Analysis and Management of Heterogeneous User Mobility in Large-scale Downlink Systems

Axel Müller‡*, Emil Björnson†, Romain Couillet¶, and Mérouane Debbah§

*Intelli Mobile Communications, Sophia Antipolis, France
†ACCESS Linnaeus Centre, Signal Processing Lab, KTH Royal Institute of Technology, Sweden.
¶Department of Telecommunications and §Alcatel-Lucent Chair on Flexible Radio, Supéléc, France
{axel.mueller, romain.couillet, merouane.debbah}@supelec.fr, emil.bjornson@ee.kth.se

Abstract—Modern cellular networks need to serve user terminals with large disparities in mobility, which incurs different accuracy of the channel state information for each user. The impact of such heterogeneous mobility on the multi-cell downlink is analyzed in this paper. The base stations serve a multitude of users by coordinated beamforming. We derive deterministic equivalents for the user performance in a large scale system where the number of transmit antennas and user terminals grow large at a fixed ratio. We show that low and high mobility users can coexist and be served simultaneously, since the CSI imperfections of a user only harms the performance of this particular user. Simulations are used to verify the applicability of our large scale approximations for systems of practical dimensions. Furthermore, we show that the performance of high mobility users can be improved by explicitly managing the user priorities in the network.

I. INTRODUCTION

In order to keep up with the exponential traffic growth in cellular networks, the infrastructure is becoming increasingly dense [1]; the conventional macro base stations (BSs) are supplemented with small-cell BSs that act as hotspots inside the macro cells. The terminology of heterogeneous networks (HetNets) is commonly used to refer to these systems [2] and some of the key practical challenges are interference management and load balancing across BSs.

It is not only the infrastructure that is heterogeneous in modern networks, but also the user conditions and particularly the user mobility. While the conventional cellular networks, also known as mobile networks, were mainly deployed for outdoor usage, we can already see a large disparity in use cases: low mobility indoor users coexist with medium and high mobility outdoor users. The mobility is closely related to the quality of the channel state information (CSI) that can be acquired at the BSs; high mobility implies low-quality CSI.

In conventional networks with orthogonal allocation of time/frequency resources, heterogeneous user mobility mainly affects the scheduling decisions. However, modern cellular networks support multi-antenna transmission where a multitude of users are served simultaneously using the same time/frequency resources; inter-user interference is then controlled by spatial precoding [3]–[5]. Since CSI is a prerequisite for effective spatial interference suppression [3], this raises the question of how to take the user mobility into account in the precoding design.

The majority of the existing literature on precoding with imperfect CSI considers homogeneous CSI quality among the users, which is not the case in HetNets. A notable exception is [6] where the authors considered joint scheduling and precoding in the downlink of a single-cell system. The authors concluded that only low mobility users can be served jointly, while orthogonal transmission resources should be allocated to high mobility users. A similar result was obtained in [7], where a cellular network with joint transmission between the BSs was considered. The analysis revealed that the multiplexing gain is limited by the high mobility users, thus it appears as if all other users suffer when coexisting with high mobility users.

In this paper, we consider a cellular network where the users in each cell have different mobility, which is a relevant distinction in HetNets where the high mobility users are served by macro BSs and the low mobility users are associated with the small-cell BSs. Different from [7], we consider coordinated beamforming where each user is served by one BS but the precoding is coordinated across the cells to suppress inter-cell interference. To obtain fundamental insights, we consider the large-system regime where the number of transmit antennas and user terminals go to infinity at a fixed ratio. As a main contribution, we derive deterministic expressions for the asymptotic user rates, which also serve as accurate approximations in practical non-asymptotic regimes. These novel expressions generalize the prior work in [8] for single-cell systems and in [9] for multi-cell systems where only deterministic statistical CSI is utilized for suppression of inter-cell interference. To account for the different user mobility, the precoding is a simplified version of the optimal precoding parametrization in [10]; more precisely, different priority weights are assigned to cells with different mobility.

Our analysis reveals that the mobility of a user has a detrimental impact on its own achievable rate, but has no direct impact on the other users. In particular, the low mobility users will not suffer from coexisting with high mobility users under coordinated beamforming—which is fundamentally different from the case of joint multi-cell transmission in [7]. Furthermore, the heterogeneity can be managed in our framework by explicitly increasing/decreasing the priority of high mobility
users, which changes their rates at the cost/gain of other users.

II. SYSTEM MODEL

In the following, we analyze cellular downlink multi-user multiple input multiple output (MU-MIMO) systems of the kind illustrated in Fig. 1. Each of the $L$ cells consists of one base station (BS) associated with a number of single antenna user terminals (UTs). In more detail, the $l$th BS is equipped with $N_l$ antennas and serves $K_l$ UTs. We generally set $N_l \geq K_l$ in order to avoid scheduling complications. We assume transmission on a single narrow-band carrier, usage of perfectly synchronized TDD protocol, full transmit-buffers, and universal frequency reuse among the cells. The $l$th BS transmits a data symbol vector $x_l = [x_{l,1}, \ldots, x_{l,K_l}]^T$ intended for its $K_l$ uniquely associated UTs. This BS uses the linear precoding matrix $G_l \in \mathbb{C}^{N_l \times K_l}$, where the columns $g_{l,k} \in \mathbb{C}^{N_l}$ constitute the precoding vectors for each UT. We note that BSs do not directly interact with each other and users from other cells are explicitly not served. Thus, the received signal $y_{l,k} \in \mathbb{C}$ at the $k$th UT in cell $l$ is

$$y_{l,k} = \sqrt{\chi_{l,k}}(h_{l,k})^H g_{l,k} x_{l,k} + \sum_{k' \neq k} \sqrt{\chi_{l,k}}(h_{l,k})^H g_{l,k'} x_{l,k'} + n_{l,k}$$

where $n_{l,k} \sim \mathcal{CN}(0,1)$ is the independent circularly symmetric complex Gaussian noise. The transmission symbols are chosen from a Gaussian codebook, i.e., $x_{l,k} \sim \mathcal{C}/\mathcal{N}(0,1)$. We assume block-wise small scale Rayleigh fading, thus the channel vectors are modeled as $h_{l,k} \sim \mathcal{CN}(0, \frac{1}{N_l} I_{N_l})$. The path-loss and other large-scale fading effects are incorporated in the $\chi_{l,k}$ factors. The scaling factor $\frac{1}{N_l}$ in the fading variances is of technical nature and utilized in the asymptotic analysis. It is canceled for a given arbitrarily sized system by modifying the transmission power accordingly.

A. Channel State Information and Effect of User Mobility

The UTs are assumed to perfectly estimate the respective channels to their serving BS, which enables coherent reception. This is reasonable at most user speeds, if proper downlink reference signals are alternated with data symbols. Concerning the downlink, the transmitters in our model can only obtain imperfect channel state information (CSI), for example due to imperfect pilot-based channel estimation, delays in the acquisition protocols, and user mobility. High user mobility naturally implies large CSI imperfections, because the coherence time of the channel diminishes with the speed. If the system allocates a constant fraction of the coherence time for pilot signaling, the accuracy of the channel estimation reduces with a shorter coherence time. If the system, on the other hand, keeps the pilot resources fixed, then a higher mobility induces outdated CSI.

To model imperfect CSI without making explicit assumptions on the acquisition protocol, we define the estimated channel vectors $\hat{h}_{l,k} \in \mathbb{C}^{N_l}$ to be

$$\hat{h}_{l,k} = \sqrt{\chi_{l,k}} \left[ (1 - (\tau_{l,k}^m)^2)h_{l,k} + \tau_{l,k}^m \hat{y}_{l,k} \right]$$

where $\hat{y}_{l,k} \sim \mathcal{CN}(0, \frac{1}{N_l} I_{N_l})$ is the normalized independent estimation error. Using this formulation, we can set the accuracy of the channel acquisition between the UTs of cell $l$ and the BS of cell $m$ by selecting $\tau_{l,k}^m \in [0,1]$; a small number on $\tau_{l,k}^m$ implies low mobility and vice versa. Furthermore, we remark that these choices imply $\hat{h}_{l,k} \sim \mathcal{CN}(0, \frac{1}{N_l} I_{N_l})$. For convenience later on we define the aggregated estimated channel matrices as $\hat{H}_{m} = [\hat{h}_{l,m}(1), \ldots, \hat{h}_{l,K_l,m}] \in \mathbb{C}^{N_l \times K_l}$.

B. Precoding and Power Constraints

Non-linear precoding schemes are infeasible in large scale systems due to their high complexity [9]. It is therefore important to find linear precoding schemes that perform well under imperfect CSI and heterogeneous mobility. We consider precoding matrices $G_m$, $m = 1, \ldots, L$ that have a generalized regularized zero forcing (RZF) structure. The generalizations are selected to closely resemble an optimal linear precoding structure derived in [10]:

$$G_m = \left( \sum_{l=1}^{L} \alpha_{l,m}^r \hat{H}_l^m (\hat{H}_l^m)^H + \gamma_m I_{N_m} \right)^{-1} \hat{H}_l^m \beta_m^r$$

$$\pm \left( \alpha_{m}^r \hat{H}_l^m (\hat{H}_l^m)^H + Z_m + \gamma_m I_{N_m} \right)^{-1} \hat{H}_l^m \beta_m^r$$

where $Z_m = \sum_{l \neq m} \alpha_{m}^r \hat{H}_l^m (\hat{H}_l^m)^H$. The $\alpha_{m}^r$ are optimization variables and the classic regularization parameter $\gamma_m$ is chosen to be the inverse of the transmit signal to noise ratio (SNR). In [11] it is argued that such additional weights can be used to balance the precoding orthogonality to different (interference-) subspaces. This offers an intuitive explanation for the effectiveness of the generalized precoding observed in Subsection IV-C.

The factors $\beta_m$ are used to fulfill the average (w.r.t. the UTs) transmit power constraint $P_m$ of each BS $m$:

$$\frac{1}{K_m} \text{tr} \left[ G_m^H G_m^m \right] = P_m$$

The truly optimal precoding matrices under perfect CSI would be parallel to $G_m = \left( \sum_{l=1}^{L} \sum_{k=1}^{K_l} \alpha_{l,m}^r h_{l,k}^m (h_{l,k}^m)^H + \gamma_m I_{N_m} \right)^{-1} h_{l,K_l,m}^H \beta_m^r$ [5].
C. Performance Measure

Many performance measures in cellular systems are functions of the signal to interference and noise ratios (SINRs) at each UT; e.g., (weighted) sum rate, outage probability and user geometry. Under the system model, the received signal power at the \( k \)th UT of cell \( l \) (\( \text{UT}_l,k \)) is

\[
\text{Sig}_{l,k}^{(i)} = \chi_{l,k}^j(h_{l,k}^j)^*g_{l,k}^j(g_{l,k}^j)^*h_{l,k}^j.
\]

Similarly, the interference power is

\[
\text{Int}_{l,k}^{(i)} = \sum_{m \neq l} \chi_{l,k}^m(h_{l,k}^m)^*G_m^j(G_m^j)^*h_{l,k}^m + \chi_{l,k}^j(h_{l,k}^j)^*G_l^j(G_l^j)^*h_{l,k}^j.
\]

where

\[
G_l^j = \left( \alpha_l^j \beta_l^j + Z^j + \gamma_l^j I_{N_l} \right)^{-1} \beta_l^j
\]

and \( \hat{H}_l^j \) with its \( k \)th column removed. Hence, the SINR at \( \text{UT}_l,k \) can be expressed as

\[
\frac{\gamma_{l,k}^{(i)}}{\gamma_{l,k}^{(i)}} = \frac{\text{Sig}_{l,k}^{(i)}}{\text{Int}_{l,k}^{(i)}}.
\]

In the following we focus on the sum rate (SRs), which is a commonly used performance measure utilizing the SINR values and straightforward to interpret. Under the assumptions that interference is treated as noise, the SR is deterministic and compact form that can easily be interpreted. Also, the large system approximations accurate even in massive MIMO systems of modest size, as will be evidenced later via simulations (see Section IV).

We will derive a deterministic equivalent (DE) of the SINR values that allow for a large scale approximation to the sum rate expression. The DE is based on the following technical assumptions. Introducing the ratio \( c_i = N_i/K_i \), we assume

**A.1**. \( N_i, K_i \to \infty \), such that for all \( i \) we have

\[ 0 < \lim \inf c_i \leq \lim \sup c_i < \infty. \]

This asymptotic regime is denoted \( N \to \infty \) for brevity.

By adapting the analytical approach in [8] to the SINR expression in (8), we obtain a DE of the SINR, which is denoted \( \tau_{l,k}^{(i)} \) in the following.

**Proposition 1** (Deterministic Equivalent of the SINR). Under \( \text{A.1} \), we have

\[
\lim_{N \to \infty} \tau_{l,k}^{(i)} = 0
\]

where

\[
\tau_{l,k}^{(i)} = \frac{\text{Sig}_{l,k}^{(i)}}{\text{Int}_{l,k}^{(i)}} + 1
\]

with

\[
\text{Sig}_{l,k}^{(i)} = \beta_l(\chi_{l,k}^j)^2(\gamma_{l,k}^{(i)})^2(\gamma_{l,k}^{(i)})^2
\]

\[
\text{Int}_{l,k}^{(i)} = \sum_{l=1}^{L} \beta_m \left( \chi_{l,k}^m \chi_{l,k}^m(\gamma_{l,k}^{(i)})^2 \right) \chi_{l,k}^m \gamma_{l,k}^m \gamma_{l,k}^m.
\]

where \( \gamma_{l,k}^m = \alpha_l^m, \chi_{l,k}^m(\gamma_{l,k}^{(i)})^2. \) The parameter \( \beta_m \), the abbreviations \( g_m \) and \( f_m^{(i)} \), as well as the corresponding fixed-point equation \( e_m \) and \( e_m' \) are given in the following.

First, we define \( e_m \) to be the unique positive solution of the fixed-point equation

\[
e_m = \frac{\gamma_m + \frac{1}{N_m} \sum_{j=1}^{K_m} \sum_{j=1}^{K_m} \chi_{l,k}^m \chi_{l,k}^m f_m^{(i)} + \frac{1}{N_m} \sum_{l \neq m}^{K_m} \sum_{j=1}^{K_m} \chi_{l,k}^m \chi_{l,k}^m f_m^{(i)}}{N_m g_m}
\]

where

\[
g_m = -\frac{1}{N_m} \sum_{j=1}^{K_m} \sum_{j=1}^{K_m} c_{l,k}^{(i)} \left( f_m^{(i)} \right)^2
\]

and \( e_m' \) can be found directly, once \( e_m \) is known:

\[
e_m' = \frac{1}{N_m} \sum_{j=1}^{K_m} \chi_{l,k}^m \chi_{l,k}^m f_m^{(i)} + \frac{1}{N_m} \sum_{l \neq m}^{K_m} \sum_{j=1}^{K_m} \chi_{l,k}^m \chi_{l,k}^m f_m^{(i)}
\]

**Sketch of proof:** The start of the proof conditions that \( Z_m \) is fixed and follows the steps given in [8, Appendix II] for each term from Subsection II-C and the power normalization \( \beta_m \). After applying [8, Theorem 1] we obtain the fundamental equations \( e_m \). We now allow \( Z_m \) to be random and apply [12, Theorem 3.13] to obtain (12). It is admissible to apply the two theorems one after the other, as \( Z_m \) is a bounded sequence with probability one. The DEs of the treated terms are found by, again, following [8, Appendix II] and (for the interference term) [8, Lemma 7].

By employing dominated convergence arguments and the continuous mapping theorem (e.g., [12]), we see that Proposition 1 implies

\[
R_{l,k} = \log_2(1 + \gamma_{l,k}^{(i)})\xrightarrow{\text{a.s.}} 0.
\]
IV. SIMPLIFIED MODEL

A. Simplified Simulation Scenario

In the following we introduce a (geometrically symmetric) simplified simulation scenario (see Fig. 2). It enables us to demonstrate the possibilities of interference analysis and mitigation offered by the general DEs introduced in Section III. Simultaneously, the simple structure allows us to draw intuitively evident conclusions that are not obscured by complications from more realistic scenarios.

Fig. 2 illustrates the simplified simulation scenario. It is composed of one macro cell (MC) BS, BS1, in the center, that generally occupies itself with serving high mobility (HM) UTs. Furthermore, we have four small cell (SC) BSs arranged equiangular on a circle of radius $a$. These SC BSs are serving low mobility (LM) UTs. BS1 is equipped with $N$ antennas and all others utilize $N/2$ antennas. All BSs have the same number of associated UTs ($K$). To simplify analysis further, the UTs are equally distributed and co-located according to the markings in Fig. 2, i.e., at the edge of the cells defined by the SC BSs. The radius of the SC cells is assumed to be one. SCs are often placed within buildings, whose walls shield UTs that are not in the immediate vicinity of the building from interference. Thus, we assume that each UT only receives signals from the MC BS and the closest SC BS. The SCs employ traditional zero-forcing precoding, which is achieved by choosing $\forall l, k, \alpha_{l,k}^m = 1$, if $m = l$ and $\alpha_{l,k}^m = 0$ otherwise, in our precoder (3). Furthermore, the value of $\gamma_m, m = 2, \ldots, 5$ is set to be very small. The distance-dependent pathloss factors $\chi_{l,k}^m$ are calculated via $1/d^n$, where $d$ is the distance between BS$_m$ and UT$_l,k$ and $n$ is the pathloss-coefficient. For intuitively clear reasons, we introduce the definitions $\forall m : \tau_{l,m}^H = \tau_{l,m}^{HM}$, $\forall m, l \neq 1 : \tau_{l,m}^L = \tau_{l,m}^{LM}$ for the mobility parameters and $l \neq 1 : \alpha_{l}^{\Delta} = \alpha_{l}^{LM}$, $\alpha_{l}^{\Delta} = \alpha_{l}^{HM}$ for the precoding weights.

Unless otherwise noted, the following simulation parameters values are used in this section: $L = 5$, $N = 64$, $K = 16$, $n = 2$, $a = 1.5$, $P_1 = 20$ dB, $l \neq 1 : P_l = 3$ dB, $\tau_{LM} = 0.1$.

B. Numerical Analysis of Mobility

The first step of the mobility analysis is to verify the usability of the system model and the accuracy of the derived DE. For this reason, we compare in Fig. 3 the per UT rate, obtained with $\tau_{LM} = 0.1$ and $\tau_{HM} \in \{0.1, 0.4\}$ (solid lines). Furthermore, $\alpha_{LM}$ and $\alpha_{HM}$ are set to 1. As expected, we see that the rates reduce with higher mobility. In order to judge the applicability of the DE for systems of the given size, the results of Monte-Carlo (MC) simulations with 1000 channel realizations are given as marker symbols. We remind that the approximation can be improved by increasing the numbers of UTs and BS antennas.

Fig. 4 introduces two additional performance indicators, namely the average rate of only the LM UTs and the average rate of only the HM UTs. This allows to assess the impact of mobility of each user group in a distinctive manner. For this figure we choose $\alpha_{HM} = 3.09$ and $\alpha_{LM} = 1.43$ for reasons that will become evident in Subsection IV-C. The dashed lines in Fig. 4 show the optimum sum rate performance given perfect CSIT, as a reference. One can observe that performance of the LM UTs is unaffected by the CSI quality of the HM UTs. In other words, precoding with the bad channel estimates for the LM UTs of one cell does not cause additional interference to the LM UTs in the other cells. This is a rather surprising result, as others have previously found the multiplexing gain of the whole system to be limited by the high mobility UTs [7]. Yet the result makes sense if one remembers that random beamforming is optimal in large systems and also that our model of small scale fading causes all channels to be independent. The mathematical explanation of this effect is that the statistics of our estimated channels are equivalent to those of the true channels (see (2)).

C. Numerical Mobility Management

In this subsection we take full advantage of the closed form expression of the sum rate derived in Proposition 1. In order to investigate how the performance degradation caused by mobility can be managed, we employ numerical optimization
of the sum rate. More specifically, we note that one can use a single $\alpha_{LM}$ instead of four different weights for the SC UTs, since the symmetric nature of the simplified scenario makes them all equal. The numerical optimization thus corresponds to a grid search over the $\alpha_{HM}$ and $\alpha_{LM}$ weights included in the generalized RZF precoding of BS1.

Fig. 5 shows how the weights $\alpha_{HM}$ and $\alpha_{LM}$ can be used to obtain the optimal average UT rate (HM+LM), under degrading CSIT for the HM UTs. The optimization is carried out by a 2-dimensional grid search for $\alpha_{HM}$ and $\alpha_{LM}$. Recall that only the precoding at BS1 is adjusted, the other precoding matrices stay as originally defined (i.e., classical zero-forcing). We see that the optimal weights for the HM UTs ($\alpha_{HM}$) are decreasing, when its CSIT is getting worse. This indicates that less importance is placed on the corresponding estimated channels. This is reversed for the channel weights of the LM UTs ($\alpha_{LM}$). Overall, the increasing importance placed on the estimated channels (i.e., shifting from maximum ratio transmission to zero-forcing/interference cancellation) indicates that interference is becoming the dominant issue when CSIT becomes increasingly unreliable. It is especially informative to compare Fig. 5 with Fig. 4, as the latter uses the same starting values for the weights as the first, but does not adapt them to the mobility information. We observe that taking into account this information for precoding significantly improves overall system performance.

V. CONCLUSION

In this paper we analyzed the impact of heterogeneous UT mobility on the sum rate of a downlink multi-cell system employing generalized RZF precoding. We modeled the UT mobility in terms of a varying quality of the available CSIT and we derived a corresponding large scale approximation of the SINR values at each UT. Increasing UT mobility was found to reduce the sum rate of the system. In a more detailed analysis, we found that when the mobility of UTs in one cell increases, then the performance of these UTs decreases. However, the performance of UTs in other cells stays unaffected. The deterministic expression of the sum rate was successfully used in a practical mobility management scheme. Significant sum rate improvements were achieved by optimizing the weights of the generalized RZF precoding to balance between privileging served UTs and suppressing interference to other UTs. In general, the optimization prioritizes the UTs with the best CSIT. For future work, we expect large gains to be found by combining the generalized RZF precoding and proper scheduling.

REFERENCES