Postprint

Hardware Impairments in Large-scale MISO Systems: Energy Efficiency, Estimation, and Capacity Limits

Emil Björnson*†, Jakob Hoydis§, Marios Kountouris‡, and Mérouane Debbah*

*Alcatel-Lucent Chair on Flexible Radio, SUPELEC, Gif-sur-Yvette, France ({emil.bjornson, merouane.debbah}@supelec.fr)
†ACCESS Linnaeus Centre, Signal Processing Lab, KTH Royal Institute of Technology, Stockholm, Sweden
‡Department of Telecommunications, SUPELEC, Gif-sur-Yvette, France (marios.kountouris@supelec.fr)
§Bell Laboratories, Alcatel-Lucent, Stuttgart, Germany (jakob.hoydis@alcatel-lucent.com)

Abstract—The use of large-scale antenna arrays has the potential to bring substantial improvements in energy efficiency and/or spectral efficiency to future wireless systems, due to the greatly improved spatial beamforming resolution. Recent asymptotic results show that by increasing the number of antennas one can achieve a large array gain and at the same time naturally decorrelate the user channels; thus, the available energy can be focused very accurately at the intended destinations without causing much inter-user interference. Since these results rely on asymptotics, it is important to investigate whether the conventional system models are still reasonable in the asymptotic regimes. This paper analyzes the fundamental limits of large-scale multiple-input single-output (MISO) communication systems using a generalized system model that accounts for transceiver hardware impairments. As opposed to the case of ideal hardware, we show that these practical impairments create finite ceilings on the estimation accuracy and capacity of large-scale MISO systems. Surprisingly, the performance is only limited by the hardware at the single-antenna user terminal, while the impact of impairments at the large-scale array vanishes asymptotically. Furthermore, we show that an arbitrarily high energy efficiency can be achieved by reducing the power while increasing the number of antennas.

I. Introduction

The spectral efficiency in wireless communications is not only limited by the information-theoretic capacity [1], but also by practical issues such as propagation environment, channel estimation accuracy [2], transceiver hardware impairments [3], and signal processing complexity [4]. It is of profound importance to increase the spectral efficiency in 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 taking care of the aforementioned practical issues. It is known as large-scale multiple-input multiple-output (MIMO), or "massive MIMO", and is based on having a large number of antennas at the base stations and exploiting channel reciprocity in time-division duplex (TDD) mode [4]–[7]. Some key features are: 1) propagation losses are mitigated by a large array gain due to coherent beamforming; 2) impact of channel estimation errors vanishes asymptotically in the large-dimensional space; 3) low-complexity signal processing algorithms are asymptotically optimal; and 4) inter-user interference is easily mitigated by the high beamforming resolution.
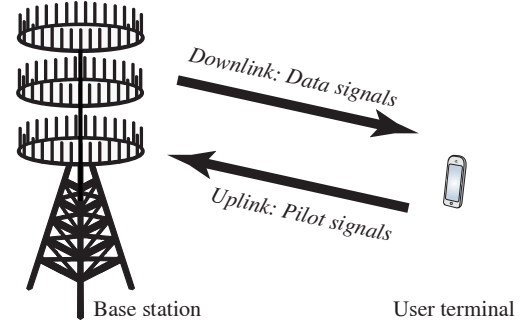
Fig. 1. Example of a large-scale MISO system where the base station has 81 antenna units and can feature $N = 162$ antennas using dual polarization.

The impact of transceiver hardware impairments on large-scale MIMO has received little attention, although large arrays might only be economically feasible if inexpensive hardware can be used at each antenna unit. Cheap hardware components are particularly prone to the impairments (e.g., amplifier non-linearities, I/Q-imbalance, and phase noise) that exist in any physical transceiver implementation [8]–[11]. Transceiver impairments fundamentally limit the capacity in the high-power regime [3], while their impact in the large-antenna regime is less known. To the best of our knowledge, the only previous work in this area is [12] which derives lower bounds on the achievable uplink sum rate for systems with phase noise.

This paper analyzes the *aggregate* impact of hardware impairments on systems with large antenna arrays, in contrast to the ideal hardware considered in [5]–[7] and the single type of impairments in [12]. Section II reviews the generalized system model with hardware impairments from [8]–[11]. Section III derives a new pilot-based channel estimator and shows that the estimation accuracy is limited by the level of impairments. Section IV derives lower and upper bounds on the downlink capacity and shows that there exists a finite capacity ceiling in large-scale MISO systems with impairments. Despite these discouraging results, Section V shows that a very high energy efficiency and resilience towards hardware impairments at the base station can be achieved. For brevity, we consider a single-user system, but extensions to multi-cell scenarios (e.g., with pilot contamination) are outlined throughout the manuscript.

II. System Model

This paper analyzes the fundamental spectral and energy efficiency limits of a communication system with an N-antenna base station (BS) and a single-antenna user terminal (UT). A main characteristic in the analysis will be that the number of antennas N can be very large. To avoid infeasible overhead signaling, we consider a TDD system where channel

state information (CSI) is obtained by pilot signaling in the uplink. The acquired CSI will be utilized for downlink data transmission, by exploiting channel reciprocity; see Fig. 1.

The flat-fading channel between the base station and the user terminal is denoted $\mathbf{h} \in \mathbb{C}^{N \times 1}$. This stochastic channel has a static realization for the duration of a data block and independent realizations between blocks. Each realization is circularly-symmetric complex Gaussian: $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$. This is known as Rayleigh fading and the positive semi-definite covariance matrix is $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\} = \mathbf{R} \in \mathbb{C}^{N \times N}$, where $\mathbb{E}\{\cdot\}$ denotes expectation. We assume that the spectral norm of $\mathbf{R}$ is uniformly bounded, irrespective of the number of antennas N. This is a common technical assumption that enables asymptotic analysis (cf. [6]), but which is also a necessary physical property that originates from the law of energy conservation.

Physical transceivers suffer from hardware impairments that 1) create a mismatch between the intended data signal and what is actually generated and emitted; and 2) distort the received signal in the reception processing. This is modeled herein by a generalized channel model from [8]–[11] with additive distortion noises at the transmitter and the receiver.

A. Generalized Downlink Channel Model

The downlink channel is used for data transmission; see Fig. 1. The received downlink signal $y \in \mathbb{C}$ in flat-fading MISO systems is conventionally modeled as

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

where s is the random zero-mean data signal with power $\mathbb{E}\{|s|^2\} = p^{\mathrm{BS}}$ and $\mathbf{w} \in \mathbb{C}^{N \times 1}$ is the unit-norm beamforming vector. The receiver noise is modeled as $n \sim \mathcal{CN}(0, \sigma_{\mathrm{UT}}^2)$ and might include interference from neighboring systems.

To model physical channels more accurately, this paper considers a generalized model [8]–[11] where the received signal is

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

The difference from the conventional model (1) is the additional distortion noise terms $\boldsymbol{\eta}_t^{\mathrm{BS}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon}_t^{\mathrm{BS}})$ and $\eta_r^{\mathrm{UT}} \sim \mathcal{CN}(0, \upsilon_r^{\mathrm{UT}})$. These are independent and describe the impact of impairments in the transmitter and receiver hardware, respectively [8]. Theoretical investigations and several measurements have suggested that

$$\boldsymbol{\Upsilon}_t^{\mathrm{BS}} = \kappa_t^{\mathrm{BS}} p^{\mathrm{BS}} \operatorname{diag}(|w_1|^2, \ldots, |w_N|^2)$$
$$\upsilon_r^{\mathrm{UT}} = \kappa_r^{\mathrm{UT}} p^{\mathrm{BS}} |\mathbf{h}^T \mathbf{w}|^2 \tag{3}$$

where w_i is the ith element of $\mathbf{w}$ and $\operatorname{diag}(\cdot, \ldots, \cdot)$ denotes a diagonal matrix. The Gaussianity is explained by the aggregate effect of many impairments and the use of compensation algorithms that remove other types of distortions [9], [10].

Remark 1. *Distortion noise is an alteration of the data signal, while the classical receiver noise models random fluctuations in the electronic circuits at the receiver. A main difference is thus that the average distortion noise power is proportional to the signal power p^{BS} and the current channel gain $\|\mathbf{h}\|^2$.[1] The proportionality constants $\kappa_t^{\mathrm{BS}}, \kappa_r^{\mathrm{UT}} \geq 0$ characterize the* level

of impairments *and are typically in the range* $[0, 0.03]$.[2] *Small values represent accurate and expensive transceiver hardware.*

B. Generalized Uplink Channel Model

The reciprocal uplink channel will be used for pilot-based channel estimation; see Fig. 1. Similar to (2), we consider a generalized model with the received signal $\mathbf{z} \in \mathbb{C}^N$ being

$$\mathbf{z} = \mathbf{h}(d + \eta_t^{\mathrm{UT}}) + \boldsymbol{\nu} + \boldsymbol{\eta}_r^{\mathrm{BS}} \tag{4}$$

where $d \in \mathbb{C}$ is a scalar deterministic pilot signal with power $p^{\mathrm{UT}} = |d|^2$. The independent additive noise $\boldsymbol{\nu} \sim \mathcal{CN}(\mathbf{0}, \mathbf{S})$ has a positive definite covariance matrix $\mathbf{S} \in \mathbb{C}^{N \times N}$. This term contains both receiver noise and potential interference from neighboring systems (e.g., pilot contamination [4]–[7]). We assume that $\mathbf{S}$ has a uniformly bounded spectral norm as $N \to \infty$, for the same physical reasons as for $\mathbf{R}$.

The transceiver impairments in the hardware used for uplink transmission are modeled by the independent distortion noises $\eta_t^{\mathrm{UT}} \sim \mathcal{CN}(0, \upsilon_t^{\mathrm{UT}})$ and $\boldsymbol{\eta}_r^{\mathrm{BS}} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon}_r^{\mathrm{BS}})$ at the transmitter and receiver, respectively. Similar to (3), their covariance matrices are model as

$$\boldsymbol{\Upsilon}_r^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}} p^{\mathrm{UT}} \operatorname{diag}(|h_1|^2, \ldots, |h_N|^2)$$
$$\upsilon_t^{\mathrm{UT}} = \kappa_t^{\mathrm{UT}} p^{\mathrm{UT}}. \tag{5}$$

Note that we might have $\kappa_t^{\mathrm{BS}} \neq \kappa_r^{\mathrm{BS}}$ at the base station and $\kappa_r^{\mathrm{UT}} \neq \kappa_t^{\mathrm{UT}}$ at the user terminal since different pieces of hardware are generally used for transmission and reception.

III. UPLINK CHANNEL ESTIMATION

This section considers estimation of the current channel realization $\mathbf{h}$ by comparing the received uplink signal $\mathbf{z}$ in (4) with the predefined pilot signal d. The classic results on pilot-based channel estimation consider Rayleigh fading channels that are observed in independent complex Gaussian noise with known statistics [13], [14]. However, this is not the case herein because the distortion noises η_t^{UT} and $\boldsymbol{\eta}_r^{\mathrm{BS}}$ depend on the unknown random channel $\mathbf{h}$. The dependence is either through the multiplication $\mathbf{h}\eta_t^{\mathrm{UT}}$ or the conditional variance of $\boldsymbol{\eta}_r^{\mathrm{BS}}$ in (5), which is essentially the same relation. Although the distortion noises are Gaussian when conditioning on a channel realization, the effective distortion is the product of Gaussian variables and, thus, has a *complex double Gaussian* distribution [15]. Consequently, an optimal channel estimator cannot be deduced from the standard results provided in [13], [14].

We will now derive the *linear* minimum mean square error (LMMSE) estimator. Note that $()^*$ denotes conjugation.

Theorem 1. *The LMMSE estimator of $\mathbf{h}$ from the observation of $\mathbf{z}$ in (4) is*

$$\widehat{\mathbf{h}} = d^* \mathbf{R} \left(p^{\mathrm{UT}}(1 + \kappa_t^{\mathrm{UT}}) \mathbf{R} + p^{\mathrm{UT}} \kappa_r^{\mathrm{BS}} \mathbf{R}_{\mathrm{diag}} + \mathbf{S} \right)^{-1} \mathbf{z} \tag{6}$$

where $\mathbf{R}_{\mathrm{diag}} = \operatorname{diag}(r_{11}, \ldots, r_{NN})$ consists of the diagonal elements of $\mathbf{R}$. The total MSE $= \mathbb{E}\{\|\widehat{\mathbf{h}} - \mathbf{h}\|^2\}$ is $\operatorname{tr}(\mathbf{C})$, where $\mathbf{C} = \mathbb{E}\{(\widehat{\mathbf{h}} - \mathbf{h})(\widehat{\mathbf{h}} - \mathbf{h})^H\}$ is the error covariance matrix:

$$\mathbf{C} = \mathbf{R} - p^{\mathrm{UT}} \mathbf{R} \left(p^{\mathrm{UT}}(1 + \kappa_t^{\mathrm{UT}}) \mathbf{R} + p^{\mathrm{UT}} \kappa_r^{\mathrm{BS}} \mathbf{R}_{\mathrm{diag}} + \mathbf{S} \right)^{-1} \mathbf{R}. \tag{7}$$

---

[1] The total transmit power is $p^{\mathrm{BS}}(1 + \kappa_t^{\mathrm{BS}})$ under hardware impairments and not the usual p^{BS}, since distortions also contribute a small amount of power. For simplicity, we will still refer to the signal power p^{BS} as the transmit power since the total power is fully characterized by and is very close to p^{BS}.

[2] The square root of each κ-parameter equals the error vector magnitude (EVM), which is a common quality measure of transceivers.

Proof: The LMMSE estimator has the form $\widehat{\mathbf{h}} = \mathbf{A}\mathbf{z}$ where $\mathbf{A}$ minimizes MSE $= \mathbb{E}\{\|\widehat{\mathbf{h}} - \mathbf{h}\|^2\}$. An expression is obtained by taking expectations over $\eta_t^{\mathrm{UT}}, \boldsymbol{\eta}_r^{\mathrm{BS}}$ for an arbitrary fixed $\mathbf{h}$ and then taking the expectation over $\mathbf{h}$. The LMMSE estimator in (6) follows from straightforward differentiation and $\mathbf{C}$ and the MSE then follow from the definitions. ∎

We remark that there might exist non-linear estimators that outperform 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 [14, Remark 1]. 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* $\mathbf{S} = \sigma_{\mathrm{BS}}^2\mathbf{I}$. *The error covariance matrix in* (7) *becomes*

$$\mathbf{C} = \lambda\left(1 - \frac{p^{\mathrm{UT}}\lambda}{p^{\mathrm{UT}}\lambda(1 + \kappa_t^{\mathrm{UT}} + \kappa_r^{\mathrm{BS}}) + \sigma_{\mathrm{BS}}^2}\right)\mathbf{I}. \quad (8)$$

In the high uplink power[3] *regime, we have*

$$\lim_{p^{\mathrm{UT}}\to\infty} \mathbf{C} = \lambda\left(1 - \frac{1}{1 + \kappa_t^{\mathrm{UT}} + \kappa_r^{\mathrm{BS}}}\right)\mathbf{I}. \quad (9)$$

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 (8). It is independent of the number of antennas N, thus letting N grow large will neither increase nor decrease the estimation error *per element*.[4] The estimation error is clearly a decreasing function of the pilot power p^{UT}, however the error variance will *not* converge to zero as the pilot power is increased. As seen in (9), there is a strictly positive error floor of $\lambda(1 - \frac{1}{1+\kappa_t^{\mathrm{UT}}+\kappa_r^{\mathrm{BS}}})$ 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 $\kappa_t^{\mathrm{UT}}, \kappa_r^{\mathrm{BS}}$ in the transmitter and receiver hardware, respectively. In terms of estimation accuracy, it is therefore equally important to have high-quality hardware at both the base station and the user terminal.

There will be an error floor also for non-diagonal $\mathbf{R}$ and $\mathbf{S}$; the general high-power limit is easily computed from (7).

The channel can be decomposed as $\mathbf{h} = \widehat{\mathbf{h}} + \boldsymbol{\epsilon}$, where $\widehat{\mathbf{h}}$ is the LMMSE estimate in Theorem 1 and $\boldsymbol{\epsilon} \in \mathbb{C}^{N\times 1}$ denotes the unknown estimation error. Contrary to conventional estimation with independent noise (cf. [13, Chapter 15.8]), $\widehat{\mathbf{h}}$ and $\boldsymbol{\epsilon}$ are neither independent nor jointly complex Gaussian, but only uncorrelated and zero-mean. The covariance matrices are $\mathbb{E}\{\widehat{\mathbf{h}}\widehat{\mathbf{h}}^H\} = \mathbf{R} - \mathbf{C}$ and $\mathbb{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H\} = \mathbf{C}$ with $\mathbf{C}$ given in (7).

A. Numerical Illustration

The estimation accuracy of the LMMSE estimator under hardware impairments is illustrated in Fig. 2 with $N = 10$

---
[3]The level of impairments in the transmitter hardware increases with the transmit power [11]. This means that $\kappa_t^{\mathrm{BS}}, \kappa_t^{\mathrm{UT}}$ are almost constant within the dynamic range of the power amplifier, but increase rapidly when moving outside this range. This is not taken into account, since the high-power regime is not our main focus. Consequently, the high-power limits derived herein are optimistic and might not be achievable in practice. The results when $N \to \infty$ are however accurate since this means that less power is allocated per antenna.

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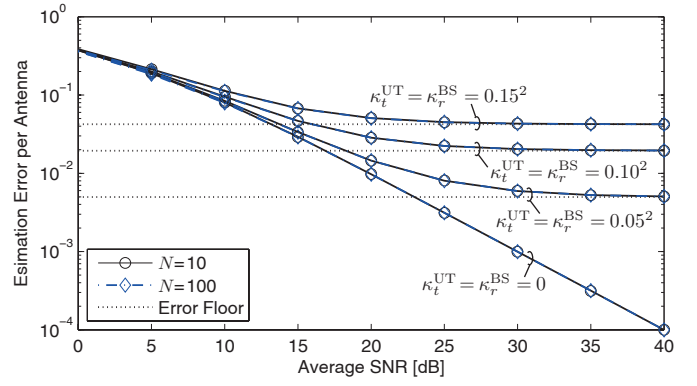

Fig. 2. Estimation error per antenna element for the LMMSE estimator in Theorem 1. Transceiver hardware impairments create a non-zero error floor.

and $N = 100$ antennas. The channel covariance matrices $\mathbf{R}$ are generated using the exponential model with correlation coefficient 0.7 [16], while the noise covariance is $\mathbf{S} = \mathbf{I}$. Fig. 2 shows the (normalized) estimation error per channel element as a function of the average SNR, defined as $p^{\mathrm{UT}}\frac{\mathrm{tr}(\mathbf{R})}{\mathrm{tr}(\mathbf{S})}$.

Four hardware setups with different levels of impairments are considered: $\kappa_t^{\mathrm{UT}} = \kappa_r^{\mathrm{BS}} \in \{0, 0.05^2, 0.10^2, 0.15^2\}$. Fig. 2 confirms that there are non-zero error floors at high SNRs, as proved by Corollary 1. The error floor increases when increasing the level 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. Moreover, the curves for $N = 10$ and $N = 100$ are virtually the same, showing that the error *per* antenna is almost independent of N even under high spatial channel correlation.

IV. DOWNLINK DATA TRANSMISSION

This section analyzes the capacity of the downlink channel in (2), where a linear beamforming vector $\mathbf{w}$ is applied to transform the MISO channel $\mathbf{h}$ into an effective single-input single-output (SISO) channel $\mathbf{h}^T\mathbf{w} \in \mathbb{C}$. The beamforming vector is a function of the estimate $\widehat{\mathbf{h}}$ obtained at the base station and the statistical properties of other channel parameters.

The receiver has a statistical characterization of the effective channel $\mathbf{h}^T\mathbf{w}$, obtained from the long-term statistics and possibly some additional downlink pilot signaling. We leave the extent of CSI at the user terminal undefined since the bounds derived in this section hold for anything between no downlink pilot signaling and perfect CSI estimation.

Under these conditions, the capacity of the generalized memoryless fading downlink channel in (2) is

$$\mathsf{C} = \mathbb{E}\left\{\max_{\mathbf{w}(\widehat{\mathbf{h}}): \|\mathbf{w}\|=1} \quad \max_{f(s): \mathbb{E}\{|s|^2\}\leq p^{\mathrm{BS}}} \mathcal{I}(s; y)\right\} \quad (10)$$

where $\mathcal{I}(s; y)$ denotes the mutual information between the received signal y and data signal s. The beamforming vector $\mathbf{w}(\widehat{\mathbf{h}})$ can be any deterministic function of $\widehat{\mathbf{h}}$, while $f(s)$ denotes the probability distribution of the data signal. The expression in (10) combines the well-known capacity formulas from [17] with a maximization over $\mathbf{w}(\widehat{\mathbf{h}})$. The expectation is taken over different data blocks with static values on $\mathbf{h}, \widehat{\mathbf{h}}$.

We will characterize the capacity by deriving lower and upper bounds on (10) and study their behavior in the limit of

infinitely many transmit antennas N. This brings insights on how hardware impairments affect large-scale MISO systems.

A. Upper Bound on Channel Capacity

An upper bound on (10) can be achieved by adding side information at the transmitter and receiver. We assume that the channel realization $\mathbf{h}$ is known perfectly at the transmitter and receiver in each fading block. Since the receiver and distortion noise in (2) are complex Gaussian and independent of the useful signal under perfect CSI, we deduce that Gaussian signaling, $s \sim \mathcal{CN}(0, p^{\mathrm{BS}})$, is optimal [1], [3] and obtain

$$\mathsf{C} \leq \mathbb{E}\left\{\max_{\mathbf{w}(\mathbf{h}) \,:\, \|\mathbf{w}\|=1} \log_2(1 + \mathrm{SINR}(\mathbf{w}))\right\} \quad \text{where} \quad (11)$$

$$\mathrm{SINR}(\mathbf{w}) = \frac{|\mathbf{h}^T \mathbf{w}|^2}{\kappa_t^{\mathrm{BS}} \sum\limits_{i=1}^{N} |h_i w_i|^2 + \kappa_r^{\mathrm{UT}} |\mathbf{h}^T \mathbf{w}|^2 + \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}}}}. \quad (12)$$

To find an upper bound in closed form, we first need to find the optimal beamforming vector in the current upper bound.

Lemma 1. *The SINR in* (12) *is maximized as*

$$\max_{\mathbf{w} \,:\, \|\mathbf{w}\|=1} \mathrm{SINR}(\mathbf{w}) = \mathbf{h}^T\left(\kappa_t^{\mathrm{BS}} \mathbf{D_h} + \kappa_r^{\mathrm{UT}} \mathbf{h}^* \mathbf{h}^T + \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}}} \mathbf{I}\right)^{-1} \mathbf{h}^* \quad (13)$$

for $\mathbf{D_h} = \mathrm{diag}(|h_1|^2, \ldots, |h_N|^2)$. *The optimum is attained by*

$$\mathbf{w} = \frac{(\kappa_t^{\mathrm{BS}} \mathbf{D_h} + \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}^*}{\left\|(\kappa_t^{\mathrm{BS}} \mathbf{D_h} + \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}^*\right\|}. \quad (14)$$

Proof: The SINR can be rewritten as a generalized Rayleigh quotient and is therefore maximized by (14). ∎

Note that the beamforming vector $\mathbf{w}$ in (14) only depends on impairments at the base station. Lemma 1 enables us to derive a closed-form upper bound on the channel capacity.

Theorem 2. *The capacity in* (10) *is upper bounded as*

$$\mathsf{C} \leq \mathsf{C}_{\mathrm{upper}} = \log_2\left(1 + \frac{G}{1 + \kappa_r^{\mathrm{UT}} G}\right) \quad (15)$$

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

$$G = \sum_{i=1}^{N} \frac{1}{\kappa_t^{\mathrm{BS}}}\left(1 - \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}} \kappa_t^{\mathrm{BS}} r_{ii}} E_1\left(\frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}} \kappa_t^{\mathrm{BS}} r_{ii}}\right) e^{\frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}} \kappa_t^{\mathrm{BS}} r_{ii}}}\right), \quad (16)$$

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

Proof: Using Lemma 1 and the Woodbury identity, the upper bound in (11) can be expressed as $\mathbb{E}\{m(\psi)\} = \mathbb{E}\{\log_2(1 + \frac{\psi}{1 + \kappa_r^{\mathrm{UT}} \psi})\}$, where $\psi = \mathbf{h}^T(\kappa_t^{\mathrm{BS}} \mathbf{D_h} + \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}}} \mathbf{I})^{-1} \mathbf{h}^*$. Since $m(\cdot)$ is a concave function, we apply Jensen's inequality to obtain a new upper bound $\mathbb{E}\{m(\psi)\} \leq m(\mathbb{E}\{\psi\})$. Finally, (15) and (16) are obtained by evaluating $\mathbb{E}\{\psi\}$. ∎

This closed-form upper bound provides important insights on the achievable performance under hardware impairments.

Corollary 2. *The upper capacity bound in* (15) *has the following asymptotic properties:*

$$\lim_{p^{\mathrm{BS}} \to \infty} \mathsf{C}_{\mathrm{upper}} = \log_2\left(1 + \frac{N}{\kappa_t^{\mathrm{BS}} + \kappa_r^{\mathrm{UT}} N}\right) \quad (17)$$

$$\lim_{N \to \infty} \mathsf{C}_{\mathrm{upper}} = \log_2\left(1 + \frac{1}{\kappa_r^{\mathrm{UT}}}\right). \quad (18)$$

Proof: Note that $r_1 > 0, \ldots, r_N > 0$ since $\mathbf{R}$ is positive definite. This implies $G \to \sum_{i=1}^{N} \frac{1}{\kappa_t^{\mathrm{BS}}} = \frac{N}{\kappa_t^{\mathrm{BS}}}$ as $p^{\mathrm{BS}} \to \infty$ for fixed N, giving (17). It also implies $\frac{1}{N} G > 0$ as $N \to \infty$, thus $\frac{G}{1 + \kappa_r^{\mathrm{UT}} G} - \frac{G}{\kappa_r^{\mathrm{UT}} G} \to 0$ as $N \to \infty$ giving (18). ∎

This corollary shows that the capacity has finite ceilings both when the downlink transmit power p^{BS} grows large and when the number of antennas N grows large. The ceilings depend on the impairment parameters $\kappa_t^{\mathrm{BS}}, \kappa_r^{\mathrm{UT}}$ and the UT 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 perfect transceiver hardware (cf. [4]–[7]) is not easily achieved in practice.

Interestingly, (18) indicates that only the quality of the terminal's receiver hardware limits the capacity as $N \to \infty$. This means that the detrimental effect of hardware impairments in the base station disappears when the number of base station antennas grow large. This is, simply speaking, since the distortion noise is spread isotropically in space while the increased spatial resolution of the array enables very exact transmit beamforming. This is a very promising result since large arrays are more prone to impairments, due to economic and implementation limitations. To verify this observation, we also need a lower bound on the channel capacity.

B. Lower Bound on Channel Capacity

We obtain a lower capacity bound by making the potentially limiting assumptions of Gaussian codebook, no downlink pilot signaling, and Gaussian CSI uncertainty at the terminal. The next theorem is obtained by an approach from [2], which has been applied to massive MIMO in [5]–[7] (among others).

Theorem 3. *The capacity in* (10) *is lower bounded as*

$$\mathsf{C} \geq \mathsf{C}_{\mathrm{lower}} = \log_2\left(1 + \widetilde{\mathrm{SINR}}\right) \quad (19)$$

where $\mathbf{v} = [v_1 \ \ldots \ v_{K_r}]^T$ *with* $\|\mathbf{v}\| = 1$ *is a function of* $\hat{\mathbf{h}}$ *and*

$$\widetilde{\mathrm{SINR}} = \frac{\left|\mathbb{E}\{\mathbf{h}^T \mathbf{v}\}\right|^2}{(1 + \kappa_r^{\mathrm{UT}})\mathbb{E}\left\{|\mathbf{h}^T \mathbf{v}|^2\right\} - |\mathbb{E}\{\mathbf{h}^T \mathbf{v}\}|^2 + \kappa_t^{\mathrm{BS}} \sum\limits_{i=1}^{N} \mathbb{E}\{|h_i|^2 |v_i|^2\} + \frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{BS}}}}. \quad (20)$$

Proof: Any heuristic beamforming $\mathbf{w}(\hat{\mathbf{h}}) = \mathbf{v}$ gives a lower bound. This transforms (10) into a SISO channel and we use the approach in [2, Section III] to find a lower bound on capacity with only statistics of $\mathbf{h}^T \mathbf{v}$ at the receiver. ∎

The lower bound in (19) can be computed numerically for any channel distribution and any way of selecting the beamforming vector based on the channel estimate. To bring some insight on the behavior when the number of antennas, N, grows large, we have the following result for the case of (approximate) maximum ratio transmission (MRT).

Corollary 3. *Suppose* $\mathbf{v} = \frac{\hat{\mathbf{h}}^*}{\|\hat{\mathbf{h}}\|}$, *then the lower bound in* (19) *can be expressed as in* (21) *at the top of the next page. The big-$\mathcal{O}$ terms* $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ *vanish as* $\frac{1}{\sqrt{N}}$ *when* $N \to \infty$, *while the other terms (except the noise term) remain strictly positive.*

Proof: The equivalence between (19) and (21) is proved by straightforward but lengthy use of Hölder's inequality and the trace lemma in [18, Lemma B.26]. ∎

$$\mathsf{C} \geq \mathsf{C}_{\mathrm{lower}} = \log_2 \left(1 + \frac{\left| \mathbb{E}\left\{ \frac{(1+d^{-1}\eta_t^{\mathrm{UT}})\sqrt{\mathrm{tr}(\mathbf{R}-\mathbf{C})}}{\sqrt{\mathrm{tr}\left(\mathbf{A}(|d+\eta_t^{\mathrm{UT}}|^2\mathbf{R}+\boldsymbol{\Psi})\mathbf{A}^H\right)}} \right\} \right|^2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)}{(1+\kappa_r^{\mathrm{UT}})\mathbb{E}\left\{ \frac{|1+d^{-1}\eta_t^{\mathrm{UT}}|^2 \mathrm{tr}(\mathbf{R}-\mathbf{C})}{\mathrm{tr}\left(\mathbf{A}(|d+\eta_t^{\mathrm{UT}}|^2\mathbf{R}+\boldsymbol{\Psi})\mathbf{A}^H\right)} \right\} - \left| \mathbb{E}\left\{ \frac{(1+d^{-1}\eta_t^{\mathrm{UT}})\sqrt{\mathrm{tr}(\mathbf{R}-\mathbf{C})}}{\sqrt{\mathrm{tr}\left(\mathbf{A}(|d+\eta_t^{\mathrm{UT}}|^2\mathbf{R}+\boldsymbol{\Psi})\mathbf{A}^H\right)}} \right\} \right|^2 + \frac{1}{N}\frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{UT}}p^{\mathrm{BS}}} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)} \right) \quad (21)$$

Combining the upper bound in Corollary 2 with the lower bound in Corollary 3, we have a clear characterization of the capacity behavior when $N \to \infty$. Contrary to the upper bound, the hardware impairments at the base station are present in (21) through $\mathbf{A}$ and $\boldsymbol{\Psi}$. However, these variables vanish if we set $\eta_t^{\mathrm{UT}} = 0$, in which case the lower bound actually equals the upper bound. Consequently, the capacity limit is mainly determined by the level of impairments at the user terminal, both in the uplink estimation (κ_t^{UT}) and the downlink transmission (κ_r^{UT})—although the former connection was not visible in the upper bound since it assumed perfect CSI.

Remark 2. *It is implicitly assumed in Corollaries 2 and 3 that $\frac{1}{N}\frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{UT}}p^{\mathrm{BS}}} \to 0$ as $N \to \infty$, meaning that the noise variance σ_{UT}^2 cannot scale as fast as $\mathcal{O}(N)$. This assumption generally holds, but not under pilot contamination of the type considered in [4]–[7]. In those cases there will be additional terms in the denominator of the SINRs that further limits the capacity.*

C. Numerical Illustration

This section illustrates our lower and upper bounds on the capacity. The average SNR is defined as $p^{\mathrm{UT}}\frac{\mathrm{tr}(\mathbf{R})}{\mathrm{tr}(\mathbf{S})}$ and $p^{\mathrm{BS}}\frac{\mathrm{tr}(\mathbf{R})}{\mathrm{tr}(\mathbf{S})}$ for pilot and data transmission, respectively, and is fixed at 20 dB, while we vary the number of antennas N and the levels of impairments. For simplicity, $\kappa^{\mathrm{BS}} \triangleq \kappa_t^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}}$ at the base station and $\kappa^{\mathrm{UT}} \triangleq \kappa_t^{\mathrm{UT}} = \kappa_r^{\mathrm{UT}}$ at the user terminal.

Fig. 3 considers a spatially uncorrelated scenario where $\mathbf{R} = \mathbf{S} = \mathbf{I}$. While the capacity with ideal hardware grows without bound as $N \to \infty$, the lower and upper bounds converge to finite limits under hardware impairments. Recall that these bounds hold under any CSI conditions at the user terminal; for example, the dashed lines in Fig. 3 show a lower bound for the practical case when a downlink pilot signal is transmitted for estimation of the effective channel $\mathbf{h}^T \frac{\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}$.

Observe that the asymptotic capacity limits in Fig. 3 are characterized by the level of impairments, thus the hardware quality has a fundamental impact on the practical spectral efficiency. 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.

Next, Fig. 4 considers the same scenario but with a fixed level of impairments $\kappa^{\mathrm{UT}} = 0.05^2$ at the user terminal and different values at the base station. As expected from the analysis, the lower and upper bounds at small N decrease as κ^{BS} is increased, but the curves converge to the same value—thus showing that the impact of impairments at the base station vanishes as N grows large.

V. ENERGY EFFICIENCY

This section analyzes the energy efficiency (in bits/Joule), which is defined as the ratio of capacity (in bits/channel use) to radiated power (in Joules/channel use) [7].
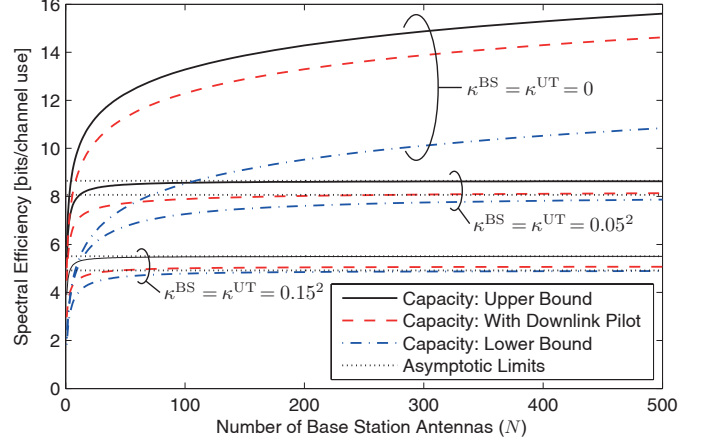

Fig. 3. Lower and upper bounds on the capacity. Hardware impairments have a fundamental impact on the asymptotic behavior as N grows large.

Definition 1. *The downlink energy efficiency is*

$$\mathrm{EE} = \frac{\mathsf{C}}{p^{\mathrm{BS}} + \alpha_1 p^{\mathrm{BS}} + \alpha_2 p^{\mathrm{UT}}} \quad (22)$$

where $\alpha_1, \alpha_2 \geq 0$ are some constants related to the overhead signaling in the downlink and uplink, respectively.

Recall from Corollary 2 that the capacity has a finite ceiling as $N \to \infty$ or $p^{\mathrm{BS}} \to \infty$, thus high energy efficiency cannot be achieved by only increasing N or p^{BS}. However, we have the following important result, which has counterparts for ideal hardware [6], [7] and the phase noise-impaired uplink [12].

Corollary 4. *Suppose the downlink transmit power p^{BS} and uplink pilot power p^{UT} are scaled proportional to $1/N^{t_{\mathrm{BS}}}$ and $1/N^{t_{\mathrm{UT}}}$, respectively. If $t_{\mathrm{BS}} \geq 0, 0 < t_{\mathrm{UT}} < \frac{1}{2}$ and $t_{\mathrm{BS}} + t_{\mathrm{UT}} < 1$, the lower bound in Corollary 3 gives asymptotically*

$$\lim_{N \to \infty} \mathsf{C} \geq \log_2\left(1 + \frac{1}{\kappa_r^{\mathrm{UT}} + \kappa_t^{\mathrm{UT}} + \kappa_r^{\mathrm{UT}}\kappa_t^{\mathrm{UT}}}\right). \quad (23)$$

Proof: The dominated convergence theorem and $t_{\mathrm{UT}} > 0$ gives $|\mathbb{E}\{\cdot\}|^2 \to 1$ in the numerator and $\mathbb{E}\{\cdot\} \to 1 + \kappa_t^{\mathrm{UT}}$ in the denominator of (23). If $t_{\mathrm{BS}} + t_{\mathrm{UT}} < 1$: $\frac{1}{N}\frac{\sigma_{\mathrm{UT}}^2}{p^{\mathrm{UT}}p^{\mathrm{BS}}} \to 0$. ∎

Corollary 4 shows that one can reduce the transmit and pilot powers (e.g., roughly as $1/\sqrt{N}$) and still obtain a non-zero capacity. The asymptotic capacity is lower bounded by (23). As expected from previous results, the lower bound in (23) only depends on the level of impairments at the user terminal. This implies a great robustness to base station impairments. The power scaling in Corollary 4 has the extraordinary consequence

$$\lim_{N \to \infty} \mathrm{EE} = \lim_{N \to \infty} \frac{\mathsf{C}}{(1+\alpha_1)\frac{p^{\mathrm{BS}}}{N^{t_{\mathrm{BS}}}} + \alpha_2\frac{p^{\mathrm{UT}}}{N^{t_{\mathrm{UT}}}}} = \infty. \quad (24)$$

Observe that this unbounded asymptotic energy efficiency is achieved by using the LMMSE estimator in Theorem 1 for uplink channel estimation and approximate MRT for downlink
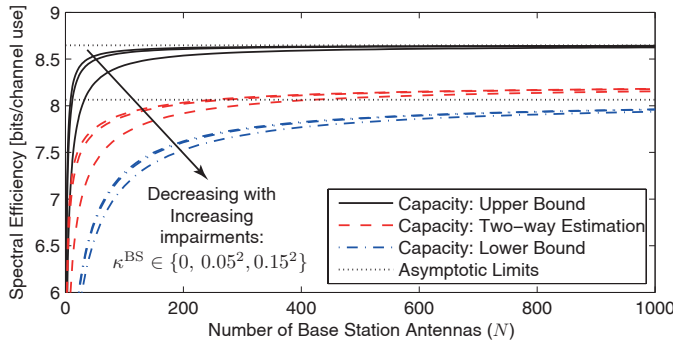
Fig. 4. Lower and upper bounds on the capacity for $\kappa^{\mathrm{UT}} = 0.05^2$. The impact of hardware impairments at the base station vanishes asymptotically.
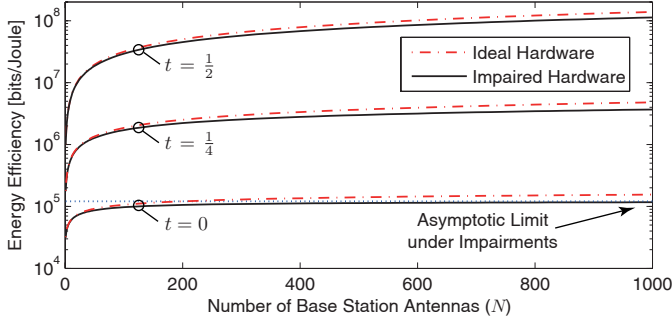

Fig. 5. Energy efficiency of ideal and impaired hardware when the transmit power and pilot powers are reduced with N as $1/N^t$ for $t \in \{0, \frac{1}{4}, \frac{1}{2}\}$.

transmission. Large-scale MISO systems can thus obtain an immense energy efficiency even under hardware impairments.

Remark 3. *Although the ratio of capacity to radiated power in* (22) *can grow unboundedly as shown in* (24)*, the overall energy efficiency (capacity divided by total power consumption) will not. Specifically, each antenna requires dedicated circuits with a non-zero power consumption $\varrho > 0$. This contributes an additional term $N\varrho$ in the denominator of* (22) *and implies that the overall energy efficiency is maximized at some finite N. This refined efficiency notion was analyzed in* [19] *and it was observed that large-scale MIMO offers great energy savings also when the circuit power is taken into account.*

A. Numerical Illustration

The energy efficiency (using the lower bound in Corollary 3) is illustrated in Fig. 5 when the transmit and pilot powers are $p^{\mathrm{BS}} = p^{\mathrm{UT}} = 30$ dBm at $N = 1$ and then reduced by a factor $1/N^t$ for $t \in \{0, \frac{1}{4}, \frac{1}{2}\}$. The channel covariance matrices $\mathbf{R}$ are generated using the exponential model with correlation coefficient 0.7 [16], while the noise covariance $\mathbf{S}$ is a scaled identity matrix such that $\mathrm{tr}(\mathbf{R})/\mathrm{tr}(\mathbf{S}) = -10$ dBm at a bandwidth of 15 kHz. The impairment parameters are set to $\kappa_t^{\mathrm{BS}} = \kappa_r^{\mathrm{BS}} = \kappa_t^{\mathrm{UT}} = \kappa_r^{\mathrm{UT}} = 0.05^2$.

Fig. 5 verifies that the energy efficiency grows unboundedly with N for $t = \frac{1}{4}$ and $t = \frac{1}{2}$, while it converges to a finite value for $t = 0$ (i.e., when the transmit/pilot powers are fixed). Note that $\alpha_1 = \alpha_2 = 0$ in the figure, but changing these parameters only result in a common vertical shift of all the curves. Fig. 5 also shows the case of ideal hardware. Despite the fundamental differences when $N \to \infty$ or $p^{\mathrm{BS}}, p^{\mathrm{UT}} \to \infty$ (i.e., there is no capacity ceiling or estimation error floor under ideal hardware), the difference in terms of energy efficiency is remarkably small between ideal and practical hardware.

VI. Conclusion

This paper analyzed how transceiver hardware impairments impact the capacity and estimation accuracy of large-scale MISO systems. We proved analytically that the impairments of physical hardware create a finite capacity ceiling and non-zero estimation error floor—irrespective of the SNR and number of base station antennas N. This stands in contrast to the very optimistic asymptotic results previously reported for ideal hardware. Interestingly, only the hardware impairments at the user terminal limits the performance as N grows large. Despite these discouraging results, we showed that large-scale MISO systems can achieve an arbitrarily high energy efficiency.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] H. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, to appear.

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

[9] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *Proc. ITG/IEEE WSA*, 2010.

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

[11] E. Björnson, P. Zetterberg, and M. Bengtsson, "Optimal coordinated beamforming in the multicell downlink with transceiver impairments," in *Proc. IEEE GLOBECOM*, 2012.

[12] A. Pitarokoilis, S. Mohammed, and E. Larsson, "Effect of oscillator phase noise on uplink performance of large MU-MIMO systems," in *Proc. Allerton*, 2012.

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

[14] 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.

[15] 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.

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

[17] 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.

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

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