r^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{BS}}^2}{p^{\mathrm{UE}}} \mathbf{I})^{-1} \mathbf{h} \right\|_2}. \tag{25}$$

in the uplink.[8]

Proof: The proof is given in Appendix C-A. ∎

Note that the beamforming vector in (24) and receive combining vector in (25) only depend on the channel vector $\mathbf{h}$, hardware impairments at the BS, and the receiver noise. Hardware impairments at the UE have no impact on $\mathbf{w}_{\mathrm{upper}}^{\mathrm{DL}}$ and $\mathbf{w}_{\mathrm{upper}}^{\mathrm{UL}}$ since their distortion noise essentially act as an interferer with the same channel as the data signal; thus filtering cannot reduce it.

The bounds in Lemma 1 are not amenable to simple analysis, but the lemma enables us to derive further bounds on the channel capacities that are expressed in closed form.

Theorem 2. *The downlink and uplink capacities in* (20) *and* (21)*, respectively, are bounded as*

$$\mathtt{C}^{\mathrm{DL}} \leq \mathtt{C}_{\mathrm{upper}}^{\mathrm{DL}} = \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \log_2 \left(1 + \frac{G^{\mathrm{DL}}}{1 + \kappa_r^{\mathrm{UE}} G^{\mathrm{DL}}} \right) \tag{26}$$

$$\mathtt{C}^{\mathrm{UL}} \leq \mathtt{C}_{\mathrm{upper}}^{\mathrm{UL}} = \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \log_2 \left(1 + \frac{G^{\mathrm{UL}}}{1 + \kappa_t^{\mathrm{UE}} G^{\mathrm{UL}}} \right) \tag{27}$$

[8]A receive combining vector $\mathbf{w}$ is a linear filter $\mathbf{w}^H \mathbf{z}$ that transforms the system into an effective single-input single-output (SISO) system.

where $r_{11}, \ldots, r_{NN}$ *are the diagonal elements of* $\mathbf{R}$*,*

$$G^{\mathrm{DL}} = \sum_{i=1}^{N} \frac{1}{\kappa_t^{\mathrm{BS}}} \left(1 - \frac{\sigma_{\mathrm{UE}}^2 e^{\frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}} \kappa_t^{\mathrm{BS}} r_{ii}}}}{p^{\mathrm{BS}} \kappa_t^{\mathrm{BS}} r_{ii}} E_1 \left(\frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}} \kappa_t^{\mathrm{BS}} r_{ii}} \right) \right), \tag{28}$$

$$G^{\mathrm{UL}} = \sum_{i=1}^{N} \frac{1}{\kappa_r^{\mathrm{BS}}} \left(1 - \frac{\sigma_{\mathrm{BS}}^2 e^{\frac{\sigma_{\mathrm{BS}}^2}{p^{\mathrm{UE}} \kappa_r^{\mathrm{BS}} r_{ii}}}}{p^{\mathrm{UE}} \kappa_r^{\mathrm{BS}} r_{ii}} E_1 \left(\frac{\sigma_{\mathrm{BS}}^2}{p^{\mathrm{UE}} \kappa_r^{\mathrm{BS}} r_{ii}} \right) \right), \tag{29}$$

and $E_1(x) = \int_1^\infty \frac{e^{-tx}}{t} dt$ *denotes the exponential integral.*

Proof: The proof is given in Appendix C-B. ∎

These closed-form upper bounds provide important insights on the achievable DL and UL performance under transceiver hardware impairments. In particular, the following two corollaries provide some ultimate capacity limits in the asymptotic regimes of many BS antennas or large transmit powers.

Corollary 2. *The downlink upper capacity bound in* (26) *has the following asymptotic properties:*

$$\lim_{p^{\mathrm{BS}} \to \infty} \mathtt{C}_{\mathrm{upper}}^{\mathrm{DL}} = \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \log_2 \left(1 + \frac{N}{\kappa_t^{\mathrm{BS}} + \kappa_r^{\mathrm{UE}} N} \right) \tag{30}$$

$$\lim_{N \to \infty} \mathtt{C}_{\mathrm{upper}}^{\mathrm{DL}} = \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \log_2 \left(1 + \frac{1}{\kappa_r^{\mathrm{UE}}} \right). \tag{31}$$

Proof: The diagonal elements of $\mathbf{R}$ satisfy $r_{ii} > 0 \ \forall i$, by definition, thus $G^{\mathrm{DL}} \to \sum_{i=1}^{N} \frac{1}{\kappa_t^{\mathrm{BS}}} = \frac{N}{\kappa_t^{\mathrm{BS}}}$ as $p^{\mathrm{BS}} \to \infty$ for fixed N, giving (30). The positive diagonal elements also implies $\frac{1}{N} G^{\mathrm{DL}} > 0$ as $N \to \infty$, thus $\frac{G^{\mathrm{DL}}}{1 + \kappa_r^{\mathrm{UE}} G^{\mathrm{DL}}} - \frac{G^{\mathrm{DL}}}{\kappa_r^{\mathrm{UE}} G^{\mathrm{DL}}} \to 0$ as $N \to \infty$ which turns (26) into (31). ∎

This corollary shows that the DL capacity has finite ceilings when either the DL transmit power p^{BS} or the number of BS antennas N grow large. The ceilings depend on the impairment parameters κ_t^{BS} and κ_r^{UE}, but the UE impairments are clearly N times more influential. Note that even very small hardware impairments will ultimately limit the capacity. In other words, the ever-increasing capacity observed in the high-SNR and large-N regimes with ideal transceiver hardware (cf. [9]–[13]) is not easily achieved in practice.

The next corollary provides analogous results for the UL.

Corollary 3. *The uplink upper capacity bound in* (27) *has the following asymptotic properties:*

$$\lim_{p^{\mathrm{UE}} \to \infty} \mathtt{C}_{\mathrm{upper}}^{\mathrm{UL}} = \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \log_2 \left(1 + \frac{N}{\kappa_r^{\mathrm{BS}} + \kappa_t^{\mathrm{UE}} N} \right) \tag{32}$$

$$\lim_{N \to \infty} \mathtt{C}_{\mathrm{upper}}^{\mathrm{UL}} = \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \log_2 \left(1 + \frac{1}{\kappa_t^{\mathrm{UE}}} \right). \tag{33}$$

Proof: This is proved analogously to Corollary 2. ∎

As seen from Corollary 3, the UL capacity also has finite ceilings when either the UL transmit power p^{UE} or the number of antennas N grow large. Analogous to the DL, the UE impairments are N times more influential than the BS impairments and thus dominate as $N \to \infty$.

The upper bounds in Corollaries 2 and 3 show that the DL and UL capacities are fundamentally limited by the transceiver

hardware impairments. To be certain of the cause of these limits, we also need lower bounds on the channel capacities.

B. Lower Bounds on Channel Capacities

We obtain lower capacity bounds by making the potentially limiting assumptions of Gaussian codebooks, treating interference as Gaussian noise, using linear single-stream beamforming in the DL, using linear receive combining in the UL, pilot-based channel estimation as in Theorem 1, and the entropy-maximizing Gaussian distribution on the CSI uncertainty at the receiver of the DL and UL.[9] The resulting lower bound is given in the following theorem.

Theorem 3. *Let $\tilde{\mathcal{H}}^{UE}$ and $\tilde{\mathcal{H}}^{BS}$ denote the CSI available in the decoding at the receiver in the downlink and uplink, respectively. These are degraded as compared to $\mathcal{H}^{UE}$ and $\mathcal{H}^{BS}$ or equal. The downlink and uplink capacities in* (20) *and* (21)*, respectively, are then bounded as*

$$\mathtt{C}^{\mathrm{DL}} \geq \mathtt{C}^{\mathrm{DL}}_{\mathrm{lower}} = \frac{T^{\mathrm{DL}}_{\mathrm{data}}}{T_{\mathrm{coher}}} \mathbb{E}\left\{\log_2\left(1+\mathtt{SINR}^{\mathrm{DL}}_{\mathrm{lower}}(\mathbf{v}^{\mathrm{DL}})\right)\right\} \quad (34)$$

$$\mathtt{C}^{\mathrm{UL}} \geq \mathtt{C}^{\mathrm{UL}}_{\mathrm{lower}} = \frac{T^{\mathrm{UL}}_{\mathrm{data}}}{T_{\mathrm{coher}}} \mathbb{E}\left\{\log_2\left(1+\mathtt{SINR}^{\mathrm{UL}}_{\mathrm{lower}}(\mathbf{v}^{\mathrm{UL}})\right)\right\} \quad (35)$$

where the beamforming vector $\mathbf{v}^{\mathrm{DL}} = [v_1^{\mathrm{DL}} \ \ldots \ v_{K_r}^{\mathrm{DL}}]^T$ and the receive combining vector $\mathbf{v}^{\mathrm{UL}} = [v_1^{\mathrm{UL}} \ \ldots \ v_{K_r}^{\mathrm{UL}}]^T$ are functions of $\hat{\mathbf{h}}$ and have unit norms. The expectations are taken over $\tilde{\mathcal{H}}^{UE}$ and $\tilde{\mathcal{H}}^{BS}$, while the SINR expressions for DL and UL are given in (36) *and* (37)*, respectively, at the top of the next page.*

Proof: This theorem is obtained by taking lower bounds on the mutual information in the same way as was previously proposed in [5] and [41]. This bounding technique was applied to massive MIMO systems with ideal hardware in [11]–[13] (among others), by making the limiting assumptions listed in the beginning of this subsection. The distortion noises from non-ideal hardware act as additional noise sources with spatially correlated covariance matrices, thus these can easily be incorporated into the proofs used in previous works. ■

This theorem is the key to the lower capacity bounding in this paper. The lower bounds in (34) and (35) can be computed numerically for any channel distribution and any way of selecting the beamforming vector (in the DL) and receiver combining vector (in the UL) from the channel estimate $\hat{\mathbf{h}}$, provided that the conditional distribution of $(\mathbf{h}, \hat{\mathbf{h}})$ given $\tilde{\mathcal{H}}$ can be characterized.[10] To bring explicit insights on the behavior when the number of antennas, N, grows large, we have the following result for the cases of (approximate) maximum ratio transmission (MRT) in the DL and (approximate) maximum ratio combining (MRC) in the UL.

Theorem 4. *Assume that no instantaneous CSI is utilized for decoding (i.e., $\tilde{\mathcal{H}}^{BS} = \tilde{\mathcal{H}}^{UE} = \emptyset$). For $\mathbf{v} = \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}$ the terms in* (36) *and* (37) *behave as*

$$\left|\mathbb{E}\{\mathbf{h}^H\mathbf{v}\}\right|^2 = \left|\mathbb{E}\left\{\varphi\right\}\right|^2 \mathrm{tr}(\mathbf{R}-\mathbf{C}) + \mathcal{O}(\sqrt{N}) \quad (38)$$

$$\mathbb{E}\left\{|\mathbf{h}^H\mathbf{v}|^2\right\} = \mathbb{E}\left\{|\varphi|^2\right\} \mathrm{tr}(\mathbf{R}-\mathbf{C}) + \mathcal{O}(\sqrt{N}) \quad (39)$$

$$\sum_{i=1}^{N} \mathbb{E}\{|h_i|^2|v_i|^2\} = \mathcal{O}(1) \quad (40)$$

where

$$\varphi = \frac{(1+d^{-1}\eta_t^{\mathrm{UE}})\sqrt{\mathrm{tr}(\mathbf{R}-\mathbf{C})}}{\sqrt{\mathrm{tr}\big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\big)}} \quad (41)$$

is a function of the stochastic variable η_t^{UE} while $\mathbf{A} = d^*\mathbf{R}\bar{\mathbf{Z}}^{-1}$ and $\mathbf{\Psi} = p^{\mathrm{UE}}\kappa_r^{\mathrm{BS}}\mathbf{R}_{\mathrm{diag}} + \mathbf{S} + \sigma^2_{\mathrm{BS}}\mathbf{I}$ are deterministic matrices.

Proof: The proof is given in Appendix C-C. ■

Similar asymptotic behaviors were derived in [11]–[13] for the case of ideal hardware.[11] In the general case with hardware impairments, the expectations of φ and $|\varphi|^2$ must be computed numerically, because the randomness of the scalar distortion noise η_t^{UE} at the UE remains even when N grows large. In the special case of $\kappa_t^{\mathrm{UE}} = 0$ (which implies $\eta_t^{\mathrm{UE}} = 0$), (38) and (39) both reduce to $\mathrm{tr}(\mathbf{R}-\mathbf{C}) + \mathcal{O}(\sqrt{N})$. For $\kappa_t^{\mathrm{UE}} > 0$, the terms in Theorem 4 are easy to compute numerically.

Based on this result, we provide now an asymptotic characterization of the downlink capacity.

Corollary 4. *Consider the DL with beamforming vector $\mathbf{v} = \frac{\hat{\mathbf{h}}^*}{\|\hat{\mathbf{h}}\|_2}$ and $\tilde{\mathcal{H}}^{UE} = \emptyset$. If $\mathbb{E}\{I^{\mathrm{UE}}_{\mathcal{H}}\} \leq \mathcal{O}(N^n)$ for some $n < 1$, the lower capacity bound in* (34) *can be expressed as*

$$\mathtt{C}^{\mathrm{DL}} \geq \frac{T^{\mathrm{DL}}_{\mathrm{data}}}{T_{\mathrm{coher}}} \times \log_2\left(1+\frac{\left|\mathbb{E}\left\{\varphi\right\}\right|^2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)}{(1+\kappa_r^{\mathrm{UE}})\mathbb{E}\left\{|\varphi|^2\right\} - \left|\mathbb{E}\left\{\varphi\right\}\right|^2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}+\frac{1}{N^{1-n}}\right)}\right) \quad (42)$$

where φ is given in (41)*. The terms $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ and $\mathcal{O}\left(\frac{1}{N^{1-n}}\right)$ vanish when $N \to \infty$, while the other terms are strictly positive in the limit.*

Proof: The expression (42) is obtained from (34) by plugging in the expressions in Theorem 4 and multiplying each term by $\frac{1}{\mathrm{tr}(\mathbf{R}-\mathbf{C})} = \frac{1}{p^{\mathrm{UE}}\mathrm{tr}(\mathbf{R}\bar{\mathbf{Z}}^{-1}\mathbf{R})} = \mathcal{O}(N^{-1})$. The interference term becomes $\frac{\mathbb{E}\{I^{\mathrm{UE}}_{\mathcal{H}}\}}{p^{\mathrm{BS}}\mathrm{tr}(\mathbf{R}-\mathbf{C})} = \mathcal{O}\left(\frac{1}{N^{1-n}}\right)$. ■

Combining the upper bound in Corollary 2 with the lower bound in Corollary 4, we have a clear characterization of the DL capacity behavior when $N \to \infty$. Both bounds are independent of κ_t^{BS} in the limit, thus the transmitter hardware of the BS plays little role in massive MIMO systems. Contrary to the upper bound, the level of receiver hardware impairments

[9] The linear processing assumption is motivated by its asymptotic optimality as $N \to \infty$ [9].

[10] Finding such a characterization is a challenging task, except for the case $\tilde{\mathcal{H}}^{\mathrm{BS}} = \tilde{\mathcal{H}}^{\mathrm{UE}} = \emptyset$ considered in Theorem 4.

[11] We stress that the assumption in Theorem 4 that decoding is performed without instantaneous CSI is only made to enable closed-form lower bounds. The BS should certainly exploit the channel estimate $\hat{\mathbf{h}}$ and the UE might receive a downlink pilot signal that enables estimation of the effective channel $\mathbf{h}^H\mathbf{v}^{\mathrm{DL}}$. While this is relatively easy to handle with ideal hardware, where the channel estimate and estimation error are independent (cf. [12]), the extension to non-ideal hardware seems intractable due the statistical dependence between the channel estimate and estimation error.

$$\mathtt{SINR}_{\text{lower}}^{\text{DL}}(\mathbf{v}^{\text{DL}}) = \frac{\left|\mathbb{E}\{\mathbf{h}^H\mathbf{v}^{\text{DL}}\,|\tilde{\mathcal{H}}^{\text{UE}}\}\right|^2}{(1+\kappa_r^{\text{UE}})\mathbb{E}\left\{|\mathbf{h}^H\mathbf{v}^{\text{DL}}|^2\,|\tilde{\mathcal{H}}^{\text{UE}}\right\} - \left|\mathbb{E}\{\mathbf{h}^H\mathbf{v}^{\text{DL}}\,|\tilde{\mathcal{H}}^{\text{UE}}\}\right|^2 + \kappa_t^{\text{BS}}\sum_{i=1}^{N}\mathbb{E}\{|h_i|^2|v_i^{\text{DL}}|^2\,|\tilde{\mathcal{H}}^{\text{UE}}\} + \frac{\mathbb{E}\{I_{\mathcal{H}}^{\text{UE}}\,|\tilde{\mathcal{H}}^{\text{UE}}\}}{p^{\text{BS}}} + \frac{\sigma_{\text{UE}}^2}{p^{\text{BS}}}} \tag{36}$$

$$\mathtt{SINR}_{\text{lower}}^{\text{UL}}(\mathbf{v}^{\text{UL}}) = \frac{\left|\mathbb{E}\{\mathbf{h}^H\mathbf{v}^{\text{UL}}\,|\tilde{\mathcal{H}}^{\text{BS}}\}\right|^2}{(1+\kappa_t^{\text{UE}})\mathbb{E}\left\{|\mathbf{h}^H\mathbf{v}^{\text{UL}}|^2\,|\tilde{\mathcal{H}}^{\text{BS}}\right\} - \left|\mathbb{E}\{\mathbf{h}^H\mathbf{v}^{\text{UL}}\,|\tilde{\mathcal{H}}^{\text{BS}}\}\right|^2 + \kappa_r^{\text{BS}}\sum_{i=1}^{N}\mathbb{E}\{|h_i|^2|v_i^{\text{UL}}|^2\,|\tilde{\mathcal{H}}^{\text{BS}}\} + \frac{\mathbb{E}\left\{(\mathbf{v}^{\text{UL}})^H(\mathbf{Q}_{\mathcal{H}}+\sigma_{\text{BS}}^2\mathbf{I})\mathbf{v}^{\text{UL}}\,|\tilde{\mathcal{H}}^{\text{BS}}\right\}}{p^{\text{UE}}}} \tag{37}$$

at the BS (κ_r^{BS}) is present in the lower bound (42), through $\mathbf{A}$ and $\mathbf{\Psi}$ in φ. However, the numerical results in Section IV-C reveal that the asymptotic impact of BS impairments is negligible also in the lower bound. This can also be seen analytically in certain cases; if $\kappa_t^{\text{UE}} = 0$ we get $\varphi = 1$ and therefore

$$\lim_{N\to\infty} \mathtt{C}_{\text{lower}}^{\text{DL}} = \frac{T_{\text{data}}^{\text{DL}}}{T_{\text{coher}}}\log_2\left(1+\frac{1}{\kappa_r^{\text{UE}}}\right). \tag{43}$$

In this special case, the lower bound actually approaches the upper bound in (31) asymptotically, and any DL capacity can be achieved by making κ_r^{UE} sufficiently small. The opposite is not true; setting $\kappa_r^{\text{BS}} = 0$ will *not* make the impact of UE impairments vanish. We therefore conclude that the DL capacity limit is mainly determined by the level of impairments at the UE, both in the uplink estimation (κ_t^{UE}) and the downlink transmission (κ_r^{UE})—although the former connection was not visible in the upper bound since it was based on perfect CSI.

For the uplink, we have the following similar asymptotic capacity characterization.

Corollary 5. *Consider the UL with receive combining vector* $\mathbf{v} = \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}$ *and* $\tilde{\mathcal{H}}^{BS} = \emptyset$. *If* $\mathbb{E}\{\|\mathbf{Q}_{\mathcal{H}}\|_2\} \leq \mathcal{O}(N^n)$ *for some* $n < 1$, *the lower capacity bound in* (35) *can be expressed as*

$$\mathtt{C}^{\text{UL}} \geq \frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} \times \log_2\left(1+\frac{|\mathbb{E}\{\varphi\}|^2+\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)}{(1+\kappa_t^{\text{UE}})\mathbb{E}\{|\varphi|^2\}-|\mathbb{E}\{\varphi\}|^2+\mathcal{O}\left(\frac{1}{\sqrt{N}}+\frac{1}{N^{1-n}}\right)}\right) \tag{44}$$

where φ *is given in* (41). *The terms* $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ *and* $\mathcal{O}\left(\frac{1}{N^{1-n}}\right)$ *vanish when* $N \to \infty$, *while the other terms are strictly positive in the limit.*

Proof: The expression (44) is obtained from (35) by plugging in the expressions from Theorem 4 and multiplying each term by $\frac{1}{\text{tr}(\mathbf{R}-\mathbf{C})} = \frac{1}{p^{\text{UE}}\text{tr}(\mathbf{R}\bar{\mathbf{Z}}^{-1}\mathbf{R})} = \mathcal{O}(N^{-1})$. The interference term becomes $\frac{\mathbb{E}\{\mathbf{v}^H\mathbf{Q}_{\mathcal{H}}\mathbf{v}\}}{p^{\text{UE}}\text{tr}(\mathbf{R}-\mathbf{C})} = \mathcal{O}\left(\frac{1}{N^{1-n}}\right)$. ∎

The upper bound in Corollary 3 and the lower bound in Corollary 5 provide a joint characterization of the uplink capacity when N grows large. The UE impairments manifest the behavior in both bounds; the BS impairments are present in (42) since φ depends on $\mathbf{A}$ and $\mathbf{\Psi}$, but their impact vanish when $\kappa_t^{\text{UE}} \to 0$. By making κ_t^{UE} appropriately small, we can thus achieve any UL capacity as N grows large. We therefore conclude that it is of main importance to have high quality hardware at the UE, which is analog to our observations for the DL. These observations are illustrated numerically in the next subsection and are explained by the following remark.

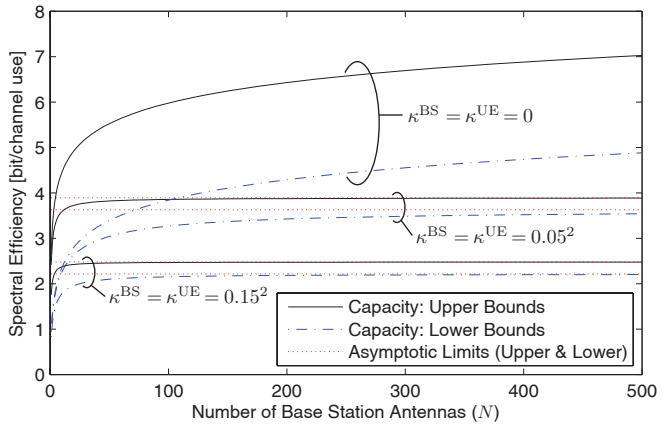
Remark 3 (BS Impairments Vanish Asymptotically)**.** *The lower and upper bounds show that it is the quality of the UE's transceiver hardware that limits the DL and UL capacities as* $N \to \infty$. *Thus, the detrimental effect of hardware impairments at the BS vanishes completely, or almost completely, when the number of BS antennas grows large. This is, simply speaking, since the BS's distortion noises are spread in arbitrary directions in the* N*-dimensional vector space while the increased spatial resolution of the array enables very exact transmit beamforming and receive combining for the useful signal. This is a very promising result since large arrays are more prone to impairments, due to implementation limitations and the will to use antenna elements of lower quality (to avoid having deployment costs that increase linearly with* N*). In contrast, the UE's distortion noises are non-vanishing since they behave as interferers with the same effective channels as the useful signals.*

Corollaries 4 and 5 assumed that the inter-user interference satisfy $\mathbb{E}\{I_{\mathcal{H}}^{\text{UE}}\} \leq \mathcal{O}(N^n)$ and $\mathbb{E}\{\|\mathbf{Q}_{\mathcal{H}}\|_2\} \leq \mathcal{O}(N^n)$, respectively, for some $n < 1$. These conditions imply that the interference terms only vanish asymptotically if the scaling with N is slower than linear. This is satisfied by *regular* interference which has constant variance (i.e., $n = 0$), but there is a special type of non-regular pilot contaminated interference in multi-cell systems that scales linearly with N. This adds an additional non-vanishing term to the denominators of (42) and (44). We detail this scenario in Section VI.

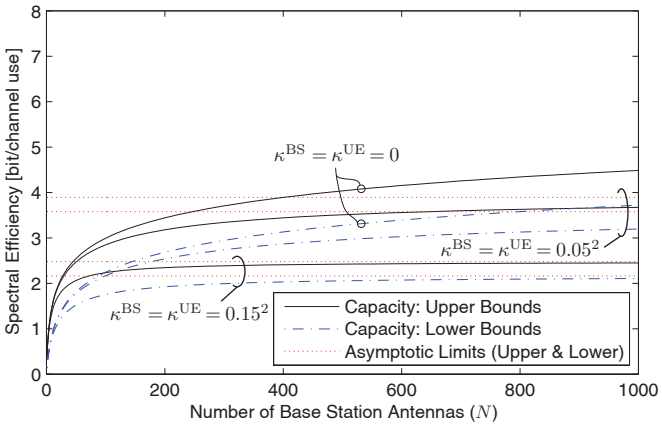
Finally, we stress that the DL and UL capacity bounds in Corollaries 2 and 3, respectively, have a very similar structure. The main difference is that the UL is only affected by UL hardware impairments (i.e., $\kappa_t^{\text{UE}}, \kappa_r^{\text{BS}}$), while the DL is affected by both DL and UL hardware impairments (i.e., all κ-parameters) due to the reverse-link channel estimation.

C. Numerical Illustrations

Next, we illustrate the lower and upper bounds on the capacity that were derived earlier in this section. We consider a scenario without interference, $\mathbf{Q}_{\mathcal{H}} = \mathbf{S} = \mathbf{0}$ and $I_{\mathcal{H}}^{\text{UE}} = 0$, and define the average SNRs as $p^{\text{UE}}\frac{\text{tr}(\mathbf{R})}{N\sigma_{\text{BS}}^2}$ and $p^{\text{BS}}\frac{\text{tr}(\mathbf{R})}{N\sigma_{\text{UE}}^2}$ in the UL and DL, respectively. We consider different fixed SNR values, while we vary the number of antennas N and the levels



(a) SNR: 20 dB



(b) SNR: 0 dB

Fig. 6: Lower and upper bounds on the capacity. Hardware impairments have a fundamental impact on the asymptotic behavior as N grows large.

of hardware impairments. We assume that the transmitter and receiver hardware of each device are of the same quality: $\kappa^{\mathrm{BS}} \triangleq \kappa_t^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}}$ at the BS and $\kappa^{\mathrm{UE}} \triangleq \kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{UE}}$ at the UE.[12] Furthermore, we assume $\frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} = \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} = 0.45$, which are the percentage of DL data and UL data. These assumptions make the bounds for the DL and UL capacities become identical, thus we can simulate the DL and UL simultaneously.

Fig. 6 considers a spatially uncorrelated scenario with $\mathbf{R} = \mathbf{I}$ for different levels of impairments: $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} \in \{0, 0.05^2, 0.15^2\}$. The meaning of these parameter values was discussed in Remark 1. Simulation results are given for SNRs of 20 dB and 0 dB. The capacity with ideal hardware grows without bound as $N \to \infty$, while the lower and upper bounds converge to finite limits under transceiver hardware impairments. The main difference between the two SNR values is the convergence speed, while the upper bounds are exactly the same and the lower bounds are approximately the same. Recall that these bounds hold under any CSI $\mathcal{H}^{\mathrm{BS}}$ at the BS and $\mathcal{H}^{\mathrm{UE}}$ at the UE; the lower bounds represent no instantaneous CSI in the decoding step and the upper bounds represent perfect CSI. Although the gap between these extremes is large for ideal hardware, the difference is remarkably small under non-ideal hardware due to the finite capacity limit (caused by distortion noise) and the channel hardening that makes stochastic inner product such as $\mathbf{h}^H\mathbf{v}$ become increasingly deterministic as N grows large. Since a main difference between the lower and upper bounds is the quality of the CSI, the small difference shows that the estimation errors have only a minor impact on the capacity; hence, the estimation error floors described in Section III has no dominating impact in the large-N regime.

[12]The transmitter and receiver hardware both involve converters, mixers, filters, and oscillators; see [30, Fig. 1] for a typical transceiver model. The main difference is the type of amplifiers, thus the assumption of identical levels of impairments makes sense when the non-linearities of the amplifiers at the transmitter are not the dominating source of distortion noise.

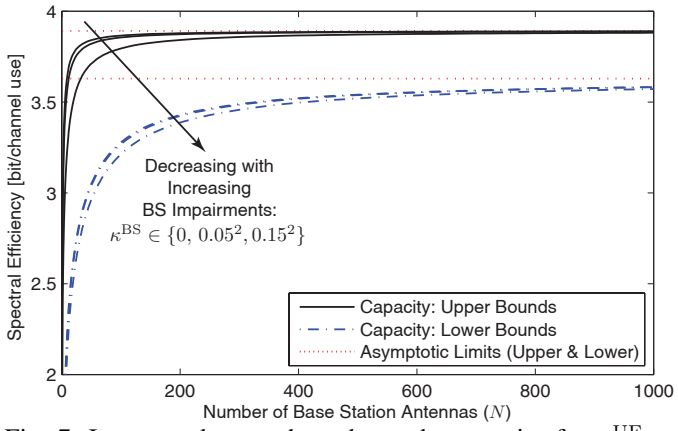


Fig. 7: Lower and upper bounds on the capacity for $\kappa^{\mathrm{UE}} = 0.05^2$. The impact of hardware impairments at the BS vanishes asymptotically.

The asymptotic capacity limits in Fig. 6 are characterized by the level of impairments, thus the hardware quality has a fundamental impact on the achievable spectral efficiency. If the SNRs are sufficiently high (e.g., 20 dB), the majority of the multi-antenna gain is achieved at relatively low N; in particular, only minor improvements can be achieved by having more than $N = 100$ antennas. Larger numbers are, however, useful for inter-user interference suppression and multiplexing; see Section VI. We need many more antennas to achieve convergence at 0 dB SNR than at 20 dB, because a 100 times larger array gain is required to compensate for the lower SNR. Hence, we conclude that the massive MIMO gains are much more attractive at higher SNRs (which matches well with the results in Section III where 20–30 dB SNR was needed to achieve a close-to-perfect channel estimate). Therefore, we only consider an SNR of 20 dB it the remainder of this section.

Fig. 7 considers the same scenario as in Fig. 6 but with a fixed level of impairments $\kappa^{\mathrm{UE}} = 0.05^2$ at the UE and different values at the BS. As expected from the analysis, the lower and upper capacity bounds increase with κ^{BS}, but the difference is only visible at small N since the curves converge to virtually the same value as $N \to \infty$. This validates that the impact of impairments at the BS vanishes as N grows large.

Finally, we consider the capacity behavior for different channel covariance models, namely the four propagation scenarios described in Section III-B. The lower and upper capacity bounds are shown in Fig. 8 for $\kappa^{\mathrm{BS}} = \kappa^{\mathrm{UE}} = 0.05$. The upper bound is identical for all the models, since it only

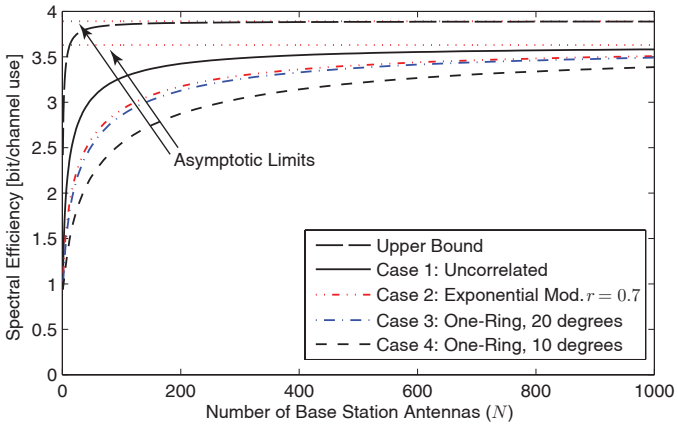


Fig. 8: Lower and upper bounds on the capacity as a function of the number of BS antennas. Four different channel covariance models are considered and the hardware impairments are characterized by $\kappa^{\mathrm{BS}} = \kappa^{\mathrm{UE}} = 0.05$.

utilizes the diagonal elements of $\mathbf{R}$. However, there are clear differences between the lower bounds. The spatially uncorrelated covariance model provides the highest performance, while the strongly spatially correlated one-ring model from [42] with 10 degrees angular spread gives the lowest performance. This stands in contrast to Section III-B, where the highly correlated channels gave the lowest estimation errors (i.e., highest estimation accuracy). However, the differences between the channel covariance models vanish asymptotically as $N \to \infty$.

V. Improving Energy Efficiency and Reducing Hardware Quality

Next, we analyze how the energy efficiency (EE) can be optimized in massive MIMO systems. The EE is measured in bit/Joule and a common EE definition is the ratio of the spectral efficiency (in bit/channel use) to the emitted power (in Joule/channel use). It has recently been shown that the array gain in massive MIMO systems can be utilized to reduce the emitted power; see [12], [13] for systems with ideal hardware and [25] for systems with phase noise from free-running oscillators. More specifically, these prior works show that one can reduce the transmit powers as $1/N^t$, for $0 < t < \frac{1}{2}$, and still achieve non-zero spectral efficiencies as $N \to \infty$. By following this power scaling law, we can achieve an *infinitely* high EE as $N \to \infty$ because the numerator has a non-zero limit and the denominator goes to zero as $1/N^t$ [1]. Although this property indicates that massive MIMO systems can be very energy efficient, the unboundedness also shows that the conventional EE metric needs to be revised when applied to massive MIMO systems. In this section, we consider a refined metric of overall EE (based on prior work in [50]–[54]) and use it to analyze the overall EE of massive MIMO systems.

Under the TDD protocol, the energy consumed in the amplifiers of the transmitters (per coherence period) is

$$E_{\mathrm{amp}} = (T_{\mathrm{pilot}}^{\mathrm{DL}} + T_{\mathrm{data}}^{\mathrm{DL}}) \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}} + (T_{\mathrm{pilot}}^{\mathrm{UL}} + T_{\mathrm{data}}^{\mathrm{UL}}) \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}} \quad \text{[Joule]} \tag{45}$$

where the parameters $\omega^{\mathrm{BS}}, \omega^{\mathrm{UE}} \in [0,1]$ are the efficiencies of the power amplifiers at the BS and UE, respectively.[13] The average power (in Joule/channel use) can then be separated as

$$\frac{E_{\mathrm{amp}}}{T_{\mathrm{coher}}} = \underbrace{\alpha_{\mathrm{DL}} \left(\frac{T_{\mathrm{pilot}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}} + \frac{T_{\mathrm{pilot}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}} \right) + \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}}}_{\text{Downlink power}} + \underbrace{\alpha_{\mathrm{UL}} \left(\frac{T_{\mathrm{pilot}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}} + \frac{T_{\mathrm{pilot}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}} \right) + \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}}}_{\text{Uplink power}} \tag{46}$$

where the ratios of DL and UL transmission are, respectively,

$$\alpha_{\mathrm{DL}} = \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{data}}^{\mathrm{DL}} + T_{\mathrm{data}}^{\mathrm{UL}}} \tag{47}$$

$$\alpha_{\mathrm{UL}} = \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{data}}^{\mathrm{DL}} + T_{\mathrm{data}}^{\mathrm{UL}}}. \tag{48}$$

In addition to the power consumed by the amplifiers, there is generally a baseband circuit power consumption which we model as $N\rho + \zeta$ [50]–[54]. The parameter $\rho \geq 0$ [Joule/channel use] describes the circuit power that scales with the number antennas; for example, hardware components that are needed at each antenna branch (e.g., converters, mixers, and filters) and computational complexity that is proportional to N (e.g., channel estimation and computing MRT/MRC). In contrast, the parameter $\zeta > 0$ [Joule/channel use] is a static circuit power term that is independent of N (but might scale with the number UEs); for example, it models baseband processing at the BS and circuit power at the UE.[14]

Based on the power consumption model described above, and inspired by the seminal work in [55], we define the overall energy efficiency (in bit/Joule) as follows.

Definition 1. *The downlink energy efficiency is*

$$\mathtt{EE}^{\mathrm{DL}} = \frac{\mathtt{C}^{\mathrm{DL}}}{\alpha_{\mathrm{DL}} \left(\frac{T_{\mathrm{pilot}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}} + \frac{T_{\mathrm{pilot}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}} + N\rho + \zeta \right) + \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}}} \tag{49}$$

and the uplink energy efficiency is

$$\mathtt{EE}^{\mathrm{UL}} = \frac{\mathtt{C}^{\mathrm{UL}}}{\alpha_{\mathrm{UL}} \left(\frac{T_{\mathrm{pilot}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{BS}}}{\omega^{\mathrm{BS}}} + \frac{T_{\mathrm{pilot}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}} + N\rho + \zeta \right) + \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \frac{p^{\mathrm{UE}}}{\omega^{\mathrm{UE}}}}. \tag{50}$$

The EE of any suboptimal transmission scheme is obtained by replacing the capacities $\mathtt{C}^{\mathrm{DL}}$ *and* $\mathtt{C}^{\mathrm{UL}}$ *with the corresponding achievable spectral efficiencies.*

This definition considers a single link, which can be any of the links in a massive MIMO system—the parameters ζ and ρ should then be interpreted as the energy per channel

[13] The efficiency of a specific amplifier depends on the transmit power, but to facilitate analysis we assume that the amplifier is optimized jointly with the transmit power to give a specific efficiency at the particular power level. The efficiency also depends on the PAPR and acceptable distortion noise, which are two properties that we also keep fixed when optimizing the EE.

[14] This term can also model the overhead power consumption of the network as a whole, which enables comparison of network architectures with different BS density, amounts of backhaul signaling, etc.

use *per user*. In the process of maximizing the EE metric in Definition 1, we first extend the power scaling laws from [12], [13], [25] to our general system model with non-ideal hardware.

Theorem 5. *Suppose the downlink transmit power p^{BS} and uplink pilot power p^{UE} are reduced with N proportionally to $1/N^{t_{\mathrm{BS}}}$ and $1/N^{t_{\mathrm{UE}}}$, respectively. If $t_{\mathrm{BS}} + t_{\mathrm{UE}} < 1$, $t_{\mathrm{BS}} \geq 0, 0 < t_{\mathrm{UE}} < \frac{1}{2}$, and $\mathbb{E}\{I_{\mathcal{H}}^{\mathrm{UE}}\} = \mathcal{O}(1)$, we have*

$$\lim_{N\to\infty} \mathtt{C}^{\mathrm{DL}} \geq \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \log_2\left(1 + \frac{1}{\kappa_r^{\mathrm{UE}} + \kappa_t^{\mathrm{UE}} + \kappa_r^{\mathrm{UE}}\kappa_t^{\mathrm{UE}}}\right). \tag{51}$$

Similarly, suppose the uplink transmit/pilot power p^{UE} is reduced with N proportionally to $1/N^{t_{\mathrm{UE}}}$. If $0 < t_{\mathrm{UE}} < \frac{1}{2}$ and $\mathbb{E}\{\|\mathbf{Q}_{\mathcal{H}}\|_2\} = \mathcal{O}(1)$, we have

$$\lim_{N\to\infty} \mathtt{C}^{\mathrm{UL}} \geq \frac{T_{\mathrm{data}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} \log_2\left(1 + \frac{1}{2\kappa_t^{\mathrm{UE}} + (\kappa_t^{\mathrm{UE}})^2}\right). \tag{52}$$

Proof: The proof is given in Appendix C-D. ∎

Theorem 5 shows that one can reduce the downlink and uplink transmit powers as N grows large (e.g., roughly proportionally to $1/\sqrt{N}$) and converge to non-zero spectral efficiencies. The asymptotic DL capacity is lower bounded by (51) and the UL capacity by (52). As expected from Section IV, these lower bounds only depend on the levels of impairments at the UE. The conditions $\mathbb{E}\{I_{\mathcal{H}}^{\mathrm{UE}}\} = \mathcal{O}(1)$ and $\mathbb{E}\{\|\mathbf{Q}_{\mathcal{H}}\|_2\} = \mathcal{O}(1)$ in Theorem 5 are stronger than the ones in Corollaries 4 and 5, thus the interfering transmissions might have to reduce their transmit powers as well if their impact should vanish asymptotically. We note that the lower bounds in Theorem 5 are achieved by using the LMMSE estimator in Theorem 1 for channel estimation and simple linear processing at the BS (approximate MRT in the DL and MRC in the UL).

Based on Theorem 5 and the upper capacity bounds in Section IV, the following corollary describes how to maximize the EE.

Corollary 6. *Suppose we want to maximize the EE metrics with respect to the transmit powers and the number of antennas. Let $\mathbb{E}\{I_{\mathcal{H}}^{\mathrm{UE}}\} = \mathcal{O}(1)$ and $\mathbb{E}\{\|\mathbf{Q}_{\mathcal{H}}\|_2\} = \mathcal{O}(1)$. If $\rho = 0$, the maximal EEs are bounded as*

$$\frac{\log_2\left(1+\frac{1}{\kappa_r^{\mathrm{UE}}+\kappa_t^{\mathrm{UE}}+\kappa_r^{\mathrm{UE}}\kappa_t^{\mathrm{UE}}}\right)}{\frac{T_{\mathrm{coher}}}{T_{\mathrm{data}}^{\mathrm{DL}}}\alpha_{\mathrm{DL}}\zeta} \leq \max_{\substack{p^{\mathrm{BS}},p^{\mathrm{UE}}\geq 0\\ N\geq 0}} \mathtt{EE}^{\mathrm{DL}} \leq \frac{\log_2\left(1+\frac{1}{\kappa_r^{\mathrm{UE}}}\right)}{\frac{T_{\mathrm{coher}}}{T_{\mathrm{data}}^{\mathrm{DL}}}\alpha_{\mathrm{DL}}\zeta} \tag{53}$$

$$\frac{\log_2\left(1+\frac{1}{2\kappa_t^{\mathrm{UE}}+(\kappa_t^{\mathrm{UE}})^2}\right)}{\frac{T_{\mathrm{coher}}}{T_{\mathrm{data}}^{\mathrm{UL}}}\alpha_{\mathrm{UL}}\zeta} \leq \max_{p^{\mathrm{UE}},N\geq 0} \mathtt{EE}^{\mathrm{UL}} \leq \frac{\log_2\left(1+\frac{1}{\kappa_t^{\mathrm{UE}}}\right)}{\frac{T_{\mathrm{coher}}}{T_{\mathrm{data}}^{\mathrm{UL}}}\alpha_{\mathrm{UL}}\zeta} \tag{54}$$

where the lower bounds are achieved as $N \to \infty$ using the power scaling law in Theorem 5.

If $\rho > 0$, the upper bounds in (53) *and* (54) *are still valid, but the asymptotic EEs are*

$$\lim_{N\to\infty} \max_{p^{\mathrm{BS}},p^{\mathrm{UE}}\geq 0} \mathtt{EE}^{\mathrm{DL}} = \lim_{N\to\infty} \max_{p^{\mathrm{UE}}\geq 0} \mathtt{EE}^{\mathrm{UL}} = 0 \tag{55}$$

and, consequently, the EEs are maximized at some finite N.

Proof: The lower bounds for $\rho = 0$ are achieved as described in corollary, while the upper bounds follow from neglecting the transmit power term in the denominator and applying the capacity upper bounds from Corollaries 2 and 3. In the case of $\rho > 0$, we note that the EE is non-zero for $N = 1$ for any non-zero transmit power, while the EE goes to zero as $N \to \infty$ since the denominators of the EE metrics grow to infinity and the numerators are bounded. ∎

This corollary reveals that the maximal overall EE is finite, also in massive MIMO systems. If the circuit power consumption does not scale with N, such that $\rho = 0$, we can achieve an EE very close to the upper bounds in (53) and (54) by having very many antennas. This changes completely when there is a non-zero circuit power per antenna: $\rho > 0$. The maximal EE is then achieved at some finite N, which naturally depends on the parameters ρ, ζ, ω^{BS}, and ω^{UE}. We illustrate this dependence numerically in the next subsection.

Since ρ has a dominating impact on the maximal EE in massive MIMO systems, one would like to find a way to reduce ρ. Generally speaking, the hardware power consumption depends on the circuit architecture and the hardware resolution [7], [21]; by tolerating larger hardware impairments we can also reduce the power dissipation in the corresponding circuits. Now recall from Section IV that the impact of hardware impairments at the BS vanishes as $N \to \infty$. This fact raises the important question: Can we increase the levels of impairments at the BS as N grows and still obtain non-zero capacities? The answer is given by the following corollary.

Corollary 7. *Suppose the levels of impairments $\kappa_t^{\mathrm{BS}}, \kappa_r^{\mathrm{BS}}$ are increased with N proportionally to N^{τ_t} and N^{τ_r}, respectively. The lower capacity bounds in Corollaries 4 and 5 (for $n \leq \frac{1}{2}$) converge to non-zero quantities as $N \to \infty$ if $\tau_r < \frac{1}{2}$ in the UL and $\tau_t + \tau_r < 1$ and $\tau_r < \frac{1}{2}$ in the DL.*

Proof: The proof is given in Appendix C-E. ∎

This corollary shows that we can indeed increase the levels of impairments, κ_t^{BS} and κ_r^{BS}, at the BS roughly proportionally to $\sqrt{N}$ and still have a non-zero asymptotic capacity. The numerical results in the next subsection shows that only minor degradations of the lower capacity bounds appear when the impairment scaling law in Corollary 7 is followed.

Recall from (5) that the conventional EVM measure of transceiver quality equals the square root of the κ-parameters, thus Corollary 7 shows that the EVMs can be increased proportionally to $N^{1/4}$. A high-quality BS antenna element with an EVM of 0.03 can thus be replaced by 256 low-quality antenna elements with an EVM of 0.12, while the loss in capacity is negligible. This is a very encouraging result, since it indicates that massive MIMO can be deployed with BS hardware components that are inexpensive, have lower quality and thus low power consumption than conventional ones (i.e., ρ is smaller). If the hardware components are treated as optimization variables, the maximal EE is achieved by jointly reducing the transmit power and the circuit power consumption with N. This optimization is, however, strongly dependent on the practical hardware setup (e.g., how an increased EVM maps to a smaller circuit power dissipation) and is outside the scope of this paper. Finally, we note that the ability to degrade the hardware quality with N comes in addition to all other benefits of massive MIMO, such as the array gain and

the decorrelation of user channels; see the multi-cell results in Section VI.

A. Numerical Illustrations

Next, we illustrate how the overall EE depends on the number of antennas, transmit powers, and circuit power parameters ζ and ρ. Based on the power consumption numbers reported in [56, Table 7], we consider two setups: $\zeta + \rho = 2\,\mu$J/channel use and $\zeta + \rho = 0.02\,\mu$J/channel use.[15] These represent the total circuit power consumptions in a system with $N = 1$ antenna. Since the total circuit power for arbitrary N is $\zeta + N\rho$, we consider three different splittings between ρ and ζ: $\frac{\rho}{\zeta+\rho} \in \{0, 0.01, 0.1\}$. From an EE optimization perspective, any scaling of the power amplifier efficiencies is equivalent to an inverse scaling of ζ and ρ. Hence, we can set the efficiencies to $\omega^{\mathrm{BS}} = \omega^{\mathrm{UE}} = 0.3$, which corresponds to 30%, without limiting the generality of our numerical results.

The transmit powers in the UL and DL are assumed to be equal and upper bounded by $p_{\max} = 0.0222\,\mu$J/channel use.[16] We consider a scenario without interference (i.e., $\mathbf{Q}_{\mathcal{H}} = \mathbf{S} = \mathbf{0}$ and $I_{\mathcal{H}}^{\mathrm{UE}} = 0$), the channel covariance matrix $\mathbf{R}$ is generated by the exponential correlation model in (17) with correlation coefficient $r = 0.7$. We let $\frac{N\sigma_{\mathrm{UE}}^2}{\mathrm{tr}(\mathbf{R})} = \frac{N\sigma_{\mathrm{BS}}^2}{\mathrm{tr}(\mathbf{R})} = \frac{p_{\max}}{100}\,\mu$J/channel use which gives an SNR of 20 dB if the maximal transmit power $p_{\max}$ is used; recall from Section IV-C that this SNR is desirable if one should operate close to the asymptotic capacity limits. To make the EE of the DL and UL equal, we consider a symmetric scenario with $\alpha_{\mathrm{DL}} = \alpha_{\mathrm{UL}} = 0.5$ and $\frac{T_{\mathrm{pilot}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} = \frac{T_{\mathrm{pilot}}^{\mathrm{UL}}}{T_{\mathrm{coher}}} = 0.05$.

Fig. 9 shows the achievable DL/UL energy efficiencies using the lower capacity bounds in Theorem 3. The levels of impairments are set to $\kappa_t^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}} = \kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{UE}} = 0.05^2$; see Remark 1 for the interpretation of these parameter values. The transmit powers are either optimized numerically for maximal EE at each N, fixed at the value that is optimal for $N = 1$, or reduced from this value according to the power scaling law in Theorem 5 with $t = \frac{1}{2}$. Fig. 9 shows that the EE is almost the same in all three cases, but varies a lot with the circuit power parameters ζ and ρ. If $\rho = 0$, the EE increases monotonically with N and eventually converge according to (53) and (54) in Corollary 6. On the contrary, the EE has a unique maximum when $\rho > 0$ and then decreases towards zero. The maximum is in the range of $5 \leq N \leq 50$ in the figure, but the exact position depends on the circuit power parameters. If $\frac{\rho}{\zeta+\rho}$ is sufficiently small we can use a larger N without losing much in EE. Hence, it is important to make the power that scales with N (e.g., the number of extra hardware components and the computational complexity) as low as possible if massive MIMO systems should excel in terms of energy efficiency. The EE with ideal hardware is also shown in Fig. 9, which reveals that the difference in EE between ideal and non-ideal hardware is small. This is because the performance loss from hardware impairments is relatively small at reasonable number of antennas.

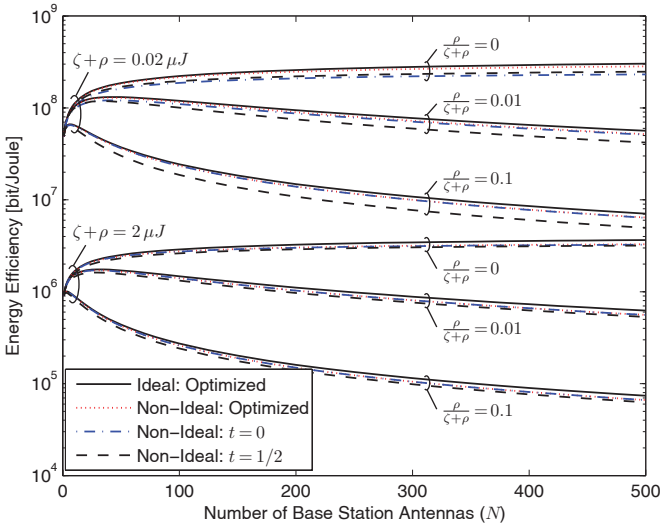


Fig. 9: Achievable energy efficiency with ideal and non-ideal hardware for fixed transmit power ($t = 0$), transmit power that decreases as $1/N^t$ for $t = \frac{1}{2}$, and the transmit power that maximizes the EE. The EE is computed using the lower bounds in Theorem 3 and are valid for both DL and UL transmissions.

In order to compare the three different power allocations, we show the corresponding transmit powers in Fig. 10. Recall that the power is either fixed, reduced with N according to the scaling law in Theorem 5, or optimized for maximal EE at each N. Despite the very similar EEs in Fig. 9, these three power allocations behave very differently. If $\rho = 0$, the optimal transmit power decreases with N but at a clearly slower pace than $1/\sqrt{N}$ (which is the fastest power scaling that gives a non-zero asymptotic rate according to Theorem 5). However, the optimal transmit for $\rho > 0$ only decreases until the maximal EE is achieved (which is in the range of $5 \leq N \leq 50$) and then *increases* with N. This makes much sense, because when the circuit power increases we can also afford using more transmit power to get a higher spectral efficiency. In the case when the circuit power is large (i.e., $\zeta + \rho = 2\,\mu J$), we see that it is often optimal to use full transmit power, as represented by the upper straight line. To summarize, the transmit power in massive MIMO systems can be decreased monotonically with N, but this is generally not the way to maximize the EE since we have $\rho > 0$ in most practical systems. The loss in EE by decreasing the power appears to be small, but the loss in spectral efficiency is naturally larger due to the definition of the EE. If we want a simple design rule, it is better to keep the total power fixed for all N than to decrease it with N.

Finally, the ability to increase the levels of impairments at the BS with N is illustrated in Fig. 11. We consider the same symmetric scenario as in the previous two figures, but the average SNR is set to 20 dB in the DL and UL. We have $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{UE}} = 0.05^2$ at the UE, while the levels of impairments at the BS are scaled as $\kappa_t^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}} = 0.05^2 N^\tau$ for different τ-values: $\tau \in \{0, \frac{1}{4}, \frac{1}{2}, 1, 2\}$. The lower capacity bounds are shown in Fig. 11 as a function of N. The

[15] As a reference, these numbers correspond to 18 W and 0.18 W, respectively, for a system with an effective bandwidth of 9 MHz, since then there are $9 \cdot 10^6$ channel uses per second.

[16] As a reference, this number corresponds to 200 mW, or 23 dBm, for a system with an effective bandwidth of 9 MHz.

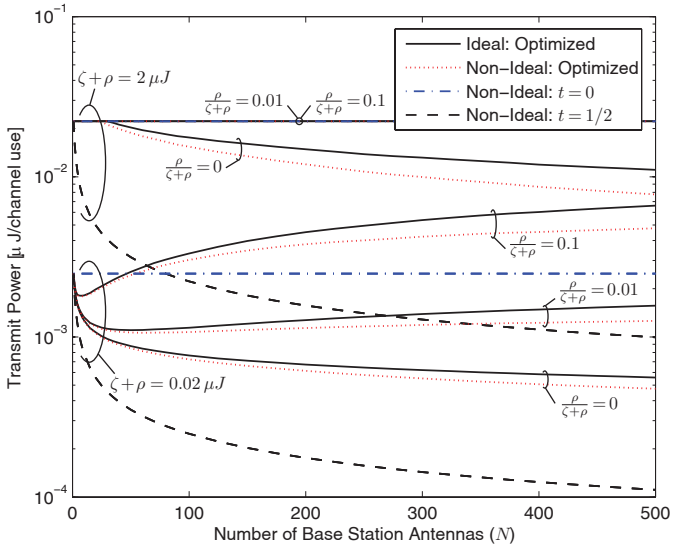


Fig. 10: The transmit powers that correspond to the curves in Fig. 9.

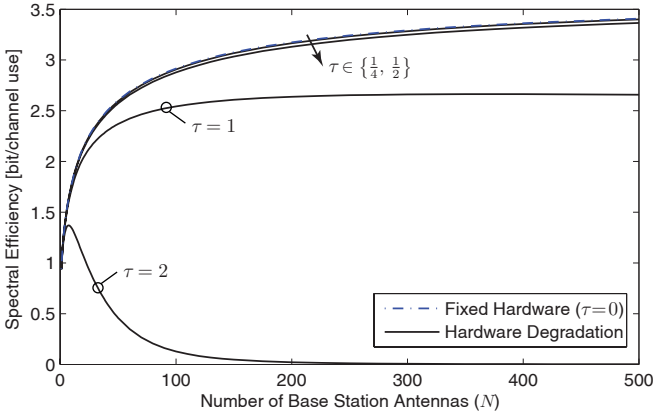


Fig. 11: Lower bounds on the capacity when the levels of impairments at the BS are increased with N as N^τ for $\tau \in \{0, \frac{1}{4}, \frac{1}{2}, 1, 2\}$. The results are valid for both DL and UL transmissions.

simulation confirms that the performance degradation is small when the impairment scaling law in Corollary 7 is followed (e.g., for $\tau = \frac{1}{4}$ and $\tau = \frac{1}{2}$). A larger performance loss is observed for $\tau = 1$ and the curve begins to bend downwards at $N \approx 350$. In the extreme case of $\tau = 2$, the lower bound goes quickly to zero. While these asymptotic results are based on the lower capacity bounds, we note that whenever $\tau > 1$ the upper capacity bounds in Corollaries 2 and 3 converge to zero as well.

VI. Extensions to Multi-Cell Scenarios

The previous sections focused on a single link of a massive MIMO system, which experiences interference from other concurrent transmissions. Recall that $I_{\mathcal{H}}^{\mathrm{UE}} \in \mathbb{C}$ and $\mathbf{Q}_{\mathcal{H}} \in \mathbb{C}^{N\times N}$ are the conditional covariances of these transmissions in the DL and UL, respectively, for a given set of channel realizations $\mathcal{H}$. The asymptotic capacity analysis in Section IV, particularly the lower capacity bounds in Corollaries 4 and 5, are based on the assumptions that $\mathbb{E}\{I_{\mathcal{H}}^{\mathrm{UE}}\} \leq \mathcal{O}(N^n)$ and $\mathbb{E}\{\|\mathbf{Q}_{\mathcal{H}}\|_2\} \leq \mathcal{O}(N^n)$ for some $n < 1$. However, the interference variance can actually increase faster than this, particularly under the so-called pilot contamination which gives rise to terms that scale linearly with N [9]–[13]. This section investigates the impact of inter-user interference on massive MIMO systems with non-ideal hardware. The BS and UE from the previous sections are referred to as the ones *under study*.

A. Inter-User Interference in the Uplink

To exemplify the impact of inter-user interference, we assume that there is a set $\mathcal{U}$ of co-users that are scheduled for UL transmission in the current coherence period. Each co-user is served by the BS under study or any of the neighboring BSs, thus the total number of co-users $|\mathcal{U}|$ is generally large. The association of UEs to BSs is arbitrary since the association has no impact on the UE under study in the UL. The block-fading channel from UE $l \in \mathcal{U}$ to the BS under study is modeled as $\mathbf{h}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_l)$, where $\mathbf{R}_l$ has bounded spectral norm and the channel is ergodic and block fading. Recall that $\mathcal{H}$ is the set of channel realizations for all channels in the system, thus we have $\mathbf{h}_l \in \mathcal{H}$ for all $l \in \mathcal{U}$. The co-user channels are assumed to be independent, which in practice means that users are selected to have no common scatterers [3], [4]—this is a basic criterion of spatial user separability in the scheduler. A more refined scheduling criteria would be the one in [48], where the coverage area is divided into location bins. The users in a bin are roughly equivalent in terms of channel statistics and should not be served simultaneously. Users in different bins have independent channels and sufficiently different spatial properties, thus selecting one user per location bin for parallel transmission is a reasonable scheduling decision.

The UL pilot signaling is limited to $T_{\mathrm{pilot}}^{\mathrm{UL}}$ channel uses in the TDD protocol depicted in Fig. 2. Since the number of active co-users generally satisfies $|\mathcal{U}| > T_{\mathrm{pilot}}^{\mathrm{UL}}$, each pilot channel use must be allocated to multiple users. We divide $\mathcal{U}$ into two disjoint sets: $\mathcal{U}_{\|}$ are the users that transmit in parallel with the pilot of the UE under study, while $\mathcal{U}_{\perp}$ are the remaining users.[17] The co-users in the same cell as the UE under study are usually in $\mathcal{U}_{\perp}$, but this is not necessary. The interference vector during UL pilot signaling is

$$\boldsymbol{\nu}_{\mathrm{interf}}^{\mathrm{pilot}} = \sum_{l\in\mathcal{U}_{\|}} d_l \mathbf{h}_l \tag{56}$$

where d_l is the signal transmitted by UE $l \in \mathcal{U}_{\|}$. These signals can be either deterministic or stochastic, thus some of the interfering transmissions can in principle carry data instead of pilot signals (cf. Remark 5 in [13]). Assuming $\mathbb{E}\{|d_l|^2\} = p^{\mathrm{UE}}$, the interference covariance matrix during pilot signaling is

$$\mathbf{S} = \mathbb{E}\left\{\boldsymbol{\nu}_{\mathrm{interf}}^{\mathrm{pilot}}(\boldsymbol{\nu}_{\mathrm{interf}}^{\mathrm{pilot}})^H\right\} = p^{\mathrm{UE}} \sum_{l\in\mathcal{U}_{\|}} \mathbf{R}_l. \tag{57}$$

[17] Only one pilot channel use is allocated per UE in this section. For other pilot lengths $B > 1$, one can construct up to B parallel pilot signals that are orthogonal in space. This increases the pilot power per UE, but does not increase the total number of orthogonal pilots.

The LMMSE estimator in Theorem 1 and the corresponding analysis in Section III holds for any covariance matrix $\mathbf{S}$, thus the explicit ways of computing $\boldsymbol{\nu}_{\text{interf}}^{\text{pilot}}$ and $\mathbf{S}$ in (56)–(57) can be plugged in directly. The channel estimate $\hat{\mathbf{h}}$ is now correlated with the co-user channels $\mathbf{h}_l$ for $l \in \mathcal{U}_{\|}$, which has an important impact on the spectral efficiency. More specifically, the interference vector during UL data transmission becomes

$$\boldsymbol{\nu}_{\text{interf}}^{\text{data}} = \sum_{l \in \mathcal{U}} d_l \mathbf{h}_l \tag{58}$$

where d_l is the independent zero-mean stochastic data signal sent by UE $l \in \mathcal{U}$ and has power $\mathbb{E}\{|d_l|^2\} = p^{\text{UE}}$. The conditional interference covariance matrix during data transmission is

$$\mathbf{Q}_{\mathcal{H}} = \mathbb{E}\left\{\boldsymbol{\nu}_{\text{interf}}^{\text{data}} (\boldsymbol{\nu}_{\text{interf}}^{\text{data}})^H | \mathcal{H}\right\} = p^{\text{UE}} \sum_{l \in \mathcal{U}} \mathbf{h}_l \mathbf{h}_l^H. \tag{59}$$

Note that $\mathbf{Q}_{\mathcal{H}}$ depends on the channel realizations in $\mathcal{H}$ and has *not* bounded spectral norm; in fact, there are $|\mathcal{U}|$ eigenvalues of $\mathbf{Q}_{\mathcal{H}}$ that grow without bound as $N \to \infty$. This property affects the lower capacity bound in (35) of Theorem 3 where the conditional interference term now becomes

$$\mathbb{E}\left\{ (\mathbf{v}^{\text{UL}})^H \mathbf{Q}_{\mathcal{H}} \mathbf{v}^{\text{UL}} \, | \tilde{\mathcal{H}}^{\text{BS}} \right\} = p^{\text{UE}} \sum_{l \in \mathcal{U}} \mathbb{E}\left\{ |\mathbf{h}_l^H \mathbf{v}^{\text{UL}}|^2 \, | \tilde{\mathcal{H}}^{\text{BS}} \right\}. \tag{60}$$

The following theorem shows how interference terms of the type $\mathbb{E}\left\{ |\mathbf{h}_l^H \mathbf{v}^{\text{UL}}|^2 \, | \tilde{\mathcal{H}}^{\text{BS}} \right\}$ in (60) behave as N grows large.

Theorem 6. *Assume that no instantaneous CSI is utilized for decoding (i.e., $\tilde{\mathcal{H}}^{BS} = \tilde{\mathcal{H}}^{UE} = \emptyset$), interference is treated as noise, and the receive combining vector $\mathbf{v} = \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}$. Under the interference model in* (56)–(59)*, the terms in* (60) *are*

$$\mathbb{E}\left\{|\mathbf{h}_l^H \mathbf{v}|^2\right\} \tag{61}$$
$$= \begin{cases} \mathbb{E}\left\{ \frac{p^{\text{UE}} (\text{tr}(\mathbf{A}\mathbf{R}_l))^2}{\text{tr}\left(\mathbf{A}(|d+\eta_t^{\text{UE}}|^2 \mathbf{R} + \boldsymbol{\Psi})\mathbf{A}^H\right)} \right\} + \mathcal{O}(\sqrt{N}), & l \in \mathcal{U}_{\|}, \\ \mathcal{O}(1), & l \in \mathcal{U}_{\perp}, \end{cases}$$

where η_t^{UE} is stochastic and $\mathbf{A}, \boldsymbol{\Psi}$ are given in Theorem 4.

The lower capacity bound in (44) *is generalized by replacing the term $\mathcal{O}(\frac{1}{N^{1-n}})$ in the denominator by*

$$\mathbb{E}\left\{ \frac{\text{tr}(\mathbf{R} - \mathbf{C})}{\text{tr}\left(\mathbf{A}(|d+\eta_t^{\text{UE}}|^2 \mathbf{R} + \boldsymbol{\Psi})\mathbf{A}^H\right)} \right\} \sum_{l \in \mathcal{U}_{\|}} \left(\frac{\text{tr}(\mathbf{A}\mathbf{R}_l)}{\text{tr}(\mathbf{A}\mathbf{R})} \right)^2 + \mathcal{O}\left(\frac{1}{N}\right). \tag{62}$$

Proof: The proof is given in Appendix C-F. ∎

This theorem shows that the effective interference from a co-user depends strongly on whether it interfered with the pilot transmission of the UE under study or not. The interference from co-users in $\mathcal{U}_{\perp}$, which were silent when the UE under study sent its pilot, vanishes asymptotically since the user channels decorrelate with N. This is the classical type of interference and is called *regular* interference in this section. In contrast, the interference from co-users in $\mathcal{U}_{\|}$, which were active during the pilot transmission remains and even scales with N. This is the very essence of *pilot contaminated interference* and Theorem 6 generalizes previous results from [9]–[13] (among others) to non-ideal hardware. The explanation to the diverse behavior in Theorem 6 is that the channel estimate $\hat{\mathbf{h}}$ used in the receive combining is independent of the co-user channels $\mathbf{h}_l$ for $l \in \mathcal{U}_{\perp}$, but correlated with $\mathbf{h}_l$ for $l \in \mathcal{U}_{\|}$ since these vectors appeared in the interference term (56) during pilot transmission. Note that Theorem 6 was derived using MRC, while minimum mean squared error (MMSE) receive combining is generally a better choice in multi-cell multi-user scenarios since it actively suppresses interference [12]. Nevertheless, the theorem establishes the baseline behavior: only the pilot contaminated interference may have a substantial impact when N is large (if a judicious receive combining is used). The severity of the pilot contamination depends on how the sets $\mathcal{U}_{\perp}$ and $\mathcal{U}_{\|}$ are chosen [43].

B. Inter-User Interference in the Downlink

The downlink transmission can also suffer from pilot contamination, especially if the numbers of antennas at neighboring BSs also grow linearly with N. The conditional interference variance in the DL takes a similar form as in (58)–(60):

$$I_{\mathcal{H}}^{\text{UE}} = p^{\text{BS}} \sum_{l \in \mathcal{U}} \mathbb{E}\left\{ |\tilde{\mathbf{h}}_l^H \mathbf{v}_l^{\text{DL}}|^2 \, | \tilde{\mathcal{H}}^{\text{UE}} \right\} \tag{63}$$

where $\mathbf{v}_l^{\text{DL}}$ is the beamforming vector for DL transmission to UE $l \in \mathcal{U}$ from its (arbitrary) serving BS and $\tilde{\mathbf{h}}_l$ is the channel from that BS to the UE under study. For brevity, we will not dive into the details since these require assumptions on the decision making at other BSs. The general behavior is however the same: UEs with parallel UL pilots cause non-vanishing interference to each other in the DL, while the impact of all other interfering DL transmissions vanish as N grows large.

C. Numerical Illustrations

The impact of inter-user interference and pilot contamination on multi-cell systems with non-ideal hardware is now studied numerically. We consider UL scenarios with spatially uncorrelated channels, define the average SNR as $p^{\text{UE}} \frac{\text{tr}(\mathbf{R})}{N \sigma_{\text{BS}}^2}$, and let $\frac{T_{\text{data}}^{\text{UL}}}{T_{\text{coher}}} = 0.45$ be the fraction of UL data transmission.

In Fig. 12 we consider the two types of inter-user interference from Theorem 6: regular interference from a UE whose pilot is orthogonal to the UE under study and pilot contaminated interference from a UE with an overlapping pilot. We want to investigate how the achievable per-user spectral efficiency in massive MIMO systems depends on the strength of the pilot contaminated interference, thus we consider a scenario where we operate close to the asymptotic limits: the SNR is 20 dB and the number of antennas is set to $N = 200$ (see Fig. 6). We consider three levels of impairments: $\kappa_t^{\text{UE}} = \kappa_r^{\text{BS}} \in \{0, 0.05^2, 0.1^2\}$. The lower capacity bounds are shown without interference, with only pilot contaminated interference, and with both types of interference. The horizontal axis in Fig. 12 shows the performance as a function of the relative channel gain of the pilot contaminated interference (with respect to the useful channel).

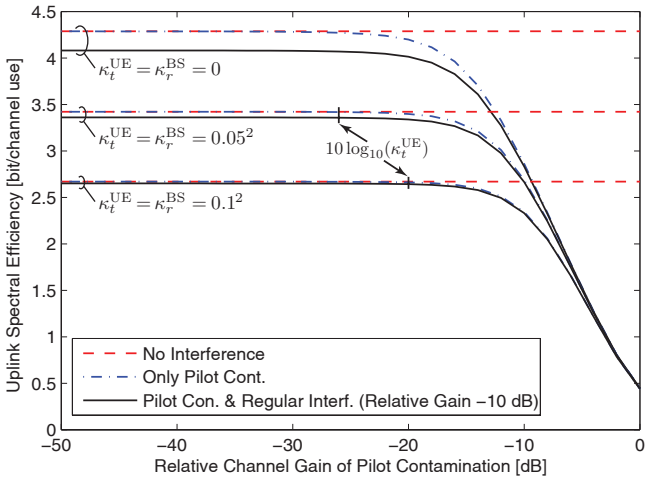


Fig. 12: Lower capacity bounds of a user that experiences pilot contaminated interference of varying strength and possibly regular inter-user interference that is -10 dB weaker than the useful channel. The interference drowns in the distortion noise if it is weaker than the level of impairments at the UE.

We make several observations. Firstly, the ideal hardware case is more sensitive to interference than the non-ideal hardware case. This is particularly evident when it comes to regular inter-user interference, which gives a much larger performance gap in the ideal case. With non-ideal hardware, the regular interference (from a channel that is only -10 dB weaker than the useful channel) has little impact. This is due to the large number of antennas, which decorrelate the user channels. Secondly, the figure shows that pilot contaminated interference has a negligible impact when it arrives over a channel that is much weaker than the useful channel, but there are breaking points where the degradation effect suddenly becomes immense. Interestingly, the breaking points are close to $10\log_{10}(\kappa_t^{\mathrm{UE}})$; that is, how much weaker the distortion noise caused by the UE is compared to the useful signal. This is very intuitive if we compare the size of the distortion term $\kappa_t^{\mathrm{UE}}\mathbb{E}\left\{|\varphi|^2\right\}$ in lower capacity bound in (44) with the interference term in (62). This is formalized as follows.

Corollary 8. *The pilot contaminated interference is negligible, when N grows large, if*

$$\kappa_t^{\mathrm{UE}} \gg \sum_{l \in \mathcal{U}_{\|}} \left(\frac{\mathrm{tr}(\mathbf{A}\mathbf{R}_l)}{\mathrm{tr}(\mathbf{A}\mathbf{R})} \right)^2. \tag{64}$$

This corollary shows that pilot contaminated interference drowns in the distortion noise under certain conditions, which are independent of the absolute SNRs but depend on relative SNR differences of the type $\mathrm{tr}(\mathbf{A}\mathbf{R}_l)/\mathrm{tr}(\mathbf{A}\mathbf{R})$. Since the distortion noise typically is 20–30 dB weaker than the useful signal, the same is needed for the pilot contaminated interference to make its impact negligible. This is not a big deal in cellular deployments; the scheduler should simply allocate different pilots within each cell and to cell-edge users

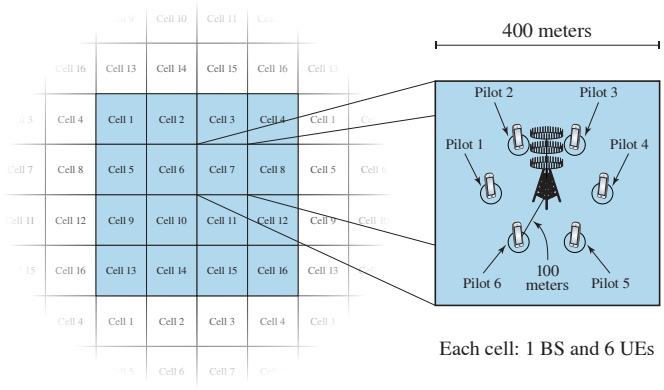


Fig. 13: Illustration of a multi-cell scenario consisting of 16 square cells with wrap-around to avoid edge effects. Each cell is $400\,\mathrm{m} \times 400\,\mathrm{m}$ and contains of 6 UEs equally spaced on a circle of radius $100\,\mathrm{m}$.

of neighboring cells.[18] This can be achieved by the pilot allocation algorithm in [43], but also by simple predefined cell sectorization as illustrated next.

Fig. 13 shows an illustration of the realistic multi-cell scenario that we use validate Corollary 8. The setup consists of 16 square cells, each of size $400\,\mathrm{m} \times 400\,\mathrm{m}$. To avoid edge effects, we use wrap-around as illustrated in Fig. 13. For simplicity, six UEs are scheduled per cell using a simple angular sectorization technique; the UEs are equally spaced on a circle of radius $100\,\mathrm{m}$. We assume that orthogonal pilots are allocated to the UEs in each cell, while the same pilots are reused across cells with the same pattern. The channel covariance matrices are identity matrices that are scaled by the channel attenuations, which are based on the 3GPP propagation model in [57]: the path loss is $10^{-1.53}/D^{3.76}$ where D is the distance in meters. The transmit powers are $p^{\mathrm{UE}} = 0.0222\,\mu\mathrm{J/channel\ use}$ and the noise variance is $\sigma^2_{\mathrm{BS}} = 10^{-7.9}\,\mu\mathrm{J/channel\ use}$. This gives an SNR of 32 dB to the serving BS and 0–13 dB to the surrounding BSs.

Fig. 14 shows the average achievable rates (based on the lower capacity bounds) with MMSE receive combining, which exploits the estimated intra-cell channels to suppress intra-cell interference. We consider ideal hardware and hardware impairments with $\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} = 0.1^2$. To illustrate the impact of pilot contamination, we compare the inter-cell pilot reuse pattern described above with the ideal case when all UEs are allocated unique pilots. We observe that pilot contamination has a substantial impact on the ideal hardware case, and the relative loss will continue to increase with N since only the curve for unique pilots grows towards infinity. In contrast, there is almost no difference between the unique and reused pilots cases in the system with non-ideal hardware—particularly

[18] As an example, suppose $\mathbf{R} = \delta^{-3.7}\mathbf{I}$ and $\mathbf{R}_l = \delta_l^{-3.7}\mathbf{I}$ where 3.7 is the path loss exponent and δ, δ_l are the distances between the BS under study and the two users. The right-hand side of (64) becomes $(\mathrm{tr}(\mathbf{A}\mathbf{R}_l)/\mathrm{tr}(\mathbf{A}\mathbf{R}))^2 = (\delta_l/\delta)^{-7.4}$ which is in the range -20 to -30 dB if UE l is 1.9–2.5 times further away from the BS than the UE under study. This is the case for most UEs in neighboring cells, but to be sure one can apply a fractional reuse pattern such that adjacent cells use different pilots. All interfering UEs will then be, at least, 2 times further away from the BS than the UE under study.

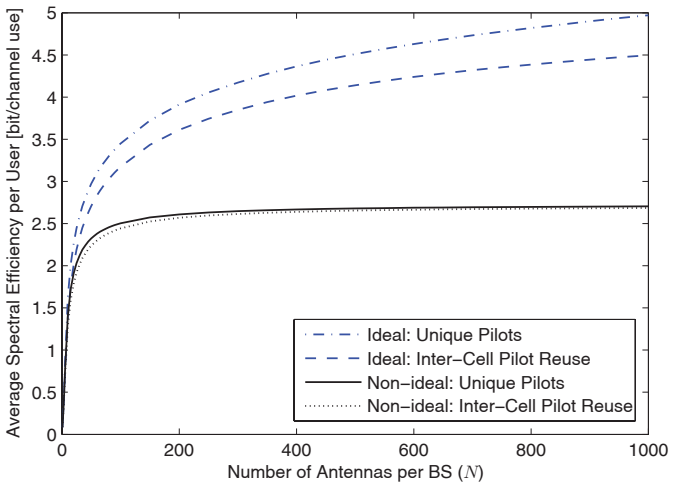


Fig. 14: Achievable UL spectral efficiency for an average user in the multi-cell scenario depicted in Fig. 13. Each UE has either a unique pilot signal or the same pilots are reused in every cell. Pilot contamination degrades the spectral efficiency under ideal hardware, while the impact on a system with hardware impairments ($\kappa_t^{\mathrm{UE}} = \kappa_r^{\mathrm{BS}} = 0.1^2$) is negligible.

not when N is large. This implies that pilot contamination might have a negligible impact on massive MIMO systems with hardware impairments, as also shown in Corollary 8. The explanation is that the distortion noise at the UE is the main limiting factor in the considered scenario, thus regular inter-user interference and the pilot contamination drowns in the distortions.

Informally speaking, the distortion noise acts as a fog that prevents the BS from seeing distant interferers in the pilot transmission phase. The number of orthogonal pilot signals is limited by $T_{\mathrm{pilot}}^{\mathrm{UL}}$ within the sight of each BS, but can otherwise be reused freely. A simple location-based pilot allocation was sufficient for the multi-cell scenario depicted in Fig. 13, but in general we believe that cell center and cell edge UEs need to be treated differently. If one can afford a fractional reuse pattern, where adjacent cells never use the same pilots, then Corollary 8 will be satisfied in most cases; see Footnote 18.

Finally, we recall from Section IV-C that the gain of increasing the number of antennas beyond $N = 100$ was small in the single-user case with non-ideal hardware. More antennas can be used in multi-cell scenarios to suppress the regular interference. The convergence to the asymptotic limit is, however, much faster with non-ideal hardware, because it is sufficient to suppress the regular interference to a level below the distortion noise.

VII. Refinements of the System Model and the Possible Implications

Using the system model defined in Section II, we have shown how the additive distortion noises from hardware impairments limit the estimation accuracy and channel capacities. The practical relevance of the system model has been verified experimentally in [15]–[17]. It can also be motivated theoretically when the impairment characteristics are static within each coherence period (e.g., due to the use of strong compensation algorithms). In this case, one can apply the Bussgang theorem which shows that any nonlinear distortion function of a Gaussian signal can be reduced to an affine function where the signal is multiplied with an effective channel and corrupted by uncorrelated Gaussian noise [7]. This results in additive distortion noise similar to the one in our system model, but not identical. For analytic tractability, we assumed in the system model of Section II that the distortion noises are *independent* of the data signals (not only uncorrelated) and Gaussian distributed (even if the data signals are not); the same assumptions were made in [14]–[19]. If one would consider an alternative model where these two assumptions are not made, then the lower capacity bounds in this paper will still hold (because the mutual information is always reduced by adding the two assumptions [35]). The upper bounds in Section IV would not hold without the two assumptions, and new upper bounds can only be derived if we impose alternative assumptions on the exact dependence between signal and distortion. In other words, the model in Section II is a tractable canonical approximation of communication systems with hardware imperfections, but it is not a perfect model of reality.

In general, the time-varying nature of hardware impairments cannot be completely mitigated, which also give rise to multiplicative distortions that vary within each coherence period. Furthermore, the covariance matrices of the additive distortion noises, given in (3), (4), (7), and (8), can be refined in several ways. This section outlines some possible refinements of the system model in Section II and how each one is expected to affect the main results—the exact analysis is not straightforward and is left for future work. Most of these model refinements will further degrade the performance, thus the upper capacity bounds in Theorem 2 is typically valid, while the lower capacity bounds need to be reduced.

A. Power Loss

It is difficult to model the total emitted power under non-ideal hardware, because some distortions are created independently in the hardware, other distortions take their power from the useful signals, and some impairment sources (e.g., non-linearities) can even reduce the emitted power. In this paper, we have implicitly assumed that the compensation algorithms scale the total emitted power such that it equals $p^{\mathrm{BS}}(1+\kappa_t^{\mathrm{BS}})$ in the DL and $p^{\mathrm{UE}}(1+\kappa_t^{\mathrm{UE}})$ in the UL. This simplification creates a small bias when comparing systems with different levels of impairments, but the simulations in [20, Section 4.3] showed that this has a negligible impact on the spectral efficiencies. Nevertheless, it is important to note that although the distortion noise caused by the BS vanishes as $N \to \infty$, there remains a *power loss* of $\frac{\kappa_t^{\mathrm{BS}}}{1+\kappa_t^{\mathrm{BS}}}$ that should be taken into account when designing massive MIMO systems.

B. High-Power Scalings

The levels of impairments in the transmitter hardware, κ_t^{BS} and κ_t^{UE}, were taken as constants in Sections II–IV. This is reasonable when operating within the dynamic/linear range

of the respective power amplifiers. Outside these ranges, the proportionality coefficients increase rapidly with the transmit powers p^{BS} and p^{UE} due non-linearities. This behavior was accurately modeled by polynomials in [19] and [58]; for example, κ_t^{BS} could have two terms: a constant term describing the low-power EVM and a term $(p^{\mathrm{BS}}/c)^q$, for some exponent q, describing the severity/order of the dominating non-linearity and a constant $c > 0$ that marks the end of the dynamic range [19]. Note that the distortion noise added by the low-noise amplifiers in radio receivers will typically not become worse with the received power, thus it is reasonable to let κ_r^{BS} and κ_r^{UE} be constants.

The consequence of having proportionality coefficients that scale with the transmit powers is that the distortion noise power increases faster than the signal power. Hence, the capacity and estimation accuracy are no longer monotonically increasing in p^{BS} and p^{UE} when using non-ideal hardware—these metrics are instead maximized at some finite transmit powers [58]. The reason that we took κ_t^{BS} and κ_t^{UE} as constants herein is that the high-power regime is not our main focus. Consequently, the high-power limits that we derived are optimistic and might not be achievable in practice—alternatively, they are the result of decreasing the propagation distance instead of increasing the actual emitted power. The results when $N \to \infty$ are however accurate since the total power (or at least the power per antenna) decreases with N; see the discussion in Section V.

C. Alternative Distortion Noise Distributions

The distortion covariance matrices in (3) and (8) are based on the assumption of independent distortion at the different BS antennas. This implies that the distortion noise has a different spatial signature than the useful signal, which is the reason why the detrimental impact of the distortion noise caused by the BS vanishes as $N \to \infty$. The underlying assumption is that the hardware chains of different antennas are decoupled. Nevertheless, there can exist cross-correlation since the same useful signal is transmitter/received over the array, thus making the hardware react similarly. Such correlation was predicted and characterized in [59] but is typically small. Thus, we believe that also in practice the distortion noise and useful signal have different spatial signatures as N grows large.

The distortion noises were assumed to be Gaussian distributed (for any fixed channel realization), but this can also be relaxed. As can be seen in the appendices, the proofs rely on that the cross-moments between the signal and the distortion are weak. The independence can, probably, be replaced with uncorrelation and that the higher-order moments are sufficiently weak, but the corresponding generalized proofs will be rather tedious and the convergence as $N \to \infty$ might be slower.

D. Multiplicative Distortions

The additive distortion model in this paper has been verified experimentally for systems that apply compensation algorithms to mitigate the main hardware impairments. It is also an accurate model for uncompensated inter-carrier interference caused by phase noise and I/Q imbalance, amplitude-amplitude nonlinearities in power amplifiers, and quantization errors [14], [31], [60]. As described in the beginning of Section VII, hardware impairments also cause channel attenuations and phase shifts that are multiplied with the channel vector $\mathbf{h}$. If these multiplicative distortions are sufficiently static (after compensation), they can be included in the channel vector $\mathbf{h}$ by an appropriate scaling of the covariance matrix $\mathbf{R}$ or by exploiting that the channel distribution is circularly symmetric. However, phase noise is a prime example of an impairment that causes multiplicative distortions that drift and accumulate within the channel coherence period [25], [61]–[63]. We now take a look closer at this type of distortions, to investigate in which ways it behaves differently from additive distortion noise. The actual channel under phase noise can be described as $\mathrm{diag}(e^{\jmath\phi_{1,t}}, \ldots, e^{\jmath\phi_{N,t}})\mathbf{h}$, where $\jmath = \sqrt{-1}$ is the imaginary unit and $\{\phi_{i,t}\}$ is the stochastic process at the ith channel element. The phase drift of free-running oscillators is commonly modeled as a Wiener process

$$\phi_{i,t} = \phi_{i,t-1} + \theta_{i,t}^{\mathrm{BS}} + \theta_t^{\mathrm{UE}} \quad \forall i \tag{65}$$

where the initial value is $\phi_{i,0} = 0$ since $t = 0$ denotes the time of the channel estimation. The innovations that occur t channel uses after the channel estimation are $\theta_{i,t}^{\mathrm{BS}} \sim \mathcal{N}(0, \Delta^{\mathrm{BS}})$ and $\theta_t^{\mathrm{UE}} \sim \mathcal{N}(0, \Delta^{\mathrm{UE}})$ at the BS and UE, respectively. Note that the single-antenna UE's hardware causes identical drifts on all channel elements, while the BS can cause identical or independent drifts depending on the use of a common oscillator (CO) or separate oscillators (SOs) at each antenna element. The phase drifts are temporally white, thus $\phi_{i,t} \sim \mathcal{N}(0, t\Delta^{\mathrm{BS}} + t\Delta^{\mathrm{UE}})$ which shows that the variance increases with time.

To comprehend the impact of phase noise, we note that the signal part of the received UL signal under MRC, $\mathbf{v} = \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}$, is

$$\begin{aligned} &\mathbf{v}^H \mathrm{diag}(e^{\jmath\phi_{1,t}}, \ldots, e^{\jmath\phi_{N,t}})\mathbf{h}d \\ &\quad \approx \underbrace{\mathbf{v}^H\mathbf{h}d}_{\text{Ideal signal}} + \underbrace{\jmath\mathbf{v}^H \mathrm{diag}(\phi_{1,t}, \ldots, \phi_{N,t})\mathbf{h}d}_{\text{Distortion from phase noise}} \end{aligned} \tag{66}$$

using the Taylor approximation $e^{\jmath\phi_{i,t}} \approx 1 + \jmath\phi_{i,t}$ because the drifts are small[19] [64], [65]. The first term in (66) is the same as without phase noise, while the second term characterizes the mismatch from the phase drift. Since $\phi_{i,t}$ has zero mean, the two terms are uncorrelated (irrespective of if $\mathbf{h}$ and d are deterministic or stochastic). We can therefore obtain a lower bound on the mutual information by treating the uncorrelated second term of (66) as independent Gaussian noise [35, Theorem 1]. By taking the average over channel realizations, data signals, and phase drifts, the variance of this

[19]Since the variance of $\phi_{i,t}$ increases linearly with t, the Taylor approximation is only valid for a certain time. The time dependence can however be mitigated by tracking the phase noise within each coherence period [64].

distortion is

$$\begin{aligned}
&\mathbb{E}\left\{\left|j\mathbf{v}^H \mathrm{diag}(\phi_{1,t},\ldots,\phi_{N,t})\mathbf{h}d\right|^2\right\} \\
&= \sum_{i_1=1}^{N}\sum_{i_2=1}^{N} \mathbb{E}\{v_{i_1}^* v_{i_2} h_{i_1} h_{i_2}^*\}\mathbb{E}\{\phi_{i_1,t}\phi_{i_2,t}^*\}\mathbb{E}\{|d|^2\} \\
&= p^{\mathrm{UE}}\mathbb{E}\{|\mathbf{v}^H\mathbf{h}|^2\}t\Delta^{\mathrm{UE}} \\
&\quad + \begin{cases} p^{\mathrm{UE}}t\Delta^{\mathrm{BS}}\mathbb{E}\{|\mathbf{v}^H\mathbf{h}|^2\}, & \text{if CO,} \\ p^{\mathrm{UE}}t\Delta^{\mathrm{BS}}\sum_{i=1}^{N}\mathbb{E}\{|h_i|^2|v_i|^2\}, & \text{if SO,} \end{cases}
\end{aligned} \tag{67}$$

where the first term originates from the UE and the second term is due to the BS having either a CO or SOs at each antenna. Recall from Theorem 4 that $\mathbb{E}\{|\mathbf{v}^H\mathbf{h}|^2\} = \mathcal{O}(N)$ and $\mathbb{E}\{|h_i|^2|v_i|^2\} = \mathcal{O}(1)$. This means that a BS with a CO causes distortion that scales as $\mathcal{O}(tN)$, while it only scales as $\mathcal{O}(t)$ when having SOs.[20] In other words, it appears to be preferable to have independent oscillators at each antenna element in massive MIMO systems, which was also noted in [25]. This is also consistent with our results for additive distortion noise: impairments at the UE are N times more influential on the capacity, thus we can degrade the quality of the BS's oscillators with N and only get a minor loss in performance. This property is also positive for distributed massive MIMO deployments where the antenna separation is large and prevents the use of a CO. A major difference from the additive distortion noise in Section II is that the distortion in (67) increases linearly with the time t, thus it eventually grows large and it becomes necessary to send pilot/calibration signals more often to mitigate it [25].

The narrowband phase-noise analysis above only considered a lower capacity bound, thus it is possible to achieve higher rates. In particular, the analysis assumed uncompensated free-running oscillators, while it might be better to track the phase noise process at the BS; for example, by using previous received signals, extra calibration signals (see [64] and references therein), and utilizing correlation between subcarriers in multi-carrier systems. The tracking might be more accurate for a CO since there are $O(N)$ observations of a single phase drift parameter, instead of $O(N)$ observations for N parameters as with SOs. Another important aspect of phase noise is that the standard deviations $\sqrt{\Delta^{\mathrm{BS}}}$ and $\sqrt{\Delta^{\mathrm{UE}}}$ are typically proportional to the carrier frequency [62], thus phase noise might be a major challenge in higher frequency bands (e.g., mmWave) [66]—unless the symbol time is also sufficiently reduced by increasing the bandwidth.

E. Imperfect Channel Reciprocity

The downlink beamforming in massive MIMO TDD systems relies on channel reciprocity; that is, if $\mathbf{h}$ is the uplink channel then $\mathbf{h}^T$ is the downlink channel. This property holds for the radio-frequency propagation channels, but the end-to-end channels are also affected by the hardware since different transceiver chains are used for transmission/reception at the BS and the UE. The actual downlink channel is $\mathbf{h}^T\mathbf{D_b}$ where the diagonal matrix $\mathbf{D_b} = \mathrm{diag}(b_1,\ldots,b_N)$ contains N calibration parameters. These are $b_i = 1\ \forall i$ for ideal hardware but we generally have $b_i \neq 1\ \forall i$ due to non-ideal hardware. The mismatch is fully specified by $b_1,\ldots,b_N$ and fortunately these parameters change slowly with time, thus one can compute estimates $\hat{b}_1,\ldots,\hat{b}_N$ using a negligible amount of overhead signaling [16] (even in massive MIMO systems [67]). Since the transmit beamforming mainly depends on the channel direction, it is often sufficient for the BS to compute the downlink channel up to an unknown scaling factor; see [16], [67]–[69] for different techniques that exploit uplink pilot transmissions. The estimates are naturally imperfect, thus $b_i = c(\hat{b}_i + e_i)$ where e_i is the estimation error and c is the unknown common scaling factor.

Imperfect channel reciprocity has no impact on the UL and is not expected to change anything fundamentally in the DL. There is a loss in received signal power since the beamforming direction is perturbed, but there is no extra self-interference since all the CSI available at the receiving UE is estimated in the downlink and thus reflects the actual downlink channel $\mathbf{h}^T\mathbf{D_b}$. In other words, the lower capacity bound in (34) is still valid if we replace $\mathbf{h}$ by $\mathbf{D_b}\mathbf{h}$ everywhere and compute the expectations with respect to the actual distributions. The beamforming vector $\mathbf{v}^{\mathrm{DL}}$ is now a function of $\mathrm{diag}(\hat{b}_1,\ldots,\hat{b}_N)\hat{\mathbf{h}}$. This perturbation of $\mathbf{v}^{\mathrm{DL}}$, as compared to having perfect reciprocity, behaves like a channel estimation error and its impact is expected to vanish as N grows large. Moreover, it should only have a minor impact on the inter-user interference in multi-cell scenarios, since the reciprocity calibration errors are independent of the co-user channels.

VIII. Conclusion

This paper analyzed the capacity and estimation accuracy of massive MIMO systems with non-ideal transceiver hardware. The analysis was based on a new system model that models the hardware impairment at each antenna by an additive distortion noise that is proportional to the signal power at this antenna. This model has several attractive features: it is mathematically tractable, it has been verified experimentally in previous works, and it can be motivated theoretically in systems that apply compensation algorithms to mitigate the hardware impairments.

We proved analytically that hardware impairments create non-zero estimation error floors and finite capacity ceilings in the uplink and downlink—irrespective of the SNR and the number of base station antennas N. This stands in contrast to the very optimistic asymptotic results previously reported for ideal hardware. Despite these discouraging results, we showed that massive MIMO systems can still achieve a huge array gain, in the sense that relatively high spectral efficiency and energy efficiency can be obtained. Furthermore, we proved that only the hardware impairments at the UEs limit the capacities as N grows large. This implies that the hardware quality at the BS can be decreased as N grows, which is an important insight and might become a key enabler for future network deployments.

In multi-cell scenarios, we proved that the detrimental effect of inter-user interference and pilot contamination drowns in

[20]Since the useful signal power $p^{\mathrm{UE}}\mathbb{E}\{|\mathbf{v}^H\mathbf{h}|^2\}$ also scales as $\mathcal{O}(N)$, the relative distortion power behaves as $\mathcal{O}(t)$ with CO and $\mathcal{O}(t/N)$ with SOs.

the distortion noise if a simple pilot allocation algorithm is used to avoid the strongest forms of pilot contaminated interference. Many quantitative conclusions can be drawn from the numerical results in Sections III–VI; for example, that there is little gain in having more than 100 antennas for a single-user link, but additional antennas are useful to suppress inter-user interference in multi-cell scenarios. The asymptotic limits under non-ideal hardware are generally reached at much fewer antennas than the asymptotic limits for ideal hardware, which implies that we can expect practical systems to benefit from the asymptotic results. We also gave a brief description of how the system model considered in this paper can be refined to model hardware impairments in even greater detail and how such refinements would affect our results.

Appendix A
New and Old Results on Random Vectors

Lemma 2. *[70, Eq. (2.2)] For invertible matrices $\mathbf{B}$ and $\tau \geq 0$, it holds that*

$$(\mathbf{B}+\tau\mathbf{x}\mathbf{x}^H)^{-1}\mathbf{x} = \frac{\mathbf{B}^{-1}\mathbf{x}}{1+\tau\mathbf{x}^H\mathbf{B}^{-1}\mathbf{x}}. \tag{68}$$

Lemma 3. *Suppose $h \sim \mathcal{CN}(0,r)$ and $a,b>0$, then*

$$\mathbb{E}\left\{\frac{|h|^2}{a|h|^2+b}\right\} = \frac{1}{a}\left(1-\frac{b}{ar}E_1\left(\frac{b}{ar}\right)e^{\frac{b}{ar}}\right) \tag{69}$$

where $E_1(x)=\int_1^\infty \frac{e^{-tx}}{t}dt$ denotes the exponential integral.

Proof: Since $\varrho = |h|^2$ has the exponential distribution with mean value r, the expectation in (69) equals

$$\int_0^\infty \frac{\varrho}{a\varrho+b}\frac{e^{-\varrho/r}}{r}d\varrho = \frac{b}{a^2r}e^{\frac{b}{ar}}\int_1^\infty\left(1-\frac{1}{x}\right)e^{-\frac{\varrho b}{ar}}dx \tag{70}$$

where the equality follows from a change of variable $x = \frac{a}{b}\varrho+1$. Straightforward integration and identification of the exponential integral yield the right-hand side of (69). ∎

Lemma 4. *For any $a,b\in\mathbb{C}$ and non-zero $c,d\in\mathbb{C}$, we have*

$$\left|\frac{a}{c}-\frac{b}{d}\right| \leq \frac{|b|\,|c-d|}{|c|\,|d|}+\frac{|a-b|}{|c|}. \tag{71}$$

Proof: This is straightforward to prove by using that $|ad-bc| = |ad-bc+bd-bd| \leq |b|\,|c-d|+|d|\,|a-b|$. ∎

Lemma 5. *Consider M arbitrary matrices $\mathbf{M}_1,\ldots,\mathbf{M}_M \in \mathbb{C}^{N\times N}$ and an Hermitian positive semi-definite matrix $\mathbf{B}\in\mathbb{C}^{N\times N}$. It follows that*

$$|\mathrm{tr}(\mathbf{M}_1\cdots\mathbf{M}_M\mathbf{B})| \leq \mathrm{tr}(\mathbf{B})\prod_{i=1}^M\|\mathbf{M}_i\|_2. \tag{72}$$

If $\mathbf{M}_1,\ldots,\mathbf{M}_M,\mathbf{B}$ have uniformly bounded spectral norms, then

$$|\mathrm{tr}(\mathbf{M}_1\cdots\mathbf{M}_M\mathbf{B})| = \mathcal{O}(N). \tag{73}$$

Proof: The bound in (72) follows from that $\mathbf{B}$ has non-negative eigenvalues and each matrix $\mathbf{M}_i$ cannot amplify these by more than $\|\mathbf{M}_i\|_2$. The $\mathcal{O}(N)$-scaling follows from (72) by using the assumptions and $\mathrm{tr}(\mathbf{B}) \leq N\|\mathbf{B}\|_2$. ∎

Lemma 6. *[71, Lemma B.26] Let $\mathbf{B}\in\mathbb{C}^{N\times N}$ be deterministic and $\mathbf{x}=[x_1\ \ldots\ x_N]^T\in\mathbb{C}^N$ be a stochastic vector of independent entries. Assume that $\mathbb{E}\{x_i\}=0$, $\mathbb{E}\{|x_i|^2\}=1$, and $\mathbb{E}\{|x_i|^\ell\}=\chi_\ell<\infty$ for $\ell\leq 2q$. Then, for any $q\geq 1$,*

$$\mathbb{E}\left\{\left|\mathbf{x}^H\mathbf{B}\mathbf{x}-\mathrm{tr}(\mathbf{B})\right|^q\right\} \leq C_q\left(\mathrm{tr}(\mathbf{B}\mathbf{B}^H)\right)^{\frac{q}{2}}\left(\chi_4^{\frac{q}{2}}+\chi_{2q}\right) \tag{74}$$

where C_q is a constant depending on q only.

Appendix B
Application-Related Random Vector Results

Lemma 7. *The channel estimate $\hat{\mathbf{h}}$ can be decomposed as*

$$\hat{\mathbf{h}} = \mathbf{A}\left(\left((d+\eta_t^{\mathrm{UE}})\mathbf{I}+\mathbf{D}_r\right)\mathbf{h}+\boldsymbol{\nu}\right) \tag{75}$$

where $\mathbf{A}$ is defined in (9) and the diagonal matrix $\mathbf{D}_r$ has independent $\mathcal{CN}(0,\kappa_r^{\mathrm{BS}}p^{\mathrm{UE}})$-entries such that $\boldsymbol{\eta}_r^{\mathrm{BS}}=\mathbf{D}_r\mathbf{h}$.

For any realizations of η_t^{UE} and $\mathbf{D}_r$, the conditional distribution is

$$\hat{\mathbf{h}}|\eta_t^{\mathrm{UE}},\mathbf{D}_r \sim \mathcal{CN}\left(\mathbf{0},\mathbf{A}(\boldsymbol{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\right) \tag{76}$$

where $\boldsymbol{\Phi}=((d+\eta_t^{\mathrm{UE}})\mathbf{I}+\mathbf{D}_r)\mathbf{R}((d+\eta_t^{\mathrm{UE}})\mathbf{I}+\mathbf{D}_r)^H$.

Proof: This characterization follows directly from Theorem 1 and the system model defined in Section II. ∎

Lemma 8. *For the channel $\mathbf{h}$ and its estimate $\hat{\mathbf{h}}$ it holds that*

$$\mathbb{E}\left\{\left|\mathbf{h}^H\hat{\mathbf{h}}-(1+d^{-1}\eta_t^{\mathrm{UE}})\mathrm{tr}(\mathbf{R}-\mathbf{C})\right|^2\right\} = \mathcal{O}(N) \tag{77}$$

$$\mathbb{E}\left\{\left||\mathbf{h}^H\hat{\mathbf{h}}|-|1+d^{-1}\eta_t^{\mathrm{UE}}|\mathrm{tr}(\mathbf{R}-\mathbf{C})\right|^2\right\} = \mathcal{O}(N). \tag{78}$$

Proof: Recall that $\hat{\mathbf{h}} = \mathbf{A}\big(\mathbf{h}(d+\eta_t^{\mathrm{UE}})+\boldsymbol{\nu}+\boldsymbol{\eta}_r^{\mathrm{BS}}\big)$ for $\mathbf{A}=d^*\mathbf{R}\bar{\mathbf{Z}}^{-1}$. To prove (77), we expand the argument as

$$\left|\mathbf{h}^H\hat{\mathbf{h}}-(1+d^{-1}\eta_t^{\mathrm{UE}})\mathrm{tr}(\mathbf{R}-\mathbf{C})\right|^2 \leq 4|\mathbf{h}^H\mathbf{A}\boldsymbol{\nu}|^2 \tag{79}$$
$$+4|\mathbf{h}^H\mathbf{A}\boldsymbol{\eta}_r^{\mathrm{BS}}|^2+2|d+\eta_t^{\mathrm{UE}}|^2\left|\mathbf{h}^H\mathbf{A}\mathbf{h}-d^{-1}\mathrm{tr}(\mathbf{R}-\mathbf{C})\right|^2$$

by using the rule $|a+b|^q\leq 2^{q-1}(|a|^q+|b|^q)$ (from Hölder's inequality) twice. Next, we observe that

$$\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\boldsymbol{\nu}|^2\} = \mathrm{tr}\left(\mathbf{A}(\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\mathbf{R}\right) = \mathcal{O}(N) \tag{80}$$

$$\mathbb{E}\{|\mathbf{h}^H\mathbf{A}\boldsymbol{\eta}_r^{\mathrm{BS}}|^2\} = \kappa_r^{\mathrm{BS}}p^{\mathrm{UE}}\mathrm{tr}\left(\mathbf{A}\mathbf{R}_{\mathrm{diag}}\mathbf{A}^H\mathbf{R}\right) + \kappa_r^{\mathrm{BS}}p^{\mathrm{UE}}\sum_{i=1}^N|\mathbf{e}_i^H\mathbf{R}\mathbf{A}\mathbf{e}_i|^2 = \mathcal{O}(N) \tag{81}$$

where $\mathbf{e}_i$ is the ith column of an $N\times N$ identity matrix. The expression (80) follows from the independence of $\mathbf{h},\boldsymbol{\nu}$ and (81) follows by straightforward computation using the characterization $\boldsymbol{\eta}_r^{\mathrm{BS}}=\mathbf{D}_r\mathbf{h}$ in Lemma 7. The $\mathcal{O}(N)$-properties follows from Lemma 5 since $\mathbf{R},\mathbf{S},\mathbf{A}$ have uniformly bounded spectral norms (by assumption).

Since $\mathbf{h}\sim\mathbf{R}^{1/2}\tilde{\mathbf{h}}$ for $\tilde{\mathbf{h}}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$ and $d^{-1}\mathrm{tr}(\mathbf{R}-\mathbf{C}) = \mathrm{tr}(\mathbf{R}^{1/2}\mathbf{A}\mathbf{R}^{1/2})$ we can apply Lemma 6 to obtain

$$\begin{aligned}&\mathbb{E}\left\{|d+\eta_t^{\mathrm{UE}}|^2\left|\mathbf{h}^H\mathbf{A}\mathbf{h}-d^{-1}\mathrm{tr}(\mathbf{R}-\mathbf{C})\right|^2\right\}\\ &\leq(1+\kappa_t^{\mathrm{UE}})2\chi_4C_2\mathrm{tr}\left((\mathbf{R}-\mathbf{C})^2\right)=\mathcal{O}(N).\end{aligned} \tag{82}$$

We obtain (77) by combining (79)–(82). Expression (78) follows directly, since it is upper bounded similarly to (79). ∎

Lemma 9. *For the channel* $\mathbf{h}$ *and its estimate* $\hat{\mathbf{h}}$ *it holds that*

$$\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2^2 - \mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)\right|^2\right\} = \mathcal{O}(N) \tag{83}$$

$$\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2 - \sqrt{\mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)}\right|^2\right\} = \mathcal{O}(1) \tag{84}$$

where $\mathbf{A}$ *is defined in* (9) *and* $\mathbf{\Psi} = p^{\mathrm{UE}}\kappa_r^{\mathrm{BS}}\mathbf{R}_{\mathrm{diag}} + \mathbf{S} + \sigma_{\mathrm{BS}}^2\mathbf{I}$.

Proof: By injecting the term $\mathrm{tr}(\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H)$ that appeared in Lemma 7 and using the rule $|a+b|^q \leq 2^{q-1}(|a|^q + |b|^q)$ (from Hölder's inequality), we bound the left-hand side of (83) as

$$\begin{aligned}
&\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2^2 - \mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)\right|^2\right\} \\
&\leq 2\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2^2 - \mathrm{tr}\Big(\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\Big)\right|^2\right\} \\
&+ 2\mathbb{E}\Bigg\{\Bigg|\mathrm{tr}\Big(\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\Big) \\
&\qquad - \mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)\Bigg|^2\Bigg\}.
\end{aligned} \tag{85}$$

The first term in (85) satisfies

$$\begin{aligned}
&\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2^2 - \mathrm{tr}(\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H)\right|^2\right\} \\
&\leq 4C_2\mathbb{E}\left\{\mathrm{tr}(\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})^H\mathbf{A}^H)\right\} \\
&\leq 4C_2\|\mathbf{A}\|_2^4\,\mathbb{E}\left\{\|\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I}\|_F^2\right\} = \mathcal{O}(N)
\end{aligned} \tag{86}$$

where the first inequality follows from applying Lemma 6 on (83) for fixed $\eta_t^{\mathrm{UE}}, \mathbf{D}_r$ (note that the fourth-order moment is $\chi_4 = 2$ for complex Gaussian variables), while the second inequality follows from applying Lemma 5 twice. The scaling $\mathcal{O}(N)$ follows since σ_{BS}^2 is constant, $\|\mathbf{A}\|_2 = \mathcal{O}(1)$, $\|\mathbf{S}\|_F^2 = \mathcal{O}(N)$, $\mathbb{E}\{\mathrm{tr}(\mathbf{\Phi}\mathbf{S})\} \leq \|\mathbf{R}\|_2\|\mathbf{S}\|_2\mathbb{E}\{\|(d+\eta_t^{\mathrm{UE}})\mathbf{I}+\mathbf{D}_r\|_F^2\} = \mathcal{O}(N)$, and $\mathbb{E}\{\mathrm{tr}(\mathbf{\Phi}\mathbf{\Phi}^H)\} \leq \|\mathbf{R}\|_2\mathbb{E}\{\|((d+\eta_t^{\mathrm{UE}})\mathbf{I}+\mathbf{D}_r)((d+\eta_t^{\mathrm{UE}})\mathbf{I}+\mathbf{D}_r)^H\|_F^2\} = \mathcal{O}(N)$ using Lemma 5.

Next, we characterize the second term in (85) as

$$\begin{aligned}
&\mathbb{E}\Bigg\{\Bigg|\mathrm{tr}\Big(\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\Big) \\
&\qquad - \mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)\Bigg|^2\Bigg\} \\
&= \mathbb{E}\Bigg\{\Bigg|\mathrm{tr}\Big(\mathbf{A}\big(\mathbf{D}_r\mathbf{R}\mathbf{D}_r^H + \mathbf{D}_r\mathbf{R}(d+\eta_t^{\mathrm{UE}})^* \\
&\qquad + (d+\eta_t^{\mathrm{UE}})\mathbf{R}\mathbf{D}_r^H\big)\mathbf{A}^H\Big) - p^{\mathrm{UE}}\kappa_r^{\mathrm{BS}}\mathrm{tr}\Big(\mathbf{A}\mathbf{R}_{\mathrm{diag}}\mathbf{A}^H\Big)\Bigg|^2\Bigg\} \\
&\leq 2\mathbb{E}\left\{\left|\mathrm{tr}\Big(\mathbf{A}(\mathbf{D}_r\mathbf{R}\mathbf{D}_r^H - p^{\mathrm{UE}}\kappa_r^{\mathrm{BS}}\mathbf{R}_{\mathrm{diag}})\mathbf{A}^H\Big)\right|^2\right\} \\
&+ 4\mathbb{E}\left\{|d+\eta_t^{\mathrm{UE}}|^2\right\}\mathbb{E}\left\{\left|\mathrm{tr}\Big(\mathbf{A}\mathbf{D}_r\mathbf{R}\mathbf{A}^H\Big)\right|^2\right\} = \mathcal{O}(N).
\end{aligned} \tag{87}$$

where the equality follows from plugging in the expressions for $\mathbf{\Phi}$ and $\mathbf{\Psi}$ and noting that the terms $|d+\eta_t^{\mathrm{UE}}|^2\mathrm{tr}(\mathbf{A}\mathbf{R}\mathbf{A}^H)$, $\mathrm{tr}(\mathbf{A}\mathbf{S}\mathbf{A}^H)$, and $\sigma_{\mathrm{BS}}^2\mathrm{tr}(\mathbf{A}\mathbf{A}^H)$ cancel out. The inequality follows again from the rule $|a+b|^q \leq 2^{q-1}(|a|^q + |b|^q)$. The $\mathcal{O}(N)$-scaling follows since the first term in (87) is upper bounded by $(p^{\mathrm{UE}}\kappa_r^{\mathrm{BS}})^2\mathrm{tr}(\mathbf{A}\mathbf{R}_{\mathrm{diag}}\mathbf{A}^H\mathbf{A}\mathbf{R}_{\mathrm{diag}}\mathbf{A}^H) = \mathcal{O}(N)$ using Lemma 6 and some algebra, while $\mathbb{E}\{|\mathrm{tr}(\mathbf{A}\mathbf{D}_r\mathbf{R}\mathbf{A}^H)|^2\} \leq \|\mathbf{R}\mathbf{A}^H\mathbf{A}\|_2^2\mathbb{E}\{|\mathrm{tr}(\mathbf{D}_r)|^2\} = \mathcal{O}(N)$ using Lemma 5. The expression (83) now follows from combining (85)–(87).

Finally, the expression (84) is proved as

$$\begin{aligned}
&\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2 - \sqrt{\mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)}\right|^2\right\} \\
&\leq \mathbb{E}\left\{\frac{\left|\|\hat{\mathbf{h}}\|_2^2 - \mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)\right|^2}{\left|\|\hat{\mathbf{h}}\|_2 + \sqrt{\mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)}\right|^2}\right\} \\
&\leq \frac{\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2^2 - \mathrm{tr}\Big(\mathbf{A}(|d+\eta_t^{\mathrm{UE}}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\Big)\right|^2\right\}}{\mathrm{tr}\Big(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H\Big)} = \mathcal{O}(1)
\end{aligned} \tag{88}$$

where the first inequality follows from the rule $(a-b)(a+b) = a^2 - b^2$ and the second inequality is due to removal of non-zero terms from the denominator. The numerator scales as $\mathcal{O}(N)$ and the denominator scales at least linearly with N because $\mathrm{tr}\Big(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H\Big) \geq \lambda_{\min}(\mathbf{\Psi})\mathrm{tr}\Big(\mathbf{A}\mathbf{A}^H\Big) \geq p^{\mathrm{UE}}\lambda_{\min}(\mathbf{\Psi})\lambda_{\max}(\bar{\mathbf{Z}})\mathrm{tr}\Big(\mathbf{R}\mathbf{R}^H\Big)$, where $\mathrm{tr}\Big(\mathbf{R}\Big)$ grows linearly with N (by assumption). Here, $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denotes the largest and smallest eigenvalues of a matrix, respectively. This shows that (84) is bounded and finalizes the proof. ∎

Lemma 10. *For the estimated channel* $\hat{\mathbf{h}}$ *in* (9) *it holds that*

$$\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^k}{\|\hat{\mathbf{h}}\|_2^2}\right\} \leq \frac{2^k + 2^k\left(\frac{k}{2}\right)!(\kappa_t^{\mathrm{UE}})^{\frac{k}{2}}}{\lambda_{\min}^+(\mathbf{B})(N_B-1)} = \mathcal{O}(N^{-1}) \tag{89}$$

$$\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^k}{\|\hat{\mathbf{h}}\|_2^4}\right\} \leq \frac{2^k + 2^k\left(\frac{k}{2}\right)!(\kappa_t^{\mathrm{UE}})^{\frac{k}{2}}}{\lambda_{\min}^+(\mathbf{B})^2(N_B-1)(N_B-2)} = \mathcal{O}(N^{-2}) \tag{90}$$

for any even integer k*, where* $\lambda_{\min}^+(\mathbf{B}) > 0$ *denotes the smallest non-zero eigenvalue of* $\mathbf{B} = \sigma_{\mathrm{BS}}^2\mathbf{A}\mathbf{A}^H$ *and* $N_B = \mathrm{rank}(\mathbf{B})$.

Proof: Using the conditional distribution of the channel estimate in (76), it holds for any integer $q > 0$ that

$$\begin{aligned}
\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^k}{\|\hat{\mathbf{h}}\|_2^{2q}}\right\} &= \mathbb{E}\left\{\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^k}{\|\hat{\mathbf{h}}\|_2^{2q}}\middle|\eta_t^{\mathrm{UE}}, \mathbf{D}_r\right\}\right\} \\
&\leq \mathbb{E}\left\{\frac{2^k(1+|d^{-1}\eta_t^{\mathrm{UE}}|^k)}{\lambda_{\min}^+(\sigma_{\mathrm{BS}}^2\mathbf{A}\mathbf{A}^H)^q}\right\}\mathbb{E}\left\{\frac{1}{\|\mathbf{U}_B^H\mathbf{v}\|_2^{2q}}\right\}
\end{aligned} \tag{91}$$

where the inequality follows from $\|\hat{\mathbf{h}}\|_2^2 = \mathbf{v}^H\mathbf{A}(\mathbf{\Phi}+\mathbf{S}+\sigma_{\mathrm{BS}}^2\mathbf{I})\mathbf{A}^H\mathbf{v} \geq \sigma_{\mathrm{BS}}^2\mathbf{v}^H\mathbf{A}\mathbf{A}^H\mathbf{v} \geq \lambda_{\min}^+(\sigma_{\mathrm{BS}}^2\mathbf{A}\mathbf{A}^H)\|\mathbf{U}_B^H\mathbf{v}\|_2^2$

where $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{U}_B \in \mathbb{C}^{N \times N_B}$ is an orthogonal basis of the span of $\mathbf{A}\mathbf{A}^H$ (note that this matrix is generally rank-deficient). The expectations were separated since $\|\mathbf{U}_B^H \mathbf{v}\|_2^2$ is independent of the smallest non-zero eigenvalue. Furthermore, the rule $|a+b|^q \leq 2^q(|a|^q+|b|^q)$ from Hölder's inequality was applied on $|1 + d^{-1}\eta_t^{\mathrm{UE}}|^k$. Note that $\mathbf{U}_B^H \mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ with a dimension reduced from N to N_B.

The final result in (89)–(90) follows from $\mathbb{E}\{|d^{-1}\eta_t^{\mathrm{UE}}|^k\} = \left(\frac{k}{2}\right)!(\kappa_t^{\mathrm{UE}})^{\frac{k}{2}}$ and that [72, Lemma 2.10] with $m = 1$ and $n = N_B$ gives

$$\mathbb{E}\left\{\frac{1}{\|\mathbf{U}_B^H \mathbf{v}\|_2^{2q}}\right\} = \begin{cases} \frac{1}{N_B - 1}, & \text{if } q = 1, \\ \frac{1}{(N_B-1)(N_B-2)}, & \text{if } q = 2, \end{cases} \tag{92}$$

and that N_B scales linearly with N (see Section II). ∎

Appendix C
Collection of Proofs

A. Proof of Lemma 1

The DL capacity in (20) is upper bounded as

$$\mathtt{C}^{\mathrm{DL}} \leq \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \mathbb{E}\left\{\max_{\mathbf{w}(\mathbf{h}):\, \|\mathbf{w}\|_2 = 1} \log_2(1 + \mathtt{SINR}(\mathbf{w}))\right\} \tag{93}$$

where

$$\mathtt{SINR}(\mathbf{w}) = \frac{|\mathbf{h}^T \mathbf{w}|^2}{\kappa_t^{\mathrm{BS}} \sum_{i=1}^{N} |h_i w_i|^2 + \kappa_r^{\mathrm{UE}} |\mathbf{h}^T \mathbf{w}|^2 + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}}} \tag{94}$$

by assuming that the interference part of n is somehow canceled, perfect CSI is available, and exploiting the corresponding optimality of single-stream Gaussian signaling [2], [6]. We can write (94) as

$$\mathtt{SINR}(\mathbf{w}) = \frac{\mathbf{w}^H \mathbf{h}^* \mathbf{h}^T \mathbf{w}}{\mathbf{w}^H \left(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \kappa_r^{\mathrm{UE}} \mathbf{h}^* \mathbf{h}^T + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I}\right) \mathbf{w}} \tag{95}$$

by utilizing $\mathbf{w}^H \mathbf{w} = 1$. Since the logarithm is a monotonically increasing function, the maximization in (93) can be applied onto $\mathtt{SINR}(\mathbf{w})$. Using (95), this optimization is a generalized Rayleigh quotient problem and thus solved by

$$\mathbf{w} = \frac{(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \kappa_r^{\mathrm{UE}} \mathbf{h}^* \mathbf{h}^T + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}^*}{\left\| \left(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \kappa_r^{\mathrm{UE}} \mathbf{h}^* \mathbf{h}^T + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I}\right)^{-1} \mathbf{h}^* \right\|_2} \tag{96}$$

which is equivalent to (24) by using Lemma 2. The DL capacity bound in (22) follows from plugging (96) into (95) (we also took the complex conjugate of the real-valued SINR expression to make it more consistent with the UL).

The UL capacity bound in (23) follows from [6] and by assuming that the interference part of $\boldsymbol{\nu}$ is somehow canceled. We note that the uplink SINR with a receive combining vector $\mathbf{w}$ is

$$\frac{\mathbf{w}^H \mathbf{h} \mathbf{h}^H \mathbf{w}}{\mathbf{w}^H \left(\kappa_t^{\mathrm{UE}} \mathbf{h}\mathbf{h}^H + \kappa_r^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{BS}}^2}{p^{\mathrm{UE}}} \mathbf{I}\right) \mathbf{w}}. \tag{97}$$

The receiver combining vector in (25) maximizes (97) and achieves the upper bound in (23).

B. Proof of Theorem 2

The DL capacity bound in (22) can be rewritten as

$$\frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \mathbb{E}\left\{\log_2\left(1 + \frac{\mathbf{h}^H \left(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I}\right)^{-1} \mathbf{h}}{1 + \kappa_r^{\mathrm{UE}} \mathbf{h}^H \left(\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I}\right)^{-1} \mathbf{h}}\right)\right\} \tag{98}$$

using Lemma 2. This expression has the structure $m(\psi) = \log_2(1 + \frac{\psi}{1+\kappa_r^{\mathrm{UE}}\psi})$ where $\psi = \mathbf{h}^H (\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}$. Since $m(\psi)$ is a concave function of ψ, we apply Jensen's inequality to achieve a new upper bound

$$\mathtt{C}^{\mathrm{DL}} \leq \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} \mathbb{E}\{m(\psi)\} \leq \frac{T_{\mathrm{data}}^{\mathrm{DL}}}{T_{\mathrm{coher}}} m(\mathbb{E}\{\psi\}). \tag{99}$$

The upper bound in (26) follows from evaluating $\mathbb{E}\{\psi\}$ as

$$\begin{aligned} \mathbb{E}\{\psi\} &= \mathbb{E}\left\{\mathbf{h}^H (\kappa_t^{\mathrm{BS}} \mathbf{D}_{|\mathbf{h}|^2} + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}\right\} \\ &= \sum_{i=1}^{N} \mathbb{E}\left\{\frac{|h_i|^2}{\kappa_t^{\mathrm{BS}} |h_i|^2 + \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}}}\right\} = G^{\mathrm{DL}} \end{aligned} \tag{100}$$

where the expression for G^{DL} is obtained from Lemma 3 using $a = \kappa_t^{\mathrm{BS}}$ and $b = \frac{\sigma_{\mathrm{UE}}^2}{p^{\mathrm{BS}}}$.

The closed-form upper bound on the UL capacity in (27) is derived analogously to the DL capacity bound.

C. Proof of Theorem 4

We introduce the notation $\sqrt{\mathrm{tr}(\mathbf{R} - \mathbf{C})}\varphi = \frac{\vartheta}{\sqrt{\gamma}}$ where

$$\vartheta = (1 + d^{-1}\eta_t^{\mathrm{UE}}) \mathrm{tr}(\mathbf{R} - \mathbf{C}) \tag{101}$$

$$\gamma = \mathrm{tr}\left(\mathbf{A}(|d + \eta_t^{\mathrm{UE}}|^2 \mathbf{R} + \boldsymbol{\Psi}) \mathbf{A}^H\right). \tag{102}$$

Starting with the equivalence in (38), we use the rule $a^2 - b^2 = (a+b)(a-b)$ to obtain

$$\begin{aligned} &\left| \left| \mathbb{E}\left\{\frac{\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}\right\} \right|^2 - \left| \mathbb{E}\left\{\frac{\vartheta}{\sqrt{\gamma}}\right\} \right|^2 \right| \\ &= \left| \mathbb{E}\left\{\frac{\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2} - \frac{\vartheta}{\sqrt{\gamma}}\right\} \right| \left| \mathbb{E}\left\{\frac{\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2} + \frac{\vartheta}{\sqrt{\gamma}}\right\} \right| \\ &\leq \mathbb{E}\left\{\left| \frac{\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2} - \frac{\vartheta}{\sqrt{\gamma}} \right|\right\} \left(\left| \mathbb{E}\left\{\frac{\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}\right\} \right| + \left| \mathbb{E}\left\{\frac{\vartheta}{\sqrt{\gamma}}\right\} \right| \right). \end{aligned} \tag{103}$$

In order to prove the first part of the theorem, we must show that right-hand side of (103) behaves as $\mathcal{O}(\sqrt{N})$. Using Cauchy-Schwartz inequality and that $\gamma \geq \mathrm{tr}(\mathbf{A}\boldsymbol{\Psi}\mathbf{A}^H)$, we have

$$\begin{aligned} &\left| \mathbb{E}\left\{\frac{\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2}\right\} \right| + \left| \mathbb{E}\left\{\frac{\vartheta}{\sqrt{\gamma}}\right\} \right| \\ &\leq \mathbb{E}\{\|\mathbf{h}\|_2\} + \left| \mathbb{E}\left\{\frac{\vartheta}{\sqrt{\mathrm{tr}\left(\mathbf{A}\boldsymbol{\Psi}\mathbf{A}^H\right)}}\right\} \right| = \mathcal{O}(\sqrt{N}) \end{aligned} \tag{104}$$

where $\mathbb{E}\{\|\mathbf{h}\|_2\} = \mathcal{O}(\sqrt{N})$ and the second term is bounded in the same way since $\mathbb{E}\{\vartheta\} = \mathrm{tr}(\mathbf{R} - \mathbf{C}) = \mathcal{O}(N)$

and $\mathrm{tr}\big(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H\big)$ grows at least linearly with N (see the proof of Lemma 9). Hence, it remains to prove that the $\mathbb{E}\left\{\left|\mathbf{h}^H\hat{\mathbf{h}}/\|\hat{\mathbf{h}}\|_2 - \vartheta/\sqrt{\gamma}\right|\right\} = \mathcal{O}(1)$. To this end, we expand the expression using Lemma 4:

$$\begin{aligned}\mathbb{E}\left\{\left|\frac{\mathbf{h}^H\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|_2} - \frac{\vartheta}{\sqrt{\gamma}}\right|\right\} &\leq \mathbb{E}\left\{\frac{\left|\mathbf{h}^H\hat{\mathbf{h}} - \vartheta\right|}{\|\hat{\mathbf{h}}\|_2}\right\} \\ &+ \mathrm{tr}(\mathbf{R}-\mathbf{C})\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|\left|\|\hat{\mathbf{h}}\|_2 - \sqrt{\gamma}\right|}{\|\hat{\mathbf{h}}\|_2\sqrt{\gamma}}\right\}.\end{aligned} \tag{105}$$

The first term in (105) is asymptotically bounded since

$$\mathbb{E}\left\{\frac{|\mathbf{h}^H\hat{\mathbf{h}} - \vartheta|}{\|\hat{\mathbf{h}}\|_2}\right\} \leq \sqrt{\underbrace{\mathbb{E}\left\{|\mathbf{h}^H\hat{\mathbf{h}} - \vartheta|^2\right\}}_{\overset{(a)}{=}\mathcal{O}(N)}\underbrace{\mathbb{E}\left\{\frac{1}{\|\hat{\mathbf{h}}\|_2^2}\right\}}_{\overset{(b)}{=}\mathcal{O}(N^{-1})}} = \mathcal{O}(1) \tag{106}$$

where the expectation of the numerator and denominator are separated using Hölder's inequality, (a) follows from Lemma 8, and (b) from Lemma 10 with $k=0$. The second term of (105) is upper bounded by

$$\begin{aligned}&\underbrace{\frac{\mathrm{tr}(\mathbf{R}-\mathbf{C})}{=\mathcal{O}(N)}}\underbrace{\sqrt{\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^2}{\|\hat{\mathbf{h}}\|_2^2\gamma}\right\}}}_{=\mathcal{O}(N^{-1})}\underbrace{\sqrt{\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2 - \sqrt{\gamma}\right|^2\right\}}}_{=\mathcal{O}(1)} \\ &= \mathcal{O}(1)\end{aligned} \tag{107}$$

where Hölder's inequality was used to separate the expectations. The scaling of the first square root follows from Lemma 10 and that $\frac{1}{\gamma} \leq \frac{1}{\mathrm{tr}(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H)} = \mathcal{O}(N^{-1})$ for any realization of η_t^{UE}. The scaling of the second square root follows from Lemma 9. By plugging these scaling expressions into (103), we have proved (38).

Similarly, the equivalence in (39) follows if

$$\begin{aligned}&\mathbb{E}\left\{\left|\frac{|\mathbf{h}^H\hat{\mathbf{h}}|^2}{\|\hat{\mathbf{h}}\|_2^2} - \frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^2(\mathrm{tr}(\mathbf{R}-\mathbf{C}))^2}{\gamma}\right|\right\} \\ &\leq \big(\mathrm{tr}(\mathbf{R}-\mathbf{C})\big)^2\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^2\left|\|\hat{\mathbf{h}}\|_2^2 - \gamma\right|}{\|\hat{\mathbf{h}}\|_2^2\gamma}\right\} \\ &+ \mathbb{E}\left\{\frac{\left||\mathbf{h}^H\hat{\mathbf{h}}|^2 - |1+d^{-1}\eta_t^{\mathrm{UE}}|^2(\mathrm{tr}(\mathbf{R}-\mathbf{C}))^2\right|}{\|\hat{\mathbf{h}}\|_2^2}\right\}\end{aligned} \tag{108}$$

scales as $\mathcal{O}(\sqrt{N})$, where the inequality follows from Lemma 4. By applying Hölder's inequality on the first term, we obtain

$$\begin{aligned}&\underbrace{\frac{(\mathrm{tr}(\mathbf{R}-\mathbf{C}))^2}{\mathrm{tr}(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H)}}_{\overset{(a)}{=}\mathcal{O}(N)}\sqrt{\underbrace{\mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\mathrm{UE}}|^4}{\|\hat{\mathbf{h}}\|_2^4}\right\}}_{\overset{(b)}{=}\mathcal{O}(N^{-2})}}\sqrt{\underbrace{\mathbb{E}\left\{\left|\|\hat{\mathbf{h}}\|_2^2 - \gamma\right|^2\right\}}_{\overset{(c)}{=}\mathcal{O}(N)}} \\ &= \mathcal{O}\left(\sqrt{N}\right)\end{aligned} \tag{109}$$

where (a) follows from $\frac{1}{\gamma} \leq \frac{1}{\mathrm{tr}(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H)} = \mathcal{O}(N^{-1})$ which is a deterministic upper bound, (b) is characterized by Lemma 10, and (c) follows from Lemma 9. The second term behaves as

$$\begin{aligned}&\sqrt{\underbrace{\mathbb{E}\left\{\left||\mathbf{h}^H\hat{\mathbf{h}}| - |1+d^{-1}\eta_t^{\mathrm{UE}}|\mathrm{tr}(\mathbf{R}-\mathbf{C})\right|^2\right\}}_{\overset{(d)}{=}\mathcal{O}(N)}} \\ &\times \sqrt{\underbrace{\mathbb{E}\left\{\frac{2|\mathbf{h}^H\hat{\mathbf{h}}|^2}{\|\hat{\mathbf{h}}\|_2^4}\right\}}_{\overset{(e)}{=}\mathcal{O}(1)} + \underbrace{\mathbb{E}\left\{\frac{2|1+d^{-1}\eta_t^{\mathrm{UE}}|^2(\mathrm{tr}(\mathbf{R}-\mathbf{C}))^2}{\|\hat{\mathbf{h}}\|_2^4}\right\}}_{\overset{(f)}{=}\mathcal{O}(1)}} \\ &= \mathcal{O}(\sqrt{N})\end{aligned} \tag{110}$$

by using the rule $a^2 - b^2 = (a+b)(a-b)$ and Hölder's inequality. (d) is characterized by Lemma 8 and (f) by Lemma 10. Moreover, (e) follows since $\frac{|\mathbf{h}^H\hat{\mathbf{h}}|^2}{\|\hat{\mathbf{h}}\|_2^4} \leq \frac{\|\mathbf{h}\|_2^2}{\|\hat{\mathbf{h}}\|_2^2} = \mathcal{O}(1)$ by Cauchy-Schwartz inequality and that $\|\mathbf{h}\|_2^2, \|\hat{\mathbf{h}}\|_2^2$ have same asymptotic scaling. By plugging these scaling expressions into (108), we have proved (39).

Finally, the equivalence in (40) follows since

$$\begin{aligned}&\left|\sum_{i=1}^N \mathbb{E}\{|h_i|^2|v_i|^2\} - 0\right| \overset{(a)}{=} \left|\mathbb{E}\left\{\frac{\|\hat{\mathbf{h}}\|_4^4}{\|\hat{\mathbf{h}}\|_2^2}\sum_{i=1}^N |h_i|^2\frac{|\hat{h}_i|^2}{\|\hat{\mathbf{h}}\|_4^2}\right\}\right| \\ &\overset{(b)}{\leq} \left|\mathbb{E}\left\{\frac{\|\hat{\mathbf{h}}\|_4^4}{\|\hat{\mathbf{h}}\|_2^2}\sqrt{\sum_{i=1}^N |h_i|^4}\sqrt{\sum_{i=1}^N \frac{|\hat{h}_i|^4}{\|\hat{\mathbf{h}}\|_4^4}}\right\}\right| \\ &= \left|\mathbb{E}\left\{\frac{\|\hat{\mathbf{h}}\|_4^4}{\|\hat{\mathbf{h}}\|_2^2}\|\mathbf{h}\|_4^2\right\}\right| \\ &\overset{(c)}{\leq} \sqrt{\mathbb{E}\left\{\|\mathbf{h}\|_4^8\right\}\mathbb{E}\left\{\|\hat{\mathbf{h}}\|_4^8\right\}}\mathbb{E}\left\{\frac{1}{\|\hat{\mathbf{h}}\|_2^4}\right\} \overset{(d)}{=} \mathcal{O}(1)\end{aligned} \tag{111}$$

where (a) follows from $|v_i|^2 = \frac{|\hat{h}_i|^2}{\|\hat{\mathbf{h}}\|_2^2}$ and by inserting the L_4-norms $\|\hat{\mathbf{h}}\|_4^2$. The reason for this is that the vector $[|\hat{h}_1|^2 \ \ldots \ |\hat{h}_N|^2]^T/\|\hat{\mathbf{h}}\|_4^2$ now has unit L_2-norm, thus we apply Cauchy-Schwarz inequality in (b) to bound the sum by $\|\mathbf{h}\|_4^2$. Next, (c) is obtained by applying Hölder's inequality twice and (d) follows from that $\mathbb{E}\{\|\mathbf{h}\|_4^8\} = \mathcal{O}(N^2)$ and $\mathbb{E}\{\|\hat{\mathbf{h}}\|_4^8\} = \mathcal{O}(N^2)$ and $\mathbb{E}\left\{\frac{1}{\|\hat{\mathbf{h}}\|_2^4}\right\} = \mathcal{O}(N^{-2})$ from Lemma 10.

D. Proof of Theorem 5

The bounds in this theorem are derived using the capacity lower bounds in Corollaries 4 and 5. We begin with the DL and note that the arguments of the expectations in (42) have deterministic upper bounds since

$$|\varphi| \leq \sqrt{\frac{\mathrm{tr}(\mathbf{R}-\mathbf{C})}{p^{\mathrm{UE}}\mathrm{tr}(\mathbf{A}\mathbf{R}\mathbf{A}^H)}} \tag{112}$$

for any realization of η_t^{UE}. The dominated convergence theorem implies that we can take the limit $N \to \infty$ inside the expectations.[21] Next, we observe that scaling the pilot power p^{UE} proportionally to $1/N^{t_{\mathrm{UE}}}$ for some $t_{\mathrm{UE}} > 0$ means that

[21]To be strict, we first should multiply all terms in (42) by p^{UE}.

$N^{t_{\rm UE}}p^{\rm UE} \to B$ as $N \to \infty$ for some $0 < B < \infty$. Therefore, we have

$$\begin{aligned} N^{t_{\rm UE}}{\rm tr}(\mathbf{R}-\mathbf{C}) &= N^{t_{\rm UE}}p^{\rm UE}{\rm tr}\left(\mathbf{R}\bar{\mathbf{Z}}^{-1}\mathbf{R}\right) \\ &\to B{\rm tr}(\mathbf{R}(\mathbf{S}+\sigma_{\rm BS}^2\mathbf{I})^{-1}\mathbf{R}), \quad (113) \\ N^{t_{\rm UE}}{\rm tr}(\mathbf{A}(|d+\eta_t^{\rm UE}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H) &\\ \to B{\rm tr}(\mathbf{R}(\mathbf{S}+\sigma_{\rm BS}^2\mathbf{I})^{-1}\mathbf{R}), \quad (114) \end{aligned}$$

as $N \to \infty$. Using (113) and (114) and the dominated convergence theorem we obtain

$$\begin{aligned} &\lim_{N\to\infty} |\mathbb{E}\{\varphi\}|^2 \quad (115) \\ &= \left|\mathbb{E}\left\{\frac{(1+d^{-1}\eta_t^{\rm UE})\sqrt{B{\rm tr}\left(\mathbf{R}(\mathbf{S}+\sigma_{\rm BS}^2\mathbf{I})^{-1}\mathbf{R}\right)}}{\sqrt{B{\rm tr}\left(\mathbf{R}(\mathbf{S}+\sigma_{\rm BS}^2\mathbf{I})^{-1}\mathbf{R}\right)}}\right\}\right|^2 = 1 \\ &\lim_{N\to\infty} \mathbb{E}\{|\varphi|^2\} \quad (116) \\ &= \mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\rm UE}|^2 B{\rm tr}\left(\mathbf{R}(\mathbf{S}+\sigma_{\rm BS}^2\mathbf{I})^{-1}\mathbf{R}\right)}{B{\rm tr}\left(\mathbf{R}(\mathbf{S}+\sigma_{\rm BS}^2\mathbf{I})^{-1}\mathbf{R}\right)}\right\} = 1+\kappa_t^{\rm UE} \end{aligned}$$

which holds for any $t_{\rm UE} > 0$. The goal is to make the interference term $\frac{\mathbb{E}\{I_{\mathcal{H}}^{\rm UE}\}}{p^{\rm BS}{\rm tr}(\mathbf{R}-\mathbf{C})} = \frac{\mathbb{E}\{I_{\mathcal{H}}^{\rm UE}\}}{p^{\rm UE}p^{\rm BS}{\rm tr}(\mathbf{R}\bar{\mathbf{Z}}^{-1}\mathbf{R})}$ vanish asymptotically under the assumption that $\mathbb{E}\{I_{\mathcal{H}}^{\rm UE}\} = \mathcal{O}(1)$, which is achieved if the denominator grows to infinity with N. We note that ${\rm tr}(\mathbf{R}\bar{\mathbf{Z}}^{-1}\mathbf{R}) \geq \frac{\|\mathbf{R}\|^2}{\|\bar{\mathbf{Z}}\|_2}$ scales at least linearly with N. Hence, the product $p^{\rm UE}p^{\rm BS}$ must reduce at a slower pace than linear with N, which implies $t_{\rm BS}+t_{\rm UE} = t_{\rm sum} < 1$.

Finally, we need the $\mathcal{O}(1/\sqrt{N})$ terms in (51) to still vanish as $N \to \infty$. Some careful but lengthy algebra reveals that the $\mathcal{O}(N)$ properties in Lemmas 9–10 become $\mathcal{O}(N^{1-t_{\rm UE}})$. The term $\mathcal{O}(1/\sqrt{N})$ in the numerator of (51) becomes $\mathcal{O}(1/N^{\frac{1}{2}-\frac{t_{\rm UE}}{2}})$ while the $\mathcal{O}(1/\sqrt{N})$ in the denominator becomes $\mathcal{O}(1/N^{\frac{1}{2}-t_{\rm UE}})$. These terms vanish if $t_{\rm UE} < \frac{1}{2}$, which finishes the proof for the DL.

The proof for the UL is analogous since the uplink capacity bound in (44) has the same structure and contains the same expectations as the downlink capacity.

E. Proof of Corollary 7

Recall from the proof of Theorem 5 that the dominated convergence theorem can be applied, which means that we can take the limit $N \to \infty$ inside the expectations in the DL capacity bound of (42) and UL capacity bound of (44). If $\kappa_t^{\rm BS}$ and $\kappa_r^{\rm BS}$ grow with N, we obtain

$$\begin{aligned} \lim_{N\to\infty} |\mathbb{E}\{\varphi\}|^2 &= \left|\mathbb{E}\left\{\frac{(1+d^{-1}\eta_t^{\rm UE})\sqrt{{\rm tr}(\mathbf{R}\mathbf{R}_{\rm diag}^{-1}\mathbf{R})}}{\sqrt{{\rm tr}(\mathbf{R}\mathbf{R}_{\rm diag}^{-1}\mathbf{R})}}\right\}\right|^2 \\ &= 1 \quad (117) \\ \lim_{N\to\infty} \mathbb{E}\{|\varphi|^2\} &= \mathbb{E}\left\{\frac{|1+d^{-1}\eta_t^{\rm UE}|^2{\rm tr}(\mathbf{R}\mathbf{R}_{\rm diag}^{-1}\mathbf{R})}{{\rm tr}(\mathbf{R}\mathbf{R}_{\rm diag}^{-1}\mathbf{R})}\right\} \\ &= 1+\kappa_t^{\rm UE}. \quad (118) \end{aligned}$$

In the DL, we further note that

$$\frac{\kappa_t^{\rm BS}\sum_{i=1}^{N}\mathbb{E}\{|h_i|^2|v_i|^2\}}{{\rm tr}(\mathbf{R}-\mathbf{C})} = \mathcal{O}(\frac{\kappa_t^{\rm BS}\kappa_r^{\rm BS}}{N}) \quad (119)$$

since $\kappa_r^{\rm BS}{\rm tr}(\mathbf{R}-\mathbf{C}) \to {\rm tr}(\mathbf{R}\mathbf{R}_{\rm diag}^{-1}\mathbf{R}) = \mathcal{O}(N)$ as $N \to \infty$. If this term should vanish asymptotically, it is sufficient that $\frac{\kappa_t^{\rm BS}\kappa_r^{\rm BS}}{N} \to 0$ which corresponds to the condition in the corollary. The corresponding condition for the UL is obtained analogously and gives $\frac{(\kappa_r^{\rm BS})^2}{N} \to 0$.

Finally, we note that the noise terms (for $n \leq \frac{1}{2}$) and the $\mathcal{O}(\frac{1}{\sqrt{N}})$ terms in (42) and (44) all behave as $\mathcal{O}(\frac{\kappa_r^{\rm BS}}{\sqrt{N}})$ or smaller, after some straightforward but lengthy algebra. These terms thus vanish under the condition $\tau_r < \frac{1}{2}$ stated in the corollary.

F. Proof of Theorem 6

The interference expressions in (61) are proved similar to Theorem 4. For the case $l \in \mathcal{U}_{\|}$ we have

$$\begin{aligned} \mathbb{E}\left\{\left|\frac{|\mathbf{h}_l^H\hat{\mathbf{h}}|^2}{\|\hat{\mathbf{h}}\|_2^2} - \frac{p^{\rm UE}a_l^2}{\gamma}\right|\right\} &\leq \mathbb{E}\left\{\frac{\left||\mathbf{h}_l^H\hat{\mathbf{h}}|^2 - p^{\rm UE}a_l^2\right|}{\|\hat{\mathbf{h}}\|_2^2}\right\} \\ &+ \mathbb{E}\left\{\frac{p^{\rm UE}a_l^2\left|\|\hat{\mathbf{h}}\|_2^2-\gamma\right|}{\|\hat{\mathbf{h}}\|_2^2\gamma}\right\} = \mathcal{O}(\sqrt{N}) \end{aligned} \quad (120)$$

where $\gamma = {\rm tr}\left(\mathbf{A}(|d+\eta_t^{\rm UE}|^2\mathbf{R}+\mathbf{\Psi})\mathbf{A}^H\right)$ and $a_l = {\rm tr}(\mathbf{A}\mathbf{R}_l)$. This follows since the first term in (120) equals

$$\begin{aligned} &\mathbb{E}\left\{\frac{\left||\mathbf{h}_l^H\hat{\mathbf{h}}| - \sqrt{p^{\rm UE}}a_l\right|\left||\mathbf{h}_l^H\hat{\mathbf{h}}| + \sqrt{p^{\rm UE}}a_l\right|}{\|\hat{\mathbf{h}}\|_2^2}\right\} \\ &\leq \underbrace{\sqrt{\mathbb{E}\left\{\left||\mathbf{h}_l^H\hat{\mathbf{h}}| - \sqrt{p^{\rm UE}}a_l\right|^2\right\}}}_{=\mathcal{O}(\sqrt{N})}\underbrace{\sqrt{\mathbb{E}\left\{\frac{\|\mathbf{h}_l\|_2^2}{\|\hat{\mathbf{h}}\|_2^2} + \frac{p^{\rm UE}a_l^2}{\|\hat{\mathbf{h}}\|_2^4}\right\}}}_{=\mathcal{O}(1)} \\ &= \mathcal{O}(\sqrt{N}) \end{aligned} \quad (121)$$

by using Hölder's inequality, Lemma 6, Cauchy-Schwartz inequality, and Lemma 10. The second term in (120) is also upper bounded by $\mathcal{O}(\sqrt{N})$ by using Hölder's inequality and that $a_l = \mathcal{O}(N)$, $\mathbb{E}\{\left|\|\hat{\mathbf{h}}\|_2^2 - \gamma\right|^2\} = \mathcal{O}(N)$ from Lemma 9, $\mathbb{E}\{\|\hat{\mathbf{h}}\|_2^{-4}\} = \mathcal{O}(N^{-2})$ from Lemma 10, and $\frac{1}{\gamma} \leq \frac{1}{{\rm tr}(\mathbf{A}\mathbf{\Psi}\mathbf{A}^H)} = \mathcal{O}(N^{-1})$.

Next, the case $l \in \mathcal{U}_{\perp}$ in (61) follows from $\mathbb{E}\left\{|\mathbf{h}_l^H\mathbf{v}^{\rm UL}|^2\right\} = \mathbb{E}\left\{(\mathbf{v}^{\rm UL})^H\mathbf{R}_l\mathbf{v}^{\rm UL}\right\} \leq \|\mathbf{R}_l\|_2 = \mathcal{O}(1)$ since $\mathbf{h}_l$ and $\mathbf{v}^{\rm UL}$ are independent.

Finally, we note that the noise term in the denominator of (44) would be

$$\frac{\mathbb{E}\{\mathbf{v}^H\mathbf{Q}_{\mathcal{H}}\mathbf{v}\}}{p^{\rm UE}{\rm tr}(\mathbf{R}-\mathbf{C})} = \sum_{l\in\mathcal{U}_{\|}}\frac{p^{\rm UE}\mathbb{E}\left\{|\mathbf{h}_l^H\mathbf{v}|^2\right\}}{p^{\rm UE}{\rm tr}(\mathbf{R}-\mathbf{C})} + \mathcal{O}\left(\sqrt{\frac{1}{N}}\right) \quad (122)$$

where the first term is equal to (62) by exploiting (61) and ${\rm tr}(\mathbf{R}-\mathbf{C}) = \sqrt{p^{\rm UE}}{\rm tr}(\mathbf{A}\mathbf{R})$.

ACKNOWLEDGMENT

The authors would like to thank Romain Couillet and the anonymous reviewers for indispensable feedback on this paper.

REFERENCES

[1] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Hardware impairments in large-scale MISO systems: Energy efficiency, estimation, and capacity limits," in *Proc. Int. Conf. Digital Signal Process. (DSP)*, 2013.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

[3] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE VTC Fall*, 2011.

[4] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *Int. Symp. Wireless Commun. Systems (ISWCS)*, 2012.

[5] 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, 2000.

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

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

[8] A. Müller, A. Kammoun, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO," *IEEE Trans. Inf. Theory*, submitted. [Online]. Available: http://arxiv.org/abs/1310.1806

[9] F. Rusek, D. Persson, B. Lau, E. Larsson, T. 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, 2013.

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

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

[12] 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, 2013.

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

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

[15] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *Proc. ITG/IEEE Workshop on Smart Antennas (WSA)*, 2010.

[16] P. Zetterberg, "Experimental investigation of TDD reciprocity-based zero-forcing transmit precoding," *EURASIP J. on Adv. in Signal Process.*, Jan. 2011.

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

[18] B. Göransson, S. Grant, E. Larsson, and Z. Feng, "Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA," in *Proc. IEEE Int. W. Signal Process. Adv. Wireless Commun. (SPAWC)*, 2008.

[19] E. Björnson, P. Zetterberg, and M. Bengtsson, "Optimal coordinated beamforming in the multicell downlink with transceiver impairments," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2012.

[20] 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.

[21] A. Mezghani, N. Damak, and J. A. Nossek, "Circuit aware design of power-efficient short range communication systems," in *Proc. Int. Symp. Wireless Commun. Systems (ISWCS)*, 2010, pp. 869–873.

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

[23] B. Maham and O. Tirkkonen, "Transmit antenna selection OFDM systems with transceiver I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 865–871, 2012.

[24] T. Koch, A. Lapidoth, and P. Sotiriadis, "Channels that heat up," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3594–3612, 2009.

[25] A. Pitarokoilis, S. Mohammed, and E. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise," *IEEE Trans. Wireless Commun.*, submitted. [Online]. Available: http://arxiv.org/abs/1306.4495

[26] C. Studer and E. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 303–313, 2013.

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

[28] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[29] E. Björnson, M. Kountouris, M. Bengtsson, and B. Ottersten, "Receive combining vs. multi-stream multiplexing in downlink systems with multi-antenna users," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3431–3446, 2013.

[30] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, 2005.

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

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

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

[34] 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, 2010.

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

[36] N. O'Donoughue and J. Moura, "On the product of independent complex Gaussians," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1050–1063, 2012.

[37] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, 2011.

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

[39] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.

[40] E. Björnson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4027–4041, 2009.

[41] 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, 2006.

[42] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, 2000.

[43] 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, 2013.

[44] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[45] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[46] R. Couillet, F. Pascal, and J. W. Silverstein, "Robust estimates of covariance matrices in the large dimensional regime and application to array processing," *IEEE Trans. Inf. Theory*, to appear. [Online]. Available: http://arxiv.org/abs/1204.5320

[47] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, 2014, to appear.

[48] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving "massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.

[49] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, 1999.

[50] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, 2012.

[51] J. Xu, L. Qiu, and C. Yu, "Improving energy efficiency through multimode transmission in the downlink MIMO systems," *EURASIP J. Wirel. Commun. Netw.*, 2011.

[52] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, 2012.

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

[54] E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. Int. Conf. Telecommun. (ICT)*, 2013.

[55] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, 1990.

[56] G. Auer, O. Blume, V. Giannini, I. Godor, M. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, and W. Wajda, *D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown*. INFSO-ICT-247733 EARTH, ver. 2.0, 2012.

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

[58] C. Studer, M. Wenk, and A. Burg, "System-level implications of residual transmit-RF impairments in MIMO systems," in *Proc. European Conf. Antennas and Propagation (EuCAP)*, 2011.

[59] N. Moghadam, P. Zetterberg, P. Händel, and H. Hjalmarsson, "Correlation of distortion noise between the branches of MIMO transmit antennas," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2012.

[60] A. Mezghani, M.-S. Khoufi, and J. Nossek, "A modified MMSE receiver for quantized MIMO systems," in *Proc. ITG/IEEE Workshop on Smart Antennas (WSA)*, 2007.

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

[62] 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, 2007.

[63] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of MIMO Wiener phase-noise channels," in *Proc. Information Theory and Applications Workshop (ITA)*, 2013.

[64] H. Mehrpouyan, A. Nasir, S. Blostein, T. Eriksson, G. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, 2012.

[65] Y. Nasser, M. Noes, L. Ros, and G. Jourdain, "On the system level prediction of joint time frequency spreading systems with carrier phase noise," *IEEE Trans. Commun.*, vol. 58, no. 3, pp. 839–850, 2010.

[66] 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, 2013.

[67] C. Shepard, H. Yu, N. Anand, L. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM MobiCom*, 2012.

[68] K. Nishimori, K. Cho, Y. Takatori, and T. Hori, "Automatic calibration method using transmitting signals of an adaptive array for TDD systems," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1636–1640, 2001.

[69] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, and A. Molisch, "Downlink beamforming avoiding DOA estimation for cellular mobile communications," in *Proc. Information Theory and Applications Workshop (ITA)*, 2013.

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

[71] Z. Bai and J. Silverstein, *Spectral analysis of large dimensional random matrices*, 2nd ed. Springer Series in Statistics, 2009.

[72] A. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.