Massive MIMO for Maximal Spectral Efficiency: How Many Users and Pilots Should Be Allocated?

Emil Björnson, Member, IEEE, Erik G. Larsson, Senior Member, IEEE, and Mérouane Debbah, Fellow, IEEE

Abstract

Massive MIMO is a promising technique to increase the spectral efficiency (SE) of cellular networks, by deploying antenna arrays with hundreds or thousands of active elements at the base stations and performing phase-coherent transceiver processing. A common rule-of-thumb is that these systems should have an order of magnitude more antennas, $M$, than scheduled users, $K$, because the users’ channels are likely to be near-orthogonal when $M/K > 10$. However, it has not been proved that this rule-of-thumb actually maximizes the SE. In this paper, we analyze how the optimal number of scheduled users, $K^*$, depends on $M$ and other system parameters. To this end, new SE expressions are derived to enable efficient system-level analysis with power control, arbitrary pilot allocation, and random user locations. The value of $K^*$ in the large-$M$ regime is derived in closed form, while simulations are used to show what happens at finite $M$, in different interference scenarios, with different pilot reuse factors, and for different processing schemes. Up to half the coherence block should be dedicated to pilots and the optimal $M/K$ is less than 10 in many cases of practical relevance. Interestingly, $K^*$ depends strongly on the processing scheme and hence it is unfair to compare different schemes using the same $K$.

I. INTRODUCTION

Cellular communication networks are continuously evolving to keep up with the rapidly increasing demand for wireless data services. Higher area throughput (in bit/s per km$^2$) has
traditionally been achieved by a combination of three multiplicative factors [1]: more frequency spectrum (Hz), higher cell density (more cells per km$^2$), and higher spectral efficiency (bit/s/Hz/cell). This paper considers the latter and especially the massive multiple-input multiple-output (MIMO) concept, proposed in [2], which has been identified as the key to increase the spectral efficiency (SE) by orders of magnitude over contemporary systems [3]–[5].

The massive MIMO concept is based on equipping base stations (BSs) with hundreds or thousands of antenna elements which, unlike conventional cellular technology, are operated in a phase-coherent fashion. This can provide unprecedented array gains and a spatial resolution that allows for multi-user MIMO communication to tens or hundreds of user equipments (UEs) per cell, while maintaining robustness to inter-user interference. The research on massive MIMO has so far focused on establishing the fundamental physical (PHY) layer properties; in particular, that the acquisition of channel state information (CSI) is limited by the channel coherence block (i.e., the fact that channel responses are only static in limited time/frequency blocks) and how this impacts the SEs and the ability to mitigate inter-cell interference [2], [6], [7]. In addition, the aggressive multiplexing in massive MIMO has been shown to provide major improvements in the overall energy efficiency [8]–[11], while [12]–[14] have shown that the hardware impairments of practical transceivers have smaller impact on massive MIMO than contemporary systems. In contrast, the research community has only briefly touched on resource allocation problems in the media access control (MAC) layer (e.g., user scheduling)—although the truly achievable SEs can only be understood if the PHY and MAC layers are jointly optimized.

The importance of resource allocation for massive MIMO was described in [15], where initial guidelines were given. A main insight is that the limited number of orthogonal pilot sequences needs to be allocated intelligently among the UEs to reduce interference, which can be done by capitalizing on pathloss differences [16], [17] and spatial correlation [15], [18], [19].

In this paper, we consider a related resource allocation question: how many UEs should be scheduled per cell to maximize the spectral efficiency? This question has, to the best of our knowledge, not been answered for multi-cell systems. We show how the coherence block length, number of antennas, pilot allocation, hardware impairments, and other system parameters determine the answer. To this end, we derive new SE expressions which are valid for both

---

1 A few results for single-cell systems are available in the literature; for example, in [8].
uplink (UL) and downlink (DL) transmission, with random user locations and power control that yields uniform UE performance. We consider both conventional linear processing schemes such as maximum ratio (MR) combining/transmission and zero-forcing (ZF), and a new full-pilot zero-forcing (P-ZF) scheme that actively suppresses inter-cell interference in a fully distributed coordinated beamforming fashion. The following are the main contributions in each section if the paper:

- Section II presents the UL/DL massive MIMO system model, where the unique features are the power control and random UE locations.
- Section III provides new analytic results for channel estimation with arbitrary pilot signals and new tractable SE expressions for the UL and DL with random UE locations and power control. MR and ZF processing are considered, as well as the new P-ZF scheme.
- Section IV provides extensive simulation results on the maximal SE, where the impact of all system parameters are explained. The expected massive MIMO gains are illustrated.
- Section V extends the previous results to systems with hardware impairments.
- Finally, Section VI summarizes the main results and insights obtained in the paper.

II. SYSTEM MODEL

This paper considers a cellular network with universal time and frequency reuse. Each cell is assigned an index in the set $L$, where the cardinality $|L|$ is the number of cells. The BS in each cell is equipped with an array of $M$ antennas and communicates with $K$ single-antenna UEs in the cell. We are mainly interested in massive MIMO topologies where $M$ is large and fixed, while $K$ is a design parameter and all UEs have unlimited demand for data.

Each UE moves around randomly in its serving cell according to some ergodic cell-specific distribution. The geographical position $z_{lk} \in \mathbb{R}^2$ of UE $k \in \{1, \ldots, K\}$ in cell $l \in L$ is thus a random variable. This assumption will allow us study average performance, without conditioning on an instantaneous set of UE positions. The time-frequency resources are divided into frames consisting of $T_c$ seconds and $W_c$ Hz, as illustrated in Fig. I. This leaves room for $S = T_c W_c$ transmission symbols per frame. We assume that the frame dimensions are such that $T_c$ is smaller or equal to the coherence time of all UEs, while $W_c$ is smaller or equal to the coherence bandwidth of all UEs. Hence, all the channels are static within the frame; $h_{jlk} \in \mathbb{C}^N$ denotes the channel response between BS $j$ and UE $k$ in cell $l$ in a given frame. These channel responses
Fig. 1: The transmission is divided into frames of $S = T_c W_c$ symbols, whereof $B$ symbols are dedicated to pilot transmission. The remaining $S - B$ symbols are used for payload data, where $\zeta^{(ul)}$ and $\zeta^{(dl)}$ are respectively the fractions of UL and DL transmission.

are drawn as realizations from zero-mean circularly symmetric complex Gaussian distributions,

$$h_{jlk} \sim \mathcal{CN}\left(0, d_j(z_{lk})I_M\right),$$

(1)

where $I_M$ is the $M \times M$ identity matrix. This is a reasonable model for non-line-of-sight propagation with both few and many BS antennas (see recent channel measurements reported in [20]). The deterministic function $d_j(z)$ gives the variance of the channel attenuation from BS $j$ to any UE position $z$. We assume that the value of $d_j(z_{lk})$ is known at BS $j$ for all $l$ and $k$ (it is measured over frequency and tracked over time), while the exact UE positions are unknown.

We consider the time-division duplex (TDD) protocol shown in Fig. 1, where $B \geq 1$ out of the $S$ symbols in each frame are reserved for UL pilot signaling. There is no DL pilot signaling and no feedback of CSI, because the BSs can process both UL and DL signals using the UL channel measurements due to the channel reciprocity in TDD systems. The remaining $S - B$ symbols are allocated for payload data and are split between UL and DL transmission. We let $\zeta^{(ul)}$ and $\zeta^{(dl)}$ denote the fixed fractions allocated for UL and DL, respectively. These fractions can be selected arbitrarily, subject to the constraint $\zeta^{(ul)} + \zeta^{(dl)} = 1$ and that $\zeta^{(ul)}(S - B)$ and $\zeta^{(dl)}(S - B)$ are positive integers. Below, we define the system models for the UL and DL.

The base stations are not exchanging any short-term information in this work, but we will see
how the pilot allocation and transmission processing can be coordinated in a distributed fashion.

A. Uplink

The received UL signal \( y_j \in \mathbb{C}^M \) at BS \( j \) in a frame is modeled, similar to [7], [8], as
\[
y_j = \sum_{l \in L} \sum_{k=1}^{K} \sqrt{p_{lk}} h_{jk} x_{lk} + n_j
\]
(2)
where \( x_{lk} \in \mathbb{C} \) is the symbol transmitted by UE \( k \) in cell \( l \). This signal is normalized as \( \mathbb{E}\{ |x_{lk}|^2 \} = 1 \), while the corresponding UL transmit power is defined by \( p_{lk} \geq 0 \). The additive noise \( n_j \in \mathbb{C}^M \) is modeled as \( n_j \sim \mathcal{CN}(0, \sigma^2 I_M) \), where \( \sigma^2 \) is the noise variance.

Contrary to most previous works on massive MIMO, which assume fixed UL power, we consider statistics-aware power control\(^2\); the symbols from UE \( k \) in cell \( l \) have the transmit power \( p_{lk} = \rho d_l(z_{lk}) \), where \( \rho > 0 \) is a design parameter. This power-control policy inverts the average channel attenuation \( d_l(z_{lk}) \) and has the merit of making the average effective channel gain the same for all UEs: \( \mathbb{E}\{ |p_{lk}||h_{lk}|^2 \} = M \rho \). Hence, this policy guarantees a uniform user experience, saves valuable energy at UEs, and avoids near-far blockage where weak signals drown in stronger signals due to the finite dynamic range of analog-to-digital converters (ADCs).

B. Downlink

Building on the UL/DL channel reciprocity in calibrated TDD systems, the received DL signal \( z_{jk} \in \mathbb{C} \) at UE \( k \) in cell \( j \) in a frame is modeled as
\[
z_{jk} = \sum_{l \in L} \sum_{m=1}^{K} h_{lk}^T w_{lm} s_{lm} + \eta_{jk}
\]
(3)
where \((\cdot)^T\) denotes the transpose, \( s_{lm} \) is the symbol intended for UE \( m \) in cell \( l \), \( w_{lm} \in \mathbb{C}^M \) is the corresponding precoding vector, and \( ||w_{lm}||^2 \) is the allocated DL transmit power. Any power control can be considered in the DL since the BS has access to the estimated CSI. We will show how to select the transmit power to achieve the same SEs in the DL as in the UL. The additive noise at UE \( k \) in cell \( j \) is modeled as \( \eta_{jk} \sim \mathcal{CN}(0, \sigma^2) \), with the same variance as in the UL\(^3\).

\(^2\)Channel-aware power control was considered in [21] and [9], but it requires a rapid feedback mechanism where UEs are provided with instantaneous CSI. Since the small-scale fading average out in massive MIMO systems [2], statistical power control policies are expected to be almost equal to channel-aware policies [22], but are considerably easier to implement.

\(^3\)The noise variance is conventionally lower in the UL, due to better hardware characteristics at the BS, but since massive MIMO has an inherent robustness to noise amplification [13] it is possible to use handset-like hardware at the BSs. In any case, any disparity in noise power between the UL and DL can be absorbed into the transmit powers without loss of generality.
III. AVERAGE PER-CELL SPECTRAL EFFICIENCY

In this section, we derive and analyze the SE for multi-cell systems with random UE positions.

A. Pilot-Based Channel Estimation

BS $j$ can use its multitude of antennas for phase-coherent receive combining in the UL and transmit precoding in the DL, which can adaptively amplify desired signals and suppress interfering signals. This requires, however, some knowledge of the UEs’ channels; for example, $\sqrt{p_{lk}}h_{jlk}$ in the UL, for all $l$ and $k$. Such CSI is typically acquired by pilot signaling, where the UEs send known signals in a predefined manner. Accurate CSI acquisition is a challenging task in multi-cell systems, where the transmission resources are reused across cells, because the pilot signals are inevitably affected by inter-cell interference. This so-called pilot contamination limits the quality of the acquired CSI and the ability to reject inter-cell interference (unless intricate subspace methods can be used for decontamination, as suggested in [17]).

The impact of pilot contamination is usually studied under the assumption that exactly the same pilot signals are used in all cells. In contrast, this section derives the main properties of massive MIMO systems (with power control) for arbitrary pilot allocation, where each cell might only use a subset of the pilots. As shown in Fig. 1, the pilot signals are assumed to span $B$ symbols of each frame, where $1 \leq B \leq S$.

Each pilot signal can be represented by a deterministic vector $v \in \mathbb{C}^B$ and the fixed per-symbol power implies that all entries have unit magnitude: $|v_s| = 1$, where $[\cdot]_s$ denotes the $s$th element for $s \in \{1, \ldots, B\}$. We assume that all pilot signals originate from a fixed pilot book $\mathcal{V}$, defined as

$$\mathcal{V} = \{v_1, \ldots, v_B\} \text{ where } v_{b_1}^H v_{b_2} = \begin{cases} B, & b_1 = b_2, \\ 0, & b_1 \neq b_2, \end{cases}$$

where $(\cdot)^H$ denotes the conjugate transpose. Hence, the $B$ pilot signals form an orthogonal basis and can, for example, be the columns of a discrete Fourier transform (DFT) matrix $[23]$.

The pilot signals need not be synchronized across the cells as assumed herein, but there is little to gain from shifting the pilot signals and UL payload data signals between cells; this leads to a mix of deterministic pilots and stochastic data signals at each symbol transmission, but the average pilot contamination will not change in any substantial way [8, Remark 5]. The new full-pilot interference suppression concepts proposed in this paper are also harder to implement in such cases.
The pilot signal transmitted by UE $k$ in cell $l$ is denoted by $v_{i_{lk}}$, where $i_{lk} \in \{1, \ldots, B\}$ is the index in the pilot book. By transmitting these pilot signals over $B$ symbols in the UL system model of (5), the collective received UL signal at BS $j$ is denoted as $Y_j \in \mathbb{C}^{M \times B}$ and given by

$$Y_j = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} \sqrt{p_{lk}} h_{jlk} v_{i_{lk}}^* + N_j,$$

where $N_j \in \mathbb{C}^{M \times B}$ contains the additive noise at the receiver during the pilot signaling.

The following lemma derives the minimum mean-squared error (MMSE) estimator of the effective power-controlled UL channels, which are defined as $h_{jlk}^{\text{eff}} = \sqrt{p_{lk}} h_{jlk}$.

**Lemma 1.** The MMSE estimate at BS $j$ of the effective power-controlled UL channel $h_{jlk}^{\text{eff}}$, for any UE $k \in \{1, \ldots, K\}$ in any cell $l \in \mathcal{L}$, is

$$\hat{h}_{jlk}^{\text{eff}} = \frac{d_j(z_{lk})}{d_l(z_{lk})} Y_j \Psi_j^{-1} v_{i_{lk}}^*$$

where $(\cdot)^*$ denotes the complex conjugate and the normalized covariance matrix $\Psi_j \in \mathbb{C}^{B \times B}$ of the received signal is

$$\Psi_j = \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \frac{d_j(z_{\ell m})}{d_l(z_{\ell m})} v_{i_{\ell m}, l}^H v_{i_{\ell m}} + \frac{\sigma^2}{\rho} I_B.$$

The estimation error covariance matrix $C_{jlk} \in \mathbb{C}^{M \times M}$ is given by

$$C_{jlk} = \mathbb{E} \left\{ (h_{jlk}^{\text{eff}} - \hat{h}_{jlk}^{\text{eff}})(h_{jlk}^{\text{eff}} - \hat{h}_{jlk}^{\text{eff}})^H \right\} = \rho \frac{d_j(z_{lk})}{d_l(z_{lk})} \left( 1 - \frac{d_j(z_{\ell m})}{d_l(z_{\ell m})} \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \frac{d_j(z_{\ell m})}{d_l(z_{\ell m})} v_{i_{\ell m}, l}^H v_{i_{\ell m}} + \frac{\sigma^2}{\rho} \right) I_M$$

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

**Proof:** The proof is given in the appendix.

There are two important differences between Lemma [1] and the channel estimators that are conventionally used in the massive MIMO literature: 1) we estimate the effective channels including the UL power control; and 2) the MMSE estimator supports arbitrary pilot allocation.

The covariance matrix in (8) reveals the causes of estimation errors; it depends on the inverse signal-to-noise ratio (SNR), $\sigma^2 / \rho$, and on which UEs that use the same pilot signal (i.e., which of the products $v_{i_{lk}}^H v_{i_{\ell m}}$ that are non-zero). The ratio $d_j(z_{\ell m}) / d_l(z_{\ell m})$ describes the relative strength of the interference received at BS $j$ from UE $m$ in cell $\ell$; it is almost one for cell-edge UEs of neighboring cells, while it is almost zero when cell $\ell$ is very distant from BS $j$. 

December 23, 2014 DRAFT
Although Lemma 1 allows for estimation of all channel vectors in the whole cellular network, each BS can only resolve \( B \) different spatial dimensions since there are only \( B \) orthogonal pilot signals. To show this explicitly, we define the \( M \times B \) matrix

\[
\hat{H}_{V,j} = Y_j \Psi_j^{-1} [v_1^*, \ldots, v_B^*]
\]

using each of the \( B \) pilot signals from \( V \). The channel estimate in (6) for UE \( k \) in cell \( l \), which use the pilot \( v_{i_{lk}} \), is parallel to the \( i_{lk} \)th column of \( \hat{H}_{V,j} \); more precisely, we have

\[
\hat{h}_{jlk}^{\text{eff}} = \frac{d_j(z_{ik})}{d_l(z_{lk})} \hat{H}_{V,j} e_{i_{lk}}
\]

where \( e_i \) denotes the \( i \)th column of the identity matrix \( I_B \). This is the essence of pilot contamination; BSs cannot tell apart UEs that use the same pilot signal and cannot reject the corresponding interference since the estimated channels are parallel. In some cases (e.g., for slow changes in the user scheduling and high spatial channel correlation), statistical prior knowledge can be used to partially separate the UEs [18], but this is not considered herein since we aim at establishing fundamental properties that can be exploited in any propagation environment.

**Remark 1** (Mobility and Pilot Sharing). Each UE might have a different dimension of its coherence block, defined by some coherence time \( \tilde{T}_c \) and coherence bandwidth \( \tilde{W}_c \), depending on the propagation environment and the UE’s mobility. Suppose that \( \tilde{T}_c = aT_c \) and \( \tilde{W}_c = bW_c \) for a certain UE, where \( a \geq 1 \) and \( b \geq 1 \) since the frame structure was defined to fit into the coherence block of all UEs. Then, \( \tau = \lfloor a \rfloor \lfloor b \rfloor \) is the total number of frames that fits into the coherence block of this particular UE, where \( \lfloor \cdot \rfloor \) stands for truncation. If \( \tau > 1 \), there is no need to send pilots in every frame; it is sufficient with \( 1/\tau \) of the frames. Hence, multiple UEs with \( \tau > 1 \) can share a pilot signal, without disturbing one another, by using it in different frames.

**B. Achievable UL Spectral Efficiencies**

The channel estimates in Lemma 1 enable each BS to (semi-)coherently detect the data signals from its UEs. In particular, we assume that BS \( j \) applies a linear receive combining vector \( g_{jk} \in \mathbb{C}^M \) to the received signal, as \( g_{jk}^H y_j \), to amplify the signal from its \( k \)th UE and reject interference from other UEs in the spatial domain. We want to derive ergodic achievable SEs for the UEs, where the codewords span over both Rayleigh fading and random locations of the interfering UEs—specific UE distributions are considered in Section IV. The following lemma shows how the SE depends on the receive combining, for Gaussian codebooks where \( x_{jk} \sim \mathcal{CN}(0, 1) \).
Lemma 2. In the UL, an ergodic achievable SE at UE $k$ in cell $j$ is

$$
\zeta^{(ul)} \left( 1 - \frac{B}{S} \right) \mathbb{E}\{z\} \left\{ \log_2(1 + \text{SINR}^{(ul)}_{jk}) \right\} \text{[bit/s/Hz]} \tag{11}
$$

with the effective signal-to-interference-and-noise ratio (SINR) given by

$$
\text{SINR}^{(ul)}_{jk} = \frac{p_{jk} |\mathbb{E}\{h\} \{g^h_{jk} h_{jjk}\}|^2}{\sum_{l \in L} \sum_{m=1}^K p_{lm} |\mathbb{E}\{h\} \{g^h_{jk} h_{jlm}\}|^2 - p_{jk} |\mathbb{E}\{h\} \{g^h_{jk} h_{jjk}\}|^2 + \sigma^2 |\mathbb{E}\{h\}\{g_{jk}\}|^2}. \tag{12}
$$

The expectations $\mathbb{E}\{z\}\{\cdot\}$ and $\mathbb{E}\{h\}\{\cdot\}$ are with respect to UE positions and channel realizations, respectively.

**Proof:** By coding over variations in the channel realizations $\{h\}$ and UE positions $\{z\}$, an achievable SE is given by $\mathbb{E}\{z,h\}\{I(x_{lk},y_j)\}$, where $I(x_{lk},y_j)$ is the mutual information between the transmitted and received signal in $\text{I}$. The lemma follows from computing a lower bound on $I(x_{lk},y_j)$, similar to [2], [6], [7], [13], [24], by treating $\mathbb{E}\{h\}\{g^h_{jk} h_{jjk}\}$ as the desired channel and treating interference and channel variations as worst-case Gaussian noise.

To compute the expectations in the SE expression of Lemma 2, we need to specify the receive combining. The combining schemes for massive MIMO can have either passive or active interference rejection. The canonical example of passive rejection is maximum ratio (MR) combining, defined as

$$
g^{\text{MR}}_{jk} = \hat{H}_{v,j} e_{i_{jk}} = \hat{h}^{\text{eff}}_{jjk}, \tag{13}
$$

which maximizes the gain of the desired signal and relies on that interfering signals are rejected automatically since the co-user channels are quasi-orthogonal to $\hat{h}^{\text{eff}}_{jjk}$ when $M$ is large.

In contrast, active rejection is achieved by making the receive combining as orthogonal to the interfering channels as possible. This is conventionally achieved by zero-forcing (ZF) combining, where the combining is selected to orthogonalize the $K$ intra-cell channels:

$$
g^{\text{ZF}}_{jk} = \hat{H}_{v,j} E_j (E^H_j \hat{H}^H_{v,j} \hat{H}_{v,j} E_j)^{-1} e_{i_{jk}}, \tag{14}
$$

where $E_j = [e_{i_{j1}} \ldots e_{i_{jK}}] \in \mathbb{C}^{B \times K}$ and all the UEs in cell $j$ are required to use different pilots.

The next theorem provides closed-form expressions for the per-cell SEs with MR and ZF.

---

5With quasi-orthogonality we mean that two vectors $a, b \in \mathbb{C}^M$ satisfy $\frac{a^H b}{M} \rightarrow 0$ as $M \rightarrow \infty$, although $a^H b$ will not converge to zero and might even go to infinity, e.g., proportionally to $\sqrt{M}$ as with Rayleigh fading channel vectors.
Theorem 1. In the UL, an achievable SE in cell $j$ is
\[
\text{SE}^{(ul)}_j = \sum_{k=1}^{K} \zeta^{(ul)}(j) \left( 1 - \frac{B}{S} \right) \log_2 \left( 1 + \frac{B}{I_{\text{scheme}jk}} \right) \quad \text{[bit/s/Hz/cell]}
\] (15)
where the interference term $I_{\text{scheme}jk}$ for UE $k$ depends on the receive combining scheme. MR combining gives
\[
I_{\text{MR}jk} = \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \left( \mu^{(2)}_{jl} + \frac{\mu^{(2)}_{jl} - \left( \mu^{(1)}_{jl} \right)^2}{M} \right) \mathbf{v}_{ijkl}^H \mathbf{v}_{ilm} + \left( \sum_{l \in \mathcal{L}} \mu^{(1)}_{jl} \right) \frac{K}{M} + \frac{\sigma^2}{M \rho} \left( \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \mu^{(1)}_{jl} \mathbf{v}_{ijkl}^H \mathbf{v}_{ilm} + \frac{\sigma^2}{\rho} \right)
\] (16)
and ZF combining gives
\[
I_{\text{ZF}jk} = \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \left( \mu^{(2)}_{jl} + \frac{\mu^{(2)}_{jl} - \left( \mu^{(1)}_{jl} \right)^2}{M - K} \right) \mathbf{v}_{ijkl}^H \mathbf{v}_{ilm} + \left( \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \mu^{(1)}_{jl} \mathbf{v}_{ijkl}^H \mathbf{v}_{ilm} + \frac{\sigma^2}{\rho} \right) \left( \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \mu^{(1)}_{jl} \mathbf{v}_{ijkl}^H \mathbf{v}_{ilm} + \frac{\sigma^2}{\rho} \right) \left( \frac{K}{M - K} \right).
\] (17)

The following notation was used:
\[
\mu^{(\omega)}_{jl} = \mathbb{E}_{z_{lm}} \left\{ \left( \frac{d_j(z_{lm})}{d_l(z_{lm})} \right)^{\omega} \right\} \quad \text{for} \quad \omega = 1, 2.
\] (18)

Proof: The proof is given in the appendix.

The closed-form SE expressions in Theorem 1 are only functions of the pilot allocation and the propagation parameters $\mu^{(1)}_{jl}$ and $\mu^{(2)}_{jl}$ from (18). The latter two are the average ratio between the channel variance to BS $j$ and the channel variance to BS $l$, for an arbitrary UE in cell $l$, and the second-order moment of this ratio, respectively. These parameters are equal to 1 for $j = l$ and otherwise go to zero as the distance between BS $j$ and cell $l$ increases. The SE expressions manifest the importance of pilot allocation, since all the interference terms in (16) and (17) contain inner products of pilot signals that are either equal to zero or $B$. The first term in (16) and (17) describes the pilot contamination, while the second term is the inter-user interference. The difference between MR and ZF is that the latter scheme cancels some interference, but at the price of reducing the array gain from $M$ to $M - K$.

ZF combining only actively suppresses intra-cell interference, while the inter-cell interference is passively suppressed just as in MR combining. Further interference rejection can be achieved...
by coordinating the combining across cells, such that both intra-cell and inter-cell interference are actively suppressed by the receive combining. We propose a new full-pilot zero-forcing (P-ZF) combining, defined as

\[ g_{jk}^{P-ZF} = \hat{H}_{V,j} \left( \hat{H}_{V,j}^H \hat{H}_{V,j} \right)^{-1} e_{ijk}. \]  

(19)

In contrast to the conventional ZF in (14), which only orthogonalize the \( K \) intra-cell channels in \( \hat{H}_{V,j} E_j \), P-ZF exploits that all the \( B \) estimated channel directions in \( \hat{H}_{V,j} \) are known at BS \( j \) and orthogonalizes all these directions to also mitigate parts of the inter-cell interference; a similar downlink concept was proposed in [15]. The cost is a loss in array gain of \( B \), instead of \( K \) as with conventional ZF. There is no signaling between BSs in this coordinated multipoint (CoMP) scheme—BS \( j \) estimates \( \hat{H}_{V,j} \) from the UL pilot signaling—and thus the P-ZF scheme is fully distributed and scalable. Achievable SEs with P-ZF are given by the following theorem.

**Theorem 2.** In the UL, an achievable SE in cell \( j \) with P-ZF combining is

\[ \text{SE}_j = \sum_{k=1}^{K} \zeta^{(ul)} \left( 1 - \frac{B}{S} \right) \log_2 \left( 1 + \frac{B}{I_{P-ZF}^{jk}} \right) \text{[bit/s/Hz/cell]} \]  

(20)

where

\[ I_{P-ZF}^{jk} = \left( \sum_{l \in \mathcal{L}_j} \sum_{m=1}^{K} \left( \frac{\mu_{jl}^{(2)} + \frac{\mu_{jl}^{(2)} - \left( \frac{\mu_{jl}^{(1)}}{M - B} \right)^2}{M - B}}{\sqrt{\sum_{l' \in \mathcal{L}_j} \sum_{m'=1}^{K} \mu_{jl'} \left( h_{i,m}^H v_{i,m} + \frac{\sigma^2}{\rho} \right)} + \sigma^2 \rho \left( \frac{1}{M - B} \right)} \right) \right) \]  

(21)

**Proof:** The proof is given in the appendix.

\[ \square \]

**C. Achievable DL Spectral Efficiencies**

The channel estimates from Lemma [1] are also used for linear precoding in the DL, where the \( M \) channel inputs are utilized to make each data signal add up (semi-)coherently at its desired UE and to suppress the interference caused to other UEs. Recall from (3) that \( w_{jk} \in \mathbb{C}^M \) is the precoding vector associated with UE \( k \) in cell \( j \). We express these precoding vectors as

\[ w_{jk} = \sqrt{\frac{q_{jk}}{\mathbb{E}_h \{ \| \hat{g}_{jk} \|^2 \}}} \hat{g}_{jk}^* \]  

(22)

December 23, 2014   DRAFT
where the average transmit power $q_{jk} \geq 0$ is a function of the UE positions, but not the instantaneous channel realizations. The vector $\tilde{g}_{jk} \in \mathbb{C}^M$ defines the spatial directivity of the transmission and is based on the acquired CSI; the normalization with the average squared norm $E\{\|\tilde{g}_{jk}\|^2\}$ gives the analytic tractability that enables the following results.

**Lemma 3.** In the DL, an ergodic achievable SE at UE $k$ in cell $j$ is

$$\zeta^{(dl)} \left(1 - \frac{B}{S}\right) E\{\mathbf{z}\} \left\{\log_2\left(1 + \text{SINR}^{(dl)}_{jk}\right)\right\} \text{[bit/s/Hz]}$$

with the effective SINR is given by

$$\text{SINR}^{(dl)}_{jk} = \frac{q_{jk} |E\{h\} \{\tilde{g}_{jk}^H h_{jk}\}|^2}{\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} q_{lm} E\{h\} \{\|\tilde{g}_{lm}\|^2\} - \frac{q_{jk} |E\{h\} \{\tilde{g}_{jk}^H h_{jk}\}|^2}{E\{h\} \{\|\tilde{g}_{jk}\|^2\}} + \sigma^2}.$$  \hspace{1cm} (24)

**Proof:** This result follows from the same procedures as the proof of Lemma 2.

The precoding can be designed in a variety of ways. The next theorem shows that there is a strong connection between transmit precoding in the DL and receive combining in the UL.

**Theorem 3.** Let $\{\mathbf{g}_{jk}^{\text{scheme}}\}$ be the set of receive combining vectors used in the UL. Then, there exist a DL power control policy $\{q_{jk}\}$, with $\sum_{j \in \mathcal{L}} \sum_{k=1}^{K} q_{jk} = \sum_{j \in \mathcal{L}} \sum_{k=1}^{K} p_{jk}$, for which

$$\text{SINR}^{(dl)}_{jk} = \text{SINR}^{(ul)}_{jk}$$  \hspace{1cm} (25)

by using $\mathbf{g}_{jk} = \mathbf{g}_{jk}^{\text{scheme}}$ for all $j$ and $k$. Consequently, an achievable SE in the DL of cell $j$ is

$$\text{SE}^{(dl)}_j = \sum_{k=1}^{K} \zeta^{(dl)} \left(1 - \frac{B}{S}\right) \log_2 \left(1 + \frac{B}{I_{jk}^{\text{scheme}}}ight) \left[\text{bit/s/Hz/cell}\right]$$

where the interference term $I_{jk}^{\text{scheme}}$ is the same as in the UL (for MR, ZF, or P-ZF).

**Proof:** The proof is given in the appendix.

This theorem shows that the SINRs that are achieved in the UL are also achievable in the DL, by selecting the power control coefficients $\{q_{jk}\}$ properly. The total transmit power is the same, but is allocated differently over the UEs. This is a consequence of the uplink-downlink duality \[^6\] , conventionally considered for single-cell systems with perfect CSI, which is applicable also

[^6]: Usually, the power is normalized by $\|\tilde{g}_{jk}\|^2$ instead of $E\{h\} \{\|\tilde{g}_{jk}\|^2\}$ in multi-user MIMO systems \[^5\] , but the difference is small in massive MIMO since $|E\{h\} \{\|\tilde{g}_{jk}\|^2\} - \|\tilde{g}_{jk}\|^2|/M \to 0$ as $M \to \infty$, for most precoding schemes.
in our general multi-cell massive MIMO setup with estimated CSI. The exact expression for the power control coefficients is only given in the proof, since the main purpose of Theorem 3 is the fact that equal UL/DL performance is possible, which allows for joint analysis in what follows.

Motivated by Theorem 3, this paper considers three types of linear precoding vectors: MR precoding which amplifies the desired signal by setting \( \bar{g}_{jk} = g_{jk}^{MR} \); ZF precoding that actively rejects intra-cell interference by setting \( \bar{g}_{jk} = g_{jk}^{ZF} \); and P-ZF precoding that actively rejects both intra- and inter-cell interference by setting \( \bar{g}_{jk} = g_{jk}^{P,ZF} \). We stress that P-ZF precoding is a fully distributed coordinated beamforming scheme tailored to massive MIMO systems, since each BS only uses locally estimated CSI.

D. Finite and Asymptotic Analysis

Based on Theorems 1–3, the sum of the per-cell achievable SEs in the UL and DL are given by the following corollary.

**Corollary 1.** Looking jointly at the UL and DL, an achievable SE in cell \( j \) is

\[
\text{SE}_{j} = \text{SE}_{j}^{(ul)} + \text{SE}_{j}^{(dl)} = \sum_{k=1}^{K} \left( 1 - \frac{B S}{B} \right) \log_{2} \left( 1 + \frac{B}{I_{\text{scheme}}^{jk}} \right) \quad \text{[bit/s/Hz/cell]} \quad (27)
\]

where the interference term \( I_{\text{scheme}}^{jk} \) for UE \( k \) is given in (16) for MR, in (17) for ZF, and in (21) for P-ZF. This SE can be divided between the UL and DL arbitrarily using positive fractions \( \zeta^{(ul)} \) and \( \zeta^{(dl)} \), with \( \zeta^{(ul)} + \zeta^{(dl)} = 1 \).

This is a convenient result that allows us to analyze and optimize the SE of the network as a whole, without having to separate the UL and DL. Since it is hard to gain any further quantitative insights from the structure of the SE expression in Corollary 1, we will analyze it for a particular network topology in Section IV. In the remainder of this section, we analyze the asymptotic limit for large number of antennas.

**Corollary 2.** When \( M \to \infty \) (with \( K, B \leq S < \infty \)), the effective SINRs with MR, ZF, and P-ZF converge to the same limit:

\[
\frac{B}{I_{jk}^{MR}}, \frac{B}{I_{jk}^{ZF}}, \frac{B}{I_{jk}^{P-ZF}} \to \frac{B}{\sum_{l \in L} \sum_{m=1}^{K} \mu_{jl}^{(2)} v_l^{H} v_{lm}}. \quad (28)
\]
The effect of pilot contamination is very clear in (28), since only the UEs that interfered during pilot transmission (i.e., \( \mathbf{v}_{ijk}^H \mathbf{v}_{ilm} \neq 0 \)) affect the asymptotic limit. To maximize the asymptotic SINR in (28), we should allocate the pilot signals such that \( \mathbf{v}_{ijk}^H \mathbf{v}_{ilm} = 0 \) whenever \( \mu_{jl}^{(2)} \) is large. If \( \beta = \frac{B}{K} \) is an integer, this amounts to allocating orthogonal pilots among the UEs in each cell and making sure that only \( \frac{1}{\beta} \) of the interfering cells reuse these pilots. We refer to this as fractional pilot reuse and let \( \beta \geq 1 \) denote the pilot reuse factor. An explicit example is provided in the next section for hexagonal cells. The following result holds for any network topology.

**Corollary 3.** Let \( B = K \beta \) for some integer \( \beta \), assume orthogonal intra-cell pilot signals, and let \( \mathcal{L}_j \subset \mathcal{L} \) be the subset of cells that use the same pilots as cell \( j \). The SE in cell \( j \) approaches

\[
\text{SE}_j^\infty = K \left( 1 - \frac{K\beta}{S} \right) \log_2 \left( 1 + \frac{1}{\sum_{l \in \mathcal{L}_j \setminus \{j\}} \mu_{jl}^{(2)}} \right)
\]

when \( M \to \infty \). This SE is maximized jointly for all cells when the number of scheduled UEs is either \( K^* = \left\lfloor \frac{S}{2\beta} \right\rfloor \) or \( K^* = \left\lceil \frac{S}{2\beta} \right\rceil \) (i.e., one of the closest integers to \( \frac{S}{2\beta} \)).

**Proof:** The logarithmic part of (29) is independent of \( K \), while the concave pre-log factor \( K \left( 1 - \frac{K\beta}{S} \right) \) is maximized by \( K = \frac{S}{2\beta} \). The concavity implies that the optimal integer \( K^* \) is one of the closest integers to \( \frac{S}{2\beta} \).

Corollary 3 is a main contribution of this paper and proves that the number of scheduled UEs should be proportional to the frame length \( S \) (when \( M \) is large enough); for example, we get \( K^* = \frac{S}{2} \) for \( \beta = 1 \) and \( K^* = \frac{S}{6} \) for \( \beta = 3 \). Since both \( S = 200 \) and \( S = 10000 \) are reasonable coherence block lengths in practice, depending on the UE mobility and propagation environment, this means that we should schedule between tens and several thousands of UEs for simultaneous transmission in order to be optimal. This is only possible if the UE selection policy is scalable and there is a high load of UEs. If \( K^* = \frac{S}{2\beta} \) is an integer, the asymptotically optimal SE is

\[
\text{SE}_j^\infty = \frac{S}{4\beta} \log_2 \left( 1 + \frac{1}{\sum_{l \in \mathcal{L}_j \setminus \{j\}} \mu_{jl}^{(2)}} \right)
\]

and increases linearly with the frame length \( S \) (in the large-\( M \) regime).

Interestingly, the asymptotically optimal scheduling gives \( B = \frac{S}{2} \) for any \( \beta \), which means that half the frame is allocated to pilot transmission. This extraordinary fact was initially conjectured in [2] for \( \beta = 1 \). At first sight, it bears some similarity with the results in [27] and [28] for block-fading noncoherent point-to-point (P2P) MIMO channels, where the maximal degrees of
freedom (DoF) are $\frac{S}{4}$ and are achieved by having $\frac{S}{2}$ transmit/receive antennas and using pilot signals of the same length. The fundamental difference is that the DoF concept for P2P MIMO channels, where unbounded SE is achieved at high SNRs, does not apply to cellular networks [29]. Instead, the pre-log factor $\frac{S}{4\beta}$ in (30) may be interpreted as the relative improvement in SE that can be achieved by aggressive scheduling of UEs in massive MIMO systems.

We have now established the asymptotically optimal number of scheduled UEs, as $M \rightarrow \infty$. Next, we investigate the impact on practical systems with finite $M$ for a certain network topology.

IV. Optimizing Number of UEs in Hexagonal Networks

The concept of cellular communications has been around for decades [30]. Although practical deployments have irregular cells, it is common practice to establish general properties by analyzing symmetric networks where the cells are regular polygons [31]; in particular, hexagons.

In this section, we consider the symmetric network topology depicted in Fig. 2 with hexagonal cells. All the time/frequency resources allocated to payload data transmission are used in all the cells. However, inspired by [15], we consider pilot books of size $B = \beta K$ to allow for fractional pilot reuse that mitigates the pilot contamination from neighboring cells.

The hexagonal grid is infinitely large, to avoid edge effects and to give all cells the same properties. The cell radius is denoted by $r > 0$ and is the distance from the cell center to the corners. Each cell can be uniquely indexed by a pair of integers $\alpha_{j}^{(1)}, \alpha_{j}^{(2)} \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. This integer pair specifies the location of BS $j$ [30]:

$$b_{j} = \sqrt{3} \begin{bmatrix} \frac{\sqrt{3}r}{2} \\ \frac{r}{2} \end{bmatrix} \alpha_{j}^{(1)} + \begin{bmatrix} 0 \\ \sqrt{3}r \end{bmatrix} \alpha_{j}^{(2)} \in \mathbb{R}^2. \quad (31)$$

The coordinate system imposed by $\alpha_{j}^{(1)}$ and $\alpha_{j}^{(2)}$ is illustrated in Fig. 3.
Every cell on the hexagonal grid has 6 interfering cells in the first surrounding tier, 12 in the second tier, etc. As shown in the early works on hexagonal networks \cite{30}, \cite{31}, this limits which pilot reuse factors that give symmetric reuse patterns: \( \beta \in \{1, 3, 4, 7, 9, 12, 13, \ldots \} \).

Our simulations consider a classic pathloss model where the variance of the channel attenuation in (1) is \( d_j(z) = \frac{C}{\|z - b_j\|^\kappa} \), where \( \| \cdot \| \) is the Euclidean norm, \( C > 0 \) is a reference value, and \( \kappa \geq 2 \) is the pathloss exponent. These assumptions allow us to compute \( \mu_{jl}^{(\omega)} \) in (18) as

\[
\mu_{jl}^{(\omega)} = \mathbb{E}_{z_{lm}} \left\{ \left( d_j(z_{lm}) \right)^\omega \right\} = \mathbb{E}_{z_{lm}} \left\{ \left( \frac{\|z_{lm} - b_j\|}{\|z_{lm} - b_l\|} \right)^{\kappa \omega} \right\}
\]

for any UE distributions in the cells. We notice that \( C \) and \( r \) cancel out in (32), if the UE distribution is invariant to the radius. Since the power control makes the SEs in Theorems 1–3 independent of the UEs’ positions, we only need to define the parameter ratio \( \rho/\sigma^2 \); that is, the average SNR (over fading) between any UE and any antenna at its serving BS.

Fig. 4: Simulation of optimized SE, as a function of \( M \), with average inter-cell interference.
A. Optimizing SE for different interference levels

We simulate the SE in an arbitrary cell on the hexagonal grid in Fig. 2 and take all non-negligible interference into account. The UEs can be anywhere in the cells, but at least 0.14r from the serving BS (this makes the analysis independent of r). Since the SE expressions in Section III are the same for the UL and DL, except for the fractions $\zeta^{(ul)}$ and $\zeta^{(dl)}$, we simulate the sum of these SEs and note that it can be divided arbitrarily between the UL and DL. The same linear processing schemes are used in both directions. The simulations consider MR, ZF, and P-ZF precoding/combining, and all results are obtained by computing the closed-form expressions from Section III for different parameter combinations.

For each number of antennas, $M$, we optimize the SE with respect to the number of UEs $K$.
and the pilot reuse factor $\beta$ (which jointly determine $B = \beta K$). We set the coherence block length to $S = 400$ (e.g., achieved by 2 ms coherence time and 200 kHz coherence bandwidth), set the SNR to $\rho/\sigma^2 = 5$ dB, and pick $\kappa = 3.7$ as the pathloss exponent. The impact of changing the different system parameters is investigated in Section IV-B.

We consider three propagation environments with different severity of inter-cell interference:

1) Average case: Averaging over uniform UE locations in all cells.

2) Best case: All UEs in other cells are at the cell edge furthest from BS $j$ (for each $j$).

3) Worst case: All UEs in other cells are at the cell edge closest to BS $j$ (for each $j$).

The corresponding values on the parameters $\mu_1^{(1)}$ and $\mu_2^{(2)}$ were computed by Monte-Carlo simulations with $10^6$ UE locations in each cell.

The best case is overly optimistic since the desirable UE positions in the interfering cells are different with respect to different cells. However, it gives an upper bound on what is achievable by coordinated scheduling across cells. The worst case is overly pessimistic since the UEs cannot all be at the worst locations, with respect to all other cells, at the same time. The average case is probably the most applicable in practice, where the averaging comes from UE mobility, scheduling, and (pseudo-)random switching of pilot sequences between the UEs in each cell. Results for the average case are shown in Fig. 4, the best case in Fig. 5, and the worst case in Fig. 6. The optimized SE and the corresponding $K^*$ are shown in (a) and (b), respectively.

The achievable SEs (per cell) are very different between the best case interference and the two other cases—this confirms the fact that results from single-cell analysis of massive MIMO is often not applicable to multi-cell cases (and vice versa). ZF brings much higher SEs than MR under the best case inter-cell interference, since then the potential gain from mitigating intra-cell interference is very high. P-ZF is equivalent to ZF in the best case, but excels under worst case inter-cell interference since it can actively suppress also inter-cell interference. In the realistic average case, the optimized SEs are rather similar for MR, ZF, and P-ZF; particularly in the practical range of $10 \leq M \leq 200$ antennas. In all cases, the largest differences appear when the number of antennas is very large (notice the logarithmic $M$-scales). At least $M = 10^5$ antennas are needed to come close to the asymptotic limit in (30), and many more antennas are required under best case interference. Clearly, the asymptotic limits should generally not be

---

A higher pathloss exponent reduces the inter-cell interference, but requires more signal power to maintain a certain SNR.
used as performance indicators since unrealistically many antennas are needed for convergence.

As seen from Figs. 4–6, the main difference between MR, ZF, and P-ZF is not the values of the optimized SE but how they are achieved; that is, which number of UEs $K^*$ and which pilot reuse factor $\beta$ that are used. The general behavior is that larger $M$ implies a higher $K^*$ and a smaller $\beta$, because the channels become more orthogonal with $M$. Since the reuse factor is an integer, $K^*$ changes non-continuously when $\beta$ is changed; smaller $\beta$ allows for larger $K^*$, and vice versa. MR schedules the largest number of UEs and switches to a smaller reuse factor at fewer antennas than the other schemes. In contrast, P-ZF schedules the smallest number of UEs and has the highest preference of large reuse factors, since this it can suppress more inter-cell interference in these cases. Simply speaking, MR gives low per-user SEs to many UEs (sometimes more than $M$), while ZF and P-ZF give higher per-user SEs to fewer UEs.

Recall from Corollary 3 that $K = \frac{S}{2\beta}$ becomes the optimal number of UEs as $M \to \infty$. This property is confirmed by Figs. 4–6 since $K^* \to 67$ in the average case (where $\beta = 3$), $K^* \to 200$ in the best case (where $\beta = 1$), and $K^* \to 50$ in the worst case (where $\beta = 4$).

B. Impact of System Parameters

We now focus on the average case of inter-cell interference, due to its practical relevance, and investigate how each system parameter affects the simulation results. We focus on the range $10 \leq M \leq 1000$ antennas, and when other system parameters than $M$ are varied we only consider $M = 100$ (medium massive MIMO setup) and $M = 500$ (large massive MIMO setup).

We begin by studying the impact of the pilot reuse factor $\beta$. Fig. 7 shows the per-cell SE for $\beta = 1$ and $\beta = 3$, which provide the highest SEs for $M \leq 1000$. The curves are smooth and there are wide regions around the $\beta$-switching points where both values provide almost equal SEs. This robustness simplifies cell planning and scheduling based on user load.

Changes in the pilot reuse factor have major impact on the optimal number of UEs and their achievable performance. The SE per UE is shown in Fig. 8 for the operating points that maximize the SE in the cell; this is basically the ratio $SE/K^*$, where $SE$ was given in Fig. 4(a) and $K^*$ was given in Fig. 4(b). We notice that MR gives the lowest SE per scheduled UE, while P-ZF gives the highest SE per scheduled UE. The numbers are around 1 bit/s/Hz for MR, in the range 1–2.5 bit/s/Hz for ZF, and in the range 1–3 bit/s/Hz for P-ZF. Since the pilot signaling consumes between 2 and 40 percent of the frame in this simulation, the payload data need to be encoded.
with up to 4.5 bit/symbol, which can be achieved by conventional 64-QAM with a 3/4 coding rate. Hence, all the per-user SEs in Fig. 8 are straightforward to implement in practice.

Fig. 9 shows the ratio $M/K^*$ for the same scenario as in the previous figures. This ratio can be interpreted as the number of BS antennas per UE [7]. There is a common rule of thumb which says that massive MIMO systems should have an order of magnitude more BS antennas than UEs. The operating points that satisfy this guideline are above the horizontal dotted line. This simulation indicates that an optimized system might not follow this guideline; in fact, there is a few occasions where MR even prefers to have $M/K^* < 1$. Generally speaking, it seems that having 2–8 times more BS antennas than UEs is the range to aim at for practical deployments.

Since the cells might not be fully loaded at every time instant, Fig. 10 shows the per-cell SE as a function of the number of scheduled UEs. As noted before, the peak numbers (which are star marked) are at different $K$ for each scheme. If MR, ZF, and P-ZF are compared for a given $K$, the differences between the schemes can either be larger or smaller than at the peak numbers. Although ZF and P-ZF often provide better SE than MR, it is interesting to note that MR is competitive when $K$ is large—both in terms of SE and since its computational complexity scales as $O(MK^2)$, while the complexity of ZF and P-ZF scales as $O(MK^2)$ [9].

Next, Fig. 11 investigates how the average SNR $\rho/\sigma^2$ affects the results. The SE saturates already at an SNR of 5 dB due to the array gain from coherent processing—this is why 5 dB was used in the previous figures. Massive MIMO can operate also at lower SNRs, but with a performance loss. ZF and P-ZF are particularly sensitive to the SNR level, since the active interference suppression requires a higher CSI estimation quality than simple MR processing.
Finally, Fig. 12 investigates how the length of the coherence block, $S$, affects the per-cell SE. In the case of $M = 100$ antennas, the gain of increasing $S$ above 500 is relatively small—the system cannot schedule more UEs since the ratio $M/K$ would then be too small, so the gain mainly comes from reducing the prelog factor $\left(1 - \frac{B}{S}\right)$. However, in the case of $M = 500$, the system can utilize an increasing $S$ to schedule more UEs and achieve major improvements in SE. As the number of UEs increases, the part of the intra-cell interference that cannot be rejected due to imperfect CSI becomes the main limiting factor. The benefit of P-ZF then diminishes.

V. SPECTRAL EFFICIENCIES WITH HARDWARE IMPAIRMENTS

The analytic and numeric analysis in the previous sections have focused on cellular networks where the BSs and UEs are equipped with ideal transceiver hardware, which can radiate any waveform without distortions and which can receive any waveform with infinite resolution. However, practical transceivers inevitably operate with non-linearities in amplifiers, clock drifts
in local oscillators, finite-precision ADCs, I/Q imbalance in mixers, and finite-order analog filters [25], [32]–[35]. In this section, we provide a prediction of how these hardware impairments affect the achievable SEs in multi-cell massive MIMO systems. It was recently shown in [12] that the hardware impairments caused by the BS array are negligible in massive MIMO systems, since the desired signals are amplified by the array gain from coherent processing while the distortions add non-coherently. Hence, the hardware impairments in the UE hardware are expected to be the main hardware limitation [12] and is thus the one considered in this section.

Similar to [33]–[35], we model the hardware impairments as a reduction of the original signals by a factor $\sqrt{1-\epsilon^2}$ and replacing it with Gaussian distortion noise that carries the removed power. More precisely, the UL system model in (2) is generalized as

$$y_j = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} h_{jlk} \left( \sqrt{(1-\epsilon^2)p_{lk}x_{lk}} + \varepsilon_{lk} \right) + n_j, \tag{33}$$

where $\varepsilon_{lk} \sim \mathcal{CN}(0, \epsilon^2 p_{lk})$ is the UL distortion noise caused at UE $k$ in cell $l$, and the DL system model in (3) is generalized as

$$z_{jk} = \sqrt{1-\epsilon^2} \left( \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} h_{lk}^T w_{lm} s_{lm} + \eta_{jk} \right) + \epsilon_{jk}, \tag{34}$$

where $\epsilon_{jk} \sim \mathcal{CN}(0, \epsilon^2 (\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \| h_{jk}^T w_{lm} \|^2 + \sigma^2))$ is the DL distortion noise caused at UE $k$ in cell $j$. Notice that $\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \| h_{jk}^T w_{lm} \|^2 + \sigma^2$ is the power of the term in parenthesis in (34). The parameter $\epsilon$ determines the level of impairments and can be interpreted as the error vector magnitude (EVM) [33]; typical values in LTE are in the range $0 \leq \epsilon \leq 0.17$ [36]. Based on these generalized system models, the following counterpart of Lemmas 2 and 3 is obtained.
Lemma 4. Under hardware impairments, a jointly achievable SE in the UL and DL of UE $k$ in cell $j$ is

$$
(1 - \frac{B}{S}) \mathbb{E}_z \left\{ \log_2 (1 + \widetilde{\text{SINR}}_{jk}) \right\} \quad \text{[bit/s/Hz]} \tag{35}
$$

with the effective SINR given by

$$\widetilde{\text{SINR}}_{jk} = \frac{(1 - \epsilon^2)p_{jk}|\mathbb{E}_h \{g^H_{jk}h_{jk}\}|^2}{\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} p_{lm} \mathbb{E}_h \{|g^H_{jlm}h_{jlm}|^2\} - (1 - \epsilon^2)p_{jk}|\mathbb{E}_h \{g^H_{jk}h_{jk}\}|^2 + \sigma^2 \mathbb{E}_h \{||g_{jk}||^2\}}. \tag{36}
$$

**Proof:** The proof is given in the appendix. ■

The SE expression in Lemma 4 resembles our previous results in Section III with the only differences that there is a loss in desired signal power by a factor $(1 - \epsilon^2)$ and that this power is turned into self-interference in the denominator of the SINR. Under the assumption of MR, ZF, or P-ZF processing in the UL and DL, we have the following closed-form SE expression.

**Theorem 4.** Looking jointly at the UL and DL, an achievable SE in cell $j$ under hardware impairments is

$$\text{SE}_j = \sum_{k=1}^{K} \left( 1 - \frac{B}{S} \right) \log_2 \left( 1 + \frac{B(1 - \epsilon^2)}{I_{jk}^{\text{scheme}} + B\epsilon^2} \right) \quad \text{[bit/s/Hz/cell]} \tag{37}
$$

where the interference term $I_{jk}^{\text{scheme}}$ for UE $k$ is given in (16) for MR, in (17) for ZF, and in (21) for P-ZF. If $M \to \infty$ (with $K, B \leq S < \infty$), the effective SINRs with these processing schemes approach the upper limit

$$\frac{B(1 - \epsilon^2)}{\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \mu_{jl}^{(2)} v^H_{ijkl} v_{ilm} + B\epsilon^2}. \tag{38}
$$

**Proof:** This result follows straightforwardly from Theorems 1–3 since the SINR expressions in (36) only differ from those in Section III by the $(1 - \epsilon^2)$-factors. ■

Using the tractable SE expression in Theorem 4 for simulation, Fig. 13 shows the per-cell SE in the average inter-cell interference. This figure shows results for ideal hardware with $\epsilon = 0$ (as in Fig. 4(a)) and for hardware impairments with $\epsilon = 0.1$, which is a large EVM number in these contexts [36]. Interestingly, there is only a tiny difference in SE for $M < 5000$, mainly because the SE per UE is relatively small at the optimized operating points and thus the distortion noise is only a minor limiting factor. For higher number of antennas, the difference is substantial.
because of the asymptotic limits for ideal hardware in (28) and for hardware impairments in (38) are different. We conclude that hardware impairments seem to have a small impact on practical massive MIMO systems, which have been optimized for high SE.

VI. CONCLUSION

This paper investigated how many UEs, $K$, that should be scheduled in massive MIMO systems to maximize the SE per cell for a fixed $M$. Conventional SE expressions are strongly dependent on the UE positions, which makes it hard to optimize $K$. In contrast, we derived new SE expressions that are independent of the instantaneous UE positions, due to power control and averaging over random UE locations. In fact, the new expressions are the same for the UL and DL, which allows for joint network optimization. When applied to symmetric network topologies, where each cell is representable for any cell, these expressions can directly provide the network-wide performance—which otherwise would require extensive Monte-Carlo simulations.

The analytic results treat MR and ZF processing and a new distributed cooperation scheme, coined P-ZF, that suppresses inter-cell interference by listening to the pilot transmissions from neighboring cells. The asymptotic analysis shows that the SE-optimal $K^*$ approaches $\frac{S}{\beta}$ as $M \to \infty$, irrespective of the processing scheme. Hence, $B = \beta K^* \to \frac{S}{2}$ which means that half the frame should be spent on pilot signaling when $M$ is large enough. The corresponding asymptotic SE limit is not reached for practical $M$, but an unconventionally large fraction of the frame should still be allocated for pilots: 5% to 40% were observed in simulations for $M \leq 1000$.

Generally speaking, high per-cell SEs are achieved by scheduling many UEs for simultaneous transmission, while the SE per UE might only be 1–4 bit/s/Hz. P-ZF gives the highest performance per UE, while MR gives the lowest SE per UE. In contrast, MR schedules the largest number of UEs and P-ZF the smallest number. ZF processing is often the best choice in terms of per-cell SE, thus showing that the inter-cell interference suppression offered by P-ZF is only needed in special cases with strong inter-cell interference. The extensive simulations show that massive MIMO prefers an SNR of 0–5 dB, that a fractional pilot reuse of $\beta = 3$ is often a decent choice, and that the technology is very robust to distortion noise from hardware impairments. Based on the simulations, we notice that massive MIMO with $M = 100$ can easily achieve a $10\times$ gain in SE over the IMT-Advanced requirement of 3 bit/s/Hz/cell. For large arrays with $M = 500$ antennas, massive MIMO can even provide a $40\times$ gain over IMT-Advanced.
APPENDIX: COLLECTION OF PROOFS

Proof Lemma[1]: As shown in [37], Theorem 11.1, the expression for an MMSE estimator with jointly Gaussian channels and (colored) noise is

\[ \hat{h}_{jlk} = \mathbb{E}_h \{ h_{jlk} \vec{\text{vec}}(Y_j) \} (\mathbb{E}_h \{ \vec{\text{vec}}(Y_j) \vec{\text{vec}}(Y_j)^H \})^{-1} \vec{\text{vec}}(Y_j), \]  

where \( \vec{\text{vec}}(\cdot) \) denotes vectorization. Direct algebraic computation using the vectorization rule \( (C^T \otimes A)\vec{\text{vec}}(B) = \vec{\text{vec}}(ABC) \), where \( \otimes \) is the Kronecker product, shows that

\[ \mathbb{E}_h \{ h_{jlk} \vec{\text{vec}}(Y_j)^H \} = \mathbb{E}_h \{ h_{jlk} \hat{h}_{jik}^H (\mathbf{v}_{ik}^H \otimes I_M) \} = \left( \mathbf{v}_{ik}^H \otimes \frac{d_j(z_{lk})}{d_l(z_{ik})} \mathbf{I}_M \right) \]  

since the channels are independent. Similarly, the mutual independence of the UE channels implies that

\[ \mathbb{E}_h \{ \vec{\text{vec}}(Y_j) \vec{\text{vec}}(Y_j)^H \} = \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \mathbb{E}_{h_{\ell m}} \{ \vec{\text{vec}}(h_{j\ell m}^H \mathbf{v}_{\ell m}) \vec{\text{vec}}(h_{j\ell m}^H \mathbf{v}_{\ell m}^H) \} + \sigma^2 \mathbf{I}_{MB} \]

\[ = \left( \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} d_j(z_{\ell m}) d_l(z_{\ell m}) \mathbf{v}_{\ell m} \mathbf{v}_{\ell m}^H + \sigma^2 \mathbf{I}_B \right) \otimes \mathbf{I}_M. \]

The expression (6) is obtained by substituting (40) and (41) into (39), normalizing by \( \rho \) and using the vectorization rule above. According to the definition in (8), the error covariance matrix \( \mathbf{C}_{jlk} \) is given by

\[ \rho \frac{d_j(z_{lk})}{d_l(z_{ik})} \mathbf{I}_M - \mathbb{E}_h \{ h_{jlk} \vec{\text{vec}}(Y_j)^H \} (\mathbb{E}_h \{ \vec{\text{vec}}(Y_j) \vec{\text{vec}}(Y_j)^H \})^{-1} \mathbb{E}_h \{ h_{jlk} \vec{\text{vec}}(Y_j)^H \} \]

\[ = \rho \frac{d_j(z_{lk})}{d_l(z_{ik})} \left( 1 - \frac{d_j(z_{lk})}{d_l(z_{ik})} \mathbf{v}_{ik}^H \Psi_j^{-1} \mathbf{v}_{ik} \right) \mathbf{I}_M = \rho \frac{d_j(z_{lk})}{d_l(z_{ik})} \left( 1 - \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \frac{d_j(z_{\ell m}) B}{d_l(z_{\ell m})} \mathbf{v}_{\ell m} \mathbf{v}_{\ell m}^H + \sigma^2 \right) \mathbf{I}_M, \]

where the last equality follows from the fact that the pilot signals form an orthogonal basis.

Proof Theorem[1]: The first step for MR combining is to compute the expectations in (12) with respect to the channel realizations. These are obtained from [13], Corollary 2 by setting \( \kappa = \delta = 0, \xi = \sigma^2, \) and \( \lambda_{jlm} = \frac{d_j(z_{lm})}{d_l(z_{lm})}. \) Plugging these expressions into (12) yields, for MR,

\[ \text{SINR}_{jlk}^{(ul)} = \frac{\mathbf{v}_{ijk}^H \Psi_j^{-1} \mathbf{v}_{ijk}}{\sum_{l \in \mathcal{L}} \sum_{m=1}^{M} \left( \frac{d_j(z_{lm})}{d_l(z_{lm})} \right) \left( \frac{d_j(z_{lm})}{d_l(z_{lm})} \right)^2 \mathbf{v}_{ijk}^H \Psi_j^{-1} \mathbf{v}_{ilm} - \mathbf{v}_{ijk}^H \Psi_j^{-1} \mathbf{v}_{ijk} + \frac{\sigma^2}{M \rho} \right), \]

by multiplying each term by \( \frac{1}{M^2 \rho^2 \mathbf{v}_{ijk}^H \Psi_j^{-1} \mathbf{v}_{ijk}}. \) The expression in (16) is now obtained by considering an achievable lower bound \( \mathbb{E}_z \{ \log_2(1 + \frac{1}{f(z)}) \} \geq \log_2(1 + \frac{1}{\mathbb{E}_z(1/(f(z)))}) \) where the
expectation with respect to user positions are moved to the denominator of the SINRs. This leads to expectations of the following types:

\[
\mathbb{E}_z \left\{ \frac{1}{v_{ij}^H \Psi_{j}^{-1} v_{ij}} \right\} = \mathbb{E}_z \left\{ \frac{\sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \frac{d_{j}(z_{lm})}{d_{i}(z_{lm})} v_{ij}^H v_{im} + \sigma^2}{B} \right\}
\]

\[
= \frac{\sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \mu_{j\ell}^{(1)} v_{ij}^H v_{im} + \sigma^2}{B}
\]

\[
= \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \frac{d_{j}(z_{lm})}{d_{i}(z_{lm})} \mu_{j\ell}^{(1)} v_{ij}^H v_{im} + \sigma^2 \rho
\]

\[
= \frac{\mu_{j}^{(2)} + (\mu_{j}^{(1)})^2}{B} v_{ij}^H v_{im}
\]

\[
(44)
\]

where we have utilized the definition in (18) to identify all expectations.

The expectations in (12) with respect to the channel realizations for ZF are

\[
\mathbb{E}_h \{ ||g_{jk}^{ZF}||^2 \} = \frac{1}{(M-K)\rho v_{ij}^H \Psi_{j}^{-1} v_{ij}}
\]

\[
p_{jk}|\mathbb{E}_h \{ (g_{jk}^{ZF})^H h_{ij} \}|^2 = 1
\]

\[
p_{jm}|\mathbb{E}_h \{ (g_{jk}^{ZF})^H h_{jm} \}|^2 = \left( \frac{d_{j}(z_{jm})}{d_{i}(z_{jm})} \right)^2 v_{ij}^H v_{im} + \frac{\rho d_{j}(z_{jm})}{d_{i}(z_{jm})} \left( 1 - A_{jm}^{ZF} \frac{d_{j}(z_{jm})}{d_{i}(z_{jm})} \right) v_{ij}^H \Psi_{j}^{-1} v_{im}
\]

\[
= \frac{\mu_{j}^{(2)} + (\mu_{j}^{(1)})^2}{B} v_{ij}^H v_{im}
\]

\[
(45)
\]

\[
(46)
\]

\[
(47)
\]

\[
(48)
\]

with \( A_{jm}^{ZF} = \frac{1}{B} \sum_{k=1}^{K} v_{jk}^H v_{im} \), where (46) follows from the definition of ZF and by utilizing well-known properties of Wishart matrices (see e.g., [8, Proof of Proposition 2]) and (47) is a consequence of the ZF principle. The first term in (48) follows from (47) whenever \( v_{ij}^H v_{im} \neq 0 \) (i.e., when the same pilot signal is used). The second term is the product between \( \mathbb{E}_h \{ ||g_{jk}^{ZF}||^2 \} \) and the variance of the estimation error of the effective channel \( \sqrt{P_{lm}} h_{jm} \) if \( A_{jm}^{ZF} \neq 0 \) (i.e., if UE \( m \) in cell \( l \) uses any of the pilots that are used in cell \( j \) so that the ZF applies to it) or the original variance of \( \sqrt{P_{lm}} h_{jm} \) if \( A_{jm}^{ZF} = 0 \). Using (46)–(48), we obtain the expression

\[
\text{SINR}_{jk}^{(ul)} = \frac{1}{\sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \left( \frac{d_{j}(z_{lm})}{d_{i}(z_{lm})} \right)^2 v_{ij}^H v_{im} + \frac{\rho d_{j}(z_{jm})}{d_{i}(z_{jm})} \left( 1 - A_{jm}^{ZF} \frac{d_{j}(z_{jm})}{d_{i}(z_{jm})} \right) v_{ij}^H \Psi_{j}^{-1} v_{im}} - 1 + \frac{\sigma^2}{(M-K)\rho v_{ij}^H \Psi_{j}^{-1} v_{ij}}
\]

\[
(49)
\]

for ZF. Finally, the achievable SE in the theorem is obtained by using Jensen’s inequality in the
same way as for MR, where the expectation in (44) and (45) reappear along with

\[
\mathbb{E}_z \left\{ -\left( \frac{d_j(z_{lm})}{d_l(z_{lm})} \right)^2 \frac{v_{im}^H \Psi_j^{-1} v_{ilm}}{v_{jjk}^H \Psi_j^{-1} v_{jjk}} \right\} \leq -\left( \mu_j^{(1)} \right)^2 \sum_{l \in \mathcal{L}} \frac{1}{K} \sum_{m=1}^K \frac{\mu_j^{(1)} v_{ilm}^H v_{ilm}}{\rho v_{jjk}^H \Psi_j^{-1} v_{jjk}} + \frac{\sigma^2}{\rho},
\]

(50)

where the inequality is once again from Jensen’s inequality.

**Proof Theorem 2:** Similar to the proof of Theorem 1 for P-ZF we obtain

\[
\mathbb{E}_h \{ \| g_j^{\text{P-ZF}} \|^2 \} = \frac{1}{(M-B) \rho v_{jjk}^H \Psi_j^{-1} v_{jjk}}
\]

(51)

\[
P_{jk} | \mathbb{E}_h \{ (g_j^{\text{P-ZF}})^H h_{jjk} \} |^2 = 1
\]

(52)

\[
P_{lm} | \mathbb{E}_h \{ (g_j^{\text{P-ZF}})^H h_{jm} \} |^2 = \left( \frac{d_j(z_{lm})}{d_l(z_{lm})} \right)^2 \frac{v_{jjk}^H v_{ilm}}{B} + \frac{\rho d_j(z_{lm})}{(M-B) \rho v_{jjk}^H \Psi_j^{-1} v_{jjk}} \left( 1 - \frac{d_j(z_{lm})}{d_l(z_{lm})} \right) \frac{v_{ilm}^H \Psi_j^{-1} v_{ilm}}{(M-B) \rho v_{jjk}^H \Psi_j^{-1} v_{jjk}}
\]

(53)

by following the procedures used for ZF. Using (51)–(53), we obtain the expression

\[
\text{SINR}_{jk}^{(ul)} = \frac{1}{\sum_{l \in \mathcal{L}} \sum_{m=1}^K \left( \frac{d_j(z_{lm})}{d_l(z_{lm})} \right)^2 \frac{v_{jjk}^H v_{ilm}}{B} + \frac{\rho d_j(z_{lm})}{(M-B) \rho v_{jjk}^H \Psi_j^{-1} v_{jjk}} \left( 1 - \frac{d_j(z_{lm})}{d_l(z_{lm})} \right) \frac{v_{ilm}^H \Psi_j^{-1} v_{ilm}}{(M-B) \rho v_{jjk}^H \Psi_j^{-1} v_{jjk}} - 1 + \frac{\sigma^2}{(M-B) \rho v_{jjk}^H \Psi_j^{-1} v_{jjk}}}
\]

(54)

for P-ZF. The final expression is obtained by considering an achievable lower bound \( \mathbb{E}_z \{ \log_2 (1 + \frac{1}{f(x)}) \} \geq \log_2 (1 + \frac{1}{\mathbb{E}_z \{ f(x) \}}) \) using Jensen’s inequality, similar to the ZF case in Theorem 1.

**Proof Theorem 3:** Suppose that \( \gamma_{jk} = \text{SINR}_{jk}^{(ul)} \) is the UL SINR value achieved by UE \( k \) in cell \( j \) for a given receive combining scheme. The goal of the proof is to show that we can also achieve \( \gamma_{jk} = \text{SINR}_{jk}^{(dl)} \) for the DL SINR in (24). This condition can also be expressed as

\[
\frac{\gamma_{jk} \mathbb{E}_h \{ \| \mathbf{g}_j \|^2 \}}{\mathbb{E}_h \{ \| \mathbf{g}_j \|^2 \}} = q_{jk} \frac{\mathbb{E}_h \{ \| \mathbf{h}_{jjk} \|^2 \}}{\mathbb{E}_h \{ \| \mathbf{g}_j \|^2 \}} + \frac{q_{jk} \mathbb{E}_h \{ \| \mathbf{h}_{jjk} \|^2 \}}{\mathbb{E}_h \{ \| \mathbf{g}_j \|^2 \}} + \sigma^2.
\]

(55)

We define the \( K |\mathcal{L}| \times K |\mathcal{L}| \) block matrix \( \Psi \), where each block is \( K \times K \) and the \( (j,l) \)th block is denoted \( \Psi_{jl} \). Its \( (k,m) \)th element is given by

\[
[\Psi_{jl}]_{k,m} = \begin{cases} \frac{\mathbb{E}_h \{ \| \mathbf{g}_{j}^H \mathbf{h}_{jjk} \|^2 \}}{\mathbb{E}_h \{ \| \mathbf{g}_{j} \|^2 \}} - \frac{\mathbb{E}_h \{ \| \mathbf{h}_{jjk} \|^2 \}}{\mathbb{E}_h \{ \| \mathbf{g}_{j} \|^2 \}} & \text{if } k = m, j = l, \\ \frac{\mathbb{E}_h \{ \| \mathbf{g}_{j}^H \mathbf{h}_{jjk} \|^2 \}}{\mathbb{E}_h \{ \| \mathbf{g}_{j} \|^2 \}} & \text{otherwise}. \end{cases}
\]

(56)
Moreover, we define the $K|L| \times K|L|$ block diagonal matrix $D$, where the $j$th $K \times K$ block is $D_j$ and its $k$th diagonal element is 

$$[D_j]_{k,k} = \gamma_{jk} \frac{E\{h_j\} \|\hat{g}_{jk}\|^2}{|E\{h_j\} \hat{g}_{jjk}|^2}. \quad (57)$$

Using this notation, (55) can be expressed as 

$$[D_j]_{k,k} = \frac{q_{jk}}{\sum_l \sum_{m=1}^K q_{lm} [\Psi_{jl}]_{k,m}} \Leftrightarrow [D_j]_{k,k} \sigma^2 = q_{jk} - \sum_{l \in L} \sum_{m=1}^K q_{lm} [D_j]_{k,k} [\Psi_{jl}]_{k,m}. \quad (58)$$

In matrix form, the DL SINR conditions for all UEs in all cells can be expressed as $D \sigma^2 = q - D \Psi q$, where $q = [q_1^T \ldots q_{|L|}^T]^T$ and $q_j = [q_{j1} \ldots q_{jK}]^T$ contain the DL transmit powers in the $j$th cell. This expression can now be solved for $q$. The matrix $D$ depends only on the precoding vectors, thus for any choice of precoding scheme the sought SINRs are achieved by the power control policy 

$$q^* = \sigma^2 (I_{K|L|} - D \Psi)^{-1} 1$$

where $1$ is the vector with only ones. $q^*$ is a feasible power control (i.e., has positive values) if all eigenvalues of $(I_{K|L|} - D \Psi)$ are larger than zero. We need to show that this always holds. We notice that the UL SINR condition, which is satisfied by assumption, can be expressed in a similar matrix form where $\Psi$ is replaced by $\Psi^T$:

$$D \sigma^2 = p - D \Psi^T p \Leftrightarrow p = \sigma^2 (I_{K|L|} - D \Psi^T)^{-1} 1,$$ 

where $p = [p_1^T \ldots p_{|L|}^T]^T$ and $p_j = [p_{j1} \ldots p_{jK}]^T$, if $\hat{g}_{jk} = \hat{g}_{jk}^{\text{scheme}}$ for all $j$ and $k$. Since the eigenvalues of $(I_{K|L|} - D \Psi)$ and $(I_{K|L|} - D \Psi^T)$ are the same, we can always select the DL powers according to (59). It is straightforward to verify that $1^T q^* = p^T 1$, thus the total transmit power is the same in the DL and UL. Since the same SINRs as in the UL are achieved in the DL for any UE positions, the SE in (26) follows directly from Theorems 1 and 2.

**Proof Lemma 4**: The derivations of Lemmas 2 and 3 are based on the following principle: if we receive $s + n$, where $s$ is a Gaussian information signal and $n$ is uncorrelated noise, then an achievable SE is $\log_2 \left( 1 + \frac{|E\{|s|^2\}|^2}{E\{|n|^2\} + |E\{|s|^2\}|E\{|s|^2\}} \right) \quad [24]$. For the hardware impairment models in (33) and (34), the received signals (after linear processing) behave as $\sqrt{1 - \epsilon^2} (s + n) + \epsilon \eta$ instead, where $E\{|\eta|^2\} = E\{|n|^2\} + E\{|s|^2\}$. Since the distortion $\eta$ is uncorrelated with $s$ and $n$ by assumption, $n + \eta$ is also uncorrelated with $s$ and the corresponding achievable SE is
\[
\log_2 \left( 1 + \frac{(1 - \epsilon^2)\mathbb{E}\{|s|^2\}}{(1 - \epsilon^2) (\mathbb{E}\{|n|^2\} + \mathbb{E}\{|s|^2\} - |\mathbb{E}\{s\}|^2) + \epsilon^2 (\mathbb{E}\{|n|^2\} + \mathbb{E}\{|s|^2\})} \right)
\]

= \log_2 \left( 1 + \frac{(1 - \epsilon^2)\mathbb{E}\{|s|^2\}}{\mathbb{E}\{|n|^2\} + \mathbb{E}\{|s|^2\} - (1 - \epsilon^2)|\mathbb{E}\{s\}|^2} \right). \quad (61)

The only impact of the distortion is thus the \((1 - \epsilon^2)\)-factors in front of \(\mathbb{E}\{s\}\) in the numerator and denominator. The UL SINRs in (36) follow directly from this observation, while the DL SINRs are achieved by also utilizing the power control policy from the proof of Theorem 3.

REFERENCES

[1] Nokia Siemens Networks, “2020: Beyond 4G radio evolution for the Gigabit experience,” White Paper, Tech. Rep., 2011.
[2] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[3] R. Baldemair, E. Dahlman, G. Fodor, G. Mildh, S. Parkvall, Y. Selén, H. Tullberg, and K. Balachandran, “Evolving wireless communications: Addressing the challenges and expectations of the future,” IEEE Veh. Technol. Mag., vol. 8, no. 1, pp. 24–30, Mar. 2013.
[4] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Commun. Mag., vol. 52, no. 2, pp. 74–80, Feb. 2014.
[5] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.
[6] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination and precoding in multi-cell TDD systems,” IEEE Trans. Commun., vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
[7] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” IEEE J. Sel. Areas Commun., vol. 31, no. 2, pp. 160–171, Feb. 2013.
[8] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” IEEE Trans. Commun., vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
[9] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, “Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?” IEEE Trans. Wireless Commun., submitted, Available: http://arxiv.org/abs/1403.6150.
[10] D. Ha, K. Lee, and J. Kang, “Energy efficiency analysis with circuit power consumption in massive MIMO systems,” in Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC), 2013.
[11] 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.
[12] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, “Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits,” IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
[13] E. Björnson, M. Matthaiou, and M. Debbah, “Massive MIMO with arbitrary non-ideal arrays: Hardware scaling laws and circuit-aware design,” IEEE Trans. Wireless Commun., submitted, Available: http://arxiv.org/abs/1409.0875.
[14] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, “Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise,” IEEE Trans. Wireless Commun., to appear.
[15] 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, Sept. 2012.
[16] M. Li, Y.-H. Nam, B. Ng, and J. Zhang, “A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse,” in *Proc. IEEE Globecom*, 2012.

[17] R. Müller, M. Vehkaperä, and L. Cottatellucci, “Blind pilot decontamination,” in *Proc. WSA*, 2013.

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

[19] M. Li, S. Jin, and X. Gao, “Spatial orthogonality-based pilot reuse for multi-cell massive MIMO transmission,” in *Proc. WCSP*, 2013.

[20] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Massive MIMO in real propagation environments,” *IEEE Trans. Wireless Commun.*, 2014, submitted. [Online]. Available: [http://arxiv.org/abs/1403.3376](http://arxiv.org/abs/1403.3376)

[21] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, “Uplink power control with MMSE receiver in multi-cell MU-massive-MIMO systems,” in *Proc. IEEE ICC*, 2014.

[22] H. Yang and T. Marzetta, “A macro cellular wireless network with uniformly high user throughputs,” in *Proc. IEEE VTC-Fall*, 2014.

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

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

[25] 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.

[26] H. Boche and M. Schubert, “A general duality theory for uplink and downlink beamforming,” in *Proc. IEEE VTC-Fall*, 2002, pp. 87–91.

[27] L. Zheng and D. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

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

[29] A. Lozano, R. Heath, and J. Andrews, “Fundamental limits of cooperation,” *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sept. 2013.

[30] V. M. Donald, “The cellular concept,” *Bell System Technical Journal*, vol. 58, no. 15–41, pp. 113–381, 1979.

[31] D. Cox, “Cochannel interference considerations in frequency reuse small-coverage-area radio systems,” *IEEE Trans. Commun.*, vol. 30, no. 1, pp. 135–142, Jan. 1982.

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

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

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

[35] 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, Jan. 2013.

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

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