Interference Management in 5G Reverse TDD HetNets: A Large System Analysis

Luca Sanguinetti, *Member, IEEE*, Aris L. Moustakas, *Senior Member, IEEE*, and Mérouane Debbah, *Senior Member, IEEE*

Abstract

This work analyzes a heterogeneous network (HetNet), which comprises a macro base station (BS) equipped with a large number of antennas and an overlaid dense tier of small cell access points (SCAs) using a wireless backhaul for data traffic. The static and low mobility user equipments (UEs) are associated with the SCAs while those with medium-to-high mobility are served by the macro BS. A reverse time division duplexing (TDD) protocol is used by the two tiers, which allows the BS to locally estimate both the intra-tier and inter-tier channels. This knowledge is then used at the BS either in the uplink (UL) or in the downlink (DL) to simultaneously serve the macro UEs (MUEs) and to provide the wireless backhaul to SCAs. A concatenated linear precoding technique employing either zero-forcing (ZF) or regularized ZF is used at the BS to simultaneously serve MUEs and SCAs in DL while nulling interference toward those SCAs in UL. We evaluate and characterize the performance of the system through the power consumption of UL and DL transmissions under the assumption that target rates must be satisfied and imperfect channel state information is available for MUEs. The analysis is conducted in the asymptotic regime where the number of BS antennas and the network size (MUEs and SCAs) grow large with fixed ratios. Results from large system analysis are used to provide concise formulae for the asymptotic UL and DL transmit powers and precoding vectors under the above assumptions. Numerical results are used to validate the analysis in different settings and to make comparisons with alternative network architectures.

I. Introduction

The biggest challenge for next generation wireless communication systems (5G) today is to support the ever-growing demands for higher date rates and to ensure a consistent quality of service (QoS) throughout the entire network [1]. Meeting these demands requires to increase network capacity by a factor for thousand over the next years [2]. At the same time, the power

L. Sanguinetti, A. L. Moustakas and M. Debbah are with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France ({luca.sanguinetti, merouane.debbah}@supelec.fr).

L. Sanguinetti is also with the University of Pisa, Dipartimento di Ingegneria dell'Informazione, Pisa, Italy (luca.sanguinetti@iet.unipi.it). A. L. Moustakas is also with Department of Physics, National & Capodistrian University of Athens, Athens, Greece (arislm@phys.uoa.gr).

consumption of the information and communication technology industry and the corresponding energy-related pollution are becoming major societal and economical concerns [3]–[6]. Hence, more cellular network capacity on the one hand and less energy consumption on the other are the seemingly contradictory future requirements on 5G. Since spectral resources are scarce, there is a broad consensus that this can only be achieved with a substantial network densification, i.e., a significant increase in the number of antennas deployed per unit area. In general, there are two different approaches for this, namely, large-scale or “massive” MIMO systems [7]–[9] and small-cell networks [10]. The first approach relies on using arrays with a few hundred antennas simultaneously serving many tens of user equipments (UEs) in the same frequency-time resource. The basic premise behind massive MIMO is to reap all the benefits of conventional MIMO, but on a much greater scale [8], [9]. The second approach relies on a very dense deployment of operator-installed low-cost and low-power small-cell access points (SCAs) possibly equipped with cognitive and cooperative functionalities. Although promising, each technology alone is unlikely to meet the QoS and capacity requirements for 5G [11]. On the other hand, a promising solution is a two-tier heterogeneous network (HetNet) in which the two above technologies coexist and interplay with each other in order to improve network performance [1]. In particular, massive MIMO is used to ensure outdoor coverage and to serve mobile UEs (allowing to minimize handoffs), while SCAs act as the main capacity-driver for indoor and outdoor UEs with low mobility. While conventional base stations (BSs) are typically connected through a high capacity wired backhaul network, the same is not true for SCAs, which are likely to be connected via an unreliable backhaul infrastructure whose features may strongly vary from case to case, with variable characteristics of error rate, delay, capacity and especially deployment cost. For such systems, the backhaul represents one of the major bottlenecks [10]. A more economically and viable alternative is to make use of the wireless link as a backhaul [12], [13].

A. Main contributions

In this work, we characterize and analyze the power consumption of an HetNet consisting of a massive MIMO macro tier overlaid with a second tier of SCAs. The UEs are endowed with a single antenna and have different speeds. Those that are static or have low mobility are associated with the SCAs while the medium-to-high mobility ones are served by the macro BS. The excess antennas at the BS are used to serve the macro UEs (MUEs) and at the same time to play the role itself of wireless backhaul to the SCAs. The latter are divided in two groups such that the distance between SCAs belonging to the same group is maximized and the arising interference is mitigated. A similar division is performed on the MUEs on the basis of their proximity to the SCAs (see Fig. 1 of Section II). On the other hand, the interference between the macro and second tier (the so-called two-tier interference) is handled using a reverse time-division-duplexing (TDD) mode, i.e., the BS is in downlink (DL) mode when the SCAs operate in uplink (UL), and vice versa. The TDD protocol results in a channel reciprocity that enables not only the estimation of large-dimensional channels at the BS, but also an implicit coordination between the two tiers without the need of exchanging channel state information (CSI) through the wireless backhaul. A minimum-mean-square-error (MMSE) receiver is used in UL at the BS for interference mitigation. On the other hand, a concatenated linear precoding technique employing either zero-forcing (ZF) or regularized ZF (RZF) is used in DL to satisfy rate constraints and to null interference towards SCAs, thereby providing the static small cell UEs (SUEs) a high-quality UL connection with very small power. The design and analysis of the network is performed under the assumption of imperfect CSI for the MUEs (due to their mobility) and is conducted in the asymptotic regime where the number of BS antennas N and the network size (MUEs and SCAs) grow large with fixed ratio.

As we shall see, the use of BS antennas for MMSE reception and precoding allows to keep the UL and DL transmit powers of all network devices at a relatively low level for small to moderate estimation errors of MUE channels. However, we show that for a given set of target rates there is a critical value of imperfect CSI beyond which the network operation becomes infeasible as it is manifested by the divergence of all powers. In this case, MUEs with high mobility have to lower their own target rates or they have to be served using other transmission protocols that dispense from CSI (such as for example space-time coding). This might also result into a substantial reduction of the served rates.

In summary, the main contributions of this work account for: *i*) the development of a reverse TDD protocol for the coexistence of a massive MIMO macro tier and a dense tier of SCAs using a wireless backhaul for data traffic; *ii*) the asymptotically design of a concatenated RZF precoding technique for meeting rate constraints under imperfect CSI of MUEs; *iii*) the large system analysis of the power consumption in the UL and DL of each tier. Next, all the above contributions are compared to existing literature.

B. Comparison with Related literature

The main literature for the system under investigation and the proposed TDD protocol is represented by [14] wherein the authors propose a similar protocol to exploit the excess antennas at the BS for intra- and inter-tier interference reduction. In contrast to [14], a wireless backhaul is introduced here for the secondary tier and imperfect CSI is assumed for MUEs. The wireless backhaul forces us to modify the transmission protocol in [14] so as to account for reverse TDD not only between tiers but also between SCAs. Moreover, we are interested in evaluating the power consumption of the network rather than the average sum rate and conduct the analysis in the large system regime. Similar differences can be identified with respect to [15].

The wireless backhaul has also been recently considered in [13] and [16]. In [13], the authors focus on the scalability properties of a wireless backhaul network modelled as a random multi-antenna extended network. Geometric arguments are used to derive an information theoretic upper bound on the capacity of the network. In [16], a two-tier network is considered under the assumption that SCAs are full-duplex devices equipped with interference cancellation capabilities. A different line of research for wireless backhaul is in the context of mm-Wave communications. In [17], for example, the use of outdoor mm-Wave communications for backhaul networking is considered and a wind sway analysis is presented to establish a notion of beam coherence time. This highlights a previously unexplored tradeoff between array size and wind-induced movement.

The impact of imperfect CSI has been investigated in [18] wherein the authors consider the DL of a multi-cell MIMO system serving UEs with large disparities in mobility. The analysis is conducted in the asymptotical regime and shows that the mobility of a UE has a detrimental effect on its own achievable rate, but has no direct impact on the other UEs. Instead, we consider a two-tier network and evaluate the impact of imperfect CSI on the power consumption in the UL and DL of each tier, while guaranteeing requested rates. Moreover, our analysis shows that orthogonal transmission resources should be allocated to highly mobile MUEs. A similar result has been pointed out in [19] and [20].

The asymptotically optimal design of linear precoding techniques has received great attention in the last years. Some results in this context can be found in [21]–[23]. Differently from [21], [22], this work considers a two tier network and is focused on analyzing the dual problem, namely, the power consumption of the overall network subject to target service rates. On the other hand, the major differences with respect to [23] are the system under investigation and the imperfect CSI assumption at the BS. As we shall see, this leads to different formulae for the asymptotic DL transmit powers and precoding vectors.

C. Organization

The remainder of this paper is organized as follows.[1] Next section introduces the network architecture along with the transmission protocol and channel model. Section III is focused on the UL phase of the BS and aims at computing the power required by all transmitters taking into account the arising interference. In Section IV, we consider the DL and deals with the asymptotic analysis and design of ZF and RZF. In Section V, numerical results are used to validate the theoretical analysis and make comparisons among different network architectures. In Section VI, we discuss a possible solution to overcome the mobility issues along with that of some other practical aspects such as network design, channel correlation at the BS antennas and dynamic UL-DL TDD configurations. Finally, the major conclusions and implications are drawn in Section VII.

II. System model

We consider a single-cell network where a macro tier is augmented with a certain number of low range SCAs. Each SCA possesses a single antenna and devotes its available resources to its pre-scheduled SUE. The macro BS employs N transmit antennas to serve its associated single-antenna MUEs. As shown in Fig. 1, we assume that the SCAs are divided into two groups $\mathcal{S}_{\mathcal{R}}$ (red colour) and $\mathcal{S}_{\mathcal{B}}$ (blue colour) such that the distance between SCAs belonging to the same group is maximized. A similar division is performed on the MUEs on the basis of their proximity to the SCAs. We denote $\mathcal{M}_{\mathcal{R}}$ $(\mathcal{M}_{\mathcal{B}})$ the set collecting MUEs that are closest to SCAs in $\mathcal{S}_{\mathcal{R}}$ $(\mathcal{S}_{\mathcal{B}})$. For notational convenience, we call $\mathcal{R} = \mathcal{M}_{\mathcal{R}} \cup \mathcal{S}_{\mathcal{R}}$ and $\mathcal{B} = \mathcal{M}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{B}}$.

While conventional systems have large disparity between peak and average rates, we aim at designing the system so as to guarantee target rates or, equivalently, signal-to-interference-plus-noise ratio (SINR) values. The analysis is conducted in the asymptotic regime in which the number of BS antennas increases as the network size becomes large.

[1] The following notation is used throughout the paper. Matrices and vectors are denoted by bold letters. The superscript † denotes hermitian operation and $|\mathcal{S}|$ is used to denote the cardinality of the enclosed set $\mathcal{S}$. We let $\mathbf{I}_K$ denote the $K \times K$ identity matrix, whereas $\mathbf{1}_K$ is the K-dimensional unit vector. We use $\mathcal{CN}(\cdot,\cdot)$ to denote a multi-variate circularly-symmetric complex Gaussian distribution whereas $\mathcal{N}(\cdot,\cdot)$ stands for a real one. The notation $\xrightarrow{a.s.}$ stands for almost surely equivalent. The notation $[\mathbf{D}]_{k,i}$ the (k,i)th element of the enclosed matrix $\mathbf{D}$ whereas $[\mathbf{a}]_i$ stands for the ith entry of vector $\mathbf{a}$.

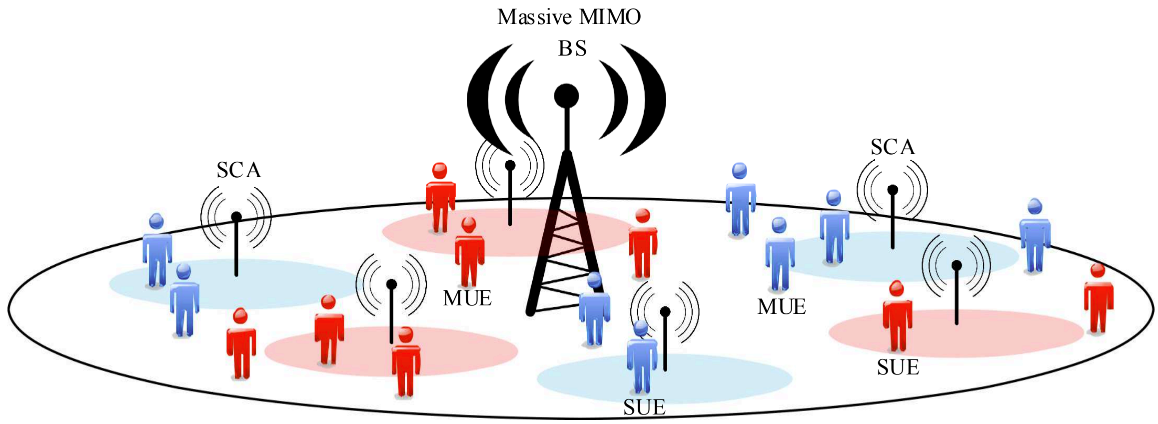

Fig. 1. Network architecture. The SCAs are divided in two different groups, namely, $\mathcal{S}_{\mathcal{B}}$ (blue colour) and $\mathcal{S}_{\mathcal{R}}$ (red colour). The same division is performed on the MUEs on the basis of their minimum distance from the SCAs.

A known problem with the asymptotic analysis is that the target rates are not guaranteed to be achieved when N is finite and relatively small (see for example [24]). This is because the approximation errors are translated into fluctuations in the resulting SINR values. However, these errors vanish rapidly also in the finite regime when N is large enough as it is envisioned for massive MIMO systems [25]. Nevertheless, most of asymptotic results are expected to remain valid also for MIMO systems equipped with a limited number of antennas in which scheduling protocols are used for serving the large number of UEs [26], [27].

A. Transmission model

The operating protocol is sketched in Fig. 2. In the frequency-time slot (W_1, T_1), the MUEs and SCAs in $\mathcal{R}$ use the frequency band W_1 for UL transmissions (BS $\leftarrow$ MUE and BS $\leftarrow$ SCA) for a time interval of length T_1 whereas the SCAs in $\mathcal{B}$ transmit to their associated SUEs in the DL (SCA $\rightarrow$ SUE). In (W_1, T_2), the converse takes place, i.e., the BS makes use of W_1 for a time interval of length T_2 to transmit in the DL to the MUEs and SCAs in $\mathcal{R}$ whereas the SUEs associated to the SCAs in $\mathcal{B}$ use W_1 for UL transmissions. The frequency-time slots (W_2, T_1) and (W_2, T_2) are used in the dual way. As seen, the exchange of information within each tier takes place in a reverse order, i.e., the BS is in the DL mode (BS $\rightarrow$ MUE) when the SCAs operate in the UL (SCA $\leftarrow$ SUE), and vice versa. We assume that transmissions across tiers are perfectly synchronized (the impact of asynchronous transmissions will be discussed in Section VI) and that the channel frequency response is flat over each frequency band. We also assume that $T_1 + T_2$ is upper bounded by the coherence time of the channel. In these circumstances, UL and DL channels

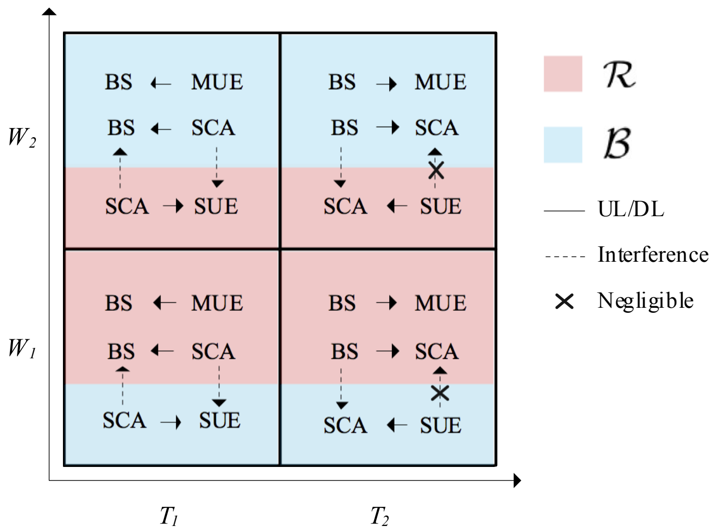

Fig. 2. Illustration of the transmission protocol. The exchange of information within each tier takes place in a reverse order, i.e., the BS is in the DL mode (BS $\rightarrow$ MUE) when the SCAs operate in the UL (SCA $\leftarrow$ SUE), and vice versa.

can be considered as reciprocal and the BS can make use of UL estimates for DL transmissions (more details on this will be given later on).

Remark 1. *In the simulations, comparisons will be made with the two transmission protocols shown in Fig. 3. In particular, the one on the left-hand-side applies to a network in which SCAs are connected to the BS through a wired backhaul [14]. On the other hand, the right-hand-side protocol is for a massive MIMO system in which only the macro-tier is present [9]. Although not necessarily needed, the division into $\mathcal{R}$ and $\mathcal{B}$ is maintained in both cases for a fair comparison.*

B. Channel Model and Assumptions

We denote $\mathbf{h}_i^{(\mathcal{M}_\mathcal{R})} = [h_i^{(\mathcal{M}_\mathcal{R})}(1), h_i^{(\mathcal{M}_\mathcal{R})}(2), \ldots, h_i^{(\mathcal{M}_\mathcal{R})}(N)]^T \in \mathbb{C}^{N \times 1}$ the vector whose entry $h_i^{(\mathcal{M}_\mathcal{R})}(n)$ accounts for the instantaneous propagation channel between the ith MUE in $\mathcal{M}_\mathcal{R}$ and the nth antenna at the BS. Accordingly, we let $\mathbf{H}^{(\mathcal{M}_\mathcal{R})} = [\mathbf{h}_1^{(\mathcal{M}_\mathcal{R})}\, \mathbf{h}_2^{(\mathcal{M}_\mathcal{R})} \cdots \mathbf{h}_{|\mathcal{M}_\mathcal{R}|}^{(\mathcal{M}_\mathcal{R})}] \in \mathbb{C}^{N \times |\mathcal{M}_\mathcal{R}|}$ be the matrix collecting the channels of all MUEs in $\mathcal{M}_\mathcal{R}$. We assume that:

$$\mathbf{h}_i^{(\mathcal{M}_\mathcal{R})} = \sqrt{l(\mathbf{x}_i)}\mathbf{w}_i \tag{1}$$

where $\mathbf{x}_i$ denotes the position of MUE i in $\mathcal{M}_\mathcal{R}$ (computed with respect to the BS), $\mathbf{w}_i \sim \mathcal{CN}(0, \mathbf{I}_N)$ accounts for the small-scale fading channel and $l(\mathbf{x}_i) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is the average channel gain due to pathloss at distance $||\mathbf{x}_i||$. Since the forthcoming analysis does not depend on a particular choice of $l(\mathbf{x}_i)$ as long as it is a decreasing function of the distance $||\mathbf{x}_i||$ and is bounded from below, we keep it generic [28]. The same model is adopted for the channels of SCAs and SUEs in $\mathcal{S}_\mathcal{R}$

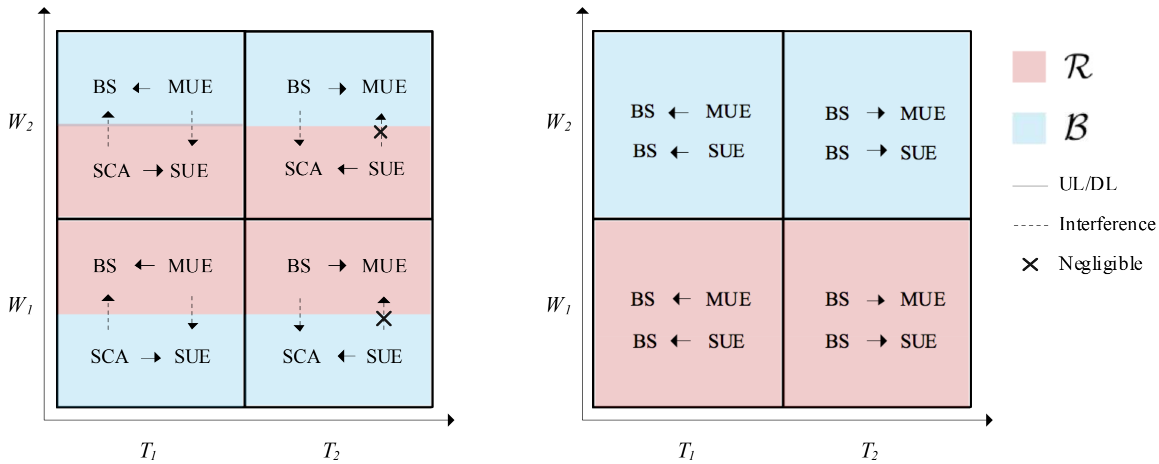

Fig. 3. Illustration of alternative transmission models and network configurations. *a*) HetNet with wired backhaul. The BS $\leftrightarrows$ SCA links are absent.; *b*) Massive MIMO. All UEs are served by the macro BS.

and $\mathcal{S}_\mathcal{B}$. In particular, we let $\mathbf{H}^{(\mathcal{S}_\mathcal{R})} \in \mathbb{C}^{N\times|\mathcal{S}_\mathcal{R}|}$ and $\mathbf{H}^{(\mathcal{S}_\mathcal{B})} \in \mathbb{C}^{N\times|\mathcal{S}_\mathcal{B}|}$ be the matrices collecting the channel gains from the BS antennas and the SCAs in $\mathcal{S}_\mathcal{R}$ and $\mathcal{S}_\mathcal{B}$, respectively.

In all subsequent discussions, we assume that imperfect CSI is available for $\mathbf{H}^{(\mathcal{M}_\mathcal{R})}$. In particular, we let $\widehat{\mathbf{H}}^{(\mathcal{M}_R)} = [\widehat{\mathbf{h}}_1^{(\mathcal{M}_R)}, \widehat{\mathbf{h}}_2^{(\mathcal{M}_R)}, \ldots, \widehat{\mathbf{h}}_{|\mathcal{M}_R|}^{(\mathcal{M}_R)}]$ denote an estimate of $\mathbf{H}^{(\mathcal{M}_\mathcal{R})}$ and assume that each vector $\widehat{\mathbf{h}}_i^{(\mathcal{M}_\mathcal{R})}$ is modelled as $\widehat{\mathbf{h}}_i^{(\mathcal{M}_\mathcal{R})} = \sqrt{l(\mathbf{x}_i)}\widehat{\mathbf{w}}_i$ with [22]

$$\widehat{\mathbf{w}}_i = \sqrt{1-\tau_i^2}\mathbf{w}_i + \tau_i \mathbf{v}_i \tag{2}$$

where $\mathbf{v}_i \sim \mathcal{CN}(0, \mathbf{I}_N)$ models the independent channel estimation errors. The parameter $\tau_i \in [0,1]$ reflects the accuracy or quality of the channel estimate $\widehat{\mathbf{h}}_i^{(\mathcal{M}_R)}$, i.e., $\tau_i = 0$ corresponds to perfect CSI, whereas for $\tau_i = 1$ the CSI is completely uncorrelated to the true channel. Observe that imperfect CSI arises naturally for MUEs as a consequence of mobility [18], [22]. Since SCAs occupy fixed positions in the network, then the propagation channels remain constant for a sufficiently large number of phases to be accurately estimated. For this reason, in all subsequent discussions we assume that $\mathbf{H}^{(\mathcal{S}_\mathcal{R})}$ and $\mathbf{H}^{(\mathcal{S}_\mathcal{B})}$ are perfectly known at the BS. The same assumption is made for the SUE channels.

III. Large System Analysis of the macro-tier interference in UL

We start dealing with the case in which the BS is in UL mode. Without loss of generality, the frequency-time slot (W_1, T_1) of Fig. 2 is considered.[2] As seen, two instances of interference appear. One comes from UL signals of MUEs and SCAs in $\mathcal{R}$ and affects the receiving SUEs in

[2]The same analysis can be performed for (W_2, T_1).

$\mathcal{B}$ whereas the other accounts for the interference that the BS experiences from the DL mode of SCAs in $\mathcal{S}_{\mathcal{B}}$. Despite the potentially large number of transmitting MUEs and SCAs, the former interference is limited due to the geographical separation of co-channel SUEs, which is one of the main advantages for spatially separating the system into $\mathcal{R}$ and $\mathcal{B}$. The large number of antennas provides an effective means to mitigate the latter interference and at the same time to simultaneously serve all transmitters in $\mathcal{R}$. For this purpose, we assume that an MMSE receiver is employed at the BS.

For notational convenience, we denote $K = |\mathcal{R}|$ the total number of transmitters (MUEs and SCAs) in $\mathcal{R}$ and call $S = |\mathcal{S}_{\mathcal{B}}|$ the number of SCAs in $\mathcal{S}_{\mathcal{B}}$. We also let

$$c = \frac{K}{N} \tag{3}$$

and

$$c_S = \frac{K+S}{N}. \tag{4}$$

We denote $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] = \left[\mathbf{H}^{(\mathcal{M}_{\mathcal{R}})}, \mathbf{H}^{(\mathcal{S}_{\mathcal{R}})}\right] \in \mathbb{C}^{N\times K}$ the matrix collecting the instantaneous UL channels of MUEs and SCAs in $\mathcal{R}$ and denote $\{p_k^{(\mathcal{R},\mathrm{ul})} \geq 0; k = 1, 2, \ldots, K\}$ the corresponding UL transmit powers. Letting $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K] \in \mathbb{C}^{N\times K}$ be the MMSE matrix, the UL achievable rate for device k in $\mathcal{R}$ is

$$R_k^{(\mathcal{R},\mathrm{ul})} = \log_2\left(1 + \mathrm{SINR}_k^{(\mathcal{R},\mathrm{ul})}\right) \tag{5}$$

with $\mathrm{SINR}_k^{(\mathcal{R},\mathrm{ul})}$ given by

$$\mathrm{SINR}_k^{(\mathcal{R},\mathrm{ul})} = \frac{p_k^{(\mathcal{R},\mathrm{ul})}\left|\mathbf{g}_k^\dagger \mathbf{h}_k\right|^2}{\sum\limits_{i=1,i\neq k}^{K} p_i^{(\mathcal{R},\mathrm{ul})}\left|\mathbf{g}_k^\dagger \mathbf{h}_i\right|^2 + \sum\limits_{s=1}^{S} p_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})}\left|\mathbf{g}_k^\dagger \mathbf{h}_s^{(\mathcal{S}_{\mathcal{B}})}\right|^2 + \sigma^2\left\|\mathbf{g}_k\right\|^2} \tag{6}$$

where σ^2 accounts for thermal noise and $p_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})} \geq 0$ is the DL transmit power of SCA s in $\mathcal{S}_{\mathcal{B}}$. As mentioned earlier, we assume that the MMSE receiver operates under the assumption of imperfect knowledge of $\mathbf{H}^{(\mathcal{M}_{\mathcal{R}})}$. This amounts to setting

$$\mathbf{G} = \widehat{\mathbf{H}}^\dagger\left(N\sigma^2\mathbf{I}_N + \sum_{i=1}^{K} p_i^{(\mathcal{R},\mathrm{ul})}\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^\dagger + \sum_{s=1}^{S} p_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})}\mathbf{h}_s^{(\mathcal{S}_{\mathcal{B}})}\mathbf{h}_s^{(\mathcal{S}_{\mathcal{B}})\dagger}\right)^{-1} \tag{7}$$

where $\widehat{\mathbf{h}}_k$ is the kth column of $\widehat{\mathbf{H}}$ given by $\widehat{\mathbf{H}} = \left[\widehat{\mathbf{H}}^{(\mathcal{M}_{\mathcal{R}})}\mathbf{H}^{(\mathcal{S}_{\mathcal{R}})}\right]$. Observe that we assume that $\{p_k^{(\mathcal{R},\mathrm{ul})}\}$ and $\{p_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})}\}$ are perfectly known at the BS. This information can be easily acquired through signalling [29].

Remark 2. *It is worth observing that in the frequency-time slot* (W_1, T_1) *under consideration* $\mathbf{H}^{(\mathcal{S}_{\mathcal{R}})}$ *can be easily acquired at the BS using UL pilots from SCAs in* $\mathcal{S}_{\mathcal{R}}$*. On the other hand, the estimation of* $\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})}$ *must be performed in a different way since the UL mode (BS* $\leftarrow$ *SCA) for* $\mathcal{S}_{\mathcal{B}}$ *takes place over the frequency band* W_2*. A possible solution might consist in using pilots that the SCAs in* $\mathcal{S}_{\mathcal{B}}$ *send in DL (SCA* $\rightarrow$ *SUE) to their associated SUEs. An alternative approach might be to periodically switch the operations of frequency bands* W_1 *and* W_2 *[30].*

The aim of this section is to compute the UL and DL transmit powers $\{p_k^{(\mathcal{R},\mathrm{ul})} \geq 0\}$ and $\{p_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})} \geq 0\}$ required to meet the corresponding rate constraints $\{r_k^{(\mathcal{R},\mathrm{ul})}; k = 1, 2, \ldots, K\}$ and $\{r_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})}; s = 1, 2, \ldots, S\}$ under imperfect CSI of MUEs. For notational convenience, the superscripts $^{(\mathrm{ul})}$ and $^{(\mathrm{dl})}$ are dropped in the sequel.

We start assuming that the downlink powers $\{p_s^{(\mathcal{S}_{\mathcal{B}})}\}$ are fixed and given. Then, the following lemma can be proved.

Lemma 1. *If the MMSE receiver in* (7) *is employed at the BS, then in the limit* $N, K, S \rightarrow \infty$ *with fixed* $c \in (0, 1]$ *and* $c_S \in (0, 1]$

$$p_k^{(\mathcal{R})} - \overline{p}_k^{(\mathcal{R})} \xrightarrow{a.s.} 0 \tag{8}$$

where $\overline{p}_k^{(\mathcal{R})}$ *is obtained through*

$$\overline{p}_k^{(\mathcal{R})} = \frac{1}{\mu\delta} \frac{\gamma_k}{l(\mathbf{x}_k)(1 - \tau_k^2)} \tag{9}$$

with $\tau_k = 0$ *if* $k \in \mathcal{S}_{\mathcal{R}}$ *and* $\gamma_k = 2^{r_k^{(\mathcal{R})}} - 1$*. The quantities* μ *and* δ *are computed as the unique solutions to the following set of equations:*

$$\mu\sigma^2 + \frac{1}{N}\sum_{k=1}^{K} \frac{\gamma_k}{\delta(1 - \tau_k^2) + \gamma_k} + \frac{1}{N}\sum_{s=1}^{S} \frac{p_s^{(\mathcal{S}_{\mathcal{B}})} l(\mathbf{x}_s)\mu}{1 + p_s^{(\mathcal{S}_{\mathcal{B}})} l(\mathbf{x}_s)\mu} = 1 \tag{10}$$

and

$$\delta = \left(1 + \frac{\frac{1}{N}\sum\limits_{k=1}^{K} \frac{\gamma_k}{\delta} \frac{\tau_k^2}{1-\tau_k^2}}{\mu\sigma^2 + \frac{1}{N}\sum\limits_{k=1}^{K} \frac{\gamma_k \delta\left(1-\tau_k^2\right)^2}{\left(\delta(1-\tau_k^2)+\gamma_k\right)^2} + \frac{1}{N}\sum\limits_{s=1}^{S} \frac{p_s^{(\mathcal{S}_{\mathcal{B}})} l(\mathbf{x}_s)\mu}{\left(1+p_s^{(\mathcal{S}_{\mathcal{B}})} l(\mathbf{x}_s)\mu\right)^2}} \right)^{-1} \tag{11}$$

Proof: The proof relies on using the same random matrix theory results of [22] since for a given set of $\{p_s^{(\mathcal{S}_{\mathcal{B}})}\}$ the SINR in (6) is in the same form of that in [22]. The details are omitted for simplicity. ■

The evaluation of $\{p_s^{(\mathcal{S_B})}\}$ requires to compute the DL SINR of the sth SUE in $\mathcal{S_B}$, which is given by

$$\mathrm{SINR}_s^{(\mathcal{S_B})} = \frac{p_s^{(\mathcal{S_B})}|h_s|^2}{\sigma^2 + \sum\limits_{k=1}^{K} p_k^{(\mathcal{R})}|h_{s,k}|^2} \tag{12}$$

where h_s is the channel propagation coefficient from its serving SCA whereas $h_{s,k}$ is the channel coefficient of the kth interfering UL transmission in $\mathcal{R}$.[3] Observe that if K is large, then the interference term in (12) can be reasonably assumed to be deterministic and equal to its mean. More specifically, under the assumption that all powers $p_k^{(\mathcal{R})}$ are finite and the cell size is fixed, using the law of large numbers yields

$$\lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} p_k^{(\mathcal{R})}|h_{s,k}|^2 = \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} p_k^{(\mathcal{R})} l(\mathbf{x}_{s,k}) \tag{13}$$

where $\mathbf{x}_{s,k}$ denotes the distance of SUE s from transmitter k in $\mathcal{R}$. In contrast, the power of the useful signal in (12) is a random quantity that depends on the fluctuations of $|h_s|^2$. To overcome this issue, we resort to the ergodic mutual information as a metric[4] and use the following asymptotic result:

$$R_s^{(\mathcal{S_B})} \;=\; \mathbb{E}_{h_s}\left[\log_2\left(1+\mathrm{SINR}_s^{(\mathcal{S_B})}\right)\right] \underset{K\to\infty}{\to} \frac{1}{\log(2)}\mathrm{Ei}(-\overline{\mathrm{SINR}}_s^{(\mathcal{S_B})}) \tag{14}$$

where $\log(2) = 0.6931$ and $\mathrm{Ei}(x) = \int_{-x}^{\infty} dt \frac{e^{-t}}{t}$ is the exponential integral whereas $\overline{\mathrm{SINR}}_s^{(\mathcal{S_B})}$ is given by

$$\overline{\mathrm{SINR}}_s^{(\mathcal{S_B})} \;=\; \frac{p_s^{(\mathcal{S_B})} l(\mathbf{x}_s)}{\sigma^2 + \sum\limits_{k=1}^{K} \bar{p}_k^{(\mathcal{R})} l(\mathbf{x}_{s,k})} \tag{15}$$

with $\bar{p}_k^{(\mathcal{R})}$ being obtained from (9). Setting $R_s^{(\mathcal{S_B})}$ in (14) equal to $r_s^{(\mathcal{S_B})}$ and inverting the exponential integral provides the target SINR γ_s. Setting $\overline{\mathrm{SINR}}_s^{(\mathcal{S_B})} = \gamma_s$ and using (9), the DL power of SCA s satisfying the target rate constraint is obtained as

$$p_s^{(\mathcal{S_B})} = \frac{\gamma_s}{l(\mathbf{x}_s)}\left(\sigma^2 + \frac{1}{\mu\delta}\sum_{k=1}^{K}\frac{\gamma_k}{1-\tau_k^2}\frac{l(\mathbf{x}_{k,s})}{l(\mathbf{x}_k)}\right) \tag{16}$$

[3]In writing (12), we have not included the interference coming from the other SCAs in $\mathcal{S_B}$ as they are assumed to be relatively spaced apart. Its contribution can be checked a posteriori to be negligible.

[4]Observe that an alternative route might be that of using the outage capacity criterion.

with $\tau_k = 0$ if $k \in \mathcal{S}_{\mathcal{R}}$ and $\gamma_k = 2^{r_k^{(\mathcal{R})}} - 1$. Plugging the above result into (10) and (11), it follows that the computation of the powers $\{p_k^{(\mathcal{R})}\}$ and $\{p_s^{(\mathcal{S}_{\mathcal{B}})}\}$ reduces to the easy task of finding μ and δ as the unique solutions of a set of equations that depend only on system parameters (such as imperfect CSI factors $\{\tau_k\}$, SINR constraints $\{\gamma_k\}$ and number of MUEs and SCAs).

Remark 3. *From the above results, it follows that the imperfect CSI coefficients $\{\tau_k\}$ impact on both μ and δ in* (10) *and* (11). *In particular, from* (9) *it follows that $1/\delta$ can be thought of as the fractional UL power increase of all transmitters (MUEs and SCAs) in $\mathcal{R}$. Interestingly, this happens even though only the MUE channels are estimated erroneously while perfect CSI is assumed for SCAs. More precisely, it turns out that $\delta \to 0$ and thus $p_k^{(\mathcal{R})} \to \infty \ \forall k$ when*

$$\frac{1}{N}\sum_{k\in\mathcal{M}_{\mathcal{R}}} \gamma_k \frac{\tau_k^2}{1-\tau_k^2} \geq 1 - c_S \tag{17}$$

meaning that there exist critical values for $\{\tau_k\}$ and $\{\gamma_k\}$ beyond which all powers diverge. If $\tau_k = \tau$ for any $k \in \mathcal{M}_{\mathcal{R}}$ one gets that τ has to be smaller than $\tau_{MMSE}^{(\max)}$ given by

$$\tau_{MMSE}^{(\max)} = \left(1 + \gamma^{(\mathcal{M}_{\mathcal{R}})}\frac{c}{1-c_S}\right)^{-1/2} \tag{18}$$

where $\gamma^{(\mathcal{M}_{\mathcal{R}})}$ stands for

$$\gamma^{(\mathcal{M}_{\mathcal{R}})} = \frac{1}{K}\sum_{k\in\mathcal{M}_{\mathcal{R}}} \gamma_k. \tag{19}$$

Remark 4. *It is worth observing that under the assumption of perfect CSI (i.e., $\tau_k = 0 \ \forall k$), the above results can be easily extended to an MMSE-based successive interference cancellation receiver. However, the imperfect CSI case makes the computations much more involved due to the presence of residual interference. Computing the transmit powers in this case is an interesting and challenging problem, which is left for future works.*

IV. Large System Analysis of the macro-tier interference in DL

We now consider the case in which the BS is in DL mode. Without loss of generality, the frequency-time slot (W_1, T_2) of Fig. 2 is considered. As for the UL, two instances of interference arise. The interference experienced by MUEs and SCAs from UL transmissions in $\mathcal{S}_{\mathcal{B}}$ can be reasonably neglected since the number of transmitting SUEs is relatively small (one per SCA) and geographically far away from the MUEs and SCAs in $\mathcal{R}$. On the other hand, the interference from BS to the SCAs in UL must be properly mitigated to avoid a severe degradation of the network performance. For this purpose, we assume that the BS makes use of linear precoding and sacrifices some of its degrees of freedom (or excess antennas) to simultaneously serve all receivers in $\mathcal{R}$ and at the same time to null the interference towards $\mathcal{S}_{\mathcal{B}}$. We let $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K] \in \mathbb{C}^{N \times K}$ be the precoding matrix and denote $p_k^{(\mathcal{R},\text{dl})}$ the DL transmit power assigned to the kth device in $\mathcal{R}$. The total DL transmit power at the BS is thus given by

$$P^{(\mathcal{R},\text{dl})} = \sum_{k=1}^{K} p_k^{(\mathcal{R},\text{dl})} \left\|\mathbf{v}_k\right\|^2. \tag{20}$$

The achievable DL rate for a generic receiver k in $\mathcal{R}$ takes the form [31]

$$R_k^{(\mathcal{R},\text{dl})} = \log_2\left(1 + \text{SINR}_k^{(\mathcal{R},\text{dl})}\right) \tag{21}$$

where $\text{SINR}_k^{(\mathcal{R},\text{dl})}$ is given by

$$\text{SINR}_k^{(\mathcal{R},\text{dl})} = \frac{p_k^{(\mathcal{R},\text{dl})} \left|\mathbf{h}_k^{\dagger}\mathbf{v}_k\right|^2}{\sum\limits_{i=1, i\neq k}^{K} p_i^{(\mathcal{R},\text{dl})} \left|\mathbf{h}_k^{\dagger}\mathbf{v}_i\right|^2 + \sigma^2}. \tag{22}$$

We denote $r_k^{(\mathcal{R},\text{dl})}$ the DL target rate of the kth receiver in $\mathcal{R}$. Imposing $R_k^{(\mathcal{R},\text{dl})} = r_k^{(\mathcal{R},\text{dl})}$ amounts to setting $\text{SINR}_k^{(\mathcal{R},\text{dl})} = \gamma_k$ with γ_k given by $\gamma_k = 2^{r_k^{(\mathcal{R},\text{dl})}} - 1$. Thanks to the reciprocity of UL and DL channels, the BS can exploit UL estimates for DL transmissions. As for the UL, we assume that perfect knowledge of $\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})}$ and $\mathbf{H}^{(\mathcal{S}_{\mathcal{R}})}$ is available while imperfect CSI is assumed for $\mathbf{H}^{(\mathcal{M}_{\mathcal{R}})}$. For notational convenience, the superscript $^{(\text{dl})}$ is dropped in the sequel.

The complete elimination of the macro-tier interference at SCAs in $\mathcal{S}_{\mathcal{B}}$ can be achieved by constraining the precoding matrix $\mathbf{V}$ to lie in the null space of $\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})}$. Under the assumption of perfect knowledge of $\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})}$, this is achieved setting $\mathbf{V} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\mathbf{F}$ where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_K] \in \mathbb{C}^{N \times K}$ is a design matrix and $\mathbf{T}^{(\mathcal{S}_{\mathcal{B}})} \in \mathbb{C}^{N \times N}$ is obtained as

$$\mathbf{T}^{(\mathcal{S}_{\mathcal{B}})} = \mathbf{I}_N - \mathbf{H}^{(\mathcal{S}_{\mathcal{B}})}\left(\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})^{\dagger}}\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})}\right)^{-1}\mathbf{H}^{(\mathcal{S}_{\mathcal{B}})^{\dagger}}. \tag{23}$$

Let $\mathbf{U} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\mathbf{H} \in \mathbb{C}^{N \times K}$ be the composite channel and denote $\widehat{\mathbf{U}}$ its corresponding estimate defined as (under the assumptions given above)

$$\widehat{\mathbf{U}} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\widehat{\mathbf{H}} \tag{24}$$

where $\widehat{\mathbf{H}}$ is given by $\widehat{\mathbf{H}} = [\widehat{\mathbf{H}}^{(\mathcal{M}_R)}\, \mathbf{H}^{(\mathcal{S}_R)}]$. The matrix $\widehat{\mathbf{U}}$ is used in the sequel to design $\mathbf{F}$ according to the RZF and ZF criteria.

A. Regularized Zero Forcing

We start assuming that the processing matrix $\mathbf{F}$ takes the following form:

$$\mathbf{F} = \left(\widehat{\mathbf{U}}\boldsymbol{\Lambda}^{-1}\widehat{\mathbf{U}}^{\dagger} + N\rho\mathbf{I}_N\right)^{-1}\widehat{\mathbf{U}} \tag{25}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\{l(\mathbf{x}_1), l(\mathbf{x}_2), \ldots, l(\mathbf{x}_K)\}$ and $\rho > 0$ is a design parameter. We refer to the concatenated precoding matrix $\mathbf{V}_{\mathrm{RZF}} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\mathbf{F}$ as RZF and denote

$$P_{\mathrm{RZF}}^{(\mathcal{R})} = \sum_{k=1}^{K} p_k^{(\mathcal{R})} \left\|\mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\mathbf{f}_k\right\|^2 \tag{26}$$

its corresponding transmit power.

Remark 5. *Observe that in the design of* $\boldsymbol{\Lambda}$ *in* (25) *we have exploited knowledge of positions* $\{\mathbf{x}_i\}$*. This information can be easily observed and estimated accurately at the BS because it changes slowly with time (relative to the small-scale fading) even for MUEs with medium-to-high mobility. This choice is motivated by the results illustrated in [23] wherein it is proved that in the downlink of a single-tier MIMO system with perfect CSI, RZF is asymptotically equivalent to the optimal linear precoder when the same rate constraints are imposed for all UEs. This is in sharp contrast to the commonly used RZF precoder where* $\boldsymbol{\Lambda} = \mathbf{I}_K$ *[22]. The latter becomes optimal only when the ratio between the SINR requirement and the average channel attenuation is the same for all UEs [23]. Due to the imperfect CSI assumption and projection into the null space of SCAs in* $\mathcal{S}_{\mathcal{B}}$*, the above optimality results does not hold in principle for the network under investigation. However, the exploitation of positions* $\{\mathbf{x}_i\}$ *is instrumental to get a closed form expression for the optimal regularization parameter* ρ*.*

For convenience, we let

$$A = \frac{1}{K}\sum_{k\in\mathcal{M}_{\mathcal{R}}} \frac{\gamma_k}{(1-\tau_k^2)l(\mathbf{x}_k)} + \frac{1}{K}\sum_{k\in\mathcal{S}_{\mathcal{R}}}\frac{\gamma_k}{l(\mathbf{x}_k)} \tag{27}$$

and

$$B = \frac{1}{K}\sum_{k\in\mathcal{M}_{\mathcal{R}}} \gamma_k \frac{\tau_k^2}{1-\tau_k^2} \tag{28}$$

and denote μ the fixed point of the following equation:

$$\mu = (1 - c_S + c)\left(\rho + \frac{c}{1+\mu}\right)^{-1}. \tag{29}$$

In addition, we define

$$\gamma^{(\mathcal{R})} = \frac{1}{K}\sum_{k=1}^{K}\gamma_k \tag{30}$$

the average SINR requirements of all devices in $\mathcal{R}$. The following lemmas are proved in Appendix.

Lemma 2. *If RZF is used and $N, K, S \to \infty$ with $c \in (0, 1]$ and $c_{\mathcal{S}} \in (0, 1]$, then*

$$P_{\textit{RZF}}^{(\mathcal{R})} - \overline{P}_{\textit{RZF}}^{(\mathcal{R})} \xrightarrow{a.s.} 0 \tag{31}$$

where $\overline{P}_{\textit{RZF}}^{(\mathcal{R})}$ is given by

$$\overline{P}_{\textit{RZF}}^{(\mathcal{R})} = cA\sigma^2 \frac{(1+\mu)^2}{\mu\left(c+\rho\,(1+\mu)^2\right) - c\left(\gamma^{(\mathcal{R})} + B(1+\mu)^2\right)}. \tag{32}$$

Also,

$$p_k^{(\mathcal{R})} - \overline{p}_k^{(\mathcal{R})} \xrightarrow{a.s.} 0 \tag{33}$$

where $\overline{p}_k^{(\mathcal{R})}$ is computed as

$$\overline{p}_k^{(\mathcal{R})} = \frac{\gamma_k}{l(\mathbf{x}_k)\mu^2}\frac{\overline{P}_{\textit{RZF}}^{(\mathcal{R})}\left(1-\tau_k^2+\tau_k^2\,(1+\mu)^2\right) + \frac{\sigma^2}{l(\mathbf{x}_k)}\,(1+\mu)^2}{1-\tau_k^2} \tag{34}$$

with $\tau_k = 0$ if $k \in \mathcal{S}_{\mathcal{R}}$.

Lemma 3. *The optimal ρ is given by*

$$\rho^{\star} = \frac{1-c_S+c}{\gamma^{(\mathcal{R})}} - \frac{c}{1+\gamma^{(\mathcal{R})}} \tag{35}$$

and $\overline{P}_{\textit{RZF}}^{(\mathcal{R})}$ reduces to

$$\overline{P}_{\textit{RZF}}^{(\mathcal{R})} = c\sigma^2\frac{A}{\rho^{\star}\gamma^{(\mathcal{R})} - cB}. \tag{36}$$

To the best of our knowledge, this is the first time that the optimal value of ρ minimizing the power consumption is given in explicit form for a concatened RZF precoder with imperfect CSI. Interestingly, $\rho^{\star}$ does not depend on $\{\tau_k\}$ and it is basically in the same form of the perfect CSI case with the exception of the term $1 - c_S + c$ that accounts for the interference nulling towards the SCAs [23].

Remark 6. *Since $\overline{P}_{\textit{RZF}}^{(\mathcal{R})}$ must be positive and finite, from* (36) *it follows that the asymptotic analysis can be applied to RZF only when the following condition is satisfied:*

$$\frac{1}{K}\sum_{k\in\mathcal{M}_{\mathcal{R}}}\gamma_k\frac{\tau_k^2}{1-\tau_k^2} < \frac{\rho^{\star}}{c}\gamma^{(\mathcal{R})} = \frac{1-c_S}{c} + \frac{1}{1+\gamma^{(\mathcal{R})}} \tag{37}$$

from which setting $\tau_k = \tau$ *for any* $k \in \mathcal{M}_{\mathcal{R}}$ *one gets* $\tau < \tau_{RZF}^{(\max)}$ *with*

$$\tau_{RZF}^{(\max)} = \left(1 + \frac{c}{\rho^\star}\frac{\gamma^{(\mathcal{M}_{\mathcal{R}})}}{\gamma^{(\mathcal{R})}}\right)^{-1/2} \tag{38}$$

where $\gamma^{(\mathcal{M}_{\mathcal{R}})}$ *is defined as in* (19).

B. Zero Forcing

Setting $\mathbf{\Lambda} = \mathbf{I}_K$ into (25) yields $\mathbf{F} = \left(\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\dagger + N\rho\mathbf{I}_N\right)^{-1}\widehat{\mathbf{U}}$ from which using the Woodbury matrix identity and imposing $\rho = 0$ the ZF precoder easily follows $\mathbf{V}_{\mathrm{ZF}} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\widehat{\mathbf{U}}\left(\widehat{\mathbf{U}}^\dagger\widehat{\mathbf{U}}\right)^{-1}$ or, equivalenty,

$$\mathbf{V}_{\mathrm{ZF}} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}^\dagger\mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\widehat{\mathbf{H}}\right)^{-1} \tag{39}$$

where we have taken into account that $\mathbf{T}^{(\mathcal{S}_{\mathcal{B}})^\dagger} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}$ and $\mathbf{T}^{(\mathcal{S}_{\mathcal{B}})^2} = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}$.

Lemma 4. *If ZF is used and* $K, N \to \infty$ *with* $c_{\mathcal{S}} \in (0, 1]$, *then*

$$P_{ZF}^{(\mathcal{R})} - \overline{P}_{ZF}^{(\mathcal{R})} \xrightarrow{a.s.} 0 \tag{40}$$

where $\overline{P}_{ZF}^{(\mathcal{R})}$ *is given by*

$$\overline{P}_{ZF}^{(\mathcal{R})} = c\sigma^2\frac{A}{1 - c_{\mathcal{S}} - cB}. \tag{41}$$

Also,

$$p_k^{(\mathcal{R})} - \overline{p}_k^{(\mathcal{R})} \xrightarrow{a.s.} 0 \tag{42}$$

where $\overline{p}_k^{(\mathcal{R})}$ *is given by*

$$\overline{p}_k^{(\mathcal{R})} = \frac{\gamma_k}{1 - \tau_k^2}\left(\sigma^2 + \tau_k^2 l(\mathbf{x}_k)\overline{P}_{ZF}^{(\mathcal{R})}\right) \tag{43}$$

with $\tau_k = 0$ *if* $k \in \mathcal{S}_{\mathcal{R}}$.

Proof: The proof follows the same steps of that for Lemmas 2 and 3 and thus is omitted for space limitations. ■

Remark 7. *Since* $\overline{P}_{ZF}$ *must be positive and finite, it follows that the asymptotic analysis can be applied to ZF only when the following condition is satisfied:* $1 - c_{\mathcal{S}} - cB > 0$ *or, equivalently,*

$$\frac{1}{K}\sum_{k \in \mathcal{M}_{\mathcal{R}}}^{K} \gamma_k \frac{\tau_k^2}{1 - \tau_k^2} < \frac{1 - c_{\mathcal{S}}}{c}. \tag{44}$$

If $\tau_k = \tau$ for any k, then we have that

$$\gamma^{(\mathcal{M}_{\mathcal{R}})} < \frac{1-c_{\mathcal{S}}}{c}\frac{1-\tau^2}{\tau^2} \tag{45}$$

from which it follows $\tau < \tau_{\mathit{ZF}}^{(\max)}$ with

$$\tau_{\mathit{ZF}}^{(\max)} = \left(1+\gamma^{(\mathcal{M}_{\mathcal{R}})}\frac{c}{1-c_{S}}\right)^{-1/2}. \tag{46}$$

Comparing $\tau_{\mathit{ZF}}^{(\max)}$ to $\tau_{\mathit{RZF}}^{(\max)}$ in (38), it is seen that $\tau_{\mathit{RZF}}^{(\max)}$ is always larger than $\tau_{\mathit{ZF}}^{(\max)}$, meaning that RZF is more robust than ZF to imperfect CSI of MUEs. Observe that the same condition must be fulfilled in the UL (see (18) in Section III).

V. Numerical results

The accuracy of the above asymptotic characterization is now validated numerically by Monte-Carlo simulations. The results are obtained for 1000 different channel realizations and UE distributions. We assume that the BS is equipped with $N = 128$ antennas and covers a square area centered at the BS with side length 500m over which 16 SCAs are distributed on a regular grid with an inter-site distance of 125 m. We assume that 128 MUEs are active in the cell and that a single SUE is uniformly distributed within a disc of radius 35 m around each SCA. The SUEs are associated with the closest SCA while the MUEs are associated with the BS. Accordingly, the two sets $\mathcal{R}$ and $\mathcal{B}$ count 64 MUEs and 8 SCAs with 8 MUEs in the proximity of each SCA. A random snapshot of the network is depicted in Fig. 4. We assume that the UL and DL wireless backhaul rates of SCAs $r_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{dl})}$ and $r_s^{(\mathcal{S}_{\mathcal{B}},\mathrm{ul})}$ are equal and fixed to 3 bit/s/Hz. The pathloss function $l(\mathbf{x})$ is modelled as [32]

$$l(\mathbf{x}) = 2L_{\bar{x}}\left(1+\frac{\|\mathbf{x}\|^{\beta}}{\bar{x}^{\beta}}\right)^{-1} \tag{47}$$

where $\beta \geq 2$ is the pathloss exponent, $\bar{x} > 0$ is some cut-off parameter and $L_{\bar{x}}$ is a constant that regulates the attenuation at distance $\bar{x}$. We assume that $\beta = 3.5$ and $L_{\bar{x}} = -86.5$ dB. The latter is such that for $f_c = 2.4$ GHz the attenuation at $\bar{x}$ is the same as that in the cellular model analyzed in [33]. Although in TDD systems the effective values of $\{\tau_k\}$ are expected to be different between UL and DL (since the channels are estimated in the UL and then used in DL), the same values of $\{\tau_k\}$ are used for both links in all subsequent simulations. In particular, we assume $\tau_k = \tau\ \forall k$ and let $\tau^2 = \underline{\tau}^2 + \varsigma^2$ where $\underline{\tau}^2$ is a constant term while ς accounts for estimation errors induced by mobility. We set $\underline{\tau}^2 = 0.08$ and compute ς^2 as follows

$$\varsigma^2 = 1 - J_0^2\left(2\pi\frac{v}{\lambda}\xi\right) \tag{48}$$

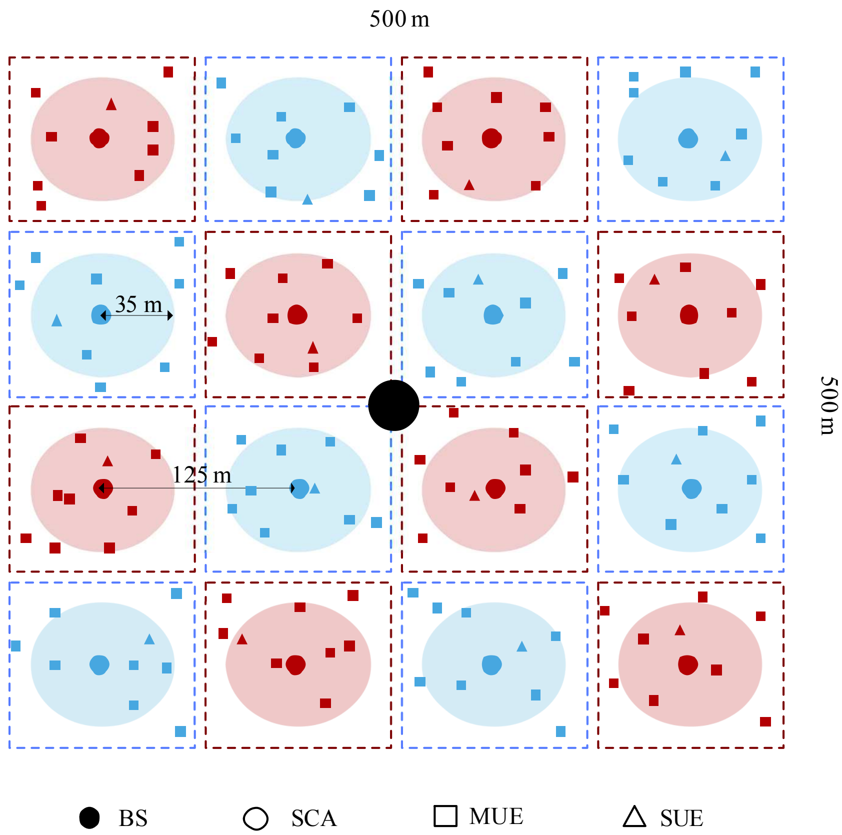

Fig. 4. A snapshot of the UE distribution in the simulated network wherein the number of SCAs is 16 and the number of MUEs is 128. The latter are distributed such that 8 of them are in the proximity of the coverage are of a given SCA. A single SUE is active for each SCA.

where v is the velocity (in m/s) of MUEs, λ is the carrier wavelength (in meter) and ξ is the UL or DL slot duration (in seconds). Since $\lambda = 0.125$ m, setting $\xi = 1$ ms and $v = 15$ or 50 km/h yields $\tau^2 = 0.1$ and 0.3, respectively. The parameter setting is summarized in Table I for simplicity. Comparisons are made with the two alternative protocols and network configurations mentioned in Remark 1. In particular, we consider a HetNet in which the SCAs use a wired backhaul infrastructure for data traffic and a massive MIMO system in which all UEs (MUEs and SUEs) are served by the macro BS.

Fig. 5 depicts the average transmit power for MUEs and SCAs over the cell as a function of the requested rate of MUEs when $\tau^2 = 0.1$. Despite the fact that the power of SCAs in (9) does not depend explicitly on the MUE rates $r_k^{(\mathcal{R},\mathrm{ul})}$ for $k \in \mathcal{M}_{\mathcal{R}}$, a mild dependence on MUE requirements is shown in the results of Fig. 5. This is due to the fact that μ and δ in (10) and (11) depend on $\{r_k^{(\mathcal{R},\mathrm{ul})}\}$ through $\{\gamma_k\}$. A similar behaviour is observed in Fig. 6 for the DL power of SCAs. From the results of Fig. 5, it also follows that the average UL power of MUEs in the proposed HetNet

TABLE I

GENERAL SYSTEM PARAMETERS

Parameter	Value	Parameter	Value
Bandwidth	$W = 10$ MHz	Total number of SCAs	16
Noise power	$W\sigma^2 = -104$ dBm	Small-cell radius	$R = 35$ m
Macro-cell side length	500 m	Inter-side distance of SCAs	$\Delta = 125$ m
Cut-off parameter	$\bar{x} = 25$ m	Pathloss coefficient	$\beta = 3.5$
Carrier frequency	$f_c = 2.4$ GHz	Average pathloss attenuation at $\bar{x}$	$L_{\bar{x}} = -86.5$ dB
Number of BS antennas	$N = 128$	Imperfect CSI for MUEs	$\tau^2 = 0, 0.1$ and 0.3
Total number of MUEs	128	Wireless Backhaul Requirements for SCAs	$r_s = 3$ bit/s/Hz

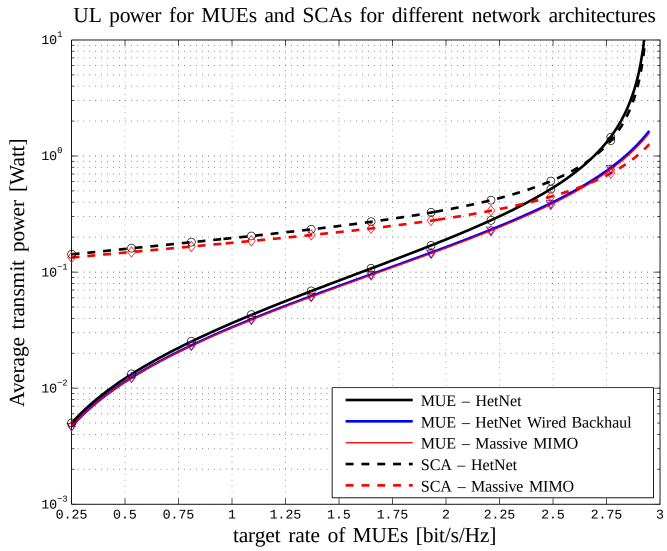

Fig. 5. Average UL transmit power for MUEs and SCAs as a function of MUE rates for different network architectures when $\tau^2 = 0.1$ and the wireless backhaul traffic is fixed to 3 bit/s/Hz.

is substantially the same of a HetNet with wired backhaul even though a wireless backhaul traffic of 3 bit/s/Hz is provided. On the contrary, Fig. 6 shows a significant power reduction for SCAs in DL mode. This is because the reception of associated SUEs is properly shielded from nearby MUEs in the proposed network architecture.

Figs. 7 and 8 provide insights on the effect of channel uncertainty on power consumption. Clearly, $\tau^2 = 0$ corresponds to the perfect CSI case. It can be seen that for $\tau^2 = 0.3$ (corresponding to a velocity of 50 km/h) the system becomes infeasible if the target rates go beyond a certain level given approximately by 1.5 bit/s/Hz (as obtained through (18)). Indeed, the power of every transmitter becomes infinite. However, this happens for a relatively narrow window of rate values, thereby allowing the system to operate at relatively low powers up to the critical values. In particular, we

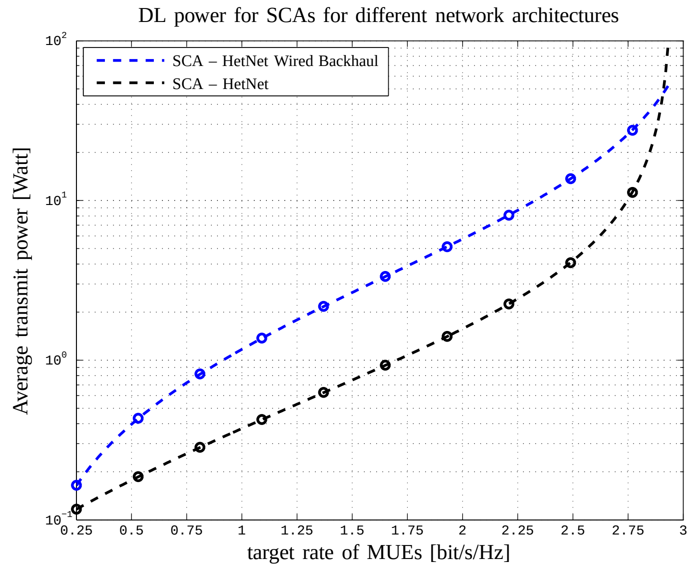

Fig. 6. Average DL transmit power for SCAs towards their respective SUEs as a function of MUE rates for the HetNet and HetNet with wired Backhaul architectures when $\tau^2 = 0.1$ and the wireless backhaul traffic is fixed to 3 bit/s/Hz.

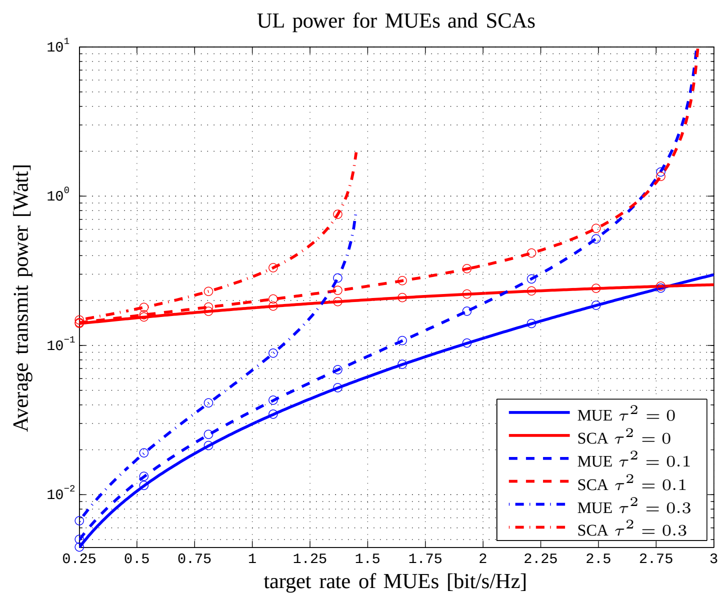

Fig. 7. Average UL transmit power for MUEs and SCAs for the proposed HetNet architecture as a function of MUE rates for different values of τ^2. The wireless backhaul traffic is set to 3 bit/s/Hz.

observe that for $\tau^2 = 0.1$ each MUE requires 80 mW to achieve a target rate of 1.5 bit/s/Hz. Taking into account that the bandwidth is $W = 10$ MHz and the number of MUEs for frequency-time slot is 64 MUEs, this corresponds to an aggregate area throughput of 3.94 Gb/s/km^2.

Figs. 9 and 10 illustrate the average DL transmit power of the BS when RZF and ZF are employed with $\tau^2 = 0.1$. As expected, RZF provides a substantial power reduction with respect to ZF. In particular, we observe that for a target rate of 2 bit/s/Hz only 0.1 W are required at the BS

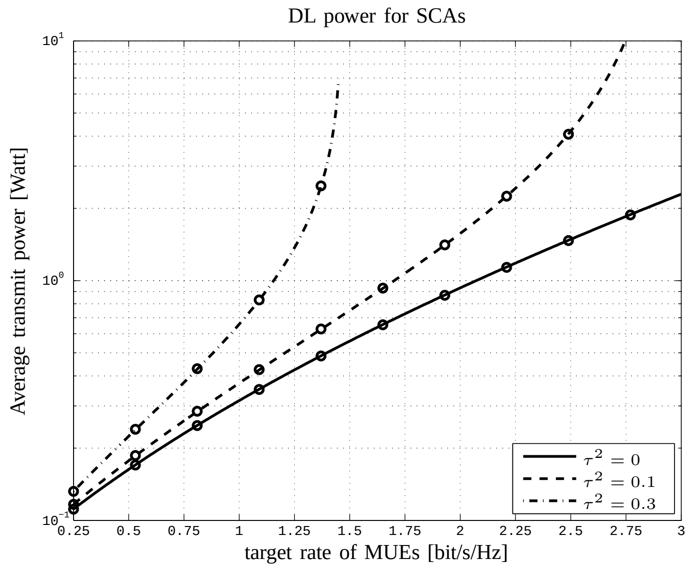

Fig. 8. Average DL transmit power for SCAs in the proposed HetNet architecture as a function of MUE rates for different values of τ^2. The wireless backhaul traffic is set to 3 bit/s/Hz.

to serve (in DL) all MUEs and SCAs. This corresponds to an area throughput of 5.12 Gbit/s/km^2. Compared to the two alternative network configurations, a marginal increase of power is required by the proposed HetNet with both precoding techniques. However, this is achieved at the price of a substantial power saving at the SUEs. Indeed, numerical results reveal that for a target rate of 3 bit/s/Hz, the required power for a SUE is 0.85 mW in the HetNet case while it is 0.2 W for a massive MIMO system.

Fig. 11 reports the average DL transmit power at the BS of RZF and ZF when $\tau^2 = 0, 0.1$ and 0.3. As seen, for $\tau^2 = 0.3$ the power required by both precoding techniques diverge when the MUE target rate increases. As expected, RZF is more robust than ZF to imperfect CSI and can handle rates up to 1.75 bit/s/Hz.

VI. Discussions and Perspectives

In this section, we discuss the impact of mobility along with some other practical aspects of the proposed HetNet such as network design, channel correlation at the BS antennas and dynamic UL-DL TDD configuration.

A. Impact of mobility

In [18], the authors show that if the network sum rate is considered then low and high mobility MUEs can coexist and be served simultaneously. This is because the imperfect CSI of each

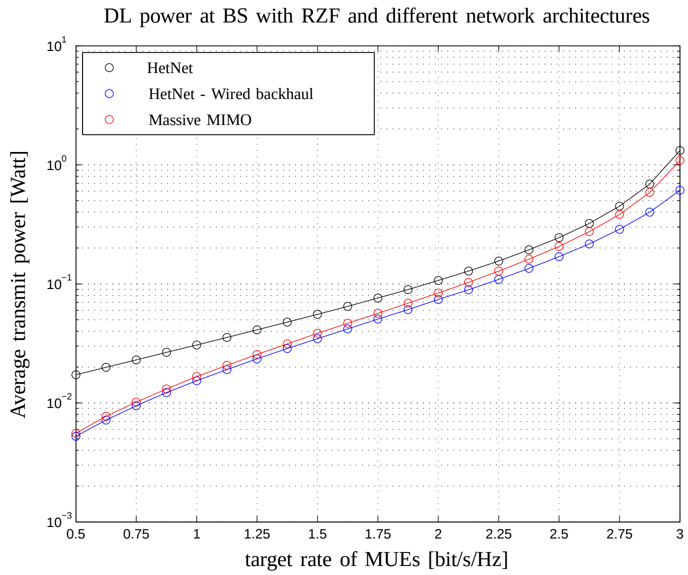

Fig. 9. Average DL transmit power at the BS when RZF is employed with $\tau^2 = 0.1$ and wireless backhaul 3 bit/s/Hz. Comparisons are made with a HetNet with wired backhaul and a single-tier massive MIMO systems operating according to the transmission protocols of Fig. 3.

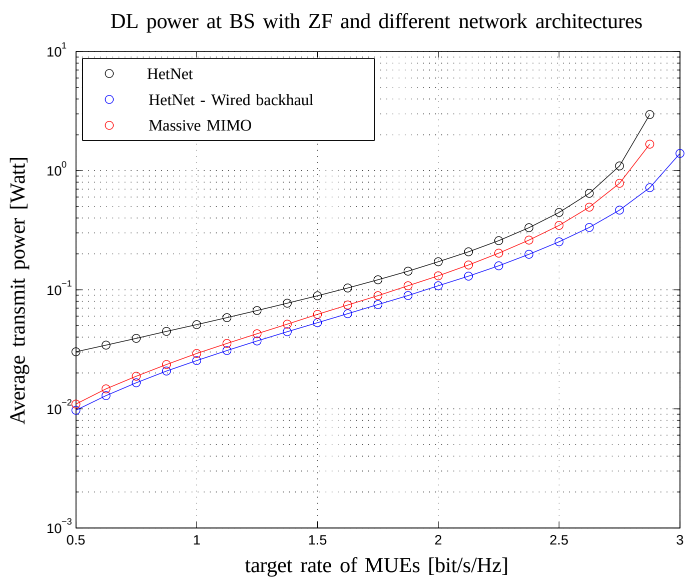

Fig. 10. Average DL transmit power at the BS when ZF is employed with $\tau^2 = 0.1$ and wireless backhaul 3 bit/s/Hz. Comparisons are made with a HetNet with wired backhaul and a single-tier massive MIMO systems operating according to the transmission protocols of Fig. 3.

given MUE has detrimental effects only on its own achievable rate while it has no impact on the performance of the others. This is in sharp contrast to the results obtained in this work where we have shown that the UL and DL transmit powers for meeting target rates depend heavily on the mobility of each MUE. In particular, a single MUE with high mobility and rate requirements might largely increase (or even make diverge) the required powers. This calls for alternative solutions.

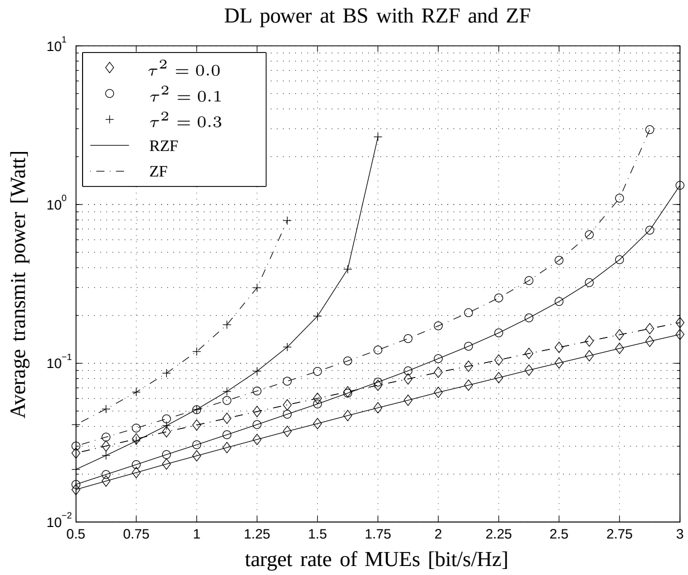

Fig. 11. Average DL transmit power at the BS when RZF and ZF are employed with different values of τ^2 and wireless backhaul 3 bit/s/Hz.

The simplest one would be to lower the target rate (and thus the corresponding SINR) for the MUEs with large channel estimation errors such that, for example, in UL the condition for $\delta \to 0$ in (17) is never met. An alternative solution for DL mode might consist in dividing the MUEs in two sets $\mathcal{A}_M$ and $\mathcal{A}_H$ (with $\mathcal{A}_M \cap \mathcal{A}_H = \emptyset$) characterized by medium and high mobility, respectively. The MUEs in $\mathcal{A}_M$ are served simultaneously while those in $\mathcal{A}_H$ are served one at a time using space-time coding (STC) techniques (that do not require any CSI at the BS). Consider for example the frequency-time slot (W_1, T_2) for which $\mathcal{M}_{\mathcal{R}} = \mathcal{A}_M \cup \mathcal{A}_H$. Let $K = |\mathcal{A}_M| + |\mathcal{S}_R|$, $K_{STC} = |\mathcal{A}_H|$ and $S = |\mathcal{S}_{\mathcal{B}}|$. The BS would first simultaneously transmit to the K MUEs and SCAs in $\mathcal{R} = \mathcal{A}_M \cup \mathcal{S}_{\mathcal{R}}$ while removing the interference towards to the S SCAs in $\mathcal{S}_{\mathcal{B}}$. Then, it would serve the K_{STC} MUEs in $\mathcal{A}_H$ (one at a time) while nulling the interference to $\mathcal{S}_{\mathcal{B}}$. In these circumstances, the signal transmitted to MUE k in $\mathcal{A}_H$ takes the form $\mathbf{x}_k = \mathbf{T}^{(\mathcal{S}_{\mathcal{B}})}\mathbf{s}_k$ with $\mathbf{s}_k$ being such that $\mathbb{E}\{\mathbf{s}_k\mathbf{s}_k^\dagger\} = p_k/N\mathbf{I}_N$ (corresponding to uniform STC). As a consequence, the deterministic equivalent of the DL SINR of MUE k in $\mathcal{A}_H$ is found to be:

$$\mathrm{SINR}_k^{(\mathcal{A}_H,\mathrm{dl})} - \left(1 - \frac{S}{N}\right)\frac{p_k l(\mathbf{x}_k)}{\sigma^2} \xrightarrow{a.s.} 0 \tag{49}$$

from which (using dominated convergence arguments and continuous mapping theorem) it follows that $p_k^{(\mathcal{A}_H,\mathrm{dl})} - \overline{p}_k^{(\mathcal{A}_H,\mathrm{dl})} \xrightarrow{a.s.} 0$ with

$$\overline{p}_k^{(\mathcal{A}_H,\mathrm{dl})} = \frac{1}{1-\frac{S}{N}}\frac{\gamma_k\sigma^2}{l(\mathbf{x}_k)}. \tag{50}$$

Let T_{STC} be the time required to serve the K_{STC} MUEs in $\mathcal{A}_H$ and call $T_{LP} = T_2 - T_{STC}$ where LP stands for linear precoding and T_2 is defined in Fig. 2. Accordingly, the average throughput R_{AVG} (in bit/s/Hz) of the network over $T_2 = T_{LP} + T_{STC}$ is

$$R_{\text{AVG}} = \frac{T_{LP}}{T_2}\sum_{k=1}^{K}\log_2\left(1+\gamma_k\right) + \frac{T_{STC}}{T_2}\frac{1}{K_{STC}}\sum_{k=K+1}^{K+K_{STC}}\log_2\left(1+\gamma_k\right) \tag{51}$$

and the corresponding energy consumption is obtained as

$$\overline{E}_{T_2} = c\sigma^2\frac{AT_{LP}}{1-c_S-cB} + T_{STC}\sum_{k=K+1}^{K+K_{STC}}\frac{1}{1-\frac{S}{N}}\frac{\gamma_k\sigma^2}{l(\mathbf{x}_k)}. \tag{52}$$

As seen, the rate of MUEs served by STC is reduced by a factor $1/K_{STC}$ compared to the other ones if $T_{STC} \approx T_{LP}$. On the other hand, if $T_{STC} \approx K_{STC}T_{LP}$ then the rates are comparable, but the energy consumption increases substantially.

B. Tradeoff between proximity effect and number of users

A close inspection of (15) reveals that the interference $I_K = \sum_{k=1}^{K} p_k^{(\mathcal{R})} l(\mathbf{x}_{sk})$ of UL signals to SUEs in $\mathcal{S}_{\mathcal{B}}$ increases with K, i.e., the number of MUEs and SCAs in $\mathcal{R}$. An order of magnitude estimation of I_K in our simulations shows that $I_K/\sigma^2 \approx 10^3$. This means that, although spatially separated (see Fig. 4 for example), the interference from UL transmitters in $\mathcal{R}$ might cause a significant DL power rise at the SCAs in $\mathcal{S}_{\mathcal{B}}$. Therefore, the SCA radius R, the inter-site distance Δ (of between SCAs) and K must be carefully designed so as to keep the interference at a tolerable level. A possible way to get some insights into this is to let K be extremely large (so that the MUEs become asymptotically dense outside the SCA coverage area) and compute the average I_K experienced at a generic SUE. Let us for simplicity set the MUEs at the minimum distance d to the edge of SCA cell, which is given by $d = \Delta/2 - R$. Assume also that the power of MUEs is constant, i.e., $p_k^{(\mathcal{R})} = p^{(\mathcal{R})}$. This is a reasonable approximation since their distances from the BS is approximately the same. In the above circumstances, one gets

$$I_K = \sum_{k=1}^{K} p_k^{(\mathcal{R})} l(\mathbf{x}_{sk}) \sim K\frac{p^{(\mathcal{R})}l(d)}{A_c}. \tag{53}$$

where A_c is the area of the cell. Plugging the above result into (15), it follows that the interference can be kept at a tolerable level if d is such that $K\frac{p^{(\mathcal{R})}l(d)}{A_c}$ is not larger than a prescribed value.

C. Impact of correlation at BS antennas

If the BS antennas are correlated, then the channel vector of the ith device in $\mathcal{A}$ is modelled as $\mathbf{h}_i^{(\mathcal{A})} = \sqrt{l(\mathbf{x}_i)}\boldsymbol{\Theta}_i^{1/2}\mathbf{w}_i$ where $\boldsymbol{\Theta}_i$ denotes the ith channel correlation at the BS [22]. As for positions $\mathbf{x}_i$, the matrices $\boldsymbol{\Theta}_i$ are usually assumed to change slowly compared to the channel coherence time and thus are supposed to be perfectly known at the BS [22], [32]. As a consequence, $\widehat{\mathbf{h}}_i^{(\mathcal{A})}$ can be reasonably modelled as $\widehat{\mathbf{h}}_i^{(\mathcal{A})} = \sqrt{l(\mathbf{x}_i)}\boldsymbol{\Theta}_i^{1/2}\widehat{\mathbf{w}}_i$ with $\widehat{\mathbf{w}}_i$ still given by (2). Assume that RZF is employed. Then, $\mathbf{V}_{\rm RZF} = \mathbf{T}^{(\mathcal{S}_\mathcal{B})}\mathbf{F}$ with

$$\mathbf{F} = \left(\sum_{k=1}^{K}\tilde{\boldsymbol{\Theta}}_i^{1/2}\widehat{\mathbf{w}}_i\widehat{\mathbf{w}}_i^{\dagger}\tilde{\boldsymbol{\Theta}}_i^{1/2} + N\rho\mathbf{I}_N\right)^{-1}\widehat{\mathbf{U}} \tag{54}$$

where $\widehat{\mathbf{U}}$ is given by (24) and $\tilde{\boldsymbol{\Theta}}_i^{1/2} = \mathbf{T}^{(\mathcal{S}_\mathcal{B})}\boldsymbol{\Theta}_i^{1/2}$. Let $\{e_k; k=1,2,\ldots,K\}$ be the unique positive solution of $e_k = \frac{1}{N}\mathrm{tr}\big(\tilde{\boldsymbol{\Theta}}_k\boldsymbol{\Omega}\big)$ with

$$\boldsymbol{\Omega} = \left(\frac{1}{N}\sum_{i=1}^{K}\frac{\tilde{\boldsymbol{\Theta}}_i}{1+e_i} + \rho\mathbf{I}_N\right)^{-1} \tag{55}$$

and call $\mathbf{e}'_k = [e'_{1,k}, e'_{2,k}, \ldots, e'_{K,k}]^T = \big(\mathbf{I}_K - \boldsymbol{\Gamma}\big)^{-1}\mathbf{v}_k$ with

$$[\boldsymbol{\Gamma}]_{i,\ell} = \frac{\mathrm{tr}\left(\tilde{\boldsymbol{\Theta}}_i\boldsymbol{\Omega}\tilde{\boldsymbol{\Theta}}_\ell\boldsymbol{\Omega}\right)}{(1+e_\ell)^2} \tag{56}$$

$$\mathbf{v}_k = \left[\mathrm{tr}\left(\tilde{\boldsymbol{\Theta}}_1\boldsymbol{\Omega}\tilde{\boldsymbol{\Theta}}_k\boldsymbol{\Omega}\right), \ldots, \mathrm{tr}\left(\tilde{\boldsymbol{\Theta}}_K\boldsymbol{\Omega}\tilde{\boldsymbol{\Theta}}_k\boldsymbol{\Omega}\right)\right]^T. \tag{57}$$

Using the results in [22], [32], the deterministic equivalent of $\overline{p}_k^{(\mathcal{R})}$ is found to be

$$\overline{p}_k^{(\mathcal{R})} = \frac{\overline{P}_{\rm RZF}^{(\mathcal{R})}\left(1-\tau_k^2+\tau_k^2\left(1+e_k\right)^2\right) + \frac{\sigma^2}{l(\mathbf{x}_k)}\left(1+e_k\right)^2}{\left(1-\tau_k^2\right)l(\mathbf{x}_k)e_k^2} \tag{58}$$

where $\tau_k = 0$ for $k \in \mathcal{S}_\mathcal{R}$ and $\overline{P}_{\rm RZF}^{(\mathcal{R})}$ is the corresponding deterministic equivalent of the transmit power [22]

$$\overline{P}_{\rm RZF}^{(\mathcal{R})} = c\sigma^2\frac{A}{1-cB} \tag{59}$$

with A and B now given by

$$A = \frac{1}{K}\sum_{k\in\mathcal{M}_\mathcal{R}}\frac{\gamma_k}{(1-\tau_k^2)l(\mathbf{x}_k)}\frac{e'_k}{e_k^2} + \frac{1}{K}\sum_{k\in\mathcal{S}_\mathcal{R}}\frac{\gamma_k}{l(\mathbf{x}_k)}\frac{e'_k}{e_k^2} \tag{60}$$

$$B = \frac{1}{K}\sum_{k\in\mathcal{M}_\mathcal{R}}\frac{\gamma_k\left(1-\tau_k^2+\tau_k^2\left(1+e_k\right)^2\right)}{(1-\tau_k^2)}\frac{e'_k}{(1+e_k^2)e_k^2} + \frac{1}{K}\sum_{k\in\mathcal{S}_\mathcal{R}}\gamma_k\frac{e'_k}{(1+e_k^2)e_k^2}. \tag{61}$$

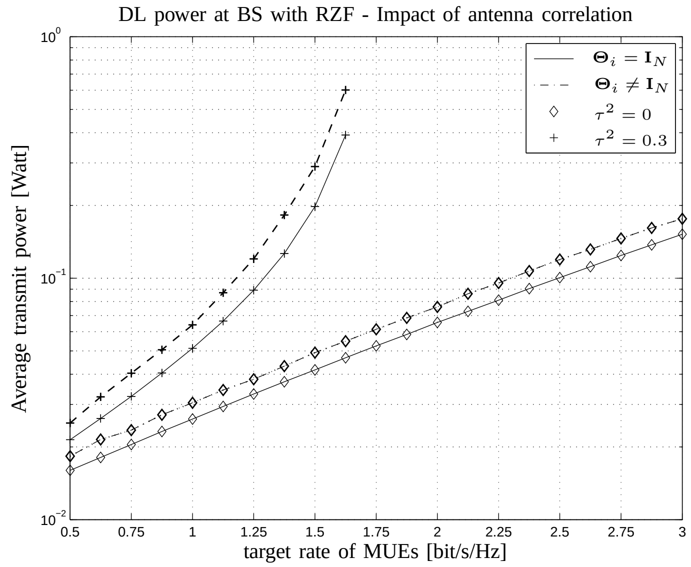

Fig. 12. Average DL transmit power at the BS when RZF is employed with different values of τ^2 and wireless backhaul 3 bit/s/Hz. Comparisons are made with the case in which BS antennas are correlated.

From the above results, it follows that (although possible) the case $\mathbf{\Theta}_i \neq \mathbf{I}_N$ requires to compute the fixed point of a set of equations and to evaluate the inverse of $\mathbf{I}_K - \mathbf{\Gamma}$ and $\mathbf{D}$. All this is not only much more involved than the case $\mathbf{\Theta}_i = \mathbf{I}_N$ but it is also less instrumental to get insights into the structure of the asymptotic transmit powers and into the interplay among the different parameters (such as imperfect CSI factors $\{\tau_k\}$, SINR constraints $\{\gamma_k\}$ and number of MUEs and SCAs). In addition to this, the optimal regularization parameter ρ can only be found through a numerical optimization procedure.

Fig. 12 reports the DL transmit power when the BS is equipped with RZF and antennas are correlated. Following [32], the entries of $\mathbf{\Theta}_k$ for $k = 1, 2, \ldots, K$ are computed as

$$[\mathbf{\Theta}_k]_{i,\ell} = \frac{1}{\Delta_\varphi} \int_{\theta_k - \Delta_\varphi/2}^{\theta_k + \Delta_\varphi/2} e^{\mathrm{i}\pi\cos(\varphi)} \partial\varphi \tag{62}$$

where Δ_φ is the angular spread and θ_k is the directional of departure of the kth signal. We assume that θ_k are uniformly distributed in $[0, 2\pi)$ while we set $\Delta_\varphi = \pi/12$. As seen, only a marginal difference is observed in terms of required power between the two cases. Moreover, imperfect CSI has the same impact in both cases. A similar behaviour (not shown for space limitations) is obtained for larger values of Δ_φ up to $\Delta_\varphi = \pi/6$.

D. Dynamic UL-DL TDD

The transmission protocol of Fig. 2 relies on the assumption that transmissions across tiers are perfectly synchronized. However, the synchronous operation with a common UL and DL configuration in multiple cells may not match the instantaneous traffic situation in a particular cell. The amount of traffic for DL and UL may vary significantly with time and between cells. This calls for the adoption of a dynamic UL-DL configuration [30]. Henceforth, we discuss some practical implications of dynamic TDD for the proposed network architecture. Consider for example the frequency band W_1 in Fig. 2 and assume that the SCAs in $\mathcal{S}_{\mathcal{R}}$ are not aligned with the UL and DL transmissions of MUEs in $\mathcal{M}_{\mathcal{R}}$. If for example the UL phase BS $\leftarrow$ SCA is shorter than BS $\leftarrow$ MUE, then the subsequent DL phase BS $\rightarrow$ SCA would partially overlap with BS $\leftarrow$ MUE. A similar situation would occur if the SCAs are in the UL for a longer time interval. In both cases, the adoption of a dynamic TDD protocol at the SCAs would require a full duplex BS. On the other hand, if the SCAs in $\mathcal{S}_{\mathcal{B}}$ are not aligned with the MUEs in $\mathcal{M}_{\mathcal{R}}$, the following two situations might occur. If SCA $\rightarrow$ SUE is longer than BS $\leftarrow$ MUE, then the linear precoder at the BS must be designed so as to also mitigate interference towards the SUEs in $\mathcal{S}_{\mathcal{B}}$. If SCA $\rightarrow$ SUE is shorter than BS $\leftarrow$ MUE, the SCAs in $\mathcal{S}_{\mathcal{B}}$ are affected by the interference due to UL transmissions in $\mathcal{M}_{\mathcal{R}}$ and $\mathcal{S}_{\mathcal{R}}$. The effect of this interference would be the same of that evaluated in Section III. In summary, the proposed network architecture and transmission protocol allow dynamic TDD transmissions within the small-cell tier (from SCAs to SUEs) while a full duplex BS would be required to handle asynchronous transmissions at the macro-tier level.

VII. Conclusions

This work has focused on the power consumption in the UL and DL of a HetNet in which a massive MIMO macro tier (serving medium-to-high mobility UEs) is overlaid with a dense tier of SCAs using a wireless backhaul for traffic. A reverse (inter-tier and intra-tier) TDD protocol has been proposed to let the BS simultaneously handle the traffic of macro UEs and SCAs without causing much interference to the overlaid tier. Linear processing has been used at the BS for data recovery and transmission while satisfying rate requirements and mitigating interference. In particular, we have considered an MMSE receiver and a concatenated linear precoding technique based on ZF and RZF. The analysis has been conducted in the asymptotic regime where the number of BS antennas and network size grow large with fixed ratio. Results from random matrix theory have been used to derive closed-form expressions for the transmit powers and beamforming vectors as well as to investigate the impact of imperfect CSI on the power consumption. It turns out that for a given set of target rates there is a critical value of imperfect CSI beyond which the power of all transmitters rapidly increases (and eventually diverges). However, analytical and numerical results have shown that when such critical values are not met the proposed architecture allows to achieve an area throughput on the order of a few Gb/s/km^2 in UL and DL on a 10 MHz band with a very limited amount of power.

There are several interesting directions for future work. One of them is to develop scheduling algorithms for serving MUEs characterized by very high mobility without lowering the served rates. The extension of the analysis to a multi-cell network in which multiple BSs (with limited cooperation) are active is also very much interesting and is currently under investigation. An interesting problem is also to develop network architectures able to exploit the gains of massive MIMO when the macro tier level operates according to a frequency division duplexing (FDD) system.

APPENDIX

The proof of Lemma 2 follows the same steps of [22] with the only exception that the projection matrix $\mathbf{T}^{(\mathcal{S}_\mathcal{B})}$ must be included in the analysis. In particular, it can be proved that if $N, K, S \to \infty$ with fixed ratios $c \in (0,1]$ and $c_\mathcal{S} \in (0,1]$ then

$$\mathrm{SINR}_k^{(\mathcal{R})} - \overline{p}_k^{(\mathcal{R})} f_k \to 0 \tag{63}$$

where f_k is such that

$$f_k = \frac{(1-\tau_k^2)\, l(\mathbf{x}_k)\mu^2}{\overline{P}_{\mathrm{RZF}}^{(\mathcal{R})}\left(1-\tau_k^2+\tau_k^2\,(1+\mu)^2\right)+\frac{\sigma^2}{l(\mathbf{x}_k)}\,(1+\mu)^2} \tag{64}$$

with $\tau_k = 0$ if $k \in \mathcal{M}_\mathcal{R}$ and μ being the solution of (29) and

$$\overline{P}_{\mathrm{RZF}}^{(\mathcal{R})} = -\frac{c\mu'}{(1+\mu)^2}\frac{1}{K}\sum_{k=1}^{K} p_k l(\mathbf{x}_k) \tag{65}$$

with $\mu' = -\frac{\mu(1+\mu)^2}{c+\rho(1+\mu)^2}$. Assume now that the power $\overline{p}_k^{(\mathcal{R})}$ is chosen such that $\mathrm{SINR}_k^{(\mathcal{R})}$ is equal to a specified γ_k in the large system limit. From (63), we obtain (43) which used in (65) yields

$$\overline{P}_{\mathrm{RZF}}^{(\mathcal{R})} = -\, c\frac{\mu'}{(1+\mu)^2}\sum_{k=1}^{K}\frac{1}{K}\frac{\gamma_k}{(1-\tau_k^2)\, l(\mathbf{x}_k)\mu^2}\left(\overline{P}_{\mathrm{RZF}}\left(1-\tau_k^2+\tau_k^2\,(1+\mu)^2\right)+\frac{\sigma^2}{l(\mathbf{x}_k)}\,(1+\mu)^2\right). \tag{66}$$

We are now left with proving Lemma 3. Solving with respect to $\overline{P}_{\mathrm{RZF}}$ leads to (32) from which one gets (omitting the computations for simplicity)

$$\overline{P}_{\mathrm{RZF}}' = \frac{\overline{P}_{\mathrm{RZF}}}{\partial\rho} = 2c^2A\sigma^2\frac{\gamma^{(\mathcal{R})}-\mu}{\left(\mu\left(c+\rho\,(1+\mu)^2\right)-c\left(\gamma^{(\mathcal{R})}+B(1+\mu)^2\right)\right)^2}. \tag{67}$$

From the above equation, it turns out that the minimum power is achieved when $\mu = \gamma^{(\mathcal{R})}$. Plugging this result into (29) yields (35) from which (36) easily follows from (32) replacing μ with $\gamma^{(\mathcal{R})}$. This concludes the proof.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. E. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5g be?" *CoRR*, vol. abs/1405.2957, 2014. [Online]. Available: http://arxiv.org/abs/1405.2957

[2] "The 1000x data challenge," Qualcomm, Tech. Rep. [Online]. Available: http://www.qualcomm.com/1000x

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

[4] "Smart 2020: Enabling the low carbon economy in the information age," The Climate Group and Global e-Sustainability Initiative (GeSI), Tech. Rep., 2008. [Online]. Available: http://www.smart2020.org

[5] Green Touch Consortium, Tech. Rep. [Online]. Available: http://www.greentouch.org

[6] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, 2011.

[7] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590 – 3600, Nov. 2010.

[8] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40 – 60, Jan 2013.

[9] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186 – 195, February 2014.

[10] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37 – 43, March 2011.

[11] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5g," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44 – 51, May 2014.

[12] B. Li and P. Liang, "Small cell in-band wireless backhaul in massive MIMO systems: A cooperation of next-generation techniques," *CoRR*, vol. abs/1402.2603, 2014. [Online]. Available: http://arxiv.org/abs/1402.2603

[13] H. S. Dhillon and G. Caire, "Wireless backhaul networks: Capacity bound, scalability analysis and design guidelines," *CoRR*, vol. abs/1406.2738, 2014. [Online]. Available: http://arxiv.org/abs/1406.2738

[14] J. Hoydis, K. Hosseini, S. t. Brink, and M. Debbah, "Making smart use of excess antennas: Massive MIMO, small cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5 – 21, 2013. [Online]. Available: http://dx.doi.org/10.1002/bltj.21602

[15] K. Hosseini, J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO and small cells: How to densify heterogeneous networks," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 5442 – 5447.

[16] B. Li and P. Liang, "Small cell in-band wireless backhaul in massive MIMO systems: A cooperation of next-generation techniques," *CoRR*, vol. abs/1402.2603, 2014.

[17] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391 – 4403, Oct. 2013.

[18] A. Muller, E. Bjornson, R. Couillet, and M. Debbah, "Analysis and management of heterogeneous user mobility in large-scale downlink systems," in *Asilomar Conference on Signals, Systems and Computers*, Nov. 2013, pp. 773 – 777.

[19] H. Shirani-Mehr, G. Caire, and M. Neely, “MIMO downlink scheduling with non-perfect channel state knowledge,” *IEEE Transactions on Communications*, vol. 58, no. 7, pp. 2055 – 2066, July 2010.

[20] P. de Kerret and D. Gesbert, “Degrees of freedom of the network MIMO channel with distributed CSI,” *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6806 – 6824, Nov. 2012.

[21] R. Zakhour and S. Hanly, “Base station cooperation on the downlink: Large system analysis,” *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.

[22] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, “Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[23] L. Sanguinetti, E. Björnson, M. Debbah, and A. Moustakas, “Optimal linear pre-coding in multi-user MIMO systems: A large system analysis,” in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2014.

[24] A. T. H. Asgharimoghaddam and N. Rajatheva, “Decentralizing the optimal multi-cell beamforming via large system analysis,” in *Proceedings of the IEEE International Conference on Communications*, Sydney, Australia, June 2014. [Online]. Available: http://arxiv.org/abs/1310.3843

[25] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[26] 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.

[27] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, “Achieving “massive MIMO” spectral efficiency with a not-so-large number of antennas,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3226 – 3239, Sept. 2012.

[28] L. Sanguinetti, A. L. Moustakas, E. Björnson, and M. Debbah, “Large system analysis of the energy consumption distribution in multi-user MIMO systems with mobility,” *CoRR*, vol. abs/1406.5988, 2014.

[29] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, “Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?” *submitted to IEEE Trans. Wireless Comm.*, April 2014. [Online]. Available: http://arxiv.org/abs/1403.6150

[30] Z. Shen, A. Khoryaev, E. Eriksson, and X. Pan, “Dynamic uplink-downlink configuration and interference management in TD-LTE,” *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 51 – 59, Nov. 2012.

[31] 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.

[32] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing – The large-scale array regime,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441 – 6463, Oct 2013.

[33] G. Calcev, D. Chizhik, B. Goransson, S. Howard, H. Huang, A. Kogiantis, A. Molisch, A. Moustakas, D. Reed, and H. Xu, “A wideband spatial channel model for system-wide simulations,” *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 389 – 403, March 2007.