# 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 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, $K^\star$, 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 reuse, and random user locations. The value of $K^\star$ 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, $K^\star$ depends strongly on the processing scheme and hence it is unfair to compare different schemes using the same $K$.**

*Index Terms*—**Coordinated multipoint, massive MIMO, multi-cell, pilot contamination, spectral efficiency, user scheduling.**

## 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 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)

E. Björnson and E. G. Larsson are with Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden ({emil.bjornson,erik.larsson}@liu.se). M. Debbah is with CentraleSupélec, Gif-sur-Yvette, France (merouane.debbah@supelec.fr) and with the Mathematical and Algorithmic Sciences Lab, Huawei, Paris, France.

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 the 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.[1] 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 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 of each section:

- 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

---

[1]A few results for single-cell systems are available in the literature; for example, in [8].

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

We consider a cellular network where payload data is transmitted with universal time and frequency reuse. Each cell is assigned an index in the set $\mathcal{L}$, where the cardinality $|\mathcal{L}|$ 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 at the time, out of a set of $K_{\max}$ UEs. We are interested in massive MIMO topologies where $M$ and $K_{\max}$ are large and fixed, while $K$ is a design parameter and all UEs have unlimited demand for data. The subset of active UEs changes over time, thus the name UE $k \in \{1, \ldots, K\}$ in cell $l \in \mathcal{L}$ is given to different UEs at different times. The geographical position $\mathbf{z}_{lk} \in \mathbb{R}^2$ of UE $k$ in cell $l$ is therefore an ergodic random variable with a cell-specific distribution. This model is used to study the average performance for a random rather than fixed set of interfering UEs. The time-frequency resources are divided into frames consisting of $T_c$ seconds and $W_c$ Hz, as illustrated in Fig. 1.[2] 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; $\mathbf{h}_{jlk} \in \mathbb{C}^N$ denotes the channel response between BS $j$ and UE $k$ in cell $l$ in a given frame. These channel responses are drawn as realizations from zero-mean circularly symmetric complex Gaussian distributions:

$$\mathbf{h}_{jlk} \sim \mathcal{CN}\Big(\mathbf{0}, d_j(\mathbf{z}_{lk})\mathbf{I}_M\Big), \qquad (1)$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix. This is a theoretical model for non-line-of-sight propagation that is known to give representative results with both few and many BS antennas (see recent channel measurements reported in [21]). The deterministic function $d_j(\mathbf{z})$ gives the variance of the channel attenuation from BS $j$ to any UE position $\mathbf{z}$. The value of $d_j(\mathbf{z}_{lk})$ varies slowly over time and frequency, thus we assume that the value is known at BS $j$ for all $l$ and $k$ and that each UE knows its value to its serving BS. The exact UE positions $\mathbf{z}_{lk}$ 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
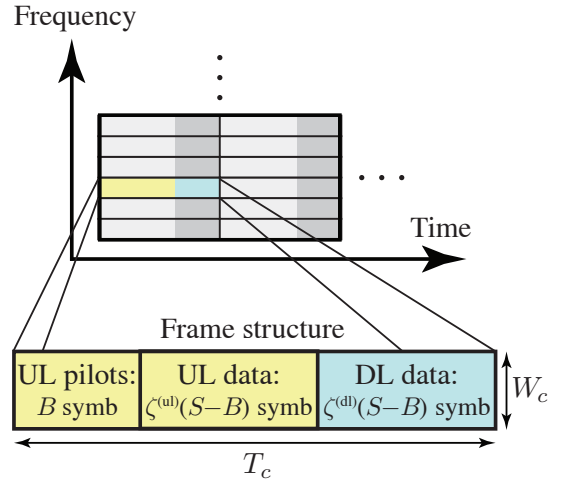


Fig. 1: The transmission is divided into frames of $S = T_c W_c$ symbols, whereof $B$ symbols are dedicated to pilot transmission. The remaining $S - B$ symbols are used for payload data, where $\zeta^{(\mathrm{ul})}$ and $\zeta^{(\mathrm{dl})}$ are respectively the fractions of UL and DL transmission.

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 $\zeta^{(\mathrm{ul})}$ and $\zeta^{(\mathrm{dl})}$ denote the fixed fractions allocated for UL and DL, respectively. These fractions can be selected arbitrarily, subject to the constraint $\zeta^{(\mathrm{ul})} + \zeta^{(\mathrm{dl})} = 1$ and that $\zeta^{(\mathrm{ul})}(S - B)$ and $\zeta^{(\mathrm{dl})}(S - B)$ are positive integers. Below, we define the system models for the UL and DL.

The BSs 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 $\mathbf{y}_j \in \mathbb{C}^M$ at BS $j$ in a frame is modeled, similar to [7] and [8], as

$$\mathbf{y}_j = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} \sqrt{p_{lk}} \mathbf{h}_{jlk} x_{lk} + \mathbf{n}_j \qquad (2)$$

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 $\mathbf{n}_j \in \mathbb{C}^M$ is modeled as $\mathbf{n}_j \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$, 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[3]; the symbols from UE $k$ in cell $l$ have the transmit

---

[2] This paper concentrates on frames that carry user-specific signals, in particular, payload data and pilots. From time to time, the network also needs special frames to transmit cell-specific control and system information and to enable random access. The design of these control frames is outside the scope of this paper, but some initial results are found in [20].

[3]Channel-aware power control was considered in [22] 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 [23], but are considerably easier to implement.

power $p_{lk} = \frac{\rho}{d_l(\mathbf{z}_{lk})}$, where $\rho > 0$ is a design parameter.[4] This power-control policy inverts the average channel attenuation $d_l(\mathbf{z}_{lk})$ and has the merit of making the average effective channel gain the same for all UEs: $\mathbb{E}\{p_{lk}\|\mathbf{h}_{llk}\|^2\} = M\rho$. 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

$$z_{jk} = \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \mathbf{h}_{ljk}^{\mathrm{T}} \mathbf{w}_{lm} s_{lm} + \eta_{jk} \tag{3}$$

where $(\cdot)^{\mathrm{T}}$ denotes transpose, $s_{lm}$ is the symbol intended for UE $m$ in cell $l$, $\mathbf{w}_{lm} \in \mathbb{C}^M$ is the corresponding precoding vector, and $\|\mathbf{w}_{lm}\|^2$ 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 show later 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.[5]

**Remark 1** (Synchronization Issues). *The UL/DL system models in* (2) *and* (3) *assume perfect synchronization across all cells, as commonly done in the massive MIMO literature; cf.* [2], [6]–[8], [15]. *Local synchronization is achievable, for example, using the cyclic prefix in OFDM-based systems, but network-wide synchronization is probably infeasible over large coverage areas. The processing techniques analyzed in this paper can thus be used to suppress the strong interference between the closest tiers of neighboring cells, while the interference from distant cells is asynchronously received and practically insuppressible. We expect that the simplified synchronization modeling used here and elsewhere has negligible impact on the system performance, since the insuppressible distant interferers are weak as compared to (partially suppressed) interference from neighboring cells.*

## 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 coherent receive combining in the UL and transmit precoding in the DL, which can adaptively amplify desired signals and suppress

---

[4] The parameter $\rho$ needs to be selected such that UEs at the cell edge do not use more transmit power than their amplifiers can handle or the spectrum regulations allow. This is not a critical limitation in massive MIMO since high SEs are provided also at low SNRs (see Fig. 12), but it might be necessary to occasionally drop severely shadowed UEs from service.

[5] 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.

interfering signals. This requires, however, some knowledge of the UEs' channels; for example, $\sqrt{p_{lk}}\mathbf{h}_{jlk}$ 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 reuse, 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 $1 \le B \le S$.[6] Each pilot signal can be represented by a deterministic vector $\mathbf{v} \in \mathbb{C}^B$ and the fixed per-symbol power implies that all entries have unit magnitude: $|[\mathbf{v}]_s| = 1$, 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

$$\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_B\} \quad \text{where} \quad \mathbf{v}_{b_1}^{\mathrm{H}} \mathbf{v}_{b_2} = \begin{cases} B, & b_1 = b_2, \\ 0, & b_1 \ne b_2, \end{cases} \tag{4}$$

where $(\cdot)^{\mathrm{H}}$ 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 [24].

The pilot signal transmitted by UE $k$ in cell $l$ is denoted by $\mathbf{v}_{i_{lk}}$, 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 $\mathbf{Y}_j \in \mathbb{C}^{M \times B}$ and given by

$$\mathbf{Y}_j = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} \sqrt{p_{lk}}\mathbf{h}_{jlk}\mathbf{v}_{i_{lk}}^{\mathrm{T}} + \mathbf{N}_j, \tag{5}$$

where $\mathbf{N}_j \in \mathbb{C}^{M \times B}$ 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 $\mathbf{h}_{jlk}^{\mathrm{eff}} = \sqrt{p_{lk}}\mathbf{h}_{jlk}$.

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

$$\hat{\mathbf{h}}_{jlk}^{\mathrm{eff}} = \frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})} \mathbf{Y}_j (\mathbf{\Psi}_j^{\mathrm{T}})^{-1} \mathbf{v}_{i_{lk}}^* \tag{6}$$

---

[6] 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.

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

---

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

$$\boldsymbol{\Psi}_j = \sum_{\ell\in\mathcal{L}}\sum_{m=1}^{K}\frac{d_j(\mathbf{z}_{\ell m})}{d_\ell(\mathbf{z}_{\ell m})}\mathbf{v}_{i_{\ell m}}\mathbf{v}_{i_{\ell m}}^{\text{H}} + \frac{\sigma^2}{\rho}\mathbf{I}_B. \tag{7}$$

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

$$\mathbf{C}_{jlk} = \mathbb{E}\left\{(\mathbf{h}_{jlk}^{\text{eff}} - \hat{\mathbf{h}}_{jlk}^{\text{eff}})(\mathbf{h}_{jlk}^{\text{eff}} - \hat{\mathbf{h}}_{jlk}^{\text{eff}})^{\text{H}}\right\}$$
$$= \rho\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\left(1 - \frac{\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}B}{\sum_{\ell\in\mathcal{L}}\sum_{m=1}^{K}\frac{d_j(\mathbf{z}_{\ell m})}{d_\ell(\mathbf{z}_{\ell m})}\mathbf{v}_{i_{lk}}^{\text{H}}\mathbf{v}_{i_{\ell m}} + \frac{\sigma^2}{\rho}}\right)\mathbf{I}_M \tag{8}$$

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

*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 $\mathbf{v}_{i_{lk}}^{\text{H}}\mathbf{v}_{i_{\ell m}}$ that are non-zero). The ratio $d_j(\mathbf{z}_{\ell m})/d_\ell(\mathbf{z}_{\ell m})$ 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 $j$.

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

$$\hat{\mathbf{H}}_{\mathcal{V},j} = \mathbf{Y}_j(\boldsymbol{\Psi}_j^{\text{T}})^{-1}[\mathbf{v}_1^* \ldots, \mathbf{v}_B^*] \tag{9}$$

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

$$\hat{\mathbf{h}}_{jlk}^{\text{eff}} = \frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\hat{\mathbf{H}}_{\mathcal{V},j}\mathbf{e}_{i_{lk}} \tag{10}$$

where $\mathbf{e}_i$ denotes the $i$th column of the identity matrix $\mathbf{I}_B$. 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 possibility is not considered herein since we want to develop methods to suppress pilot contamination that can be utilized in any propagation environment.

**Remark 2** (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 $\mathbf{g}_{jk} \in \mathbb{C}^M$ to the received signal, as $\mathbf{g}_{jk}^{\text{H}}\mathbf{y}_j$, to amplify the signal from its $k$th UE and reject interference from other UEs in the spatial domain. We want to derive the ergodic achievable SE for any UE, where codewords span over both the Rayleigh fading and random locations of the interfering UEs—specific UE distributions are considered in Section IV. For notational convenience, we assume that $\beta = \frac{B}{K}$ is an integer that we refer to as the *pilot reuse factor*. The cells in $\mathcal{L}$ are divided into $\beta \geq 1$ disjoint subsets such that the same $K$ pilot sequences are used within a set, while different pilots are used in different sets. We refer to this as *non-universal pilot reuse*. An explicit example is provided in Section IV for hexagonal cells, while the result in this section holds for any network topology. 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 of an arbitrary UE $k$ in cell $j$ is*

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

*where the effective signal-to-interference-and-noise ratio (SINR),* $\text{SINR}_{jk}^{(\text{ul})}$, *is given in (12) at the top of the page. The expectations* $\mathbb{E}_{\{\mathbf{z}\}}\{\cdot\}$ *and* $\mathbb{E}_{\{\mathbf{h}\}}\{\cdot\}$ *are with respect to UE positions and channel realizations, respectively.*

*Proof:* By coding over variations in the channel realizations $\{\mathbf{h}\}$ and positions $\{\mathbf{z}\}$ of the interfering UEs, an achievable SE is given by $\mathbb{E}_{\{\mathbf{z},\mathbf{h}\}}\{\mathcal{I}(x_{lk}, \mathbf{y}_j)\}$, where $\mathcal{I}(x_{lk}, \mathbf{y}_j)$ is the mutual information between the transmitted and received signal in (2) for fixed channel realizations and UE positions. The lemma follows from computing a lower bound on $\mathcal{I}(x_{lk}, \mathbf{y}_j)$, similar to [6], [7], [13], [25], [26], by making three limiting assumptions: 1) a Gaussian codebook

is used; 2) the signal component received over the effective channel mean $\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}$ is the only desired signal, while the interference and the signal component over the remaining uncorrelated channel $\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk} - \mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}$ are treated as noise (i.e., not exploited in the decoding); and 3) the noise is taken as worst-case Gaussian distributed in the decoding, leading to a further lower bound on the mutual information. ∎

The ergodic achievable SE in Lemma 2, for any UE in cell $j$, is a lower bound on the ergodic capacity, which is unknown for general multi-cell networks. Similar bounds are found in [6]–[8] and the bounding technique interacts with the Rayleigh fading, which is why its expectations end up inside the logarithm while the user positions are averaged at the outside. To compute these expectations 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

$$\mathbf{g}_{jk}^{\mathrm{MR}} = \widehat{\mathbf{H}}_{\mathcal{V},j}\mathbf{e}_{i_{jk}} = \hat{\mathbf{h}}_{jjk}^{\mathrm{eff}}, \tag{13}$$

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 $\hat{\mathbf{h}}_{jjk}^{\mathrm{eff}}$ when $M$ is large.[7]

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:

$$\mathbf{g}_{jk}^{\mathrm{ZF}} = \widehat{\mathbf{H}}_{\mathcal{V},j}\mathbf{E}_j\big(\mathbf{E}_j^{\mathrm{H}}\widehat{\mathbf{H}}_{\mathcal{V},j}^{\mathrm{H}}\widehat{\mathbf{H}}_{\mathcal{V},j}\mathbf{E}_j\big)^{-1}\mathbf{e}_{i_{jk}}, \tag{14}$$

where $\mathbf{E}_j = [\mathbf{e}_{i_{j1}}\dots\mathbf{e}_{i_{jK}}] \in \mathbb{C}^{B\times K}$ 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.

**Theorem 1.** *Let $\mathcal{L}_j(\beta) \subset \mathcal{L}$ be the subset of cells that uses the same pilots as cell $j$. In the UL, an achievable SE in cell $j$ is*

$$\mathrm{SE}_j^{(\mathrm{ul})} = K\zeta^{(\mathrm{ul})}\left(1 - \frac{B}{S}\right)\log_2\left(1 + \frac{1}{I_j^{\mathrm{scheme}}}\right) \text{ [bit/s/Hz/cell]} \tag{15}$$

*where the interference term*

$$I_j^{\mathrm{scheme}} = \sum_{l\in\mathcal{L}_j(\beta)\backslash\{j\}}\left(\mu_{jl}^{(2)} + \frac{\mu_{jl}^{(2)} - \left(\mu_{jl}^{(1)}\right)^2}{G^{\mathrm{scheme}}}\right)$$
$$+ \frac{\left(\sum_{l\in\mathcal{L}}\mu_{jl}^{(1)}Z_{jl}^{\mathrm{scheme}} + \frac{\sigma^2}{\rho}\right)\left(\sum_{\ell\in\mathcal{L}_j(\beta)}\mu_{j\ell}^{(1)} + \frac{\sigma^2}{B\rho}\right)}{G^{\mathrm{scheme}}} \tag{16}$$

*depends on the receive combining scheme through $G^{\mathrm{scheme}}$ and $Z_{jl}^{\mathrm{scheme}}$. MR combining is obtained by $G^{\mathrm{MR}} = M$ and*

$Z_{jl}^{\mathrm{MR}} = K$, *while ZF combining is obtained by $G^{\mathrm{ZF}} = M - K$ and*

$$Z_{jl}^{\mathrm{ZF}} = \begin{cases} K\left(1 - \dfrac{\mu_{jl}^{(1)}}{\sum_{\ell\in\mathcal{L}_j(\beta)}\mu_{j\ell}^{(1)} + \frac{\sigma^2}{B\rho}}\right) & \text{if } l\in\mathcal{L}_j(\beta), \\ K & \text{if } l\notin\mathcal{L}_j(\beta). \end{cases} \tag{17}$$

*The following notation was used:*

$$\mu_{jl}^{(\omega)} = \mathbb{E}_{\mathbf{z}_{lm}}\left\{\left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^\omega\right\} \quad \text{for } \omega = 1,2. \tag{18}$$

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

The closed-form SE expressions in Theorem 1 are lower bounds on the ergodic capacity and slightly more conservative than the non-closed-form bound in Lemma 2; see Section IV-B for a numerical comparison. We stress that the closed-form SEs are only functions of the pilot allocation and the propagation parameters $\mu_{jl}^{(1)}$ and $\mu_{jl}^{(2)}$ defined in (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 expression manifests the importance of pilot allocation, since the interference term in (16) contains summations that only consider the cells that use the same pilots as cell $j$.

The first term in (16) 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 through $Z_{jl}^{\mathrm{scheme}}$, at the price of reducing the array gain $G^{\mathrm{scheme}}$ 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

$$\mathbf{g}_{jk}^{\mathrm{P\text{-}ZF}} = \widehat{\mathbf{H}}_{\mathcal{V},j}\big(\widehat{\mathbf{H}}_{\mathcal{V},j}^{\mathrm{H}}\widehat{\mathbf{H}}_{\mathcal{V},j}\big)^{-1}\mathbf{e}_{i_{jk}}. \tag{19}$$

In contrast to the conventional ZF in (14), which only orthogonalize the $K$ intra-cell channels in $\widehat{\mathbf{H}}_{\mathcal{V},j}\mathbf{E}_j$, P-ZF exploits that all the $B$ estimated channel directions in $\widehat{\mathbf{H}}_{\mathcal{V},j}$ 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 $\widehat{\mathbf{H}}_{\mathcal{V},j}$ 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.** *Let $\mathcal{L}_l(\beta) \subset \mathcal{L}$ be the subset of cells that uses the same pilots as cell $l$. In the UL, an achievable SE in cell $j$ with P-ZF combining is given by (15) for $G^{\mathrm{P\text{-}ZF}} = M - B$*

---

[7]With *quasi*-orthogonality we mean that two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{C}^M$ satisfy $\frac{\mathbf{a}^{\mathrm{H}}\mathbf{b}}{M} \to 0$ as $M \to \infty$, although $\mathbf{a}^{\mathrm{H}}\mathbf{b}$ will not converge to zero and might even go to infinity, e.g., proportionally to $\sqrt{M}$ as with Rayleigh fading channel vectors.

*and*

$$Z_{jl}^{\text{P-ZF}} = K \left( 1 - \frac{\mu_{jl}^{(1)}}{\sum\limits_{\ell \in \mathcal{L}_l(\beta)} \mu_{j\ell}^{(1)} + \frac{\sigma^2}{B\rho}} \right). \qquad (20)$$

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

The SE expressions were derived assuming that $M$ and $K$ are the same in all cells, for notational brevity. However, the results in this section are straightforward to extend to cell-specific $M$ and $K$ values.

### 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 $\mathbf{w}_{jk} \in \mathbb{C}^M$ is the precoding vector associated with UE $k$ in cell $j$. We express these precoding vectors as

$$\mathbf{w}_{jk} = \sqrt{\frac{q_{jk}}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}}} \check{\mathbf{g}}_{jk}^* \qquad (21)$$

where the average transmit power $q_{jk} \geq 0$ is a function of the UE positions, but not the instantaneous channel realizations. The vector $\check{\mathbf{g}}_{jk} \in \mathbb{C}^M$ defines the spatial directivity of the transmission and is based on the acquired CSI; the normalization with the average squared norm $\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}$ gives the analytic tractability that enables the following results.[8]

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

$$\zeta^{(\text{dl})} \left( 1 - \frac{B}{S} \right) \mathbb{E}_{\{\mathbf{z}\}} \left\{ \log_2(1 + \text{SINR}_{jk}^{(\text{dl})}) \right\} \quad [bit/s/Hz] \tag{22}$$

*with the effective SINR, $\text{SINR}_{jk}^{(\text{dl})}$, given by*

$$\frac{q_{jk} \frac{|\mathbb{E}_{\{\mathbf{h}\}}\{\check{\mathbf{g}}_{jk}^{\text{H}} \mathbf{h}_{jjk}\}|^2}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}}}{\sum\limits_{l \in \mathcal{L}} \sum\limits_{m=1}^{K} q_{lm} \frac{\mathbb{E}_{\{\mathbf{h}\}}\{|\check{\mathbf{g}}_{lm}^{\text{H}} \mathbf{h}_{ljk}|^2\}}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{lm}\|^2\}} - q_{jk} \frac{|\mathbb{E}_{\{\mathbf{h}\}}\{\check{\mathbf{g}}_{jk}^{\text{H}} \mathbf{h}_{jjk}\}|^2}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}} + \sigma^2}. \tag{23}$$

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

Note that Lemma 3 takes into account the fact that each UE only knows the expectations in (23) and not the instantaneous channels (see [6, Theorem 1] for more details).

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*

[8] Conventionally, the power is normalized by $\|\check{\mathbf{g}}_{jk}\|^2$ instead of $\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}$ in multi-user MIMO systems [27], but the difference is small in massive MIMO since $|\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\} - \|\check{\mathbf{g}}_{jk}\|^2|/M \to 0$ as $M \to \infty$, for most precoding schemes.

*policy $\{q_{jk}\}$, with $\sum\limits_{j \in \mathcal{L}} \sum\limits_{k=1}^{K} q_{jk} = \sum\limits_{j \in \mathcal{L}} \sum\limits_{k=1}^{K} p_{jk}$, for which*

$$\text{SINR}_{jk}^{(\text{dl})} = \text{SINR}_{jk}^{(\text{ul})} \tag{24}$$

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

$$\text{SE}_j^{(\text{dl})} = K\zeta^{(\text{dl})} \left( 1 - \frac{B}{S} \right) \log_2 \left( 1 + \frac{1}{I_j^{\text{scheme}}} \right) \ [bit/s/Hz/cell] \tag{25}$$

*where the interference term $I_j^{\text{scheme}}$ 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 [28], conventionally considered for single-cell systems with perfect CSI, which is applicable also 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 $\check{\mathbf{g}}_{jk} = \mathbf{g}_{jk}^{\text{MR}}$; ZF precoding that actively rejects intra-cell interference by setting $\check{\mathbf{g}}_{jk} = \mathbf{g}_{jk}^{\text{ZF}}$; and P-ZF precoding that actively rejects both intra- and inter-cell interference by setting $\check{\mathbf{g}}_{jk} = \mathbf{g}_{jk}^{\text{P-ZF}}$. 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*

$$\begin{aligned} \text{SE}_j &= \text{SE}_j^{(\text{ul})} + \text{SE}_j^{(\text{dl})} \\ &= K \left( 1 - \frac{B}{S} \right) \log_2 \left( 1 + \frac{1}{I_j^{\text{scheme}}} \right) \ [bit/s/Hz/cell] \end{aligned} \tag{26}$$

*where the interference term $I_j^{\text{scheme}}$ for UE $k$ is given Theorem 1 for MR and ZF and in Theorem 2 for P-ZF. This SE can be divided between the UL and DL arbitrarily using any positive fractions $\zeta^{(\text{ul})}$ and $\zeta^{(\text{dl})}$, with $\zeta^{(\text{ul})} + \zeta^{(\text{dl})} = 1$.*

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 further insights from the structure of the SE expression in (26), we analyze it for a particular network topology in Section IV. In the remainder of this section, we consider the limit of a large number of antennas.

**Corollary 2.** *Let $\mathcal{L}_j(\beta) \subset \mathcal{L}$ be the subset of cells that uses the same pilots as cell $j$. When $M \to \infty$ (with $K, B \leq S < \infty$), the effective SINRs with MR, ZF, and P-ZF converge to the same limit:*

$$\frac{1}{I_j^{\mathrm{MR}}}, \frac{1}{I_j^{\mathrm{ZF}}}, \frac{1}{I_j^{\mathrm{P\text{-}ZF}}} \to \frac{1}{\displaystyle\sum_{l \in \mathcal{L}_j(\beta) \backslash \{j\}} \mu_{jl}^{(2)}}. \tag{27}$$

The ultimate effect of pilot contamination is very clear in (27), since only the cells that interfered with cell $j$ during pilot transmission (i.e., cells with indices in the set $\mathcal{L}_j(\beta) \backslash \{j\}$) affect the asymptotic limit. To maximize the asymptotic SINR in (27), one should place the cells with large $\mu_{jl}^{(2)}$ in different subsets (i.e., $\mathcal{L}_j(\beta) \cap \mathcal{L}_l(\beta) = \emptyset$) so that these cells use different pilots. The asymptotic limit can be used as follows to find the optimal $K$.

**Corollary 3.** *Let $\mathcal{L}_j(\beta) \subset \mathcal{L}$ be the subset of cells that uses the same pilots as cell $j$. The SE in cell $j$ approaches*

$$\mathrm{SE}_j^\infty = K \left(1 - \frac{K\beta}{S}\right) \log_2 \left(1 + \frac{1}{\sum_{l \in \mathcal{L}_j(\beta) \backslash \{j\}} \mu_{jl}^{(2)}}\right) \tag{28}$$

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

*Proof:* The logarithmic part of (28) 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 $K^\star = \frac{S}{2}$ for $\beta = 1$ and $K^\star = \frac{S}{6}$ 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 $K^\star = \frac{S}{2\beta}$ is an integer, the asymptotically optimal SE is

$$\mathrm{SE}_j^\infty = \frac{S}{4\beta} \log_2 \left(1 + \frac{1}{\sum_{l \in \mathcal{L}_j(\beta) \backslash \{j\}} \mu_{jl}^{(2)}}\right) \tag{29}$$

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 $\beta = 1$. The rationale is that the SE gain from adding an extra UE outweighs the pre-log loss at the existing UEs if at least half the frame is used for data (a criterion independent of $\beta$). The asymptotically optimal $\beta$ cannot be computed in closed-form, but we notice that a larger $\beta$ leads to fewer interferers in $\mathcal{L}_j(\beta)$ and also reduces the pre-log factor; hence, a larger $\beta$ brings SINR improvements until a certain point where the pre-log loss starts to dominate.
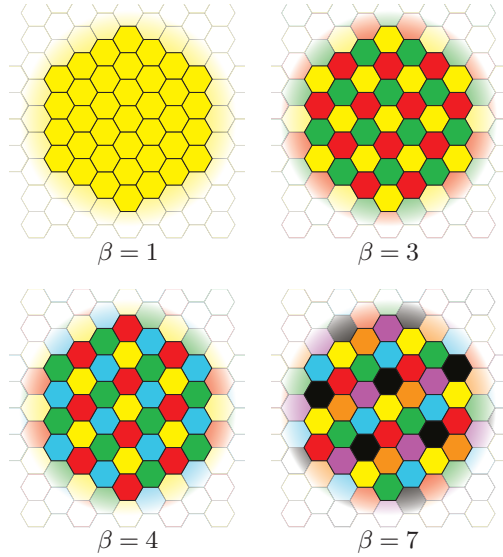


Fig. 2: Part of a hexagonal network, colored for different pilot reuse factors $\beta$.
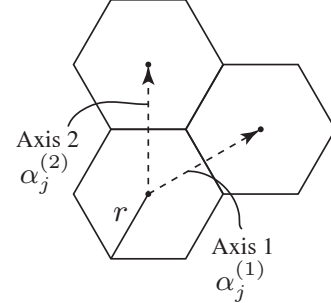


Fig. 3: The coordinate system for a hexagonal grid.

At first sight, these results bear some similarity with the results in [26] and [29] 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 [30]. Instead, the pre-log factor $\frac{S}{4\beta}$ in (29) 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 $M \to \infty$. 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 [31]. Although practical deployments have irregular cells, it is common practice to establish general properties by analyzing symmetric networks where the cells are regular polygons [32]; 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 non-universal 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 $j$ [31]:

$$\mathbf{b}_j = \sqrt{3}\begin{bmatrix} \sqrt{3}r/2 \\ r/2 \end{bmatrix}\alpha_j^{(1)} + \begin{bmatrix} 0 \\ \sqrt{3}r \end{bmatrix}\alpha_j^{(2)} \in \mathbb{R}^2. \quad (30)$$

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 [31], [32], 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 $d_j(\mathbf{z}) = \frac{C}{\|\mathbf{z}-\mathbf{b}_j\|^\kappa}$, 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

$$\mu_{jl}^{(\omega)} = \mathbb{E}_{\mathbf{z}_{lm}}\left\{\left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^\omega\right\} = \mathbb{E}_{\mathbf{z}_{lm}}\left\{\left(\frac{\|\mathbf{z}_{lm}-\mathbf{b}_l\|}{\|\mathbf{z}_{lm}-\mathbf{b}_j\|}\right)^{\kappa\omega}\right\} \quad (31)$$

for any UE distributions in the cells. We note that $C$ and $r$ cancel out in (31), if the UE distributions in each cell are independent of $C$ and $r$. 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.

### 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 $\zeta^{(\mathrm{ul})}$ and $\zeta^{(\mathrm{dl})}$, 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. The simu-

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 determine $B = \beta K$) by searching the range of all reasonable integer values. We set the coherence block length to $S = 400$ (e.g., 2 ms coherence time and 200 kHz coherence
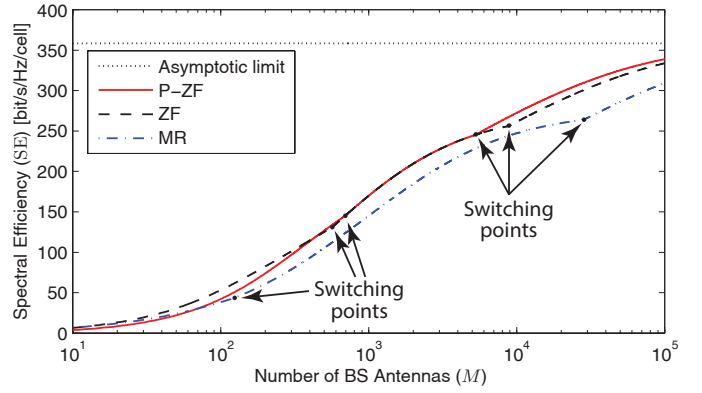


(a) Optimized SE per cell.
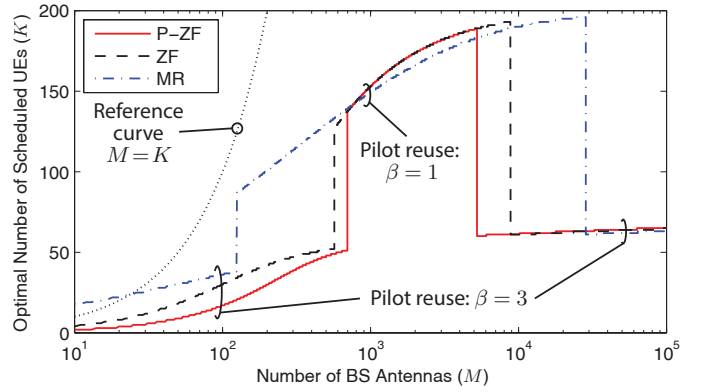


(b) Corresponding optimal number of UEs: $K^\star$.

Fig. 4: Simulation of optimized SE, as a function of $M$, with average inter-cell interference.

bandwidth), set the SNR to $\rho/\sigma^2 = 5$ dB, and pick $\kappa = 3.7$ as pathloss exponent.[9] The impact of changing the different system parameters is considered 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 of the parameters $\mu_{jl}^{(1)}$ and $\mu_{jl}^{(2)}$ 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 random switching of pilot sequences between the UEs

---

[9] A higher pathloss exponent reduces the inter-cell interference, but requires more signal power to maintain a certain SNR.
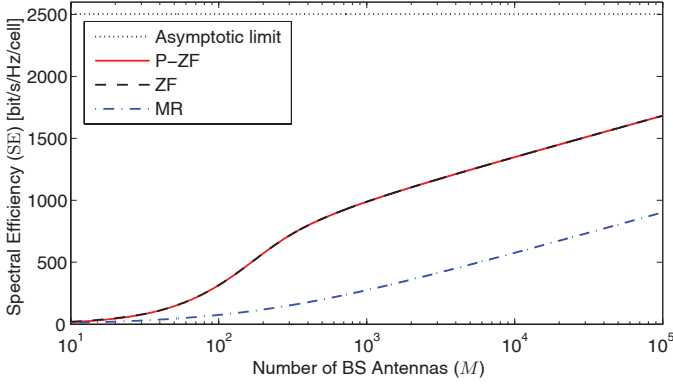
(a) Optimized SE per cell.
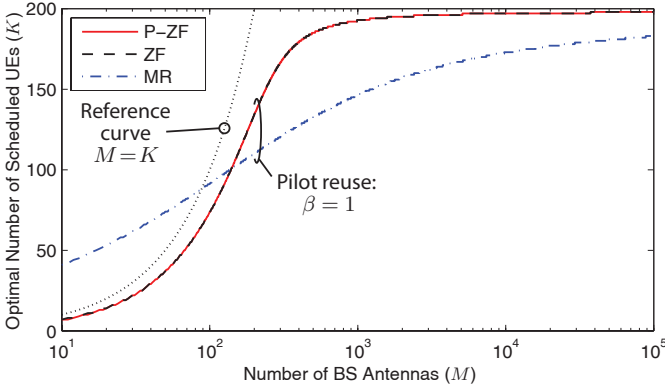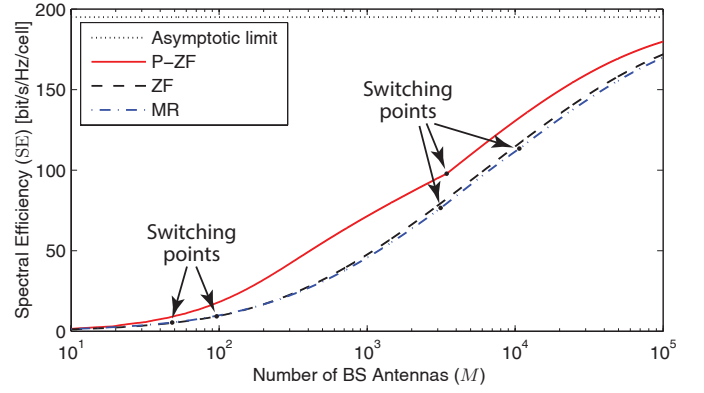


(b) Corresponding optimal number of UEs: $K^\star$.

Fig. 5: Simulation of optimized SE, as a function of $M$, with best-case inter-cell interference.



(a) Optimized SE per cell.
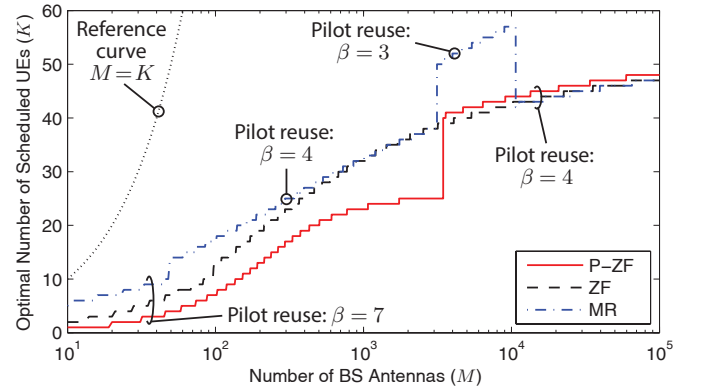


(b) Corresponding optimal number of UEs: $K^\star$.

Fig. 6: Simulation of optimized SE, as a function of $M$, with worst-case inter-cell interference.

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 $K^\star$ 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 $10 \le M \le 200$ antennas. In all cases, the largest differences appear when the number of antennas is very large (notice the logarithmic $M$-scales). At least $M = 10^5$ is needed to come close to the asymptotic limit in (29), which was proved by Corollary 3, and many more antennas are required under best case interference. Clearly, the asymptotic limits should not be 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 $K^\star$ and which pilot reuse factor $\beta$ that are used. The general behavior is that larger $M$ implies a higher $K^\star$ and a smaller $\beta$, because the channels become more orthogonal with $M$. Since the reuse factor is an integer, $K^\star$ changes non-continuously when $\beta$ is changed; smaller $\beta$ allows for larger $K^\star$, 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 $K^\star \to 67$ in the average case (where $\beta = 3$), $K^\star \to 200$ in the best case (where $\beta = 1$), and $K^\star \to 50$ 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 $10 \le M \le 1000$ antennas, and when other system parameters
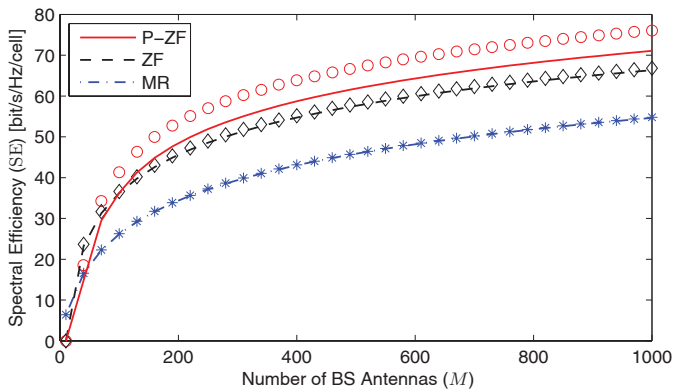
Fig. 7: Per-cell SE for $K = 10$. The lines are based on Theorems 1 and 2 while the markers are computed numerically from Lemma 2.



Fig. 8: Impact of changing the pilot reuse factor $\beta$, for a system optimized for high per-cell SE.



Fig. 9: Achievable SE per UE, for a system optimized for high per-cell SE.

than $M$ are varied we only consider $M = 100$ (medium massive MIMO setup) and $M = 500$ (large massive MIMO setup).

We begin by verifying the accuracy of the closed-form expressions in Theorems 1 and 2, by comparing the formulas to Monte-Carlo simulations based on Lemma 2. The formulas are exact in the best and worst interference cases, but Fig. 7 shows that the interference variations in the average case result in some loss in SE. The figure considers $K = 10$ UEs and Monte-Carlo simulations are represented by markers. The MR and ZF formulas in Theorem 1 are very tight. However, there is a few percent of deviation for P-ZF in Theorem 2, since a lower bound on the ability of cancel inter-cell interference is used to get a tractable formula. Hence, P-ZF will actually perform slightly better than reported in the simulations in this paper.

Next, we study the impact of the pilot reuse factor $\beta$ using the formulas from Theorems 1 and 2. Fig. 8 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. 9 for the operating points that maximize the SE in the cell; this is basically the ratio SE/$K^\star$, where SE was given in Fig. 4(a) and $K^\star$ 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. 9 are straightforward to implement in practice.

Fig. 10 shows the ratio $M/K^\star$ 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
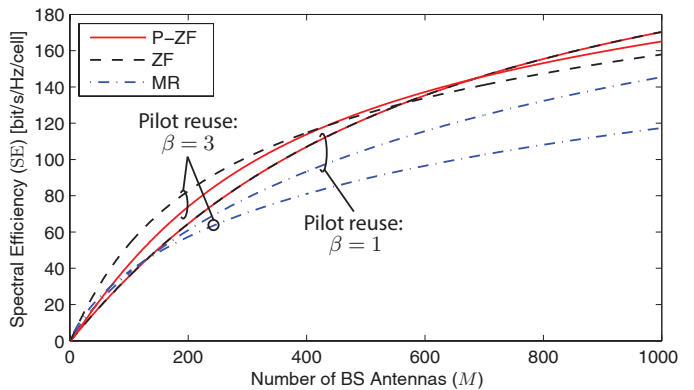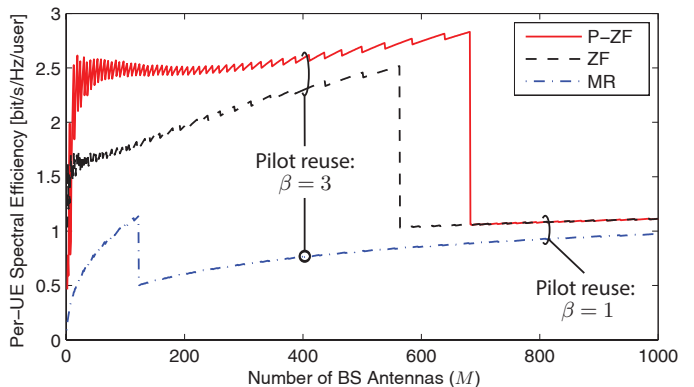
thumb which says that massive MIMO systems should have an order of magnitude more BS antennas than UEs. The operating points that satisfy this guideline are above the horizontal dotted line. This simulation indicates that an optimized system might not follow this guideline; in fact, there is a few occasions where MR even prefers to have $M/K^\star < 1$. 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. 11 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 $\mathcal{O}(MK)$, while the complexity of ZF and P-ZF scales as $\mathcal{O}(MK^2)$ [9].

Next, Fig. 12 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.
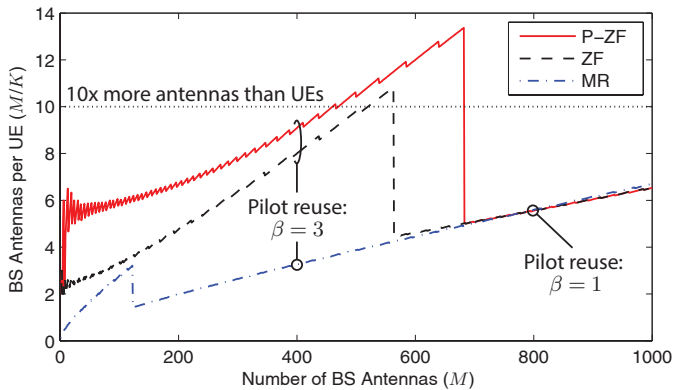
Fig. 10: Number of BS antennas per UE with different processing schemes, for a system optimized for high per-cell SE.
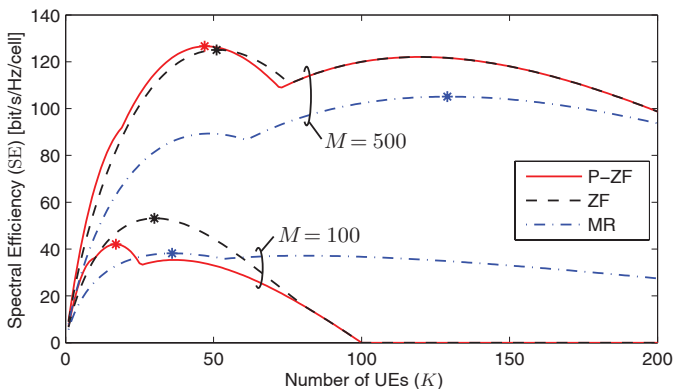


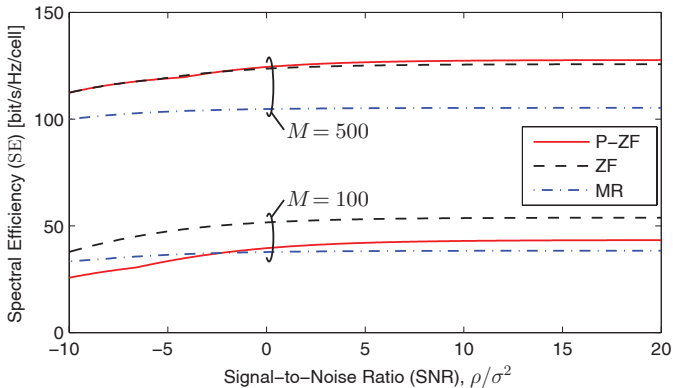Fig. 11: Achievable per-cell SE as a function of the number of scheduled UEs.



Fig. 12: Impact of SNR variations on the SE.



Fig. 13: Per-cell SE as a function of the coherence block length $S$.

## 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 non-ideal analog filters [33]–[36]. In this section, we provide a prediction of how these hardware impairments affect the achievable SEs in multi-cell massive MIMO systems. We notice that it was recently shown in [12], using impairments models developed and evaluated in [33]–[35], 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 henceforth we only consider those impairments 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

$$\mathbf{y}_j = \sum_{l \in \mathcal{L}} \sum_{k=1}^{K} \mathbf{h}_{jlk} \left( \sqrt{(1-\epsilon^2)p_{lk}} x_{lk} + \varepsilon_{lk} \right) + \mathbf{n}_j, \quad (32)$$

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

$$z_{jk} = \sqrt{1-\epsilon^2} \left( \sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \mathbf{h}_{ljk}^{\mathrm{T}} \mathbf{w}_{lm} s_{lm} + \eta_{jk} \right) + e_{jk}, \quad (33)$$

where $e_{jk} \sim \mathcal{CN}(0, \epsilon^2(\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \|\mathbf{h}_{ljk}^{\mathrm{T}} \mathbf{w}_{lm}\|^2 + \sigma^2))$ is the DL distortion noise caused at UE $k$ in cell $j$. Notice that $\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \|\mathbf{h}_{ljk}^{\mathrm{T}} \mathbf{w}_{lm}\|^2 + \sigma^2$ is the power of the term in parenthesis in (33). The parameter $\epsilon$ determines the level of impairments and can be interpreted as the error vector magnitude (EVM) [34]; typical values in LTE are in the range

Finally, Fig. 13 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 $(1 - \frac{B}{S})$. 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.
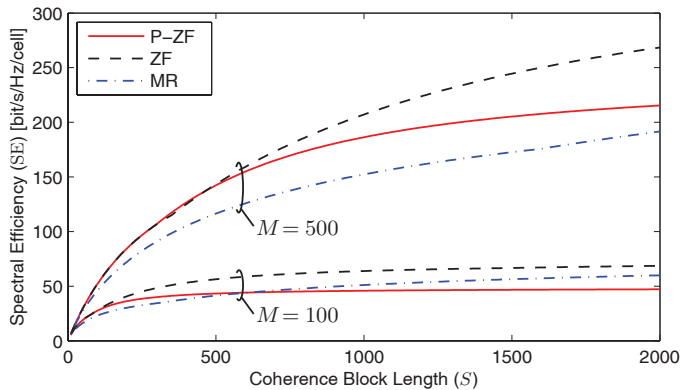
$$\widetilde{\mathrm{SINR}}_{jk} = \frac{(1-\epsilon^2)p_{jk}|\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2}{\sum_{l\in\mathcal{L}}\sum_{m=1}^{K} p_{lm}\mathbb{E}_{\{\mathbf{h}\}}\{|\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jlm}|^2\} - (1-\epsilon^2)p_{jk}|\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{g}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2 + \sigma^2\mathbb{E}_{\{\mathbf{h}\}}\{\|\mathbf{g}_{jk}\|^2\}}. \tag{35}$$

$0 \leq \epsilon \leq 0.17$ [37]. 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 an arbitrary UE $k$ in cell $j$ is*

$$\left(1 - \frac{B}{S}\right)\mathbb{E}_{\{\mathbf{z}\}}\left\{\log_2(1 + \widetilde{\mathrm{SINR}}_{jk})\right\} \quad \text{[bit/s/Hz]} \tag{34}$$

*where the effective SINR is given in (35) at the top of the page.*

*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.** *Let $\mathcal{L}_j(\beta) \subset \mathcal{L}$ be the subset of cells that uses the same pilots as cell $j$. Looking jointly at the UL and DL, an achievable SE in cell $j$ under hardware impairments is*

$$\mathrm{SE}_j = K\left(1 - \frac{B}{S}\right)\log_2\left(1 + \frac{1-\epsilon^2}{I_j^{\mathrm{scheme}} + \epsilon^2}\right) \text{ [bit/s/Hz/cell]} \tag{36}$$

*where the interference term $I_j^{\mathrm{scheme}}$ is defined in (16) and depends on $G^{\mathrm{scheme}}$ and $Z_{jl}^{\mathrm{scheme}}$. The parameter values with MR, ZF, and P-ZF are as follows:*

| Scheme | $G^{\mathrm{scheme}}$ | $Z_{jl}^{\mathrm{scheme}}$ |
|---|---|---|
| MR | $M(1-\epsilon^2)$ | $K$ |
| ZF | $(M-K)(1-\epsilon^2)$ | $\begin{cases} K\left(1 - \frac{(1-\epsilon^2)\mu_{jl}^{(1)}}{\sum_{\ell\in\mathcal{L}_j(\beta)}\mu_{j\ell}^{(1)}+\frac{\sigma^2}{B\rho}}\right) \\ \quad \text{if } l\in\mathcal{L}_j(\beta) \\ K \text{ if } l\notin\mathcal{L}_j(\beta) \end{cases}$ |
| P-ZF | $(M-B)(1-\epsilon^2)$ | $K\left(1 - \frac{(1-\epsilon^2)\mu_{jl}^{(1)}}{\sum_{\ell\in\mathcal{L}_l(\beta)}\mu_{j\ell}^{(1)}+\frac{\sigma^2}{B\rho}}\right)$ |

*If $M\to\infty$ (with $K, B \leq S < \infty$), the effective SINRs with these processing schemes approach the upper limit*

$$\frac{1-\epsilon^2}{\sum_{l\in\mathcal{L}_j(\beta)\setminus\{j\}}\mu_{jl}^{(2)} + \epsilon^2}. \tag{37}$$

*Proof:* This result follows straightforwardly from Theorems 1–3, since the SINR expressions in (35) only differ from those in Section III by the $(1-\epsilon^2)$-factors. ∎

Using the tractable SE expression in Theorem 4 for simulation, Fig. 14 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
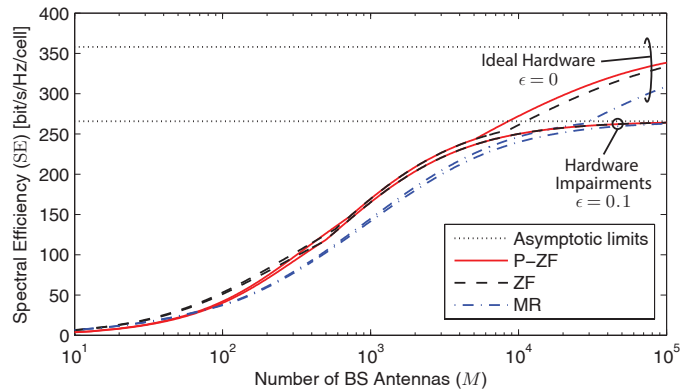


Fig. 14: Optimized per-cell SE with or without hardware impairments.

[37]. 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 (27) and for hardware impairments in (37) 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 $\frac{S}{2\beta}$ 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 to pilots: 5% to 40% were observed in simulations for $M \leq 1000$.

$$\text{SINR}_{jk}^{(\text{ul})} = \frac{\mathbf{v}_{i_{jk}}^{\text{H}} \boldsymbol{\Psi}_j^{-1} \mathbf{v}_{i_{jk}}}{\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \left( \frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})} \frac{1}{M} + \left( \frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})} \right)^2 \mathbf{v}_{i_{jk}}^{\text{H}} \boldsymbol{\Psi}_j^{-1} \mathbf{v}_{i_{lm}} \right) - \mathbf{v}_{i_{jk}}^{\text{H}} \boldsymbol{\Psi}_j^{-1} \mathbf{v}_{i_{jk}} + \frac{\sigma^2}{M\rho}} \tag{42}$$

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 non-universal 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. The results in this paper are for uncorrelated fading, while spatially correlated fading is expected to reduce the inter-user interference [18] thus leading to higher SEs and allowing for smaller $\beta$.

## APPENDIX: COLLECTION OF PROOFS

**Proof of Lemma 1:** As shown in [38, Theorem 11.1], the expression for an MMSE estimator $\hat{\mathbf{h}}_{jlk}^{\text{eff}}$ of $\mathbf{h}_{jlk}^{\text{eff}}$ with jointly Gaussian channels and (colored) noise is

$$\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{h}_{jlk}^{\text{eff}}\text{vec}(\mathbf{Y}_j)^{\text{H}}\} \left( \mathbb{E}_{\{\mathbf{h}\}}\{\text{vec}(\mathbf{Y}_j)\text{vec}(\mathbf{Y}_j)^{\text{H}}\} \right)^{-1} \text{vec}(\mathbf{Y}_j) \tag{38}$$

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

$$\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{h}_{jlk}^{\text{eff}}\text{vec}(\mathbf{Y}_j)^{\text{H}}\} = \mathbb{E}_{\{\mathbf{h}\}}\left\{\mathbf{h}_{jlk}^{\text{eff}}(\mathbf{h}_{jlk}^{\text{eff}})^{\text{H}}\left(\mathbf{v}_{i_{lk}}^{\text{H}} \otimes \mathbf{I}_M\right)\right\}$$
$$= \left(\mathbf{v}_{i_{lk}}^{\text{H}} \otimes \rho\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\mathbf{I}_M\right) \tag{39}$$

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

$$\mathbb{E}_{\{\mathbf{h}\}}\{\text{vec}(\mathbf{Y}_j)\text{vec}(\mathbf{Y}_j)^{\text{H}}\} = \sigma^2\mathbf{I}_{MB}$$
$$+ \sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \mathbb{E}_{\{\mathbf{h}_{j\ell m}\}}\left\{\text{vec}(\mathbf{h}_{j\ell m}^{\text{eff}}\mathbf{v}_{i_{\ell m}}^{\text{T}})\text{vec}(\mathbf{h}_{j\ell m}^{\text{eff}}\mathbf{v}_{i_{\ell m}}^{\text{T}})^{\text{H}}\right\}$$
$$= \left(\sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \rho\frac{d_j(\mathbf{z}_{\ell m})}{d_\ell(\mathbf{z}_{\ell m})}\mathbf{v}_{i_{\ell m}}\mathbf{v}_{i_{\ell m}}^{\text{H}} + \sigma^2\mathbf{I}_B\right) \otimes \mathbf{I}_M. \tag{40}$$

The expression (6) is obtained by substituting (39) and (40) into (38), 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

$$\rho\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\mathbf{I}_M - \mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{h}_{jlk}^{\text{eff}}\text{vec}(\mathbf{Y}_j)^{\text{H}}\}$$
$$\times \left(\mathbb{E}_{\{\mathbf{h}\}}\{\text{vec}(\mathbf{Y}_j)\text{vec}(\mathbf{Y}_j)^{\text{H}}\}\right)^{-1}\mathbb{E}_{\{\mathbf{h}\}}\{\mathbf{h}_{jlk}^{\text{eff}}\text{vec}(\mathbf{Y}_j)^{\text{H}}\}^{\text{H}}$$
$$= \rho\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\left(1 - \frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\mathbf{v}_{i_{lk}}^{\text{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lk}}\right)\mathbf{I}_M$$
$$= \rho\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}\left(1 - \frac{\frac{d_j(\mathbf{z}_{lk})}{d_l(\mathbf{z}_{lk})}B}{\sum_{\ell \in \mathcal{L}} \sum_{m=1}^{K} \frac{d_j(\mathbf{z}_{\ell m})}{d_\ell(\mathbf{z}_{\ell m})}\mathbf{v}_{i_{lk}}^{\text{H}}\mathbf{v}_{i_{\ell m}} + \frac{\sigma^2}{\rho}}\right)\mathbf{I}_M, \tag{41}$$

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

**Proof of 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 $\lambda_{jlm} = \frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}$. Plugging these expressions into (12) yields, for MR, the expression in (42) at the top of the page, by multiplying each term by $\frac{1}{M^2\rho^2\mathbf{v}_{i_{jk}}^{\text{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}$. The expression in (16) for MR is now obtained by considering an achievable lower bound $\mathbb{E}_{\{\mathbf{z}\}}\{\log_2(1 + \frac{1}{f(\{\mathbf{z}\})})\} \geq \log_2(1 + \frac{1}{\mathbb{E}_{\{\mathbf{z}\}}\{f(\{\mathbf{z}\})\}})$ where the expectation with respect to user positions is moved to the denominator of the SINRs using Jensen's inequality. This leads to expectations of the following types:

$$\mathbb{E}_{\{\mathbf{z}\}}\left\{\frac{1}{\mathbf{v}_{i_{jk}}^{\text{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}\right\}$$
$$= \mathbb{E}_{\{\mathbf{z}\}}\left\{\frac{\sum_{\ell \in \mathcal{L}} \sum_{\tilde{m}=1}^{K} \frac{d_j(\mathbf{z}_{\ell\tilde{m}})}{d_\ell(\mathbf{z}_{\ell\tilde{m}})}\mathbf{v}_{i_{jk}}^{\text{H}}\mathbf{v}_{i_{\ell\tilde{m}}} + \frac{\sigma^2}{\rho}}{B}\right\}$$
$$= \frac{\sum_{\ell \in \mathcal{L}_j(\beta)} \mu_{j\ell}^{(1)}B + \frac{\sigma^2}{\rho}}{B} \tag{43}$$

$$\mathbb{E}_{\{\mathbf{z}\}}\left\{\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^2 \frac{\mathbf{v}_{i_{jk}}^{\text{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lm}}}{\mathbf{v}_{i_{jk}}^{\text{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}\right\} = \sum_{l \in \mathcal{L}_j(\beta)} \mu_{jl}^{(2)} \tag{44}$$

$$\mathbb{E}_{\{\mathbf{z}\}}\left\{\sum_{l \in \mathcal{L}} \sum_{m=1}^{K} \frac{\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}}{\mathbf{v}_{i_{jk}}^{\text{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}\right\}$$
$$= \sum_{l \in \mathcal{L}} K\mu_{jl}^{(1)}\frac{\sum_{\ell \in \mathcal{L}_j(\beta)} \mu_{j\ell}^{(1)}B + \frac{\sigma^2}{\rho}}{B} + \sum_{l \in \mathcal{L}_j(\beta)} \mu_{jl}^{(2)} - (\mu_{jl}^{(1)})^2 \tag{45}$$

where we have utilized the definition in (18) and the non-universal pilot reuse assumption to identify the expectations.

$$\text{SINR}_{jk}^{(\text{ul})} = \cfrac{1}{\displaystyle\sum_{l\in\mathcal{L}}\sum_{m=1}^{K}\left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^2\frac{\mathbf{v}_{i_{jk}}^{\mathrm{H}}\mathbf{v}_{i_{lm}}}{B} + \cfrac{\rho\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\left(1-A_{jl}^{\text{ZF}}\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\mathbf{v}_{i_{lm}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lm}}\right)}{(M-K)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}} - 1 + \cfrac{\sigma^2}{(M-K)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}} \tag{49}$$

$$\text{SINR}_{jk}^{(\text{ul})} = \cfrac{1}{\displaystyle\sum_{l\in\mathcal{L}}\sum_{m=1}^{K}\left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^2\frac{\mathbf{v}_{i_{jk}}^{\mathrm{H}}\mathbf{v}_{i_{lm}}}{B} + \cfrac{\rho\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\left(1-\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\mathbf{v}_{i_{lm}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lm}}\right)}{(M-B)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}} - 1 + \cfrac{\sigma^2}{(M-B)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}} \tag{54}$$

---

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

$$\mathbb{E}_{\{\mathbf{h}\}}\{\|\mathbf{g}_{jk}^{\text{ZF}}\|^2\} = \frac{1}{(M-K)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}} \tag{46}$$

$$p_{jk}|\mathbb{E}_{\{\mathbf{h}\}}\{(\mathbf{g}_{jk}^{\text{ZF}})^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2 = 1 \tag{47}$$

$$p_{lm}\mathbb{E}_{\{\mathbf{h}\}}\{|(\mathbf{g}_{jk}^{\text{ZF}})^{\mathrm{H}}\mathbf{h}_{jlm}|^2\} = \left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^2\frac{\mathbf{v}_{i_{jk}}^{\mathrm{H}}\mathbf{v}_{i_{lm}}}{B}$$
$$+ \frac{\rho\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\left(1-A_{jl}^{\text{ZF}}\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\mathbf{v}_{i_{lm}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lm}}\right)}{(M-K)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}} \tag{48}$$

with $A_{jl}^{\text{ZF}} = 1$ if $l \in \mathcal{L}_j(\beta)$ and zero otherwise, 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 $\mathbf{v}_{i_{jk}}^{\mathrm{H}}\mathbf{v}_{i_{lm}} \neq 0$ (i.e., when the same pilot signal is used). The second term is the product between $\mathbb{E}_{\{\mathbf{h}\}}\{\|\mathbf{g}_{jk}^{\text{ZF}}\|^2\}$ and the variance of the estimation error of the effective channel $\sqrt{p_{lm}}\mathbf{h}_{jlm}$ if $A_{jl}^{\text{ZF}} \neq 0$ (i.e., if the UE is in a cell $l \in \mathcal{L}_j(\beta)$) or the original variance of $\sqrt{p_{lm}}\mathbf{h}_{jlm}$ if $A_{jl}^{\text{ZF}} = 0$. Using (46)–(48), we obtain the expression (49) at the of the page 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 (43)–(45) reappear along with

$$\mathbb{E}_{\{\mathbf{z}\}}\left\{-\left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^2\frac{\mathbf{v}_{i_{lm}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lm}}}{\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}}\right\}$$
$$\leq -(\mu_{jl}^{(1)})^2\frac{\displaystyle\sum_{\ell\in\mathcal{L}_j(\beta)}\mu_{j\ell}^{(1)} + \frac{\sigma^2}{B\rho}}{\displaystyle\sum_{\ell\in\mathcal{L}_l(\beta)}\mu_{j\ell}^{(1)} + \frac{\sigma^2}{B\rho}}, \tag{50}$$

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

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

$$\mathbb{E}_{\{\mathbf{h}\}}\{\|\mathbf{g}_{jk}^{\text{P-ZF}}\|^2\} = \frac{1}{(M-B)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}} \tag{51}$$

$$p_{jk}|\mathbb{E}_{\{\mathbf{h}\}}\{(\mathbf{g}_{jk}^{\text{P-ZF}})^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2 = 1 \tag{52}$$

$$p_{lm}\mathbb{E}_{\{\mathbf{h}\}}\{|(\mathbf{g}_{jk}^{\text{P-ZF}})^{\mathrm{H}}\mathbf{h}_{jlm}|^2\} = \left(\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\right)^2\frac{\mathbf{v}_{i_{jk}}^{\mathrm{H}}\mathbf{v}_{i_{lm}}}{B}$$
$$+ \frac{\rho\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\left(1-\frac{d_j(\mathbf{z}_{lm})}{d_l(\mathbf{z}_{lm})}\mathbf{v}_{i_{lm}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{lm}}\right)}{(M-B)\rho\mathbf{v}_{i_{jk}}^{\mathrm{H}}\boldsymbol{\Psi}_j^{-1}\mathbf{v}_{i_{jk}}} \tag{53}$$

by following the procedures used for ZF. Using (51)–(53), we obtain the expression (54) at the top of the page for P-ZF. The final expression is obtained by considering an achievable lower bound $\mathbb{E}_{\{\mathbf{z}\}}\{\log_2(1 + \frac{1}{f(\mathbf{z})})\} \geq \log_2(1 + \frac{1}{\mathbb{E}_{\{\mathbf{z}\}}\{f(\mathbf{z})\}})$ using Jensen's inequality, similar to the ZF case in Theorem1.

**Proof of Theorem 3:** Suppose that $\gamma_{jk} = \text{SINR}_{jk}^{(\text{ul})}$ 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 $\gamma_{jk} = \text{SINR}_{jk}^{(\text{dl})}$ for the DL SINR in (23). This condition can also be expressed as

$$\frac{\gamma_{jk}\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}}{|\mathbb{E}_{\{\mathbf{h}\}}\{\check{\mathbf{g}}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2}$$
$$= \frac{q_{jk}}{\displaystyle\sum_{l\in\mathcal{L}}\sum_{m=1}^{K}q_{lm}\frac{\mathbb{E}_{\{\mathbf{h}\}}\{|\check{\mathbf{g}}_{lm}^{\mathrm{H}}\mathbf{h}_{ljk}|^2\}}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{lm}\|^2\}} - q_{jk}\frac{|\mathbb{E}_{\{\mathbf{h}\}}\{\check{\mathbf{g}}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}} + \sigma^2}. \tag{55}$$

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

$$[\boldsymbol{\Psi}_{jl}]_{k,m}$$
$$= \begin{cases} \frac{\mathbb{E}_{\{\mathbf{h}\}}\{|\check{\mathbf{g}}_{lm}^{\mathrm{H}}\mathbf{h}_{ljk}|^2\}}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{lm}\|^2\}} - \frac{|\mathbb{E}_{\{\mathbf{h}\}}\{\check{\mathbf{g}}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}} & \text{if } k=m, j=l, \\ \frac{\mathbb{E}_{\{\mathbf{h}\}}\{|\check{\mathbf{g}}_{lm}^{\mathrm{H}}\mathbf{h}_{ljk}|^2\}}{\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{lm}\|^2\}} & \text{otherwise.} \end{cases} \tag{56}$$

Moreover, we define the $K|\mathcal{L}| \times K|\mathcal{L}|$ block diagonal matrix $\mathbf{D}$, where the $j$th $K \times K$ block is $\mathbf{D}_j$ and its $k$th diagonal element is

$$[\mathbf{D}_j]_{k,k} = \frac{\gamma_{jk}\mathbb{E}_{\{\mathbf{h}\}}\{\|\check{\mathbf{g}}_{jk}\|^2\}}{|\mathbb{E}_{\{\mathbf{h}\}}\{\check{\mathbf{g}}_{jk}^{\mathrm{H}}\mathbf{h}_{jjk}\}|^2}. \tag{57}$$

Using this notation, (55) can be expressed as

$$[\mathbf{D}_j]_{k,k} = \frac{q_{jk}}{\displaystyle\sum_{l\in\mathcal{L}}\sum_{m=1}^{K}q_{lm}[\boldsymbol{\Psi}_{jl}]_{k,m} + \sigma^2}$$
$$\Leftrightarrow \quad [\mathbf{D}_j]_{k,k}\sigma^2 = q_{jk} - \sum_{l\in\mathcal{L}}\sum_{m=1}^{K}q_{lm}[\mathbf{D}_j]_{k,k}[\boldsymbol{\Psi}_{jl}]_{k,m}. \tag{58}$$

In matrix form, the DL SINR conditions for all UEs in all cells can be expressed as $\mathbf{D}\sigma^2 = \mathbf{q} - \mathbf{D}\boldsymbol{\Psi}\mathbf{q}$, where

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

$\mathbf{q}=[\mathbf{q}_1^{\mathrm{T}}\ \ldots\ \mathbf{q}_{|\mathcal{L}|}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{q}_j=[q_{j1}\ \ldots\ q_{jK}]^{\mathrm{T}}$ contain the DL transmit powers in the $j$th cell. This expression can now be solved for $\mathbf{q}$. The matrix $\mathbf{D}$ depends only on the precoding vectors, thus for any choice of precoding scheme the sought SINRs are achieved by the power control policy

$$\mathbf{q}^\star=\sigma^2(\mathbf{I}_{K|\mathcal{L}|}-\mathbf{D}\boldsymbol{\Psi})^{-1}\mathbf{D}\mathbf{1} \qquad (59)$$

where $\mathbf{1}$ is the vector with only ones. $\mathbf{q}^\star$ is a feasible power control (i.e., has positive values) if all eigenvalues of $(\mathbf{I}_{K|\mathcal{L}|}-\mathbf{D}\boldsymbol{\Psi})$ 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 $\boldsymbol{\Psi}$ is replaced by $\boldsymbol{\Psi}^{\mathrm{T}}$:

$$\mathbf{D}\sigma^2=\mathbf{p}-\mathbf{D}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{p} \quad\Leftrightarrow\quad \mathbf{p}=\sigma^2(\mathbf{I}_{K|\mathcal{L}|}-\mathbf{D}\boldsymbol{\Psi}^{\mathrm{T}})^{-1}\mathbf{D}\mathbf{1}, \qquad (60)$$

where $\mathbf{p}=[\mathbf{p}_1^{\mathrm{T}}\ \ldots\ \mathbf{p}_{|\mathcal{L}|}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{p}_j=[p_{j1}\ \ldots\ p_{jK}]^{\mathrm{T}}$, if $\check{\mathbf{g}}_{jk}=\mathbf{g}_{jk}^{\mathrm{scheme}}$ for all $j$ and $k$. Since the eigenvalues of $(\mathbf{I}_{K|\mathcal{L}|}-\mathbf{D}\boldsymbol{\Psi})$ and $(\mathbf{I}_{K|\mathcal{L}|}-\mathbf{D}\boldsymbol{\Psi}^{\mathrm{T}})$ are the same, we can always select the DL powers according to (59). It is straightforward to verify that $\mathbf{1}^{\mathrm{T}}\mathbf{q}^\star=\mathbf{p}^{\mathrm{T}}\mathbf{1}$, 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 (25) follows directly from Theorems 1 and 2.

**Proof of 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 $\log_2\left(1+\frac{|\mathbb{E}\{s\}|^2}{\mathbb{E}\{|n|^2\}+\mathbb{E}\{|s|^2\}-|\mathbb{E}\{s\}|^2}\right)$ [25]. For the hardware impairment models in (32) and (33), the received signals (after linear processing) behave as $\sqrt{1-\epsilon^2}(s+n)+\epsilon\eta$ instead, where $\mathbb{E}\{|\eta|^2\}=\mathbb{E}\{|n|^2\}+\mathbb{E}\{|s|^2\}$. Since the distortion $\eta$ is uncorrelated with $s$ and $n$ by assumption, $n+\eta$ is also uncorrelated with $s$ and the corresponding SINR is computed in (61) at the top of the page. The only impact of the distortion is thus the $(1-\epsilon^2)$-factors in front of $|\mathbb{E}\{s\}|^2$ in the numerator and denominator. The UL SINRs in (35) 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.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[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.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[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.*, vol. 14, no. 2, pp. 711–723, Feb. 2015.

[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] M. Karlsson and E. G. Larsson, "On the operation of massive MIMO with and without transmitter CSI," in *Proc. IEEE SPAWC*, 2014.

[21] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, July 2015.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

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

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

[32] 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.

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

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

[35] 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.

[36] U. Gustavsson *et al.*, "On the impact of hardware impairments on massive MIMO," in *Proc. IEEE GLOBECOM*, 2014.

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

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

**Erik G. Larsson** received his Ph.D. degree from Uppsala University, Sweden, in 2002. Since 2007, he is Professor and Head of the Division for Communication Systems in the Department of Electrical Engineering (ISY) at Linköping University (LiU) in Linköping, Sweden. He has previously been Associate Professor (Docent) at the Royal Institute of Technology (KTH) in Stockholm, Sweden, and Assistant Professor at the University of Florida and the George Washington University, USA. In the spring of 2015 he was a Visiting Fellow at Princeton University, USA, for four months.

His main professional interests are within the areas of wireless communications and signal processing. He has published some 100 journal papers on these topics, he is co-author of the textbook *Space-Time Block Coding for Wireless Communications* (Cambridge Univ. Press, 2003) and he holds 15 issued and many pending patents on wireless technology.

He has served as Associate Editor for several major journals, including the *IEEE Transactions on Communications* (2010-2014) and *IEEE Transactions on Signal Processing* (2006-2010). He serves as chair of the IEEE Signal Processing Society SPCOM technical committee in 2015–2016 and as chair of the steering committee for the *IEEE Wireless Communications Letters* in 2014–2015. He is the General Chair of the Asilomar Conference on Signals, Systems and Computers in 2015 (he was Technical Chair in 2012). He received the *IEEE Signal Processing Magazine* Best Column Award twice, in 2012 and 2014, and he is receiving the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015.

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

His research interests lie in fundamental mathematics, algorithms, statistics, information & communication sciences research. He is an Associate Editor in Chief of the journal Random Matrix: Theory and Applications and was an associate and senior area editor for IEEE Transactions on Signal Processing respectively in 2011-2013 and 2013-2014. Dr. Debbah is a recipient of the ERC grant MORE (Advanced Mathematical Tools for Complex Network Engineering). He is a IEEE Fellow, a WWRF Fellow and a member of the academic senate of Paris-Saclay. He has managed 8 EU projects and more than 24 national and international projects. He received 14 best paper awards, among which the 2007 IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award, the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC best paper award, the 2015 ICC best paper award, the 2015 IEEE Communications Society Leonard G. Abraham Prize and 2015 IEEE Communications Society Fred W. Ellersick Prize as well as the Valuetools 2007, Valuetools 2008, CrownCom2009, Valuetools 2012 and SAM 2014 best student paper awards. He is the recipient of the Mario Boella award in 2005, the IEEE Glavieux Prize Award in 2011 and the Qualcomm Innovation Prize Award in 2012. He is the co-founder of the start-up Ximinds.

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

His research interests include multi-antenna cellular communications, radio resource allocation, energy efficiency, massive MIMO, and network topology design. He is the first author of the textbook *Optimal Resource Allocation in Coordinated Multi-Cell System* (Foundations and Trends in Communications and Information Theory, 2013). He is also dedicated to reproducible research and has made a large amount of simulation code publicly available. Dr. Björnson received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA and the 2015 Ingvar Carlsson Award. He has received 5 best paper awards for novel research on optimization and design of multi-cell multi-antenna communications: ICC 2015, WCNC 2014, SAM 2014, CAMSAP 2011, and WCSP 2009.