On the Optimal Number of Cooperative Base Stations in Network MIMO

Jakob Hoydis*, Student Member, IEEE, Mari Kobayashi, Member, IEEE, and Mérouane Debbah, Senior Member, IEEE

EDICS Categories: MSP-APPL, MSP-CAPC, WIN-CONT

Abstract

We consider a multi-cell, frequency-selective fading, uplink channel (network MIMO) where $K$ user terminals (UTs) communicate simultaneously with $B$ cooperative base stations (BSs). Although the potential benefit of multi-cell cooperation grows with $B$, the overhead related to the acquisition of channel state information (CSI) will rapidly dominate the uplink resource. Thus, there exists a non-trivial tradeoff between the performance gains of network MIMO and the related overhead in channel estimation for a finite coherence time. Using a close approximation of the net ergodic achievable rate based on recent results from random matrix theory, we study this tradeoff by taking some realistic aspects into account such as unreliable backhaul links and different path losses between the UTs and BSs. We determine the optimal training length, the optimal number of cooperative BSs and the optimal number of sub-carriers to be used for an extended version of the circular Wyner model where each UT can communicate with $B$ BSs. Our results provide some insight into practical limitations as well as realistic dimensions of network MIMO systems.

Index Terms

Coordinated Multi-Point (CoMP), network MIMO, multi-cell processing, base station cooperation, Wyner cellular uplink, channel estimation, imperfect channel state information (CSI), random matrix theory.

J. Hoydis, M. Kobayashi and M. Debbah are with the Dept. of Telecommunications, SUPELEC, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France, and the Alcatel-Lucent Chair on Flexible Radio (email: {jakob.hoydis, mari.kobayashi, merouane.debbah}@supelec.fr).

February 28, 2010 DRAFT
I. INTRODUCTION

Network MIMO has become the synonym for cooperative communications in the cellular context and is regarded as one of the most important concepts to boost the interference limited performance of today’s cellular networks. It is often also referred to as multi-cell processing (MCP) or distributed antenna systems (DAS) and corresponds to a communication system where multiple base stations (BSs), connected via high speed backhaul links to a central station (CS), jointly process data either received over the uplink or transmitted over the downlink. If the BSs could cooperate without any restrictions with regards to the backhaul capacity, processing delay and complexity, and the availability of channel state information (CSI), the multi-cell interference channel would be transformed into a multi-access (uplink) or broadcast (downlink) channel without multi-cell interference. This is the main argument for network MIMO and it has been shown in many works, e.g. [1], [2], that BS-cooperation has the potential to realize significant gains in throughput and reliability.

So far, the treatment of multi-cell cooperation in the literature has been either information-theoretic ([3], [4] and references therein) but limited to simple models or based on simulations [5], [6], [7] by accounting for more realistic and complex network structures. Most common and analytically tractable are the Wyner model [8], [9] and the soft-hand-off model [10], [11] which consider cooperation between either two or three adjacent BSs on an infinite linear or circular cellular array. Variants of both models have been studied under various assumptions on the transmission schemes and the fading characteristics. For a recent summary of related results see [4] and [12].

In practical systems, perfect BSs cooperation or global processing is very difficult, if not impossible, to achieve. The main limitations are threefold: (i) limited backhaul capacity, (ii) local connectivity and (iii) imperfect CSI at the CS\(^1\). Therefore, most of recent research targets the problem of constrained cooperation. For a more detailed overview of this topic we refer to the surveys [12], [13], [14].

Information-theoretic implications of limited backhaul capacity have been studied separately for the uplink and downlink in [15] and [16]. Recently, the optimal amount of user data sharing between the BSs for the downlink with linear beamforming and backhaul constraints was studied in [17].

The difficulties related to connecting a large number of BSs to a single CS have motivated the study of systems with only locally connected BSs [11], [18], [19]. Several distributed algorithms for the uplink [20] and downlink [21], [22] have been proposed and it was shown that even with local BS connection near-

\(^1\)Also the synchronization of the BSs as well as processing complexity and delay play an important role from an implementation perspective but are so far more or less neglected in the literature.
optimal performance can be achieved with a reasonable amount of message passing and computational complexity.

One of the most critical limitations of a practical network MIMO system, somehow overlooked compared to (i) and (ii), arises from the substantial overhead related to the acquisition of CSI, indispensable to achieve the full diversity or multiplexing gains. This overhead becomes paramount, in particular for fast fading channels, when the number of antennas, sub-carriers, UTs or BSs grows [23], [6], [7], [24]. Usually, CSI for the uplink is acquired through pilot signals sent by the UTs. This implies that a part of the coherence time of the channel needs to be sacrificed to obtain CSI with a sufficiently high quality. The inherent tradeoff between the resources dedicated to channel estimation and data transmission has been studied for the point-to-point MIMO channel [25], [26] and the multi-user downlink [27]. Recently, this problem was also addressed in the context of network MIMO systems, although with a different focus. In [24], [6], [7], the authors compare several multi-cellular system architectures and conclude that the downlink performance of network MIMO systems is mainly limited by the inevitable acquisition of CSI (rather than by limited backhaul capacity). It turns also out that a conventional cellular system might outperform a network MIMO system under some circumstances assuming that the number of coordinated antennas and the used training overhead for both systems are the same. This means in essence that simply installing more antennas per BS can lead to higher performance improvements than installing costly backhaul infrastructure.

The imperfections detailed above call for robust strategies adapted to restricted BS cooperation. Some schemes [28], [29] rely on local CSI at the BSs and statistical CSI at the CS, whereas others [30], [5] consider to serve only certain subsets of UTs with multiple BSs. Several BS-cooperation schemes have been studied in [31], [32] for the combination of limited backhaul capacity and imperfect CSI. The problem of “pilot contamination” caused by non-orthogonal training sequences in adjacent cells which can lead to significant inter-cell interference was addressed in [33] and an optimized multi-cell precoding technique has been proposed.

In this paper, we also consider limited BS cooperation by focusing in particular on the effects of (iii) imperfect CSI. More precisely, we study the performance of the multi-cell uplink with partially restricted cooperation assuming that:

1) The BSs act as oblivious relays which forward their received signals to the CS via wired, high capacity backhaul links. The CSI at the CS needs to be explicitly estimated based on pilot tones sent by the UTs.

2) The backhaul links are prone to random failures unknown to the UTs and BSs.
The second constraint is relevant to a packet-switched back haul network where a packet can be either completely lost or the transport delay renders its information useless by the time of its arrival. Thus, our model is a simple and tractable way to account for several imperfections which might occur in real systems. The “erasure model” for backhaul links was first mentioned in [34] and similar setup with a single source and multiple, randomly failing relays was studied in [35]. The main contributions of this work are

- Determination of the optimal fraction of the channel coherence time used for channel training and of the optimal number of cooperative BSs as a function of various network MIMO parameters (path loss, coherence time, backhaul link failure probability).
- Derivation of a close approximation of the net ergodic achievable rate and its derivative via a deterministic equivalent accounting for arbitrary path losses between the BSs and UTs and random backhaul link failure.
- Asymptotic analysis of the optimal training length and the optimal number of cooperative BSs in the high/low signal-to-noise-ratio (SNR) regimes.

The first point appears to be a novel result, although we limit our investigation to a simple setting where $B$ fully cooperative cells do not suffer from interference outside the network. The extension of this work to more realistic networks, such as clustered systems, is left to future investigations. Although the use of deterministic equivalents in the context of network MIMO is not new [36], [37], we show that they are suitable to treat related optimization problems despite their apparent complexity.

In essence, our results show that network MIMO is most beneficial when the path loss between all BSs and UTs is similar, the SNR is low and the channel coherence time is large. Thus, network MIMO will mainly benefit cell edge users which operate at low SNR but might be served by multiple BSs. For example, given a coherence block of 1000 channel uses and almost equally strong channels between all BSs and UTs, the optimal number of cooperative BSs is around 30. If the path loss difference between the channels from a UT to its closest and second closest BS is 10 dB, this number drops to 3. The effect of random backhaul link failure causes an almost linear rate loss but does not affect the optimal training length and the optimal number of cooperating BSs significantly. Furthermore, our results confirm the intuition that UTs with weak channels, and consequently low SNR, should focus all their power on a single sub-carrier while UTs with strong channels can afford to occupy more bandwidth. Our work allows thus to gain insight into practical limitations and realistic dimensions of cooperative networks.

The rest of this paper is organized as follows. The system model including channel training and data
transmission is described in Section II. The net ergodic achievable rate as our objective function for the optimization of the training interval and the number of cooperating BSs is defined in Section III where we also discuss an extended version of the circular Wyner model. The low and high SNR characterizations of this model are given in Section IV. Numerical results and concluding remarks are presented in Sections V and VI, respectively.

Throughout this work, we use boldface lowercase and uppercase letters to designate column vectors and matrices, respectively. For a matrix $X$, $x_{ij}$ or $(X)_{ij}$ denotes the $(i,j)$ entry of $X$. $|X|$ and $\text{tr}(X)$ denote the determinant and trace and $X^T$ and $X^H$ denote the transpose and complex conjugate transpose. For two matrices $X$ and $Y$, $X \otimes Y$ denotes the Kronecker (tensor) product. We denote an identity matrix of size $M$ as $I_M$ and $\text{diag}(x_1, \ldots, x_M)$ is a diagonal matrix of size $M$ with the elements $x_i$ on its main diagonal. We use $x \sim \mathcal{CN}(m, \mathbf{R})$ to state that the vector $x$ has a complex Gaussian distribution with mean $m$ and covariance matrix $\mathbf{R}$. The natural logarithm is denoted by $\log(\cdot)$.

II. SYSTEM MODEL

A. Channel Model

We consider a multi-cell, frequency-selective fading, uplink channel from $K$ single-antenna UTs to $B$ BSs with $M$ antennas each\(^2\). An example of the channel model for $B = K = 3$ and $M = 2$ is shown in

\(^2\)Our results can be easily extended to the case where each BS has a different number of antennas.
Fig. 1. Communication takes place simultaneously from all UTs to all BSs on $L$ parallel sub-channels assuming an OFDM transmission scheme. The received signal vector $y_b(l) \in \mathbb{C}^M$ at BS $b$ on sub-channel $l$ is given as

$$y_b(l) = \sum_{k=1}^{K} \sqrt{a_{bk}} w^b_k(l) x_k(l) + n_b(l)$$ (1)

where $a_{bk}$ is the distance dependent path loss factor between UT $k$ and BS $b$ and $w^b_k(l)$ the corresponding small-scale fading channel vector of the $l$th sub-channel. We denote $x_k(l)$ the transmitted symbol of UT $k$ on sub-channel $l$ and $n_b(l) \sim \mathcal{CN}(0, \sigma^2_{BS} \mathbb{I}_M)$ is a vector of additive white Gaussian noise (AWGN).

Denoting

$$x(l) \triangleq [x_1(l) \cdots x_K(l)]^T$$ (2)

the vector of the transmitted signals from all UTs on the $l$th sub-channel and

$$H_b(l) \triangleq \left[ \sqrt{a_{b1}} w^b_1(l) \cdots \sqrt{a_{bK}} w^b_K(l) \right]$$ (3)

the aggregated $M \times K$ channel matrix from all UTs to BS $b$ on the $l$th sub-channel, we can express the received signal vector $y_b(l)$ of BS $b$ on the $l$th sub-channel in matrix form as

$$y_b(l) = H_b(l)x(l) + n_b(l).$$ (4)

We assume that the UTs are subject to the same average transmit power constraint $P$ over all sub-channels, i.e.,

$$\mathbb{E} \left[ \sum_{l=1}^{L} |x_k(l)|^2 \right] \leq P, \quad k = 1, \ldots, K.$$ (5)

We assume a discrete-time block-fading channel where the channel remains constant in frequency and time for a coherence block of $T$ channel uses and then changes randomly from one block to the other. We let $T = T_c W_c$, where $W_c$ is the bandwidth per sub-channel in Hz and $T_c$ the channel coherence time in seconds. Presuming that the bandwidth of each sub-channel $W_c$ is on the order of the channel coherence bandwidth, that the antenna spacing at the BSs is sufficiently large and that the channels from the UTs to the BSs are uncorrelated, the channel vectors $w^b_k(l)$ can be modeled as independent and identically distributed (i.i.d.) $\sim \mathcal{CN}(0, \mathbb{I}_M)$ over $l$, $k$ and $b$.

The BSs forward their received signals via wired, high-capacity backbone links to a CS which jointly processes the signals from all BSs. We assume that the backhaul links between the BSs and the CS are delay-free but prone to random failure. That is, during each coherence block $T$, the link between a given BS and the CS on all sub-channels is either error-free or erased. We further presume that the $B$ backhaul links fail independently from each other. Whereas the UTs and BSs are oblivious to the link failures, the
CS can easily detect which backhaul links are defective. A similar model was studied in [35] for the case of a single source transmitting to a remote destination via a set of relays with randomly failing links to the destination. The received signal vector $y_{CS}(l) \in \mathbb{C}^{BM}$ at the CS on sub-channel $l$ is given as

$$y_{CS}(l) = E(H(l)x(l) + n_{BS}(l)) + n_{CS}(l)$$

where

$$H(l) \triangleq \left[ H_{1}(l)^{T} \ldots H_{B}(l)^{T} \right]^{T}$$

is the aggregated $MB \times K$ channel matrix from all UTs to all BSs on sub-channel $l$ and $n_{BS}(l) \triangleq \left[ n_{1}(l)^{T} \ldots n_{B}(l)^{T} \right]^{T}$ is the stacked noise vector from all BSs. The random failure of the backhaul links is modeled by the $BM \times BM$ matrix

$$E = \text{diag}(e_1, \ldots, e_B) \otimes I_{M}$$

where the $\{e_b\}$ are i.i.d. Bernoulli random variables with $P(e_b = 0) = \epsilon$ and $\epsilon$ denotes the backhaul link failure probability. We further assume some additional noise $n_{CS}(l) \sim \mathcal{CN}(0, \sigma_{CS}^{2}I_{BM})$ at the CS. Since all processing, i.e., channel estimation and decoding, is done at the CS, we will see later that only the total received noise variance, denoted by

$$\sigma^{2} \triangleq \sigma_{BS}^{2} + \sigma_{CS}^{2}$$

impacts the performance\(^3\). Let $\text{SNR} \triangleq P/\sigma^2$ denote the average transmit signal-to-noise ratio.

We collect the path loss factors from all UTs to all antennas of all BSs in the $(BM \times K)$ matrix\(^4\)

$$V \triangleq A \otimes 1_{M}$$

where $(A)_{bk} = a_{bk}$ and $1_{M}$ is a $M$-dimensional column vector with all entries equal to one, such that

$$\mathbb{E} \left[ |h_{ij}(l)|^{2} \right] = v_{ij}, \quad l = 1, \ldots, L$$

where $h_{ij}(l)$ and $v_{ij}$ denote the $(i, j)$ entries of the matrices $H(l)$ and $V$, respectively. In the sequel, we will refer to $V$ as the variance profile of the channel matrix $H(l)$.

\(^3\)Assuming a sufficiently high backhaul capacity, the noise variance $\sigma_{CS}^{2}$ on the wired backhaul links is negligible compared to the thermal noise at the BS antennas. However, correct modeling of the backhaul links is still an open research problem beyond the scope of this paper.

\(^4\)Note that the variance profile is independent of the sub-channel index $l$ since we assume that the path loss is identical for all sub-channels. This might be not the case for extremely large bandwidth but it is a reasonable assumption for most practical scenarios.
B. Channel Training

Similar to [25], each channel coherence block of length $T$ is split into a phase for channel training and a phase for data transmission. During the training phase of length $\tau$, all UTs transmit orthogonal sequences of known pilot symbols. The orthogonality of the training sequences imposes $\tau \geq K$. Assuming that the UTs split their transmit power equally over all sub-channels, the CS estimates the channel coefficient $h_{ij}(l)$ based on the observation

$$r_{ij}(l) = \sqrt{\frac{\tau P}{L}} h_{ij}(l) + n_{ij}(l), \quad \text{if } (\mathbf{E})_{ii} = 1$$

where $n_{ij}(l) \sim \mathcal{CN}(0, \sigma^2)$. Applying the minimum mean square error (MMSE) estimator, we can decompose $h_{ij}(l)$ into the estimate $\hat{h}_{ij}(l)$ and the independent estimation error $\tilde{h}_{ij}(l)$, such that

$$h_{ij}(l) = \hat{h}_{ij}(l) + \tilde{h}_{ij}(l).$$

Let us define the variances of the estimated channel and the estimation error as

$$\hat{v}_{ij}(\tau) = \mathbb{E} \left[ |\hat{h}_{ij}(l)|^2 \right] = \frac{\tau P}{L} v_{ij}^2, \quad \forall l$$

$$\tilde{v}_{ij}(\tau) = \mathbb{E} \left[ |\tilde{h}_{ij}(l)|^2 \right] = \frac{v_{ij} \sigma^2}{\tau P v_{ij} + \sigma^2}, \quad \forall l.$$  

We denote $\hat{V}(\tau)$ and $\tilde{V}(\tau)$ the variance profiles of the estimated channel $\hat{H}(l)$ and the estimation error $\tilde{H}(l)$, respectively. One can easily verify that

$$V = \hat{V}(\tau) + \tilde{V}(\tau).$$

C. Data Transmission

During the data transmission phase, each UT $k$ sends its data by the $L$ Gaussian inputs $x_k(l) \sim \mathcal{CN}(0, P/L)$, i.i.d. over $l$ and $k$. With the knowledge of $\mathbf{E}$ and $\hat{H}(l)$, the CS “sees” in its received signal $y_{CS}(l)$ the useful term $\mathbf{E}\hat{H}(l)x(l)$ and the overall noise term $z(l)$, i.e.,

$$y_{CS}(l) = \mathbf{E}\hat{H}(l)x(l) + z(l)$$

where

$$z(l) \triangleq \mathbf{E} \left( \hat{H}(l)x(l) + n_{BS}(l) \right) + n_{CS}(l).$$

One can easily see that the statistical distribution of $z(l)$ is independent of the sub-channel index $l$. Thus, the covariance matrix $K_z(\tau)$ of $z(l), \forall l$ is given as

$$K_z(\tau) \triangleq \mathbb{E} \left[ z(l)z(l)^H \right] = \left( \frac{P}{L} Q(\tau) + \sigma_{BS}^2 \mathbf{I}_{BM} \right) \mathbf{E} + \sigma_{CS}^2 \mathbf{I}_{BM}.$$
where
\[ Q(\tau) \triangleq \mathbb{E} \left[ \hat{H}(l)\hat{H}(l)^H \right] = \text{diag} (q_1(\tau), \ldots, q_{BM}(\tau)) \] (18)
and
\[ q_i(\tau) = \sum_{k=1}^{K} \bar{e}_{ik}(\tau), \quad i = 1, \ldots, BM. \] (19)

Since the statistical distributions of all sub-channels, signals and noise are i.i.d. with respect to the sub-channel index \( l \), we will hereafter omit the dependence on \( l \).

III. NET ERGODIC ACHIEVABLE RATE

In this section, we consider the optimization of the training length \( \tau \) with the goal of maximizing the net ergodic achievable rate. For a given channel estimate \( \hat{H} \), given \( \mathbb{E} \) and perfect knowledge of \( \mathbf{K}_n(\tau) \) at the CS, the following rate per sub-channel per cell is achievable
\[
\frac{1}{B} \log \left| I_{BM} + \frac{P}{L} \mathbf{K}_z^{-1/2}(\tau)\mathbb{E}\mathbb{H}\hat{H}^H\mathbb{E}\mathbf{K}_z^{-1/2}(\tau) \right|.
\] (20)

Denoting the effective channel
\[ \mathbf{H}(\tau) \triangleq \mathbf{K}_z^{-1/2}(\tau)\mathbb{E}\hat{H} \] (21)
the ergodic achievable rate per sub-channel per cell is given by
\[
R(\tau, B) \triangleq \frac{1}{B} \mathbb{E}_{\mathbf{H}, \mathbb{E}} \left[ \log \left| I_{BM} + \frac{P}{L} \mathbf{H}(\tau)\mathbf{H}(\tau)^H \right| \right]
\] (22)

where the expectation is with respect \( \mathbf{H} \) and \( \mathbb{E} \). In order to find the optimal training length \( \tau^* \) for a given coherence block \( T \), we wish to solve the following optimization problem
\[
\begin{align*}
\text{maximize} & \quad R_{\text{net}}(\tau, B) \triangleq L \left( 1 - \frac{\tau}{T} \right) R(\tau, B) \\
\text{subject to} & \quad K \leq \tau \leq T
\end{align*}
\] (23)

where \( R_{\text{net}}(\tau, B) \) denotes the net ergodic achievable rate per cell. We will first determine the optimal training length \( \tau^*(B) \) for a fixed \( B \) and then optimize the number of cooperative BSs, such that \( R_{\text{net}}(\tau^*(B), B) \) is maximized. Here, the difficulty consists in computing the ergodic achievable rate \( R(\tau, B) \) explicitly. Since a closed-form expression of \( R(\tau, B) \) for finite dimensions of the channel matrix \( \mathbf{H} \) would be intractable, we resort to an approximation based on the theory of large random matrices. Assuming that the numbers of BSs \( B \), UTs \( K \) and sub-channels \( L \) grow infinitely large at the same rate, the following theorem can be used to provide a close approximation to the ergodic achievable rate.
**Theorem 1:** [38, Theorem 4.1] Let $H$ be an $M \times N$ random matrix with independent elements with zero mean and variance profile $V$. Let $\rho \geq 0$ and $0 < c < \infty$ be two real constants. Denote

$$C(\rho, V) \triangleq \frac{1}{M} \mathbb{E} \left[ \log \left| I_N + \frac{\rho}{N} HH^H \right| \right]$$

then

$$C(\rho, V) - \overline{C}(\rho, V) \xrightarrow{\text{a.s.}} \frac{\rho}{N} \sum_{i=1}^{M} \log \left( \frac{\rho}{\Psi_i} \right) + \frac{1}{M} \sum_{j=1}^{N} \log \left( \frac{\rho}{\Upsilon_j} \right) \to 0$$

where

$$\overline{C}(\rho, V) \triangleq \frac{1}{M} \sum_{i=1}^{M} \log \left( \frac{\rho}{\Psi_i} \right) + \frac{1}{M} \sum_{j=1}^{N} \log \left( \frac{\rho}{\Upsilon_j} \right)$$

and where the vectors $\Psi = (\Psi_1, \ldots, \Psi_M)^T$ and $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_N)^T$ are the unique solutions to the set of $M + N$ fixed point equations

$$\Psi_i = \frac{\rho}{1 + \frac{1}{N} \sum_{j=1}^{N} v_{ij} \Upsilon_j}, \quad i = 1, \ldots, M$$

$$\Upsilon_j = \frac{\rho}{1 + \frac{1}{N} \sum_{i=1}^{M} v_{ij} \Psi_i}, \quad j = 1, \ldots, N$$

such that the $\Psi_i$ and $\Upsilon_j$ are Stieltjes transforms [39], evaluated at point $\rho$, of probability measures over the positive real line.

The next corollary adapts Theorem 1 to our channel model.

**Corollary 3.1:** For $B \to \infty$, $K/B \to \eta$ and $L/B \to \gamma$, where $0 < \eta, \gamma < \infty$, the ergodic achievable rate $R(\tau, B)$ as given in (22) converges almost surely to

$$\overline{R}(\tau, B) \triangleq \mathbb{E} \left[ M \cdot \overline{C} \left( \frac{\eta}{\gamma} P, \overline{V}(\tau, \overline{E}) \right) \right]$$

where

$$\overline{V}(\tau, \overline{E}) \triangleq \frac{1}{M} \sum_{i=1}^{M} \frac{\overline{V}_i}{\overline{E}_i}$$

$$\overline{P}(\tau) \triangleq \frac{1}{K} \sum_{k=1}^{K} \frac{P_k}{\overline{K}}$$

**Proof:** Corollary 3.1 is a straightforward application of Theorem 1 to the ergodic achievable rate (22). First, since both $K_\tau(\tau)^{-1}$ and $\overline{E}$ are diagonal matrices, the variance profile of the effective channel $\overline{H}$ is given as (27). Second, the factor $M$ in (26) stems from the fact that we are interested in the ergodic rate per cell and not in the ergodic rate per antenna per cell. Third, the corollary follows then directly from Theorem 1 by rewriting $\frac{P}{\tau} = \frac{\overline{P}}{K}$. \hfill \blacksquare
Remark 3.1: The expectation in (26) can be easily computed in the following way. Let \( n \triangleq \sum_{b=1}^{B} e_b \), \( n \in [0, B] \), denote the number of non-defective backhaul links. It follows that \( n \) is binomially distributed with probability mass function

\[
f(n, \epsilon) = \binom{B}{n} (1 - \epsilon)^n \epsilon^{B-n}, \quad 1 \leq n \leq B.
\]

(28)

Further denote \( N_n \triangleq \binom{B}{n} \) and \( \{E^n_1, \ldots, E^n_{N_n}\} \), the set of all matrices \( E = \text{diag}(e_1, \ldots, e_B) \otimes I_M \) such that the number of 1s on the diagonal is equal to \( nM \), then

\[
R(\tau, B) = B \sum_{n=1}^{N_n} f(n, \epsilon) \sum_{i=1}^{N_n} M \cdot C\left( \frac{\eta}{\gamma} P, \nabla(\tau, E^n_i) \right).
\]

(29)

In the sequel, we will use Corollary 3.1 to approximate the ergodic achievable rate \( R(\tau, B) \). As we will show later through simulations and similarly to what has been observed in [40], [37], the approximation turns out to be rather tight even for very small network dimensions, for example \( B = K = 3, M = 2 \). It is difficult to show that \( R_{\text{net}}(\tau, B) \) is a concave function in \( \tau \) apart from some special cases. Nevertheless, we assume that \( R_{\text{net}}(\tau, B) \) is concave and look thus for the value \( \tau^* \) satisfying

\[
R'_{\text{net}}(\tau^*, B) = L \left( \left( 1 - \frac{\tau^*}{T} \right) R'(\tau^*, B) - \frac{1}{T} R(\tau^*, B) \right) = 0.
\]

(30)

Instead of solving (30), we use the approximation by Corollary 3.1 and find \( \tau^* \) as the solution of

\[
\left( 1 - \frac{\tau^*}{T} \right) \overline{R}(\tau^*, B) = \frac{1}{T} \overline{R}(\tau^*, B)
\]

(31)

where

\[
\overline{R}(\tau, B) \triangleq \mathbb{E}_E \left[ M \cdot C' \left( \frac{\eta}{\gamma} P, \nabla(\tau, E) \right) \right].
\]

(32)

Taking the derivative of \( \overline{C} \left( \frac{\eta}{\gamma} P, \nabla(\tau, E) \right) \) with respect to \( \tau \), we get

\[
C' \left( \frac{\eta}{\gamma} P, \nabla(\tau, E) \right) =
- \frac{1}{BM} \left[ \sum_{i=1}^{BM} \frac{\Psi_i'}{\Psi_i} + \sum_{j=1}^{K} \frac{\Upsilon_j'}{\Upsilon_j} \right] \Psi' \nabla(\tau, E) \Upsilon + \Psi' \nabla(\tau, E) \Upsilon + \Psi \nabla(\tau, E) \Upsilon' \frac{\eta}{\gamma} PBMK
\]

(33)

\*\*\*We denote \( f'(\tau) \triangleq \frac{df(\tau)}{d\tau} \).
Fig. 2. Example of a hexagonal cellular system with three cells and three UTs. Cross and circle marks indicate the locations of the BSs and UTs, respectively. The normalized distance between UT \( k \) and BS \( b \) is denoted by \( d_{bk} \). The outer cell radius is normalized to one.

where \( \Psi \) and \( \Upsilon \) are obtained from (25) by replacing \( \rho \) by \( \frac{\gamma P}{\gamma P} \) and \( v_{ij} \) by \( \bar{v}_{ij}(\tau, E) \) and \( \Psi' \triangleq (\Psi'_1 \ldots \Psi'_{BM})^T \) and \( \Upsilon' \triangleq (\Upsilon'_1 \ldots \Upsilon'_K)^T \) are the solutions to the following set of \( BM + K \) fixed point equations

\[
\Psi'_i = -\frac{\Psi_i^2}{\gamma PK} \left[ \sum_{j=1}^{K} \bar{v}_{ij}(\tau, E) \Upsilon_j + \bar{v}_{ij}(\tau, E) \Upsilon'_j \right]
\]

\[
\Upsilon'_j = -\frac{\Upsilon_j^2}{\gamma PK} \left[ \sum_{i=1}^{BM} \bar{v}'_{ij}(\tau, E) \Psi_i + \bar{v}_{ij}(\tau, E) \Psi'_i \right].
\] (34)

The derivative of the variance profile \( \bar{v}(\tau, E) \) can be computed to be

\[
\bar{v}'_{ij}(\tau, E) = (E)_{i,i} \frac{\hat{v}'_{ij}(\tau) \left[ \frac{P}{\tau} q_i(\tau) + \sigma^2 \right] - \hat{v}_{ij}(\tau) \frac{P}{\tau} q'_i(\tau)}{\left[ \frac{P}{\tau} q_i(\tau) + \sigma^2 \right]^2}
\] (35)

where

\[
\hat{v}'_{ij}(\tau) = -\bar{v}'_{ij}(\tau) = -\frac{P}{\tau} v_{ij}^2 \sigma^2 \left( \frac{P}{\tau} v_{ij} + \sigma^2 \right)^2
\] (36)

and

\[
q'_i(\tau) = \sum_{k=1}^{K} \bar{v}'_{ij}(\tau) = -\sum_{k=1}^{K} \bar{v}'_{ij}(\tau).
\] (37)

Using these approximations, \( \tau^* \) can be easily found through a line search.

A. A Cellular Example

In order to verify the accuracy of the approximations of the last section, we consider a small hexagonal cellular system with either 3 or 7 cooperative BSs (Fig. 2). Each BS is equipped with \( M = 2 \) antennas.
and we randomly place one UT in each cell. The path loss factor from UT $k$ to BS $b$ is given as

$$a_{bk} = \left( \frac{1}{d_{bk}} \right)^\beta$$  \hspace{1cm} (38)

where $d_{bk}$ is the distance between UT $k$ and BS $b$, normalized to the maximum distance within a cell, and $\beta$ is the path loss exponent which lies usually within a range of 2 to 5 dependent on the radio environment.

We normalize the variance of the total received noise at the CS to $\sigma^2 = 1$ and let the number of sub-channels be $L = 1$. Under this assumption, the received SNR of a cell-edge UT is identical to its transmit power. Figure 3 depicts the net ergodic achievable rate as a function of the transmit power $P$ for a 3- and 7-cell system. The coherence block is assumed to be $T = 1000$ and we consider three different backhaul link failure probabilities $\epsilon = \{0, 0.4, 0.7\}$. In this plot, we fixed the training length to $\tau = 40$. The results correspond to one random system snapshot of user locations where we assume that no UT is within a normalized distance of 0.1 from its closest BS. The deterministic equivalent (Corollary 3.1) approximates the simulation results very closely over the full range of $P$ and all values of $\epsilon$. We can also observe that the 7-cell system achieves a slightly larger per-cell rate than the 3-cell system. Moreover, the backhaul link failures essentially cause a loss of the degrees of freedom of the channel. This can be seen from the reduced slopes of the rate curves for $\epsilon = \{0.4, 0.7\}$. 

Fig. 3. Net ergodic achievable rate $R_{\text{net}}(\tau, B)$ over transmission power $P$ for a cellular system layout with 3 (circle marks) or 7 (cross marks) cells and $M = 2$ antennas per BS. A coherence block $T = 1000$, training length $\tau = 40$, $L = 1$ sub-channels and path loss exponent $\beta = 4$ are assumed. The markers indicate simulation results, the corresponding lines are obtained from the deterministic equivalent (Corollary 3.1).
B. Extension of the Wyner Model

The analysis in the previous section allows us to study multi-cellular networks of arbitrary size and with arbitrary path loss profiles. However, in order to gain an insight into the impact of the network topology on the overall performance, one needs essentially to average over all possible user locations. Since the latter approach is intractable, we will consider an extended version of the circular Wyner model [9] which depends solely on one parameter $\alpha$. Although simple, the Wyner model is widely considered in the literature and known to capture the essential features of a multi-cellular system. We extend this model by assuming a circular system layout where each UT can communicate with all BSs and not only with the ones located in adjacent cells. The path loss between a UT and its $i^{th}$ closest BS is given as $\alpha^{i-1}$, $\alpha \in [0, 1]$, such that the path loss factor $a_{bk}$ between UT $k$ and BS $b$ equals

$$ a_{bk} = \alpha^{\lfloor \frac{b-1}{B} \rfloor - \lfloor \frac{k-b+\frac{B}{2}}{B} \rfloor} $$

where $\lfloor \cdot \rfloor_B$ is the modulo-$B$ operator. It is noteworthy that this model requires and equal number of BSs and UTs, i.e., $B = K$. For the case of $B = K = 5$ and $M = 1$, the variance profile of the channel matrix $H$ takes the following form

$$ V = \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^2 & \alpha \\ \alpha & 1 & \alpha & \alpha^2 & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha^2 & \alpha & 1 & \alpha \\ \alpha & \alpha^2 & \alpha^2 & \alpha & 1 \end{pmatrix}. $$

(40)

Albeit unrealistic, this simple model has the advantage to be symmetric with respect to the UTs, characterized solely by one parameter $\alpha$ and scalable with $B$. This allows us to study the dependence of the optimal training on the path loss and the number of cooperating BSs in a decoupled manner. Note that the extended Wyner model (39) includes some trivial cases. Namely, $\alpha = 0$ corresponds to $B$ non-interfering cells while $\alpha = 1$ can be interpreted as a MIMO point-to-point channel with no CSI at the transmitter which has been studied in [25]. Due to the circular symmetry of the model, the sums of the elements of every row and the sums of the elements of every column of $V$ are identical and we define

$$ K \triangleq \sum_{j=1}^{K} v_{ij} = \frac{1}{M} \sum_{i=1}^{BM} v_{ij}, \quad \forall i, j. $$

(41)

Similarly, let

$$ \bar{K}(\tau) \triangleq \sum_{j=1}^{K} \hat{v}_{1j}(\tau), \quad \bar{K}(\tau) \triangleq \sum_{j=1}^{K} \tilde{v}_{1j}(\tau). $$

(42)
be the row sums of the variance profile of the estimated channel matrix $\hat{H}$ and the estimation error $\tilde{H}$. This yields

$$K_z(\tau) = \left(\frac{P}{L}\hat{K}(\tau) + \sigma_{BS}^2\right)E + \sigma_{CS}^2I_{BM}$$

(43)

and one can easily verify that the variance profile of the effective channel (27) reduces to

$$V_{Wyner}(\tau, E) \triangleq \frac{1}{P}\hat{K}(\tau) + \sigma^2E\hat{V}(\tau).$$

(44)

We will first consider the case of no backhaul link failure, i.e., $E = I$. Due to the symmetry of the variance profile, the fixed-point equations (25) allow for a closed-form solution well-studied in the literature, e.g. [40], [36]. This provides consequently a closed-form expression for the asymptotic ergodic achievable rate which we will state in the next theorem.

**Theorem 2:** [40, Theorem 4] For $B \to \infty$ and $L/B \to \gamma$, where $0 < \gamma < \infty$, the ergodic achievable rate $R(\tau, B)$ of the extended Wyner model (39) with no backhaul link failures ($\epsilon = 0$) converges almost surely to

$$R_{Wyner}(\rho(\tau), M) \triangleq \log(1 + \rho(\tau)M - F(\rho(\tau), M))$$

$$+ M \log(1 + \rho(\tau) - F(\rho(\tau), M))$$

$$- F(\rho(\tau), M)/\rho(\tau)$$

(45)

where

$$\rho(\tau) \triangleq \frac{\hat{K}(\tau)\frac{P}{L}}{\hat{K}(\tau)\frac{P}{L} + \sigma^2}$$

(46)

and

$$F(x, y) \triangleq \frac{1}{4} \left(\sqrt{1 + x(1 + \sqrt{y})^2} - \sqrt{1 + x(1 - \sqrt{y})^2}\right)^2.$$

**Remark 3.2:** Note that Theorem 2 holds for any variance profile satisfying the symmetry constraint in (41) and is hence valid for a more general class of variance profiles than (39).

Turning to the case of backhaul link failure, first note that $E$ randomly sets some of the rows of $V$ to zero. This results in a variance profile which still has equal sums of the elements of every row (for those rows which are not equal to zero) but unequal sums of the elements of every column. Thus, the only way of obtaining an exact asymptotic capacity expression is to take the expectation over all $2^B$ possible realizations of $E$ (as done in Corollary 3.1). Since this becomes quickly computationally prohibitive as one needs to consider all $2^B$ possible realizations of $E$, we propose a heuristic approximation of the
ergodic achievable rate which is given in closed form. The approximation is based on two observations. First, the factor $M$ in (45) corresponds to the ratio of the number of rows to the number of columns of the channel matrix $H$. By the strong law of large numbers, it follows that the fraction of non-zero rows converges almost surely to $(1 - \epsilon)$ for $B \to \infty$. Second, if we make the approximation that the sums of the elements of every column of the matrix $\mathbf{E} \mathbf{V}$ are identical, we can apply Theorem 2 to get the following heuristic approximation of the ergodic achievable rate

$$R_{\text{heur}}(\rho(\tau), \epsilon) \triangleq R_{\text{Wyner}}(\rho(\tau), (1 - \epsilon)M).$$

(47)

This approximation becomes exact either when all elements of the variance profile are identical ($\alpha = 1$) or when $\epsilon = 0$. Moreover, one can verify that $R_{\text{heur}}(\rho(\tau), \epsilon)$ is indeed a concave function in $\tau$. Taking the derivative with respect to $\tau$, we can also use $R'_{\text{heur}}(\rho(\tau), \epsilon)$ to approximate $R(\tau, B)'$ in (32). In general, the heuristic approximation is rather tight for $\alpha \gg 0$ and $\epsilon \ll 1$.

IV. ASYMPTOTIC ANALYSIS

We will now study the asymptotic behavior of the extended Wyner model for $K = B$ in the high and low SNR regimes. Since the transmitted power per sub-channel $P/L$ tends towards zero with an increasing bandwidth ($L \to \infty$), the wideband-regime is equivalent to the low SNR regime. We let the number of antennas per BS to be $M = 1$ since adding more antennas at each BS will not impact the optimal training length or the optimal number of BSs.

A. High SNR

In the limit of very high transmit powers, the variance of the channel estimation error $\tilde{v}_{ij}(\tau)$ (14) becomes arbitrarily small for any value of $\tau \geq 1$. This implies that the estimated channel is asymptotically identical to the real channel, i.e., $\hat{\mathbf{H}} \to \mathbf{H}$ and $\mathbf{K}_z(\tau) \to \sigma_{\text{BS}}^2 \mathbf{E} + \sigma_{\text{CS}}^2 \mathbf{I}$. Then, recalling that $\sigma^2 = \sigma_{\text{BS}}^2 + \sigma_{\text{CS}}^2$, the effective channel (21) approaches the following limit

$$\mathbf{\Pi}(\tau) = \mathbf{K}_z^{-1/2}(\tau) \mathbf{E} \mathbf{H} \xrightarrow{P \to \infty} \frac{1}{\sigma} \mathbf{E} \mathbf{H}$$

(48)

which is independent of the training length. Thus, the net ergodic achievable rate per cell (23) becomes

$$R_{\text{net}}(\tau, B) \xrightarrow{P \to \infty} \frac{L}{B} \left(1 - \frac{\tau}{T}\right) \mathbf{E} \left[\log \left|\mathbf{I}_B + \frac{P}{L\sigma^2} \mathbf{E} \mathbf{H} \mathbf{H}^\dagger\right|\right].$$

(49)

It is obvious from (49) that the optimal training interval should be chosen as small as possible, i.e.,

$$\tau^* \xrightarrow{P \to \infty} B$$

(50)
which is imposed by the orthogonality of the training signals (23). Interestingly, \( \tau^* \) is independent of the backhaul link failure probability \( \epsilon \). Replacing \( \tau \) in (49) by \( \tau^* = B \) and realizing that 

\[
\mathbb{E} \left[ \log \left| I_B + \frac{P}{L\sigma^2} \hat{E}HH^H \right| \right] \sim (1 - \epsilon)B \log \left( \frac{PK}{L\sigma^2} \right)
\]

at high SNR, we can further approximate the net ergodic achievable rate as

\[
R_{net}(\tau^*, B) \approx L \left( 1 - \frac{B}{T} \right) (1 - \epsilon) \log \left( \frac{PK}{L\sigma^2} \right).
\]  

(51)

By noticing that the right-hand-side of (51) is decreasing in \( B \), we can conclude that the net ergodic achievable rate is maximized by the smallest possible choice of \( B \), i.e.,

\[
B^* \xrightarrow{P \to \infty} 1.
\]

(52)

Intuitively, for every BS we add to the system, each UTs sees a slight increase in its received SNR. At very high SNR, however, the rate increase due to an additional BSs does not compensate for the rate loss caused by the necessary increment of the training length. Hence, for \( P \to \infty \), one BS is optimal.

We can see from (51) that the backhaul link failure results in a loss of the degrees of freedom of the channel (pre-log factor \( (1 - \epsilon) \)) and it does not affect the optimal training length or the optimal number of cooperative BSs. Let us now assume that one wants to maximize the total net ergodic achievable rate over \( B \) cells, rather than the net rate per cell. In this case, (49) is maximized by \( \tau^* = B \approx T/2 \) (dependent on \( \alpha \) and \( T \)). This result agrees with the optimal number of transmit antennas for a given coherence time in the MIMO point-to-point link studied in [25].

**B. Low SNR**

We can approximate the net ergodic achievable rate at low SNR by

\[
R_{net}(\tau, B) \approx \frac{L}{B} \left( 1 - \frac{\tau}{T} \right) \mathbb{E} \left[ \text{tr} \left( \frac{P}{L} \hat{K}_z^{-1}(\tau) \hat{E}HH^H \right) \right]
\]

\[
\approx \frac{L}{B} \left( 1 - \frac{\tau}{T} \right) \mathbb{E} \left[ \text{tr} \left( \frac{P}{L} \hat{K}(\tau) \frac{\hat{E}\hat{H}H}{\sigma^2} \right) \right]
\]

\[
\approx \frac{L}{B} \left( 1 - \frac{\tau}{T} \right) B(1 - \epsilon) \frac{P}{L} \hat{K}(\tau)
\]

\[
\approx (1 - \epsilon) \left( 1 - \frac{\tau}{T} \right) \frac{P}{\sigma^2} \hat{K}(\tau)
\]

\[
\approx (1 - \epsilon) \left( 1 - \frac{\tau}{T} \right) \frac{\tau}{L} \left( \frac{P}{\sigma^2} \right)^2 \sum_{j=1}^{B} v_{1j}^2
\]

(53)

where (a) is based on the approximation \( \log |I + xA| \approx \text{tr} (xA) \) for \( x \ll 1 \), (b) follows directly from (44), (c) is based on the observation that 

\[
\mathbb{E} \left[ \text{tr} \left( \hat{E}HH^H \right) \right] = \hat{K}(\tau) \mathbb{E} \left[ \text{tr} (E) \right] = \hat{K}(\tau)(1 - \epsilon)
\]

, (d) is obtained by assuming \( \frac{P}{L} \hat{K}(\tau) \ll \sigma^2 \) and (e) is based on the approximation \( \hat{K}(\tau) = \sum_{j=1}^{B} \hat{v}_{1j}(\tau) = \).
\[ \sum_{j=1}^{B} \frac{\tau^P v_{1j}^2}{\tau^2 v_{1j} + \sigma^2} \approx \tau \frac{P}{\sigma^2} \sum_{j=1}^{B} v_{1j}^2. \]

The expression (53) is clearly maximized by choosing \( \tau^* = T/2 \). Consequently, we have that

\[ \tau^* \xrightarrow{P/L \to 0} \frac{T}{2}, \tag{54} \]

Since all \( v_{1j} > 0 \) for \( \alpha > 0 \), one can also see that choosing \( B \) as large as possible maximizes (53).

Recalling that \( \tau \geq B \), it follows that

\[ B^* \xrightarrow{P/L \to 0} \tau^* = \frac{T}{2}, \tag{55} \]

maximizes \( R_{\text{net}}(\tau^* = T/2, B) \) at low SNR. The intuition behind this result is as follows: For a single user in a single isolated cell, the optimal training length at low SNR equals half of the coherence block.

Now, adding one more BS to the system increases the “received SNR”, given by \( \left( \frac{P}{\sigma^2} \right)^2 \sum_{j=1}^{B} v_{1j}^2 \), without causing any increase in the optimal training length. Since this SNR-gain comes essentially for free, we can keep adding BSs to the system until \( B = T/2 \). The low SNR approximation (53) scales as \( 1/L \). This implies that a larger \( L \) decreases the net rate as more channel coefficients need to be estimated and less resources are available for data transmission. We will further discuss the optimal number of sub-channels to be used in Section V-D. Also note that \( \tau^* \) and \( B^* \) are, as for the case of high SNR, independent of the backhaul link failure probability \( \epsilon \).

V. Numerical Results

We will now apply the results developed in the previous sections to determine the optimal training interval \( \tau^* \) for a given set of parameters: the coherence block \( T \), the transmit power per UT \( P \), the number of antennas per BS \( M \), the number of sub-channels \( L \), the path loss factor \( \alpha \) and the backhaul link failure probability \( \epsilon \). We will first study how \( \tau^* \) is influenced by these parameters. Then, we address the problem of how many BSs should cooperate and how much bandwidth should be used to maximize the per-cell rate. Unless otherwise stated, we assume \( T = 1000 \), \( P = 10 \) dB, \( \sigma^2 = 1 \) and \( M = L = 1 \). All results are based on the extended circular Wyner model (39).

A. Concavity of the Objective Function

In Fig. 4, we show \( R_{\text{net}}(\tau, B) \) as a function of the training length \( \tau \) for a fixed path loss factor \( \alpha = 0.7 \). We observe that \( R_{\text{net}}(\tau, B) \) is indeed concave in the range \( 1 \leq \tau \leq T \). We can also see that each additional training symbol increases \( R_{\text{net}}(\tau, B) \) until the channel estimate becomes almost perfect (\( \tau \approx 40 \)). Increasing \( \tau \) further does not improve \( R(\tau, B) \) significantly and the influence of the pre-log
Fig. 4. Net ergodic achievable rate $R_{\text{net}}(\tau, B)$ over training length $\tau$ for the extended Wyner model with $B = 8$, $T = 1000$, $P = 10$ dB, $L = M = 1$ and $\alpha = 0.7$.

Fig. 5. Comparison of the optimal training length determined via exhaustive search and the deterministic equivalent approximation for $T = 100$, $P = 10$ dB, $\alpha = 0.5$, $\epsilon = 0$, $L = M = 1$. 
factor \((1 - \tau/T)\) in (23) becomes dominating. Furthermore, a higher backhaul link failure probability \(\epsilon\) reduces the achievable rate but does not qualitatively change the shape of \(R_{\text{net}}(\tau, B)\). In principle, the failure of backhaul links corresponds to a loss of the degrees of freedom of the channel, which leads to an almost linear rate decrease.

The validity of our optimization results is demonstrated in Fig. 5. Here, we consider a system with an increasing number of BSs for a fixed coherence time \(T = 100\), \(\alpha = 0.5\) and no backhaul link failure. We find the optimal training length \(\tau^*\) by an exhaustive search based on Monte Carlo simulations and compare these results with the optimal values obtained from a line search using the deterministic equivalent approximations (26). Note that for \(\epsilon = 0\), the heuristic approximation (47) and the deterministic equivalent are identical. Since the simulation based optimization works only in a discrete space, we also show the analytical results rounded to the closest integer. We observe a good agreement between both results.

**B. Behavior of \(\tau^*\)**

The optimal training interval \(\tau^*\) as function of the number of BSs \(B\) is depicted in Fig. 6 for different path loss factors \(\alpha = \{0.3, 0.7, 0.9\}\). We let \(\epsilon = 0\) since its impact is negligible. First, when \(B\) increases, the optimum training interval gets longer. This shows the necessity of good channel estimates to exploit the additional degrees of freedom of the channel and to mitigate the increased inter-cell interference. Second, we observe a saturation of \(\tau^*\) with increasing \(B\). This is due to the effective SNR \(\rho(\tau)\) (46)
which saturates with growing \( B \) for any \( \alpha < 1 \). Thus, above some threshold (dependent on \( \alpha \)), increasing \( B \) does not impact the effective SNR substantially and \( \tau^\ast \) remains constant.

C. The Optimal Number of Cooperative BSs

From a system architecture point of view, it is interesting to ask for the optimal number of cooperating BSs for a given coherence block and path loss profile. In Fig. 7, we show the optimal number of cooperative BSs \( B^\ast \) maximizing the ergodic achievable rate \( R_{\text{net}}(\tau^\ast, B) \) as a function of the path loss factor \( \alpha \). We consider different coherence blocks \( T = \{100, 500, 1000\} \). We see that the largest cooperation gains can be achieved when the BSs have strong links to all UTs and \( T \) is large. For a coherence block of \( T = 1000 \) channel uses and a path loss factor \( \alpha = 0.8 \), \( B^\ast = 19 \) cooperative BSs are optimal. If the path loss drops to \( \alpha = 0.2 \), this number is reduced to 5. When we also decrease the coherence block length to \( T = 100 \), only \( B^\ast = 3 \) BSs should cooperate to maximize the net ergodic achievable rate.

D. The Impact of Bandwidth

Finally, we study the impact of the number of sub-channels \( L \), or equivalently the bandwidth, on the system performance. It is well known [41] that if the channels to all UTs would be perfectly known at the CS without any cost, an infinite amount of bandwidth would maximize the per-cell rates. However, since each channel needs to be estimated, increasing the bandwidth implies that less power per sub-channel is
available for channel training under a fixed total power constraint. Thus, as \( L \) grows, the net rate tends to zero since the channels cannot be learned anymore. There exists consequently a non-trivial tradeoff between \( L \) and the resulting net throughput. Fig. 8 shows the net ergodic achievable rate \( R_{\text{net}}(\tau^*, B) \) as a function of the number of sub-channels \( L \) for a setup with \( B = \{4, 5, 6\} \) BSs, \( P = 0 \) dB and no backhaul link failure. For small \( L \), increasing the bandwidth achieves significant performance gains since the system operates in the bandwidth-limited regime [41]. However, from a particular point on (depending on \( \alpha \) and \( B \)), increasing the bandwidth further leads to a decrease in rate.

VI. DISCUSSION AND CONCLUSION

In this work, we have considered a frequency-selective fading, network MIMO uplink channel with arbitrary path losses between the UTs and BSs and randomly failing backhaul links. Using a close approximation of the net ergodic achievable rate based on random matrix theory, we have studied the optimal tradeoff between the resources used for channel training and data transmission. Our asymptotic results yield very tight approximations even for small system dimensions. For an extended version of the circular Wyner model where the UTs are assumed to communicate with all BSs in an isolated, fully cooperative, multi-cell system, we determined the optimal number of BSs and the optimal amount of bandwidth used.

Our results confirm the intuition that we gain a lot from the cooperation of a few, neighboring BSs but
may even loose if we seek to extend the cooperation to far distant BSs. Our main finding is that network MIMO is beneficial when all BSs have equally strong links to all UTs ($\alpha \gg 0$) and the received SNR is low. Since this corresponds precisely to the situation of cell edge users, we believe that approaches based on serving only a subset of UTs from multiple BSs (e.g. [30]) seem most reasonable. In summary, our results shed light onto more realistic dimensions of network MIMO systems which are also important for upcoming small-cell systems with potentially more cooperative devices. We wish to conclude the paper by pointing out some shortcomings of our system model which remain as future investigations.

1) **Channel model:** Simple Wyner-type models which account for cooperation between the BSs in adjacent cells are well suited to study classical cellular architectures. However, for future small-cell networks with a larger number of potentially cooperating BSs, there is a need for new tractable channel models still capturing the essence of these systems, possibly including line-of-sight conditions.

2) **Backhaul links and Cooperation:** The correct modeling of backhaul capacity constraints and imperfections, such as quantization errors and delay, which constitute one of the main limiting factors of network MIMO, remains an open problem. Another relevant question is how a BS should decide whether to cooperate by forwarding its received data to some central processor or to process its received signals alone. In our model, the net throughput vanishes with an increasing backhaul link failure probability although each BSs could theoretically decode a part of the received messages alone\(^6\). Future work, also motivated by the recent results in [32], [42], comprises the investigation of flexible schemes which adapt the degree of cooperation according to some statistical side-information about the channels, backhaul limitations, quality of CSI, etc.

3) **Inter-cluster interference:** We have considered a multi-cell network composed of $B$ cooperative cells. Thus, our results on the optimal number of cooperating base stations do not account for the fact that no cooperation leads in practice to increased interference. Also the effects of non-orthogonal training sequences leading to “pilot contamination” [33], [23] constitute an important issue for practical system design. This needs to be also taken into account for a more realistic performance evaluation of network MIMO systems.

**ACKNOWLEDGMENTS**

This work is partially supported by the European Commission through the FP7 project WiMAGIC and by the French cluster System@tic through the project POSEIDON.

\(^6\)Assuming the availability of some codebook information.
REFERENCES


