

Spectral Efficiency of Superimposed Pilots in Uplink Massive MIMO Systems

Daniel Verenzuela, Emil Björnson and Luca Sanguinetti

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-147602>

N.B.: When citing this work, cite the original publication.

Verenzuela, D., Björnson, E., Sanguinetti, L., (2017), Spectral Efficiency of Superimposed Pilots in Uplink Massive MIMO Systems, *GLOBECOM 2017 - 2017 IEEE GLOBAL COMMUNICATIONS CONFERENCE*. <https://doi.org/10.1109/GLOCOM.2017.8254816>

Original publication available at:

<https://doi.org/10.1109/GLOCOM.2017.8254816>

Copyright: IEEE

<http://www.ieee.org/>

©2017 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.



Spectral Efficiency of Superimposed Pilots in Uplink Massive MIMO Systems

Daniel Verenzuela*, Emil Björnson*, Luca Sanguinetti^{†‡}.

*Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden.

[†]Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy.

[‡]Large Networks and System Group (LANEAS), CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France.

Abstract—Massive multiple-input multiple-output (MIMO) is a viable technology to improve the spectral efficiency (SE) by spatially multiplexing several users. A potential limitation of Massive MIMO in multicell systems is pilot contamination, which arises from interference in the channel estimation due to the reuse of pilot sequences in neighboring cells. A standard method to reduce pilot contamination, referred to as regular pilot (RP), is to adjust the length of the pilot sequences while transmitting data and pilot symbols disjointly. Alternatively, the superimposed pilot (SP) method sends a superposition of pilot and data symbols, thereby allowing the use of longer pilots which can also reduce pilot contamination. This work considers the uplink of a general multicell Massive MIMO system with SP and maximum ratio combining and derives rigorous closed-form achievable rates, which are used to make comparisons with RP. Numerical results consider a realistic random base station deployment and show that with SP the reduction of pilot contamination is outweighed by the additional coherent and non-coherent interference from the data transmission. Moreover, it turns out that, when the pilot length is optimized, RP provides comparable SE as with SP.

I. INTRODUCTION

The development of wireless communication systems is led by the continuous increase in mobile data traffic [2]. Massive multiple-input multiple-output (MIMO) has been proposed to deal with the traffic growth by increasing the spectral efficiency (SE) [3]. Massive MIMO utilizes a large number of antennas (e.g., hundreds) at the base stations (BSs) to communicate coherently with several (e.g., tens) user equipments (UEs) by virtue of spatial multiplexing.

The acquisition of channel state information (CSI) at the BS is essential in Massive MIMO to coherently detect data from UEs in the uplink. The time and frequency interval over which the channel can be considered to remain static and frequency flat, called the coherence block, has a limited size and, in turn, there is a finite number of orthogonal pilot sequences that are available for channel estimation. Thus, in multicell systems the pilot sequences need to be reused across cells. This creates coherent interference, called pilot contamination, between UEs that share the same pilots, which reduces estimation quality. In [4]–[8], all pilot sequences are assumed to be reused in all cells and pilot contamination is mitigated by exploiting spatial channel correlation [4]–[6] or data covariance matrices [7], [8]. Another approach is to not reuse the pilots in every cell [9],

[10]. This method can effectively reduce pilot contamination at the cost of increased estimation overhead that decreases the amount of data symbols transmitted per coherence block. This tradeoff has been studied in [10] suggesting a pilot overhead between 5% and 40% of the coherence block.

In all the aforementioned works, the transmission of pilot and data symbols is done separately. This method is referred to as regular pilot (RP) transmission. In [11]–[16], the authors explore an alternative method that relies on the simultaneous transmission of pilot and data signals. This method is referred to as superimposed pilot (SP) transmission. Recently, [15], [16] have shown that SP achieves promising results in Massive MIMO. Particularly, the UEs transmit a linear combination of pilot and data symbols within the whole coherence block, thereby allowing the use of longer pilot sequences and removing the estimation overhead. Having longer pilots means that they can be reused less frequently in the network. This allows to reduce pilot contamination and potentially to improve the SE. However, sending pilot and data signals simultaneously causes interference in the channel estimation process from data symbols. This degrades the estimation quality and creates correlation between channel estimates and data.

This paper evaluates the performance of SP in the uplink of a multicell Massive MIMO system and make comparisons with RP. To this end, we rigorously derive a closed-form rate expression with SP when using maximum ratio combining (MRC). This stands in contrast to prior works, such as [15], [16], which rely on approximate signal-to-interference-plus-noise ratio (SINR) expressions. The closed-form rate expression provides valuable insights into identifying all the interference sources and their impact on the SE. Then, to properly study the effects of intercell interference in a large practical system with irregular BS deployments, we provide a numerical analysis by using the stochastic geometry framework developed in [17]. Interestingly, it turns out that RP provides comparable SE than SP when the pilot length and powers are optimized.

II. SYSTEM MODEL

We consider the uplink of a multicell Massive MIMO system where each BS has M antennas and serves K single-antenna UEs. We define the set Φ_D containing all BSs in the system, where D denotes the density of BSs per unit area (measured in BS/km²). Note that this definition does not require the BSs to be distributed in any specific manner.

An extended version of this paper has been submitted to *IEEE TWC* [1]. This paper has received funding from ELLIIT and the Swedish Foundation for Strategic Research (SSF). The work of L. Sanguinetti was supported in part by ERC Starting MORE under Grant 305123.

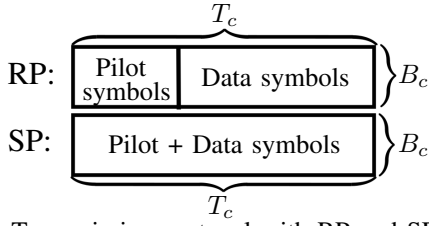


Fig. 1: Transmission protocol with RP and SP methods.

However, a stochastic geometry framework will be used later on in the numerical analysis. Without loss of generality, we focus on an arbitrary BS, denoted as BS_0 (serving UEs in cell 0), and an arbitrary UE k in cell 0, denoted as UE_{0k} . We define $\Psi_D = \Phi_D \setminus \{0\}$ as the set containing all other BSs.

We consider a system with bandwidth B_w . The communication channels are modeled as block fading where each channel is considered to be constant over a coherence block of time duration T_c and bandwidth B_c . The bandwidth is equally divided among all coherence blocks which means that B_w/B_c is an integer, and each block contains $\tau_c = B_c T_c$ complex samples. We assume uncorrelated Rayleigh fading channels since this is the first rigorous capacity analysis with SP in a multicell scenario. As done for RP, we believe that it is helpful to first develop fundamental theory for uncorrelated channels and then extend it to correlated ones (this is left for future work). Moreover, uncorrelated fading corresponds to the worst-case scenario for pilot contamination and thus gives insights into the main benefits of SP. We denote by $\mathbf{h}_{l' i} \in \mathbb{C}^{M \times 1}$ the channel between the M antennas of $\text{BS}_{l'}$ and $\text{UE}_{l' i}$ in which the small-scale fading (SSF) is modeled as $\mathbf{h}_{l' i} \sim \mathcal{CN}(\mathbf{0}, \beta_{l' i} \mathbf{I}_M)$ ¹ $\forall l', i \in \{1, \dots, K\}$ with $\beta_{l' i} \geq 0$ being the corresponding large-scale fading (LSF) coefficient. The received signal $\mathbf{Y}_0 \in \mathbb{C}^{M \times \tau_c}$ at BS_0 is

$$\mathbf{Y}_0 = \sum_{l' \in \Phi_D} \sum_{i=1}^K \mathbf{h}_{0l' i} \mathbf{x}_{l' i}^T + \mathbf{N}_0 \quad (1)$$

where \mathbf{N}_0 is the noise matrix with i.i.d. elements distributed as $[\mathbf{N}_0]_{m,j} \sim \mathcal{CN}(0, \sigma^2)$ and $\mathbf{x}_{l' i}$ represents the transmitted signal from $\text{UE}_{l' i}$ in all τ_c samples of the coherence block. The operator \cdot^T stands for transpose operation. The transmitted signal can be used for data symbols, pilot symbols or a superposition of the two depending on the employed method. We analyze the two transmission methods illustrated in Fig. 1: RP, called time-multiplexed in [15], and SP. With RP, data and pilot symbols are transmitted separately in each coherence block. Therefore, $\mathbf{x}_{l' i}$ contains only one of the two in each sample of the coherence block. With SP, pilot and data symbols are transmitted simultaneously during the whole coherence block and thus $\mathbf{x}_{l' i}$ contains a superposition of the two in each sample.

III. CHANNEL ESTIMATION

To estimate the channels, we use standard minimum mean squared error (MMSE) techniques with both RP and SP.

¹ $\mathcal{CN}(\cdot, \cdot)$ denotes a circularly symmetric complex Gaussian distribution.

We consider a set of τ_p orthogonal pilot sequences of length τ_p and, to avoid cumbersome pilot allocation between cells, each BS assigns $K \leq \tau_p$ different pilot sequences uniformly at random to the UEs served in its cell. A new pilot selection is made in each coherence block to average over pilot contamination. We denote as $\phi_{l' i} \in \mathbb{C}^{\tau_p \times 1}$, $\forall l' \in \Phi_D$, $i \in \{1, \dots, K\}$ the pilot sequence assigned to $\text{UE}_{l' i}$ with $|\phi_{l' i}|_j| = 1$, $\forall j \in \{1, \dots, \tau_p\}$. Each BS assigns the pilot sequences independently which leads to

$$\phi_{0k}^H \phi_{l' i} = \tau_p \chi_{l' i}^{(\tau_p)} \quad \text{for } l' \in \Psi_D, i \in \{1, \dots, K\} \quad (2)$$

where $\chi_{l' i}^{(\tau_p)} \in \{0, 1\}$ is a random variable being $\chi_{l' i}^{(\tau_p)} = 1$ when $\text{UE}_{l' i}$ has the same pilot as UE_{0k} and $\chi_{l' i}^{(\tau_p)} = 0$ otherwise. The operator \cdot^H denotes the Hermitian transpose. Recall that within each cell the pilot sequences assigned to the UEs are mutually orthogonal, that is $\phi_{l' i}^H \phi_{l' i'} = \tau_p$ for $i = i'$ and $\phi_{l' i}^H \phi_{l' i'} = 0$ for $i \neq i'$. Thus, for $l' \neq 0$ (where 0 is the index of the cell of interest) the random variable $\sum_{i=1}^K \chi_{l' i}^{(\tau_p)}$ is Bernoulli distributed with success probability K/τ_p . With RP, $\tau_p \in [K, \tau_c]$ samples are used for channel estimation leaving the remaining $\tau_d = \tau_c - \tau_p$ samples for data transmission. Thus, τ_p can be adjusted to find a good tradeoff between estimation accuracy and pilot overhead. On the other hand, with SP the whole coherence block is used for both data and pilot symbols. Therefore, we have that $\tau_p = \tau_d = \tau_c$. The received signal at BS_0 that is used for channel estimation with RP or SP is given by

$$\mathbf{z}_{0k} = \sum_{i=1}^K \sqrt{q_{0i}} \mathbf{h}_{00i} \phi_{0i}^T + \sum_{l' \in \Psi_D} \sum_{i=1}^K \sqrt{q_{l' i}} \mathbf{h}_{0l' i} \phi_{l' i}^T + \mathbf{C}_p \quad (3)$$

where

$$\mathbf{C}_p = \begin{cases} \bar{\mathbf{N}}_0 & \text{with RP} \\ \sum_{l' \in \Phi_D} \sum_{i=1}^K \sqrt{p_{l' i}} \mathbf{h}_{0l' i} \mathbf{s}_{l' i}^T + \mathbf{N}_0 & \text{with SP.} \end{cases} \quad (4)$$

The terms $p_{l' i}$ and $q_{l' i}$ account for the powers transmitted by $\text{UE}_{l' i}$ for data and pilot symbols, respectively. The vector $\mathbf{s}_{l' i} \in \mathbb{C}^{\tau_d \times 1}$ contains the data symbols transmitted in the whole coherence block. We assume the data symbols to be i.i.d. as $\mathbf{s}_{l' i} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{\tau_d})$. The noise matrices are defined as $\bar{\mathbf{N}}_0 \in \mathbb{C}^{M \times \tau_p}$ with RP and as $\mathbf{N}_0 = [\mathbf{n}_{01}, \dots, \mathbf{n}_{0\tau_c}] \in \mathbb{C}^{M \times \tau_c}$ with SP, where $[\bar{\mathbf{N}}_0]_{m,j} \sim \mathcal{CN}(0, \sigma^2)$ and $\mathbf{n}_{0j} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$. Let $\mathbf{z}_{0k} = \mathbf{Z}_{0k} \frac{\phi_{0k}^*}{\sqrt{\tau_p}} \in \mathbb{C}^{M \times 1}$ (where \cdot^* stands for complex conjugate) which is equivalent to a despreading operation, then

$$\mathbf{z}_{0k} = \sqrt{q_{0k} \tau_p} \mathbf{h}_{00k} + \sum_{l' \in \Psi_D} \sum_{i=1}^K \chi_{l' i}^{(\tau_p)} \sqrt{q_{l' i} \tau_p} \mathbf{h}_{0l' i} + \bar{\mathbf{c}}_p \quad (5)$$

where

$$\bar{\mathbf{c}}_p = \begin{cases} \bar{\mathbf{n}}_0 & \text{with RP} \\ \sum_{l' \in \Phi_D} \sum_{i=1}^K \sqrt{\frac{p_{l' i}}{\tau_p}} \mathbf{h}_{0l' i} \mathbf{s}_{l' i}^T \phi_{0k}^* + \sum_{j=1}^{\tau_p} \mathbf{n}_{0j} \frac{[\phi_{0k}]_j^*}{\sqrt{\tau_p}} & \text{with SP} \end{cases} \quad (6)$$

and $\bar{\mathbf{n}}_0 = \bar{\mathbf{N}}_0 \phi_{0k}^* / \sqrt{\tau_p}$ is a noise vector distributed as $\bar{\mathbf{n}}_0 \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. Recall that with SP $\tau_p = \tau_c$ in the

$$[\hat{\mathbf{s}}_{0k}]_j = \sqrt{p_{0k}} \mathbb{E} \left\{ \mathbf{v}_{00k}^H \mathbf{h}_{00k} \right\} [\mathbf{s}_{0k}]_j + \sqrt{p_{0k}} \left(\mathbf{v}_{00k}^H \mathbf{h}_{00k} - \mathbb{E} \left\{ \mathbf{v}_{00k}^H \mathbf{h}_{00k} \right\} \right) [\mathbf{s}_{0k}]_j + \sum_{\substack{l' \in \Psi_D \\ \{l', i\} \neq \{0, k\}}} \sum_{i=1}^K \sqrt{p_{l'i}} \mathbf{v}_{00k}^H \mathbf{h}_{0l'i} [\mathbf{s}_{l'i}]_j + \mathbf{v}_{00k}^H \mathbf{n}_{0j} \quad \text{with RP} \quad (13)$$

$$[\hat{\mathbf{s}}_{0k}]_j = \sqrt{\frac{p_{0k}}{M\beta_{00k}}} \mathbb{E} \left\{ \|\mathbf{h}_{00k}\|^2 \right\} [\mathbf{s}_{0k}]_j + \sqrt{\frac{p_{0k}}{M\beta_{00k}}} \left(\|\mathbf{h}_{00k}\|^2 - \mathbb{E} \left\{ \|\mathbf{h}_{00k}\|^2 \right\} \right) [\mathbf{s}_{0k}]_j \quad \text{with SP} \quad (14)$$

$$+ \underbrace{\sqrt{p_{0k}} \bar{\mathbf{v}}_{00k}^H \mathbf{h}_{00k} [\mathbf{s}_{0k}]_j + \sqrt{q_{0k}} \mathbf{v}_{00k}^H \mathbf{h}_{00k} [\phi_{0k}]_j + \sum_{\substack{l' \in \Phi_D \\ \{l', i\} \neq \{0, k\}}} \sum_{i=1}^K \left(\sqrt{q_{l'i}} [\phi_{l'i}]_j + \sqrt{p_{l'i}} [\mathbf{s}_{l'i}]_j \right) \mathbf{v}_{00k}^H \mathbf{h}_{0l'i} + \mathbf{v}_{00k}^H \mathbf{n}_{0j}}_{=n_{eff}}$$

definition of \mathbf{z}_{0k} and \bar{c}_p . Then, we use \mathbf{z}_{0k} to compute the MMSE estimate with RP and linear MMSE (LMMSE) estimate with SP of \mathbf{h}_{00k} . This is summarized in the following lemma.

Lemma 1. *For any given realization of $\chi_{l'i}^{(\tau_p)} \forall l' \in \Psi_D$, $i \in \{1, \dots, K\}$, the MMSE estimate of \mathbf{h}_{00k} with RP is*

$$\hat{\mathbf{h}}_{00k} = \frac{\bar{\gamma}_{0k}^{\text{RP}}}{\sqrt{q_{0k}\tau_p}} \mathbf{z}_{0k} \quad (7)$$

with

$$\bar{\gamma}_{0k}^{\text{RP}} = \frac{q_{0k}\tau_p\beta_{00k}}{q_{0k}\tau_p\beta_{00k} + \sum_{l' \in \Psi_D} \sum_{i=1}^K \chi_{l'i}^{(\tau_p)} q_{l'i}\tau_p\beta_{0l'i} + \sigma^2} \quad (8)$$

and has covariance matrix

$$\mathbb{E} \left\{ \hat{\mathbf{h}}_{00k} \hat{\mathbf{h}}_{00k}^H \middle| \chi_{l'i}^{(\tau_p)} \right\} = \beta_{00k} \bar{\gamma}_{0k}^{\text{RP}} \mathbf{I}_M. \quad (9)$$

With SP, the LMMSE estimate of \mathbf{h}_{00k} is

$$\hat{\mathbf{h}}_{00k} = \frac{\bar{\gamma}_{0k}^{\text{SP}}}{\sqrt{q_{0k}\tau_c}} \mathbf{z}_{0k} \quad (10)$$

with

$$\bar{\gamma}_{0k}^{\text{SP}} = \frac{q_{0k}\tau_c\beta_{00k}}{q_{0k}\tau_c\beta_{00k} + \sum_{l' \in \Psi_D} \sum_{i=1}^K \chi_{l'i}^{(\tau_c)} q_{l'i}\tau_c\beta_{0l'i} + \sum_{l' \in \Phi_D} \sum_{i=1}^K p_{l'i}\beta_{0l'i} + \sigma^2}. \quad (11)$$

The covariance matrix of $\hat{\mathbf{h}}_{00k}$ is given by

$$\mathbb{E} \left\{ \hat{\mathbf{h}}_{00k} \hat{\mathbf{h}}_{00k}^H \middle| \chi_{l'i}^{(\tau_c)} \right\} = \beta_{00k} \bar{\gamma}_{0k}^{\text{SP}} \mathbf{I}_M. \quad (12)$$

Proof: It follows from applying standard MMSE and LMMSE techniques [18, Ch. 12] to the problem at hand. ■

Remark 1. *With SP, the interference due to the received data symbols (see the first term with SP in the right-hand-side of (6)) has the following effects on the channel estimation: i) It reduces the quality of the channel estimates; ii) It changes the distribution of the channel estimates: the signal \mathbf{z}_{0k} with SP in (5) is not Gaussian. Thus, the LMMSE estimates do not coincide with the true MMSE estimates and are only uncorrelated to the channel estimation errors but not independent; iii) It creates correlation between the channel estimates and received data symbols from all UEs. These three phenomena play a key role in the achievable rate analysis with SP.*

The parameters $\bar{\gamma}_{0k}^{\text{RP}} \in [0, 1]$ and $\bar{\gamma}_{0k}^{\text{SP}} \in [0, 1]$ in Lemma 1 indicate the quality of the channel estimates. As the length τ_p of the pilot sequences increases, both $\bar{\gamma}_{0k}^{\text{RP}}$ and $\bar{\gamma}_{0k}^{\text{SP}}$ also increase

since the noise term (and data symbol interference with SP) becomes less significant and the probability of having UEs selecting the same pilot as UE_{0k} decreases. This means that as τ_p increases the variance of the channel estimate approaches the variance of the true channel and thus the estimation errors vanish. However, in practical applications $\tau_p \leq \tau_c$ with RP and $\tau_p = \tau_c$ with SP. Since τ_c is limited by the physical properties of the channel, there will always be an estimation error due to pilot contamination and noise (and the data symbol interference with SP). Observe that, for those scenarios where τ_c is much larger than K , the channel estimates with RP can be improved by letting τ_p be larger than K .

IV. ACHIEVABLE RATES WITH MRC

To evaluate the performance of the network, we derive ergodic achievable rates by applying standard lower bounding techniques on the capacity (e.g., [3]). Since we consider a fixed bandwidth, the SE is obtained simply by scaling the achievable rates with $1/B_w$. We assume that MRC is employed for data detection. Particularly, the statistics of the data symbols transmitted by UE_{0k} are obtained at BS₀ by the inner product $\mathbf{v}_{00k}^H \mathbf{Y}_0$ with $\mathbf{v}_{00k} = v_{00k} \hat{\mathbf{h}}_{00k}$, where $v_{00k} = \frac{1}{\bar{\gamma}_{0k}^{\text{RP}} \sqrt{M\beta_{00k}}}$ with RP and $v_{00k} = \frac{1}{\bar{\gamma}_{0k}^{\text{SP}} \sqrt{M\beta_{00k}}}$ with SP. These scaling factors are selected to provide an equivalent gain of $M\beta_{00k}$ for the desired signal with both methods. Note that although other detection schemes perform better with a finite number of antennas, MRC is used because with uncorrelated Rayleigh fading is close to optimal as $M \rightarrow \infty$ [3]. Moreover, the use of MRC gives a low complexity detection.

The received signal at BS₀ for an arbitrary data symbol j in the coherence block is $[\mathbf{Y}_0]_j$ (i.e., j^{th} column of \mathbf{Y}_0) where $\mathbf{x}_{l'i} = \sqrt{p_{l'i}} \mathbf{s}_{l'i}$ with RP and $\mathbf{x}_{l'i} = \sqrt{p_{l'i}} \mathbf{s}_{l'i} + \sqrt{q_{l'i}} \phi_{l'i}$ with SP. The data estimates are then computed as $[\hat{\mathbf{s}}_{0i}]_j = \mathbf{v}_{00k}^H [\mathbf{Y}_0]_j$ where the full expressions with RP and SP are given in (13) and (14), respectively, at the top of this page. With RP, we treat the term $\sqrt{p_{0k}} \mathbb{E} \left\{ \mathbf{v}_{00k}^H \mathbf{h}_{00k} \right\} [\mathbf{s}_{0k}]_j$ as the desired signal and the remaining terms in (13) as effective noise. This results into an equivalent single-input single-output (SISO) system with a deterministic channel and non-Gaussian effective noise, which is uncorrelated with the data symbol $[\mathbf{s}_{0k}]_j$. Notice that the individual terms in the effective noise (second to last terms in (13)) are also uncorrelated due to the fact that the data symbols from different UEs have zero mean

$$\text{SINR}_{0k}^{\text{SP}} = Mp_{0k}\beta_{00k} / \left(\underbrace{\frac{M}{\tau_c} \sum_{l' \in \Psi_D} \sum_{i=1}^K \frac{(p_{l'i} + (1 - \frac{1}{\tau_c})q_{l'i})q_{l'i}}{q_{0k}} \frac{\beta_{0l'i}^2}{\beta_{00k}} + \frac{M}{\tau_c} \sum_{l' \in \Phi_D} \sum_{i=1}^K \frac{(p_{l'i} + q_{l'i})p_{l'i}}{q_{0k}} \frac{\beta_{0l'i}^2}{\beta_{00k}}}_{\text{Coherent interference}} \right. \\ \left. + \frac{2}{\tau_c} p_{0k}\beta_{00k} + \frac{2}{\tau_c^2} \sum_{l' \in \Psi_D} \sum_{i=1}^K \frac{q_{l'i}p_{l'i}}{q_{0k}} \frac{\beta_{0l'i}^2}{\beta_{00k}} + \frac{1}{\tau_c^2} \sum_{l' \in \Phi_D} \sum_{i=1}^K \frac{p_{l'i}^2}{q_{0k}} \frac{\beta_{0l'i}^2}{\beta_{00k}} + \frac{1}{\gamma_{0k}^{\text{SP}}} \left(\sum_{l' \in \Phi_D} \sum_{i=1}^K (q_{l'i} + p_{l'i})\beta_{0l'i} + \sigma^2 \right) \right) \quad (22)$$

Non-coherent interference and noise

and are independent. With SP, we define

$$\bar{\mathbf{v}}_{00k} = \mathbf{v}_{00k} - v_{00k}\bar{\gamma}_{0k}^{\text{SP}}\mathbf{h}_{00k} = \mathbf{v}_{00k} - \frac{1}{\sqrt{M\beta_{00k}}}\mathbf{h}_{00k} \quad (15)$$

and we add and subtract $\sqrt{\frac{p_{0k}}{M\beta_{00k}}}\mathbb{E}\{\|\mathbf{h}_{00k}\|^2\}\{\mathbf{s}_{0k}\}_j$ from the data estimate $[\hat{\mathbf{s}}_{0k}]_j$ to obtain a desired signal with deterministic effective channel gain as depicted in (14). The term n_{eff} is defined in (14) for analytical tractability and accounts for the interference caused by received pilot and data symbols from all UEs (including self-interference from UE_{0k}), plus noise. Notice that the first term in (14) is uncorrelated with the remaining terms in (14) since the data symbols have zero mean and are independent and circularly symmetric complex Gaussian. Thus, with either RP or SP we have an equivalent SISO system with a deterministic effective channel and non-Gaussian effective noise for which an achievable rate can be obtained as a lower bound on the capacity [3, Ch. 2].

Recall that each UE in each cell has equal probability to be assigned to any pilot sequence. In particular, in an arbitrary cell $l' \in \Psi_D$ the probability of one arbitrary UE to have the same pilot as UE_{0k} is K/τ_p . Assuming that this UE is chosen uniformly at random, we have that

$$\mathbb{E}\left\{\sum_{l' \in \Psi_D} \sum_{i=1}^K \chi_{l'i}^{(\tau_p)} p_{l'i}\beta_{0l'i}\right\} = \sum_{l' \in \Psi_D} \frac{K}{\tau_p} \frac{1}{K} \left(\sum_{i=1}^K p_{l'i}\beta_{0l'i}\right). \quad (16)$$

Then, an ergodic achievable rate (defined as a lower bound on the capacity) with RP and SP is given by the next lemma and theorem, respectively.

Lemma 2. *An ergodic achievable rate for UE_{0k} with RP and MRC detection is given by*

$$R_{0k}^{\text{RP}} = B_w \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2(1 + \text{SINR}_{0k}^{\text{RP}}) \quad (17)$$

where $\text{SINR}_{0k}^{\text{RP}}$ is the effective SINR of UE_{0k} given by

$$\text{SINR}_{0k}^{\text{RP}} = \frac{Mp_{0k}\beta_{00k}}{\frac{M}{\tau_p} \sum_{l' \in \Psi_D} \sum_{i=1}^K p_{l'i}\beta_{0l'i} + \frac{1}{\gamma_{0k}^{\text{RP}}} \left(\sum_{l' \in \Phi_D} \sum_{i=1}^K p_{l'i}\beta_{0l'i} + \sigma^2\right)} \quad (18)$$

and

$$\gamma_{0k}^{\text{RP}} = \frac{q_{0k}\tau_p\beta_{00k}}{q_{0k}\tau_p\beta_{00k} + \sum_{l' \in \Psi_D} \sum_{i=1}^K q_{l'i}\beta_{0l'i} + \sigma^2}. \quad (19)$$

Proof: It follows from standard lower bounds on the capacity of the equivalent SISO system in (13) [3, Ch. 2]

with the effective channel $\mathbb{E}\{\mathbf{v}_{00k}^H \mathbf{h}_{00k}\}$. The closed form expression of the SINR follows the same approach as in [17], [3, Ch. 4] where the independence between the channel estimates and errors is used to compute the variance of the interfering terms in (13). ■

Theorem 1. *An ergodic achievable rate for UE_{0k} with SP and MRC detection is given by*

$$R_{0k}^{\text{SP}} = B_w \log_2(1 + \text{SINR}_{0k}^{\text{SP}}) \quad (20)$$

where $\text{SINR}_{0k}^{\text{SP}}$ is the effective SINR of UE_{0k} given by (22) (at the top of the page) and

$$\gamma_{0k}^{\text{SP}} = \frac{q_{0k}\tau_c\beta_{00k}}{q_{0k}\tau_c\beta_{00k} + \sum_{l' \in \Psi_D} \sum_{i=1}^K q_{l'i}\beta_{0l'i} + \sum_{l' \in \Phi_D} \sum_{i=1}^K p_{l'i}\beta_{0l'i} + \sigma^2}. \quad (21)$$

Proof: It follows from applying standard lower bounds on the capacity of the equivalent SISO system in (14) [3, Ch. 2]. The effective SINR expression in (22) follows from properties of circularly symmetric complex Gaussian vectors [1]. ■

Notice that with SP there is no pre-log factor in (20) since the whole coherence block is used for data transmission. We define coherent interference as the interference that adds constructively in the detection process due to correlation between the detection vector and the received signal. As a result, its variance scales with M . With non-coherent interference, we refer to sources of interference that are combined non-constructively and, in turn, its variance does not scale with M . There is coherent interference from pilot contamination and also from pilot and data symbols (see the first two terms in the denominator of (22)) due to the correlation between channel estimates and data symbols. Similarly, there is non-coherent interference from pilot symbols, data symbols and cross-correlation of the two (see the third and fourth terms in the denominator of (22)). The coherent gain (see the numerator of (22)) scales with M and depends on the factor γ_{0k}^{SP} (see (21)), which reflects the channel estimation quality, since it divides the noise and part of the non-coherent interference (see last term in the denominator of (22)).

In the prior works [15, Eq. (12)] and [16, Eq. (41)], approximate SINR expressions are provided with SP and MRC based on asymptotic favorable propagation and channel hardening (i.e., $\lim_{M \rightarrow \infty} \frac{\mathbf{h}_{0li}^H \mathbf{h}_{0l'i'}}{M} = 0$ if $\{l, i\} \neq \{l', i'\}$ and

$\lim_{M \rightarrow \infty} \frac{\|\mathbf{h}_{0li}\|^2}{M} = \beta_{0li}$). In contrast, the result in Theorem 1 does not rely on any asymptotic approximations. This enables

TABLE I: Simulation Parameters

Parameter	Symbol	Value
Pathloss exponent	α	3.76
Fixed propagation loss (1 km)	ω	130 dB
System bandwidth	B_w	20 MHz
BS density	D	100 BS/km ²
Noise power	$\sigma^2 B_w$	-110 dBm
Average number of BSs	N_{AV}	50

us to accurately analyze the system performance for any finite M . By comparing [15, Eq. (12)] and [16, Eq. (41)] with (22), we see that (22) contains extra interfering terms, which affect the data rates.

Remark 2. Since the pilot symbols are known to the BS, they can in principle be estimated and subtracted from the data estimate $[\hat{s}_{0k}]_j$ in (14) to increase the effective SINR. This leads to a modified effective noise

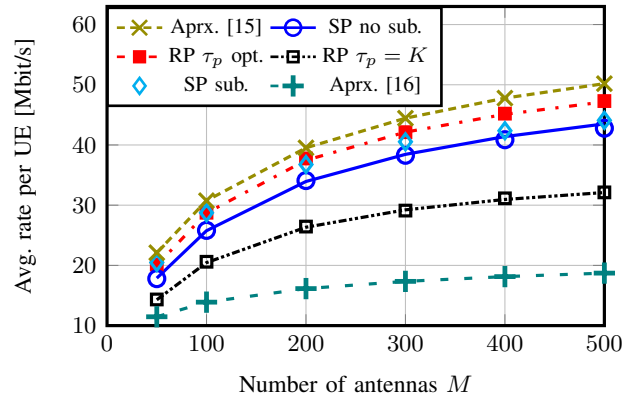
$$\bar{n}_{\text{eff}} = n_{\text{eff}} - \sum_{l' \in \Phi_D} \sum_{i=1}^K \sqrt{q_{l'i}} [\phi_{l'i}]_j \mathbf{v}_{00k}^H \hat{\mathbf{h}}_{0l'i} \quad (23)$$

which may have lower variance than n_{eff} provided that the quality of the channel estimates is good enough. When using the modified effective noise, the achievable rate can be computed as in (20) by replacing n_{eff} with \bar{n}_{eff} in (14). This is analyzed numerically in Section V.

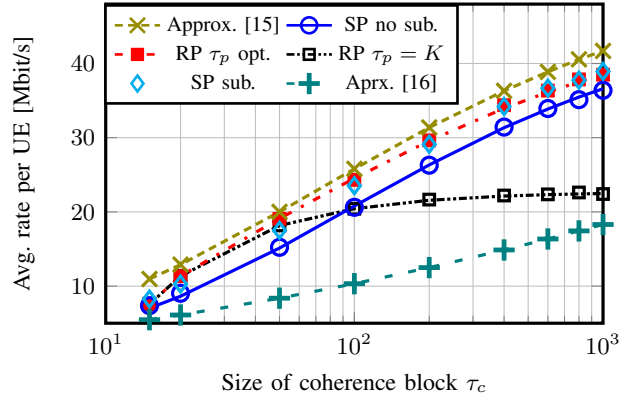
V. NUMERICAL RESULTS

To model the irregularity and large number of cells of practical system deployments, we conduct Monte Carlo (MC) simulations based on the stochastic geometry framework from [17], which has been shown to accurately model real network deployments [19]. The BSs are distributed according to a spatially homogeneous Poisson point process with density D [BS/km²]. We simulate BSs in a squared area of side length L_{SQ} [km] with an average of $N_{AV} = DL_{\text{SQ}}^2$ BSs. To avoid edge effects, the wrap around technique is used. The K UEs in each cell are assumed to be uniformly distributed within the Voronoi region around each BS. To model the LSF between UE $_{l'i}$ and BS $_l$, we define $\beta_{l'i} = \omega^{-1} d_{l'i}^{-\alpha}$ where α is the pathloss exponent and ω is the pathloss at a reference distance of 1 km. This parameter also accounts for propagation losses that are independent of the distance (e.g., wall penetration). Table I summarizes the simulation parameters. We consider statistical channel inversion power control where the transmission power of data symbols is $p_{l'i} = \rho_d / \beta_{l'i} = \rho_d \omega d_{l'i}^\alpha$ whereas that of pilot symbols is $q_{l'i} = \rho_p / \beta_{l'i} = \rho_p \omega d_{l'i}^\alpha$. The design parameters ρ_d and ρ_p are used to control the average transmit power. Moreover, we define ρ as the total average transmit power per symbol such that $\rho_d = \rho_p = \rho$ with RP and $\rho_d + \rho_p = \rho$ with SP. In the case of SP, the ratio between ρ_p and ρ_d is optimized to maximize the rate.

We evaluate the achievable rates with MRC for RP with pilot length equal to the number of users per BS, i.e., $\tau_p = K$, RP with optimal pilot length to maximize R_{0k}^{RP} (based on bisection search algorithms), standard SP as derived in Theorem 1



(a) Achievable rate per UE for $\rho = \sigma^2/4$ (SNR = -6 dB), $\tau_c = 200$ and $K = 10$.



(b) Achievable rate per UE for $\rho = \sigma^2/4$ (SNR = -6 dB), $M = 100$ and $K = 10$.

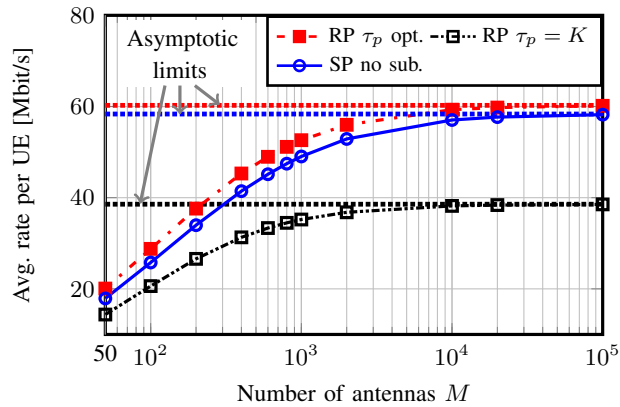
Fig. 2: Achievable rates versus M and τ_c . The lines correspond to the closed-form expressions in Lemma 2 and Theorem 1 while the markers are based on MC simulations over the SSF. All points are averaged over the LSF.

and SP when we subtract the received pilot symbols from the data estimate $[\hat{s}_{0k}]_j$ in (14), denoted as ‘‘SP sub.’’. In addition, we report the approximated data rates found in [15], [16].

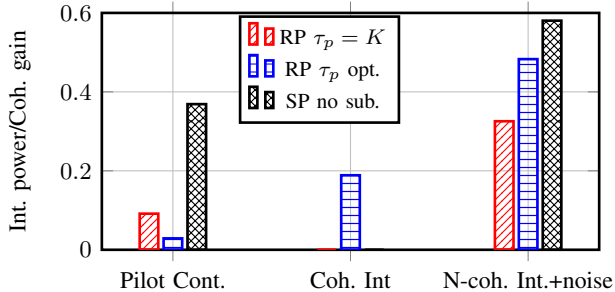
Fig. 2 depicts the achievable rates per UE versus the number of BS antennas and coherence block size. In general, we can see that SP outperforms RP with $\tau_p = K$ for most cases when τ_c is long enough to reduce pilot contamination. Otherwise, the data rates with RP, including the results when τ_p is optimized, are comparable to those of SP with pilot subtraction.² The approximation [15] is close to the MC results and therefore we can confirm its validity. However, the approximation [16] greatly underestimates the rates with SP and is discouraged.

Fig. 3a illustrates the achievable rate per UE for large values of M and make comparisons with the asymptotic results obtained by letting $M \rightarrow \infty$. We can see that more than 10^4 BS antennas are needed to converge to the limits and the relative differences among the methods vary between the finite and asymptotic number of BS antenna regime. Fig. 3b shows the strength of the interference sources with respect to the coherent

²Note that the smallest performance gap between SP and RP is found in low SNR scenarios, but these results were omitted for brevity.



(a) Achievable rate per UE for $\rho = \sigma^2/4$ (SNR = -6dB), $\tau_c = 200$ and $K = 10$.



(b) Interference sources for $M = 100$, $\tau_c = 200$ and $K = 10$. Results from Lemma 2 and Theorem 1.

Fig. 3: Achievable rates per UE versus M and interference sources received power over the coherent gain. All points are averaged over the LSF.

gain $M p_{0k} \beta_{00k}$ (see the numerators of (18) and (22)). With SP, there is a reduction of the pilot contamination contributions with respect to RP. At the same time, however, additional coherent interference appears from data transmission and, in the case of SP without pilot subtraction, that is substantial. It is important to note that the overall impact of coherent interference for $M = 100$ and $K = 10$ is lower than the impact of non-coherent interference with both RP and SP. This suggests that, in practical dense deployments, Massive MIMO may not be limited by coherent interference.

VI. CONCLUSIONS

In this paper, we derived the first rigorous achievable rate expression for a multicell Massive MIMO system with SP and MRC. We analytically and numerically compared SP and RP in a practical multicell deployment. By examining the contribution of different sources of interference, we observed that SP is able to reduce pilot contamination at the expense of creating additional coherent and non-coherent interference from the data transmission that limits the system performance. The results showed that, by optimizing the pilot length with RP, the average data rate per UE is comparable to the data rate with SP. When analyzing the large number of BS antennas regime, we encountered that asymptotic results do not always

convey accurate results to gain insights into the behavior of practical deployments. It is worth to stress that, in practical deployments, the effect of coherent interference (such as pilot contamination) can be less influential than non-coherent interference in the average data rates of UEs.

The use of SP may potentially provide better SE by using other detection schemes such as zero-forcing, multicell MMSE decoding, or iterative decoding algorithms. However, it is not clear whether the benefit will be substantial. This study is left for future work.

REFERENCES

- [1] D. Verenzuela, E. Björnson, and L. Sanguinetti, "Spectral and energy efficiency of superimposed pilots in uplink massive MIMO," *Submitted to IEEE Trans. Wireless Commun.*, 2017.
- [2] Cisco, "Visual networking index: Global mobile data traffic forecast update, 2014–2019," Tech. Rep., 2015.
- [3] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge Press, 2016.
- [4] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprasad, "Achieving "massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, 2012.
- [5] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.
- [6] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *Submitted to IEEE Trans. Wireless Commun.*, 2017. [Online]. Available: <http://arxiv.org/abs/1705.00538>
- [7] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE ICASSP*, 2012, pp. 3249–3252.
- [8] R. R. Müller, L. Cottarelli, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, 2014.
- [9] Y. Li, Y.-H. Nam, B. L. Ng, and J. Zhang, "A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse," in *Proc. IEEE GLOBECOM*, 2012, pp. 4500–4504.
- [10] E. Björnson, E. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, 2016.
- [11] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *IEEE Proc. GLOBECOM*, 1999, pp. 2162–2166.
- [12] G. T. Zhou, M. Viberg, and T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Process. Lett.*, vol. 10, no. 3, pp. 57–60, 2003.
- [13] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, 2004.
- [14] A. T. Asyhari and S. ten Brink, "Orthogonal or superimposed pilots? a rate-efficient channel estimation strategy for stationary MIMO fading channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2776–2789, 2017.
- [15] K. Upadhyay, S. A. Vorobyov, and M. Vehkaperä, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, 2017.
- [16] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multicell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, 2016.
- [17] E. Björnson, L. Sanguinetti, and M. Kountouris, "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 832–847, 2016.
- [18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [19] W. Lu and M. Di Renzo, "Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation," in *Proc. ACM MSWiM*, 2015, pp. 179–188.