

# ON OPTIMAL CHANNEL TRAINING FOR UPLINK NETWORK MIMO SYSTEMS

Jakob Hoydis, Mari Kobayashi

Dept. of Telecommunications, Supélec  
Gif-sur-Yvette, France

Mérouane Debbah

Alcatel-Lucent Chair on Flexible Radio, Supélec  
Gif-sur-Yvette, France

## ABSTRACT

We study a multi-cell frequency-selective fading uplink channel from  $K$  user terminals (UTs) to  $B$  base stations (BSs). The BSs, assumed to be oblivious of the applied encoding scheme, compress and forward their observations to a central station (CS) via capacity limited backhaul links. The CS jointly decodes the messages from all UTs. Since we assume no prior channel state information, the channel needs to be estimated during its coherence time. Based on a lower bound of the ergodic mutual information, we determine the optimal fraction of the coherence time used for channel training. We then study how the optimal training length is impacted by the backhaul capacity. Our analysis is based on large random matrix theory but shown by simulations to be tight for even small system dimensions.

**Index Terms**— Coordinated Multi-Point (CoMP), network MIMO, channel estimation, random matrix theory

## 1. INTRODUCTION

Network MIMO or multi-cell processing are seen as promising techniques to further increase the interference limited performance of today's cellular networks. In essence, a network MIMO system consists of multiple base stations (BSs) which are connected via high speed backhaul links to a central station (CS) which jointly processes their respective data. However, there are two fundamental limitations (among others) related to multi-cell processing: limited backhaul capacity and imperfect channel state information. For a detailed overview of this topic we refer to the survey [1] and references therein.

In this paper, we focus on both of these limitations, assuming that: (i) The BSs act as oblivious relays which forward compressed versions of their received signals to the CS via orthogonal error- and delay-free backhaul links, each of fixed capacity  $C$  bits/channel use, (ii) the CS estimates all channels based on pilot tones sent by the UTs, and (iii) the CS jointly processes the received signals from all BSs. We then consider a lower bound of the normalized ergodic mutual information of this channel, called the net ergodic achievable rate  $R_{\text{net}}(\tau)$ . For a given channel coherence time  $T$ , we attempt to find the optimal length  $\tau^*$  of the pilot sequences which maximizes  $R_{\text{net}}(\tau)$ . As this optimization problem is in

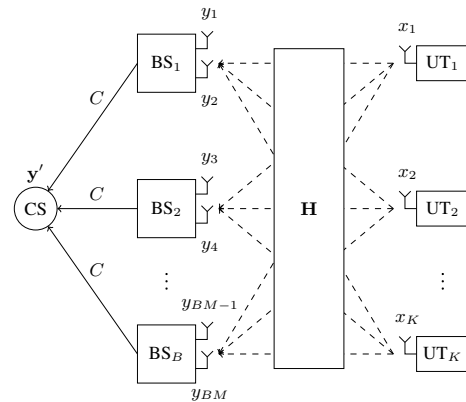


Fig. 1. System model for  $M = 2$  antennas per BS.

general intractable, we study a deterministic approximation  $\bar{R}_{\text{net}}(\tau)$  of  $R_{\text{net}}(\tau)$ , based on large random matrix theory. We then show that optimizing  $\bar{R}_{\text{net}}(\tau)$  instead of  $R_{\text{net}}(\tau)$  is asymptotically optimal but provides accurate approximations for even small system dimensions. In addition, we study the impact of the backhaul capacity on the optimal channel training length. Since we assume that the CS estimates all channels based on the compressed observations from the BSs, the estimates are impaired by thermal noise and quantization errors. As a result, increasing the backhaul capacity leads to improved channel estimates and, hence, smaller values of  $\tau^*$ .

## 2. SYSTEM MODEL

We consider a multi-cell frequency-selective fading uplink channel from  $K$  single-antenna UTs to  $B$  BSs with  $M$  antennas each, as schematically shown in Fig. 1. We assume an orthogonal frequency-division multiplexing (OFDM) transmission scheme over  $L$  sub-carriers. The stacked receive vector  $\mathbf{y}(\ell) = [y_1(\ell), \dots, y_{BM}(\ell)]^T \in \mathbb{C}^{BM}$  of all BSs on the  $\ell$ th sub-carrier at a given time reads

$$\mathbf{y}(\ell) = \mathbf{H}(\ell)\mathbf{x}(\ell) + \mathbf{n}(\ell), \quad \ell = 1, \dots, L \quad (1)$$

where  $\mathbf{x}(\ell) = [x_1(\ell), \dots, x_K(\ell)]^T \in \mathbb{C}^K$  is the vector of the transmitted signals of all UTs on sub-carrier  $\ell$ ,  $\mathbf{n}(\ell) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{BM})$  is a noise vector whose elements are

independent centered circular symmetric complex Gaussian random variables with unit variance, and  $\mathbf{H}(\ell) \in \mathbb{C}^{BM \times K}$  is the aggregated channel matrix from all UTs to all BSs on the  $\ell$ th sub-carrier. We consider a discrete-time block-fading channel model where the channel remains constant for a coherence block of  $T$  channel uses and then changes randomly from one block to the other. The elements of the matrix  $\mathbf{H}(\ell) = \{h_{ij}(\ell)\}$  are assumed to be independent and modeled as  $h_{ij}(\ell) \sim \mathcal{CN}(0, v_{ij})$ . The matrix  $\mathbf{V} = \{v_{ij}\}$  models the different path losses between the UTs and the BS-antennas and will be referred to as the *variance profile* of the channel matrix  $\mathbf{H}(\ell)$ . Since we assume no CSI at the UTs, we assume that each UT  $k$  sends its data by the  $L$  Gaussian inputs  $x_k(\ell) \sim \mathcal{CN}(0, P/L)$ , i.i.d. over  $\ell$  and  $k$ .

### 2.1. Compression at the BSs

The BSs are assumed to be oblivious to the applied encoding scheme of the UTs and forward compressed versions  $y'_i(\ell)$  of their received signal sequences  $y_i(\ell)$  to the CS via delay- and error-free backhaul links, each of capacity  $C$  bits/channel use. Since we assume that the BSs and the CS have no prior knowledge of the instantaneous channel realizations, we consider a simple, sub-optimal compression scheme which neither exploits correlations between the received signals at different antennas nor adapts the employed quantization codebook to the actual channel realization. Thus, a single quantization codebook for the compression of each sequence  $y_i(\ell)$  is used. Assuming further that each BS uses  $C/(ML)$  bits for the compression of each received complex symbol per antenna per sub-carrier, the quantization noise variance  $\sigma_i^2(\ell)$  can be upper-bounded by [2]

$$\sigma_i^2 = \sigma_i^2(\ell) = \frac{1 + \frac{P}{L} \sum_{j=1}^K v_{ij}}{2^{\frac{C}{ML}} - 1} \quad \forall \ell. \quad (2)$$

### 2.2. Channel Training

Similar to [3], 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  $K$  UTs broadcast orthogonal sequences of pilot tones of equal power  $P/L$  on all sub-carriers. The orthogonality of the training sequences imposes  $\tau \geq K$ . We assume that the CS computes the minimum mean square error (MMSE) estimate of the channels  $h_{ij}(\ell)$  from all UTs to all BSs. This allows us to decompose the channel  $h_{ij}(\ell)$  into the estimate  $\hat{h}_{ij}(\ell)$  and the independent estimation error  $\tilde{h}_{ij}(\ell)$ , such that

$$h_{ij}(\ell) = \hat{h}_{ij}(\ell) + \tilde{h}_{ij}(\ell). \quad (3)$$

The variance of the estimated channel  $\hat{v}_{ij}(\tau)$  and the estimation error  $\tilde{v}_{ij}(\tau)$  are independent of  $\ell$  and given by [2]

$$\hat{v}_{ij}(\tau) = \frac{\tau \frac{P}{L} v_{ij}^2}{\tau \frac{P}{L} v_{ij} + 1 + \sigma_i^2}, \quad \tilde{v}_{ij}(\tau) = \frac{v_{ij}(1 + \sigma_i^2)}{\tau \frac{P}{L} v_{ij} + 1 + \sigma_i^2}.$$

### 2.3. Data Transmission

In each channel coherence block, the UTs broadcast their data simultaneously during  $T - \tau$  channel uses. The CS jointly decodes the messages from all UTs, leveraging the previously computed channel estimate  $\hat{\mathbf{H}}(\ell)$ . With the knowledge of  $\hat{\mathbf{H}}(\ell)$ , the CS “sees” in its received signal  $\mathbf{y}'(\ell) = [y'_1(\ell), \dots, y'_{BM}(\ell)]^T$  the useful term  $\hat{\mathbf{H}}(\ell)\mathbf{x}(\ell)$  and the overall noise term  $\mathbf{z}(\ell) = \tilde{\mathbf{H}}(\ell)\mathbf{x}(\ell) + \mathbf{n}(\ell) + \mathbf{q}(\ell)$ , i.e.,

$$\mathbf{y}'(\ell) = \hat{\mathbf{H}}(\ell)\mathbf{x}(\ell) + \mathbf{z}(\ell) \quad (4)$$

where the quantization noise vector  $\mathbf{q}(\ell)$  has mutually independent elements  $q_i(\ell) \sim \mathcal{CN}(0, \sigma_i^2)$ ,  $i = 1, \dots, BM$ . Since the statistical distributions of all sub-carriers, signals and noise are i.i.d. with respect to the index  $\ell$ , we will hereafter omit the dependence on  $\ell$  and consider a single isolated sub-carrier.

## 3. NET ERGODIC ACHIEVABLE RATE

The capacity of the channel (4) is not explicitly known. We consider therefore an achievable lower bound of the normalized ergodic mutual information  $\frac{1}{BM} I(\mathbf{y}'; \mathbf{x} | \hat{\mathbf{H}})$ , referred to hereafter as the *ergodic achievable rate*  $R(\tau)$ . This lower bound is in essence obtained by overestimating the detrimental effect of the estimation error, treating the total noise term  $\mathbf{z}$  as independent complex Gaussian noise with covariance matrix  $\mathbf{K}_z(\tau) = \mathbb{E}[\mathbf{z}\mathbf{z}^H] \in \mathbb{R}_+^{BM \times BM}$ , given as

$$\mathbf{K}_z(\tau) = \text{diag}\left(1 + \sigma_i^2 + \frac{P}{L} \sum_{j=1}^K \tilde{v}_{ij}(\tau)\right)_{i=1}^{BM}. \quad (5)$$

Thus, the ergodic achievable rate can be written as [4, 3]

$$R(\tau) = \frac{1}{BM} \mathbb{E}_{\hat{\mathbf{H}}} \left[ \log \left| \mathbf{I}_{BM} + \frac{P}{L} \bar{\mathbf{H}}(\tau) \bar{\mathbf{H}}(\tau)^H \right| \right] \quad (6)$$

where  $\bar{\mathbf{H}}(\tau) = \mathbf{K}_z^{-\frac{1}{2}}(\tau) \hat{\mathbf{H}}$  is the effective channel matrix. Taking into account that only a fraction  $(1 - \tau/T)$  of the total coherence block length can be used for data transmission, we wish to find the optimal training length  $\tau^*$ , maximizing the *net ergodic achievable rate*

$$R_{\text{net}}(\tau) \triangleq \left(1 - \frac{\tau}{T}\right) R(\tau). \quad (7)$$

Since an explicit expression of  $R(\tau)$  for finite dimensions of the channel matrix  $\bar{\mathbf{H}}(\tau)$  seems intractable, we resort to an approximation based on the theory of large random matrices.

### 3.1. Deterministic Equivalent

In this section, we present a deterministic equivalent approximation  $\bar{R}(\tau)$  of  $R(\tau)$  in the large system limit, i.e., for  $K, BM, L \rightarrow \infty$  at the same speed. Denote  $N = BM$  the

product of the number of BSs and the number of antennas per BS. The notation  $K \rightarrow \infty$  will refer in the sequel to the following two conditions on  $K$ ,  $N$  and  $L$ :

$$0 < \liminf_{K \rightarrow \infty} \frac{N}{K} \leq \limsup_{K \rightarrow \infty} \frac{N}{K} < \infty \quad (8)$$

$$0 < \liminf_{K \rightarrow \infty} \frac{L}{K} \leq \limsup_{K \rightarrow \infty} \frac{L}{K} < \infty. \quad (9)$$

Define  $\bar{\mathbf{V}}(\tau) = \mathbf{K}_z^{-1}(\tau) \hat{\mathbf{V}}(\tau)$  the variance profile of the effective channel  $\bar{\mathbf{H}}(\tau)$  with elements

$$\bar{v}_{ij}(\tau) = \frac{\hat{v}_{ij}(\tau)}{1 + \sigma_i^2 + \frac{P}{L} \sum_{\ell=1}^K \tilde{v}_{i\ell}(\tau)} \quad (10)$$

and consider the following  $N \times N$  matrices

$$\mathbf{D}_j(\tau) = \text{diag}(\bar{v}_{1j}(\tau), \dots, \bar{v}_{Nj}(\tau)), \quad j = 1, \dots, K. \quad (11)$$

We are now in position to state the deterministic approximation  $\bar{R}(\tau)$  of  $R(\tau)$  based on a direct application of [5, Theorem 2.3] to our channel model.

**Theorem 1 (Deterministic Equivalent)** *Assume that  $K$ ,  $N$  and  $L$  satisfy (8)–(9) and  $0 \leq \bar{v}_{ij}(\tau) < v_{\max} < \infty \forall i, j$ . Then:*

(i) *The following implicit equation:*

$$\mathbf{T}(z) = \left( \frac{1}{K} \sum_{j=1}^K \frac{\mathbf{D}_j(\tau)}{1 + \frac{1}{K} \text{tr} \mathbf{D}_j(\tau) \mathbf{T}(z)} - z \mathbf{I}_N \right)^{-1}$$

*admits a unique solution  $\mathbf{T}(z) = \text{diag}(t_1(z), \dots, t_N(z))$  such that the  $t_i(z)$  are Stieltjes transforms of probability measures over  $\mathbb{R}_+$  (see e.g. [5, Proposition 2.1]).*

(ii) *Let  $P > 0$ . Denote  $\mathbf{T}_P = \mathbf{T}(-\frac{L}{KP})$ , define  $\delta_j = \frac{1}{K} \text{tr} \mathbf{D}_j(\tau) \mathbf{T}_P$ ,  $j = 1, \dots, K$ , and let*

$$\begin{aligned} \bar{R}(\tau) &= \frac{1}{N} \sum_{j=1}^K \log(1 + \delta_j) - \frac{1}{N} \sum_{j=1}^K \frac{\delta_j}{1 + \delta_j} \\ &\quad - \frac{1}{N} \log \det \left( \frac{L}{KP} \mathbf{T}_P \right). \end{aligned}$$

*Then, the following holds true:*

$$R(\tau) - \bar{R}(\tau) \xrightarrow{K \rightarrow \infty} 0.$$

### 3.2. Optimization of the training length $\tau$

In order to find the optimal training length  $\tau^*$  for a given coherence block length  $T$ , we wish to maximize  $R_{\text{net}}(\tau)$  as given by (7). As this optimization problem is intractable for finite channel dimensions, we pursue the following approach:

1. We find  $\bar{\tau}^*$  maximizing the deterministic approximation  $\bar{R}_{\text{net}}(\tau) = (1 - \frac{\tau}{T}) \bar{R}(\tau)$ .

2. We show that  $R_{\text{net}}(\tau^*) - \bar{R}_{\text{net}}(\bar{\tau}^*) \xrightarrow{K \rightarrow \infty} 0$  and  $\tau^* - \bar{\tau}^* \xrightarrow{K \rightarrow \infty} 0$ .

3. We verify by simulations that  $\bar{\tau}^*$  is very close to  $\tau^*$  for even small values of  $K$ ,  $N$  and  $L$ .

In the following, Theorem 2 provides an explicit expression of the derivative of  $\bar{R}(\tau)$  while Theorem 3 establishes concavity of  $\bar{R}_{\text{net}}(\tau)$  for matrices with a doubly-regular variance profile. Theorem 4 shows that optimizing  $\bar{R}_{\text{net}}(\tau)$  instead of  $R_{\text{net}}(\tau)$  is asymptotically optimal and provides a simple way to compute  $\bar{\tau}^*$ . All proofs can be found in [2].

**Theorem 2 (Derivative)** *Under the same conditions as for Theorem 1, the first derivative of  $\bar{R}(\tau)$  permits the explicit expression*

$$\bar{R}'(\tau) = \frac{1}{N} \sum_{j=1}^K \frac{\frac{1}{K} \text{tr} \mathbf{D}'_j(\tau) \mathbf{T}_P}{1 + \frac{1}{K} \text{tr} \mathbf{D}_j(\tau) \mathbf{T}_P} \quad (12)$$

*where  $\mathbf{T}_P = \mathbf{T}(-\frac{L}{KP})$  is given by Theorem 1 (i) and  $\mathbf{D}'_j(\tau) = \frac{d\mathbf{D}_j(\tau)}{d\tau}$ ,  $j = 1, \dots, K$ . Moreover, for any  $P, \tau > 0$ ,  $\bar{R}(\tau)$  is strictly increasing, i.e.,  $\bar{R}'(\tau) > 0$ .*

**Theorem 3 (Concavity)** *Let  $P, \tau > 0$ . Assume that  $N = K$  and that  $\bar{\mathbf{V}}(\tau)$  is a doubly regular matrix which satisfies the following regularity condition:*

$$\mathcal{K}(\tau) = \frac{1}{N} \sum_{i=1}^N \bar{v}_{ik}(\tau) = \frac{1}{N} \sum_{j=1}^N \bar{v}_{\ell j}(\tau) \quad \forall k, \ell. \quad (13)$$

*Then,  $\bar{R}_{\text{net}}(\tau)$  is a strictly concave function, i.e.,  $\bar{R}_{\text{net}}''(\tau) < 0$ .*

**Remark 3.1** *We conjecture that Theorem 3 also holds for non doubly regular variance profiles  $\bar{\mathbf{V}}(\tau)$ .*

**Theorem 4 (Convergence)** *Let  $\tau^* = \arg \max_{\tau \in [0, T]} R_{\text{net}}(\tau)$  and  $\bar{\tau}^* = \arg \max_{\tau \in [0, T]} \bar{R}_{\text{net}}(\tau)$ . Then, under the same conditions as for Theorem 1, the following holds true:*

(i)

$$R_{\text{net}}(\tau^*) - \bar{R}_{\text{net}}(\bar{\tau}^*) \xrightarrow{K \rightarrow \infty} 0. \quad (14)$$

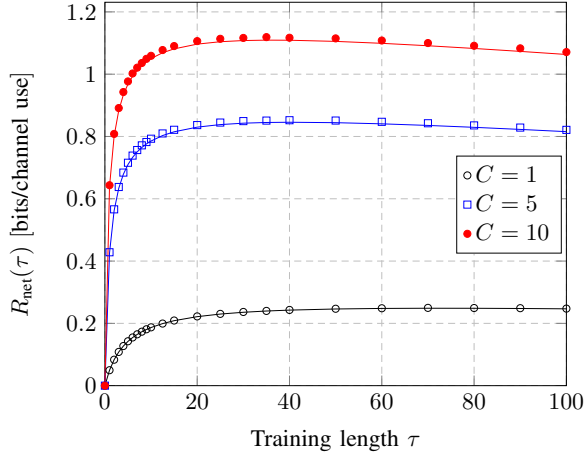
(ii) *Further assume that  $\bar{\mathbf{V}}(\tau)$  is a doubly regular matrix which satisfies the conditions of Theorem 3. Then,*

$$\tau^* - \bar{\tau}^* \xrightarrow{K \rightarrow \infty} 0 \quad (15)$$

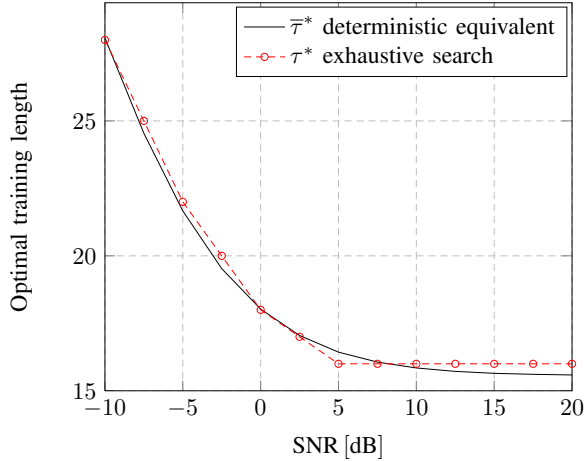
*where  $\bar{\tau}^*$  is given as the solution to*

$$\bar{R}'_{\text{net}}(\tau) = \left(1 - \frac{\tau}{T}\right) \bar{R}'(\tau) - \frac{1}{T} \bar{R}(\tau) = 0 \quad (16)$$

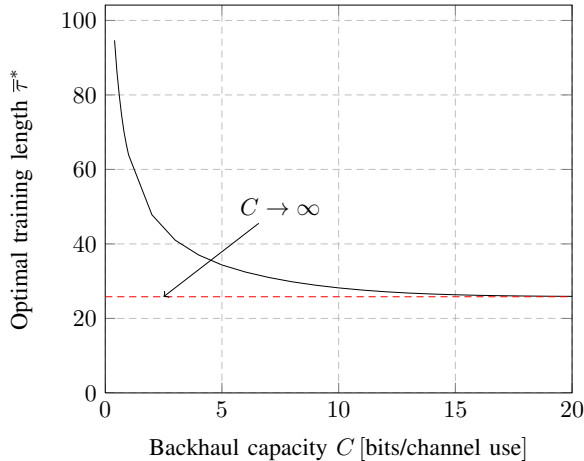
*with  $\bar{R}(\tau)$  and  $\bar{R}'(\tau)$  given by Theorem 1 (ii) and Theorem 2, respectively.*



**Fig. 2.**  $R_{\text{net}}(\tau)$  vs  $\tau$  for SNR = 0 dB and  $T = 1000$ . Markers correspond to  $R_{\text{net}}(\tau)$  as obtained by simulations, the solid line corresponds to the deterministic approximation  $\bar{R}_{\text{net}}(\tau)$ .



**Fig. 3.** Optimal training length  $\tau^*$  and  $\bar{\tau}^*$  vs SNR for  $C = 1$  bits/channel use and  $T = 100$ .



**Fig. 4.** Optimal training length  $\bar{\tau}^*$  vs backhaul capacity  $C$  for SNR = 10 dB and  $T = 1000$ .

#### 4. NUMERICAL RESULTS

In order to validate the analysis in the preceding sections, we assume  $B = K = 3$ ,  $M = 2$  and consider a randomly chosen variance profile  $\mathbf{V}$  which is not doubly-regular. We then assume  $\mathbf{V}$  fixed and average over many independent realizations of the channel matrix  $\mathbf{H}$ . The transmit signal-to-noise-ratio is defined as  $\text{SNR} = \mathbb{E}[|x_i(\ell)|^2] / \mathbb{E}[|n_i(\ell)|^2] = P/L$ . Unless otherwise stated, we assume  $T = 1000$  and  $L = 1$ .

In Fig. 2, we show  $R_{\text{net}}(\tau)$  and  $\bar{R}_{\text{net}}(\tau)$  as a function of the training length  $\tau$  for different values of  $C = \{1, 5, 10\}$ . This plot validates Theorem 3 and the conjecture in Remark 3.1 as  $\bar{R}_{\text{net}}(\tau)$  is obviously a concave function. Since the curves of  $\bar{R}_{\text{net}}(\tau)$  and  $R_{\text{net}}(\tau)$  match closely, a good approximation of  $\tau^*$  can be found by maximizing  $\bar{R}_{\text{net}}(\tau)$  instead of  $R_{\text{net}}(\tau)$ .

The validity of Theorem 4 is demonstrated in Fig. 3 which shows the optimal training length  $\tau^*$ , found by an exhaustive search based on Monte Carlo simulations, and the training length  $\bar{\tau}^*$  which maximizes  $\bar{R}_{\text{net}}(\tau)$  as a function of the SNR for  $C = 1$  bits/channel use and  $T = 100$ . The differences between both values are mainly due to the exhaustive search over a necessarily discrete set of values of  $\tau$ .

Fig. 4 shows the dependence of the optimal training length  $\bar{\tau}^*$  on the backhaul capacity  $C$  for a fixed SNR = 10 dB. One can see that  $\bar{\tau}^*$  is a decreasing function of  $C$  which converges quickly to particular value corresponding to infinite capacity backhaul links. The reason for this is the following. The CS estimates the channel coefficients based on the quantized training signals received by the BSs. The channel estimate is hence impaired by thermal noise *and* quantization errors. Therefore, increasing  $C$  results in better channel estimates and reduces the necessary training length.

#### 5. REFERENCES

- [1] D. Gesbert, S. V. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO Cooperative Networks: A New Look at Interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, Dec. 2010.
- [2] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal Channel Training in Uplink Network MIMO Systems," *IEEE Trans. Signal Process.*, 2010, submitted, available upon request.
- [3] P. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [4] T. Yoo and A. Goldsmith, "Capacity and Power Allocation for Fading MIMO Channels With Channel Estimation Error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [5] W. Hachem, P. Loubaton, and J. Najim, "A CLT for Information-Theoretic Statistics of Gram Random Matrices with a Given Variance Profile," *Annals of Applied Probability*, vol. 18, no. 6, pp. 2071–2130, 2008.