Channel Training Procedures for MIMO Interfering Point-to-Multipoint Channel

Adrian Agustin, Sandra Lagen, and Josep Vidal
Dept. Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
email: {adrian.agustin, sandra.lagen, josep.vidal}@upc.edu

***Abstract* — A precise knowledge of the MIMO channel between the serving node and the user equipment (UE) is important for attaining good data rates in downlink transmissions (DL) in cellular systems. The interfering point-to-multipoint (I-P2MP) channel, consisting of multiple transmitters coexisting in the same area, where each transmitter is intended to serve multiple users, is a model that subsumes most of the scenarios that can be found in wireless cellular networks with a dense deployment of small cells (SCs). In conventional channel estimation procedures, resources allocated to each SC for *training* tend to be orthogonal, negatively impacting in the efficiency of the whole system. Reusing resources for *training* allows releasing resources for data transmission, but at the cost of degrading the channel estimation due to interference. We propose a decentralized algorithm for interference management that enforces the coordination among SCs in the design of the training sequences for de DL. Results in this work elucidate how to reuse resources for training and significantly improve the throughput of the system.**

Keywords—Interfering point-to-multipoint; channel training; channel estimation; training sequences design.

I. Introduction

Heterogeneous networks (HetNets) are one of the solutions envisioned in current wireless networks for combating the large traffic demand coming from data hungry smartphones. Although this concept subsumes a mixture of radio technologies and cell types, here we consider just small cells (SCs) where the frequency reuse is equal to one. It turns out that this new scenario tends to be interference-dominated, which motivates the derivation of techniques for efficiently managing the created interference. Nevertheless, most of such work is focused in the data transmission, assuming that channel estimation is performed using a well-established interference-free orthogonal training-based scheme.

Pilot-assisted transmission simplifies the task at the receiver side, providing a method for acquiring the channel state information (CSI) by means of the *training*, [1]. However, since training sequences do not convey any information, the system efficiency is inversely dependent on how many resources are devoted for channel training.

In terms of point-to-point (P2P) MIMO communications, a detailed study about different linear filter estimators is provided in [2], and the optimal training sequences are derived based on the minimum mean square error (MMSE) criterion. Alternatively, a criterion based on channel differential entropy is considered in [3], taking into account the channel information gained thanks to training. Finally, a framework for deriving the training sequence is proposed in [4], based on the MMSE and a final performance metric of interest. On the other hand, the system efficiency is investigated in [5], showing how training impacts in the obtained data transmission rate, and how many resources should be reserved.

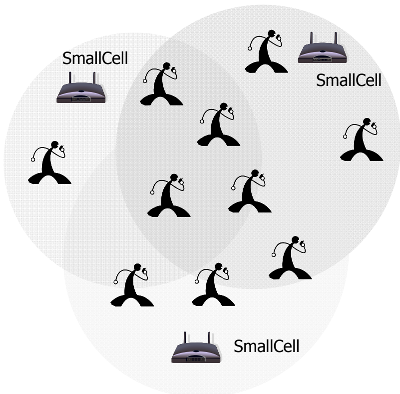

Fig. 1. Interfering P2MP MIMO deployment with 3 SCs and multiple UEs per SC.

We investigate the interfering point-to-multipoint (I-P2MP) MIMO communication system depicted in Figure 1, where each SC serves multiple UEs. Two types of impairment have been observed associated to the training sequences: *pilot contamination* and *pilot pollution*. The concept of *pilot contamination* is employed for illustrating the effect of reusing training sequences by multiple transmitters. In such a case, the measurement is disturbed by those transmitters with the same training sequence. This impairment is being considered in the massive MIMO literature for transmissions from UE to SCs, see for example [6]. On the other hand, the concept of *pilot pollution* occurs when there are many strong cells in a given area.

This work has been done in the framework of the project TROPIC FP7 ICT-2011-8-318784 funded by the European Commission. Also, by the Spanish Science and Technology Commissions funds through projects TEC2010-19171/TCM, DISNET (TEC2013-41315-R) and CONSOLIDER INGENIO CSD2008-00010 COMONSENS, and the 2014 SGR 60 AGAUR of the Catalan Government.

In contrast to the conventional approach where pilot transmission is carried out in orthogonal resources, we investigate the reuse of resources in an I-P2MP MIMO scenario with non-collocated transmitters. For designing the training sequences in such an interfering scenario, we take into account the concept of *interference cost* introduced in [7], [8] for managing the interference. However, by reducing the resources for training we are increasing the resources for data transmission, but at the cost of having a worse channel estimation. Hence, we cannot anticipate under which conditions we can improve the efficiency of the system. On the other hand, current cellular networks consider on-off SCs, so that the design of training resources should be adjusted dynamically as a function of the current topology of the network, i.e. active neighboring SCs in the given area. The contributions of this work are the following:

- A channel training procedure is provided for point-to-multipoint (P2MP) communications, i.e. each SC serving multiples UEs.
- A decentralized and scalable training sequence design is presented for I-P2MP communications, i.e. multiple SCs with multiple UEs each. Coordination among SCs is considered for improving the training sequence design, adapting it to the current interfering scenario.
- The efficiency of the system with the proposed design is analyzed with a simple data transmission phase, showing significant gains in terms of throughput.

II. SIGNAL MODEL

Consider a flat block-fading MIMO system where there are N_{cells} small cells (SCs) equipped with M antennas. The i-th SC has N_{UE}^{i} associated user equipments (UEs), each with N receiving antennas. We assume that each UE is able to estimate long-term channel state information (CSI) like the pathloss with its neighboring SCs and the channel correlation matrix.

The transmission is carried out in frames of length T symbols, L symbols are devoted for training while T-L symbols are used for data transmission. For simplicity, we assume that in the data transmission phase orthogonal resources are devoted to each SC, hence avoiding inter-cell interference. Likewise, UEs associated to a given SC are served under a TDMA (Time Division Multiple Access) scheme. Every UE only estimates the channel with its serving cell.

A. Training phase

The received signal at the k-th UE associated to the i-th SC, i.e. the (k,i)-th UE in the following, is given by,

$$\mathbf{S}_{(k,i)} = \sqrt{\rho_{(k,i),i}}\mathbf{H}_{(k,i),i}\mathbf{P}_i + \sum_{j\neq i}^{N_{cells}} \sqrt{\rho_{(k,i),j}}\mathbf{H}_{(k,i),j}\mathbf{P}_j + \mathbf{V}_{(k,i)} \quad (1)$$

where $\mathbf{H}_{(k,i),j} \in \mathbb{C}^{N\times M}$ is the channel matrix between the j-th SC and the (k,i)-th UE accounting for the fast fading effect, $\rho_{(k,i),i}$ subsumes the pathloss and shadowing impairments and it is known a priori by UE, matrix $\mathbf{P}_j \in \mathbb{C}^{M\times L}$ contains the training sequence transmitted through M antennas during L symbols and $\mathbf{V}_{(k,i)}$ is the AWGN noise of power $\sigma^2_{v(k,i)}$.

The normalized noise plus interference correlation matrix at the (k,i)-th UE is defined by,

$$\mathbf{N}_{(k,i)} = \frac{1}{\rho_{(k,i),i}}\left(\sum_{j\neq i}^{N_{cells}} \rho_{(k,i),j}\mathbf{P}_j^H\mathbf{\Theta}_{(k,i),j}\mathbf{P}_j + \sigma^2_{v(k,i)}N\mathbf{I}\right) \quad (2)$$

where $\mathbf{\Theta}_{(k,i),j} \in \mathbb{C}^{M\times M}$ is the channel correlation matrix observed by the (k,i)-th UE with respect to the j-th SC,

$$\mathbf{\Theta}_{(k,i),j} = E\left\{\mathbf{H}^H_{(k,i),j}\mathbf{H}_{(k,i),j}\right\} \quad (3)$$

being $E\{.\}$ the expectation operator.

We assume that all UEs are able to estimate the channel correlation matrices with their most harmful neighboring SCs, or it is a parameter given by the system and known by everyone. Notice that this is a long-term measurement.

The (k,i)-th UE only estimates the channel to its respective serving SC by means of a linear receive filter,

$$\hat{\mathbf{H}}_{(k,i),i} = \frac{1}{\sqrt{\rho_{(k,i),i}}}\mathbf{S}_{(k,i)}\mathbf{R}_{(k,i)} \quad (4)$$

where $\mathbf{R}_{(k,i)} \in \mathbb{C}^{L\times M}$ is the linear receive filter estimator. The error in the channel estimation can be written as

$$\begin{aligned}\xi_{(k,i),i} = E\left\{\left\|\mathbf{H}_{(k,i),i} - \hat{\mathbf{H}}_{(k,i),i}\right\|_F^2\right\} = \mathrm{tr}\left(\mathbf{\Theta}_{(k,i),i} - \mathbf{\Theta}_{(k,i),i}\mathbf{P}_i\mathbf{R}_{(k,i)}\right) \\ + \mathrm{tr}\left(\mathbf{R}^H_{(k,i)}\left(\mathbf{P}_i^H\mathbf{\Theta}_{(k,i),i}\mathbf{P}_i + \mathbf{N}_{(k,i)}\right)\mathbf{R}_{(k,i)} - \mathbf{R}^H_{(k,i)}\mathbf{P}_i^H\mathbf{\Theta}_{(k,i),i}\right)\end{aligned} \quad (5)$$

Hence, an adequate design of the training sequence ($\mathbf{P}_i$) per SC and the linear receive filter estimator ($\mathbf{R}_{(k,i)}$) at UEs influences the mean square error (MSE) of the channel estimation.

B. Data transmission phase

The received signal at the (k,i)-th UE during the *data transmission phase* is expressed as,

$$\mathbf{y}_{(k,i)} = \sqrt{\rho_{(k,i),i}}\mathbf{H}_{(k,i),i}\mathbf{x}_{(k,i)} + \mathbf{n}_{(k,i)} \quad (6)$$

where $\mathbf{x}_{(k,i)} \in \mathbb{C}^{M\times 1}$ denotes the stream of symbols to be transmitted towards the (k,i)-th UE assuming that the transmitter has not short-term channel state information at the transmitter (CSIT), and $\mathbf{n}_{(k,i)} \in \mathbb{C}^{N\times 1}$ stands for the AWGN noise of power $\sigma^2_{n(k,i)}$.

The resources employed by each SC are equally and orthogonally distributed among its associated UEs. In this regard, the average mutual information for the (k,i)-th UE is lower bounded by [5],

$$\begin{aligned}R_{(k,i)} = \mu E\left\{\log_2\det\left(\mathbf{I}_N + \rho_{(k,i),i}\frac{\hat{\mathbf{H}}_{(k,i),i}\hat{\mathbf{H}}^H_{(k,i),i}}{\sigma^2_{n(k,i)} + \xi_{(k,i),i}\rho_{(k,i),i}}\frac{P_d}{M}\right)\right\} \\ \mu = \frac{1}{N_{cells}}\frac{1}{N_{UE}^i}\frac{T-L}{T}\end{aligned} \quad (7)$$

where T is the total number of symbols in the frame, L are the symbols considered for the *training phase*, $\hat{\mathbf{H}}_{(k,i),i} \in \mathbb{C}^{N\times M}$ is the estimated channel according to (4), P_d is the power devoted for the data transmission, and finally, the value of $\xi_{(k,i),i}\rho_{(k,i),i}$ is the increase of the Gaussian noise due to the channel estimation errors that depend on the MSE in (4) and the path-loss of the (*k*,*i*)-th UE.

The average throughput served by the *i*-th SC is

$$SR_i = \sum_{k=1}^{N_{UE}^i} R_{(k,i)} \tag{8}$$

III. Training Sequence Optimization

Since the quality of the estimated channel at the receiver side influences the obtained rate performance, see (7), we have selected the minimization of the average MSE per SC as the criterion to be optimized when designing the training sequences and linear receive filter estimators:

$$\begin{aligned}(\mathrm{P0}): \quad & \underset{\{\mathbf{P}_i\},\{\mathbf{R}_{(k,i)}\}}{\text{minimize}} \quad \sum_{i=1}^{N_{cells}} \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \xi_{(k,i),i} \\ & s.t. \quad \mathrm{tr}\left(\mathbf{P}_i \mathbf{P}_i^H\right) \le \frac{P_i L}{M} \qquad \forall i\end{aligned} \tag{9}$$

where P_i is the maximum power that can be employed by the *i*-th SC, and tr(.) denotes the trace operator.

The problem (P0) in (9) is not convex with respect all optimization variables, so it might present multiple local optima. Nevertheless, if we keep fixed one set of the optimization variables, then (P0) becomes a convex problem for the remaining set of variables. This property can be employed to solve (P0) with an alternate optimization between training sequences ($\mathbf{P}_i$) and linear receive filter estimators ($\mathbf{R}_{(k,i)}$), attaining at least a local optimum. This procedures have been considered for minimizing the MMSE for data transmission in [9][10].

Accordingly, problem (P0) can be decomposed into two problems: $(\mathrm{P1})_i$ and $(\mathrm{P2})_{(k,i)}$ described below that are solved at each SC and UE, respectively.

A. Optimization at small cells

Assuming the linear receive filter estimators ($\mathbf{R}_{(k,i)}$) are fixed and given, problem (P0) is equivalent to defining the following optimization problem at the *i*-th SC,

$$\begin{aligned}(\mathrm{P1})_i: \quad & \underset{\{\mathbf{P}_i\}}{\text{minimize}} \quad f_i + g_{-i} \\ & s.t. \quad \mathrm{tr}\left(\mathbf{P}_i \mathbf{P}_i^H\right) \le \frac{P_i L}{M}\end{aligned} \tag{10}$$

where function f_i measures the impact of the training sequence $\mathbf{P}_i \in \mathbb{C}^{M \times L}$ on the UEs associated to the *i*-th SC,

$$\begin{aligned}f_i &= \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \mathrm{tr}\left(\mathbf{\Theta}_{(k,i),i} - \mathbf{\Theta}_{(k,i),i}\mathbf{P}_i\mathbf{R}_{(k,i)} - \mathbf{R}_{(k,i)}^H \mathbf{P}_i^H \mathbf{\Theta}_{(k,i),i}\right) + \\ &+ \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \mathrm{tr}\left(\mathbf{R}_{(k,i)}^H \left(\mathbf{P}_i^H \mathbf{\Theta}_{(k,i),i} \mathbf{P}_i + \frac{\sigma_{v(k,i)}^2 N}{\rho_{(k,i),i}}\right)\mathbf{R}_{(k,i)}\right)\end{aligned} \tag{11}$$

while function g_{-i} takes into account the impact of the training sequence of the *i*-th SC on UEs associated to the neighboring SCs,

$$g_{-i} = \sum_{j \ne i}^{N_{cells}} \frac{1}{N_{UE}^j} \sum_{z=1}^{N_{UE}^j} \frac{\rho_{(z,j),i}}{\rho_{(z,j),j}} \mathrm{tr}\left(\mathbf{R}_{(z,j)}^H \left(\mathbf{P}_i^H \mathbf{\Theta}_{(z,j),i} \mathbf{P}_i\right) \mathbf{R}_{(z,j)}\right) \tag{12}$$

With the objective of having a more tractable optimization problem over variable $\mathbf{P}_i$, we take advantage of the properties of the Kronecker product and vectorial operators [11]

$$\mathrm{tr}\left(\mathbf{A}^T \mathbf{X} \mathbf{B} \mathbf{C}^T\right) = \mathrm{vec}\left(\mathbf{A}\right)^T \left(\mathbf{C} \otimes \mathbf{X}\right) \mathrm{vec}\left(\mathbf{B}\right) \tag{13}$$

where $\otimes$ and vec(.) denote the Kronecker product and the vectorial operator, respectively. The vectorial operator stacks the columns of a matrix into a column vector. In this regard, we can define the following equivalent variables,

$$\mathbf{p}_i \triangleq \mathrm{vec}(\mathbf{P}_i) \qquad \mathbf{p}_i^H \triangleq \mathrm{vec}^T(\mathbf{P}_i^*) \tag{14}$$

This way, the optimization problem $(\mathrm{P1})_i$ in (10) is transformed into,

$$\begin{aligned}(\tilde{\mathrm{P}}1)_i: \quad & \underset{\{\mathbf{p}_i\}}{\text{minimize}} \quad \tilde{f}_i + \tilde{g}_{-i} \\ & s.t. \quad \mathbf{p}_i^H \mathbf{p}_i \le \frac{P_i L}{M}\end{aligned} \tag{15}$$

where

$$\begin{aligned}\tilde{f}_i &= \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \mathrm{tr}\left(\mathbf{\Theta}_{(k,i),i}\right) - \boldsymbol{\omega}_i^H \mathbf{p}_i - \mathbf{p}_i^H \boldsymbol{\omega}_i + \mathbf{p}_i^H \mathbf{\Psi}_{i,i} \mathbf{p}_i \\ &+ \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \frac{\sigma_{v(k,i)}^2 N}{\rho_{(k,i),i}} \mathrm{tr}\left(\mathbf{R}_{(k,i)}^H \mathbf{R}_{(k,i)}\right)\end{aligned} \tag{16}$$

$$\tilde{g}_{-i} = \mathbf{p}_i^H \left(\sum_{j \ne i}^{N_{cells}} \mathbf{\Psi}_{j,i}\right) \mathbf{p}_i \tag{17}$$

and

$$\begin{aligned}\boldsymbol{\omega}_i &\triangleq \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \mathrm{vec}\left(\mathbf{\Theta}_{(k,i),i} \mathbf{R}_{(k,i)}^H\right), \\ \boldsymbol{\omega}_i^H &\triangleq \frac{1}{N_{UE}^i} \sum_{k=1}^{N_{UE}^i} \mathrm{vec}^T\left(\mathbf{\Theta}_{(k,i),i}^T \mathbf{R}_{(k,i)}^T\right) \\ \mathbf{\Psi}_{j,i} &= \frac{1}{N_{UE}^j} \sum_{z=1}^{N_{UE}^j} \left(\mathbf{R}_{(z,j)} \mathbf{R}_{(z,j)}^H\right)^T \otimes \frac{\rho_{(z,j),i}}{\rho_{(z,j),j}} \mathbf{\Theta}_{(z,j),i}\end{aligned} \tag{18}$$

The optimization problem presented in (15) allows deriving a semi-closed form solution,

$$\begin{aligned}\mathbf{p}_i(\lambda_i) &= \left(\mathbf{\Psi}_{i,i} + \mathbf{\Omega}_{-i} + \lambda_i \mathbf{I}\right)^{-1} \boldsymbol{\omega}_i \\ \mathbf{\Omega}_{-i} &= \sum_{j \ne i}^{N_{cells}} \mathbf{\Psi}_{j,i}\end{aligned} \tag{19}$$

where λ_i is the Lagrange multiplier associated to the energy constraint at each SC in (15). Notice that $\mathbf{\Psi}_{i,i}$ is a measure generated by UEs served by the *i*-th SC and intended to the *i*-th SC, while $\mathbf{\Omega}_{-i}$ is generated by UEs associated to neighboring SCs and corresponds to the *interference cost*.

Finally, the training sequence is obtained by,

$$\mathbf{P}_i = \mathrm{unvec}\left(\mathbf{p}_i, M, L\right) \tag{20}$$

where the *unvec*(.) operator divides the components of the vector into L groups of M consecutive elements to compose a $M\times L$ matrix.

B. Optimization at UEs

On the other hand, if training sequences ($\mathbf{P}_i$) are assumed fixed, then problem (P0) is equivalent to define the following optimization problem at each UE

$$(\text{P2})_{(k,i)}: \underset{\{\mathbf{R}_{(k,i)}\}}{\text{minimize}} \quad \xi_{(k,i),i} \tag{21}$$

where $\xi_{(k,i),i}$ is the MSE defined in (5). The closed-form solution of the previous optimization problem is given by,

$$\mathbf{R}_{(k,i)} = \left(\mathbf{P}_i^H \boldsymbol{\Theta}_{(k,i),i} \mathbf{P}_i + \mathbf{N}_{(k,i)}\right)^{-1} \mathbf{P}_i^H \boldsymbol{\Theta}_{(k,i),i} \tag{22}$$

with $\mathbf{N}_{(k,i)} \in \mathbb{C}^{L\times L}$ being the noise plus interference covariance matrix observed by the (k,i)-th UE and defined (2).

IV. Decentralized Optimization

The two-steps optimization presented in the previous section can be tackled in a decentralized mode in case the different nodes in the system could coordinate among them. Table I presents the proposed iterative algorithm to carry out the decentralized optimization. Basically, it consists in four phases: *solving problem (P2)*, *reporting parameters*, *exchange of interference cost* and *solving problem (P1)*.

We assume a given initialization, i.e. some predefined and known training sequences at time (n-1), $\mathbf{P}_i^{n-1}$. Furthermore, each UE reports the channel covariance matrix and pathloss with all neighboring SCs ($\boldsymbol{\Theta}_{(k,i),j}$ see (3) and $\rho_{(k,i),j}$). These parameters are assumed to be constant during the optimization. First each UE computes the linear receive filter estimator according to (22), and then, we proceed to design the training sequences using (19) and (20).

For designing the training sequences at the i-th SC we need to know how that SC is interfering to the UEs associated to neighboring SCs, i.e. $\boldsymbol{\Omega}_{-i}$. This is done in two phases: *reporting parameters* and *exchange of interference cost*. In the first phase, every UE reports to its associated SC the observed noise plus interference covariance matrix $\mathbf{N}_{(k,i)}^n$. Now, the i-th SC is able to reproduce the linear receive filter estimator of each UE at the transmitter side, $\mathbf{R}_{(k,i)}^n$, and then obtain the *interference cost* of the j-th SC over the users served by the i-th SC, $\boldsymbol{\Psi}_{i,j}$. This information is forwarded to the j-th SC during the *exchange of interference cost* phase.

In the last phase, every SC has all parameters to calculate the training sequence: $\boldsymbol{\Psi}_{i,i}$, $\boldsymbol{\omega}_i$, $\boldsymbol{\Omega}_{-i}^n$, $\mathbf{N}_{(k,i)}^n$. Nevertheless, the optimal training sequence depends on the linear receive filter estimator and vice versa. In order to obtain improved training sequences we use the alternate optimization concept presented for instance in [10], assuming fixed $\boldsymbol{\Omega}_{-i}^n$ and $\mathbf{N}_{(k,i)}^n$, see lines 14-20 in Table I. Once the new training sequence is designed, we assume that each SC can inform about it to its serving UEs.

The convergence of the proposed decentralized algorithm to a local optimum can be proved following similar arguments as in [9], or [12] for the centralized solution to problem (P0) in (9). Basically, we are optimizing one variable keeping fixed the others, and at each step we monotonically decrease the objective function of the optimization problem, i.e. the sum of average MSEs.

The overhead introduced by the proposed algorithm is expected to be low because the target function is the minimization of the MSE at the long-term governed by the variation of the channel correlation matrix $\boldsymbol{\Theta}_{(k,i),j}$. Hence, the different phases of the algorithm can be performed after a large number of frames. Additionally, it is not required that all UEs either participate in the decentralized optimization or UEs in the both phases will be the same, we need just a representative number UEs because we are minimizing the sum of average MSEs per SC. In this regard, we could select static and idle UEs for designing the training sequences, while the active UEs participate in the data phase.

TABLE I. Decentralized Algorithm

1. Set n=0
2. Set $\mathbf{P}_i = \mathbf{P}_i^{ini} \quad \forall i = 1 \ldots N_{cells}$
3. Set $\boldsymbol{\Omega}_{-i} = \mathbf{0} \quad \forall i = 1 \ldots N_{cells}$
4. Report $\boldsymbol{\Theta}_{(k,i),j}, \rho_{(k,i),j} \quad \forall i,j = 1 \ldots N_{cells}, \quad k = 1 \ldots N_{UE}^i$
5. Iterate
6. Update n=n+1

#*Solving problem* $(\text{P2})_{(k,i)} \quad \forall i = 1 \ldots N_{cells}, \quad k = 1 \ldots N_{UE}^i$

7. UE computes $\mathbf{R}_{(k,i)}^n \;|\; \mathbf{N}_{(k,i)}^{n-1}, \mathbf{P}_i^{n-1}$ using (22)

#*Reporting parameters*

8. UE reports $\mathbf{N}_{(k,i)}^n \quad \forall i,j = 1 \ldots N_{cells}, \quad k = 1 \ldots N_{UE}^i$
9. SC computes $\mathbf{R}_{(k,i)}^n \;|\; \mathbf{N}_{(k,i)}^n, \mathbf{P}_i^{n-1}$ using (22) $\forall i = 1 \ldots N_{cells}, \quad k = 1 \ldots N_{UE}^i$
10. SC computes $\boldsymbol{\Psi}_{j,i}, \boldsymbol{\omega}_i \;|\; \mathbf{R}_{(k,i)}^n$ using (18) $\forall i,j = 1 \ldots N_{cells}, \quad k = 1 \ldots N_{UE}^i$

#*Exchange of interference cost*

11. Exchange of $\boldsymbol{\Psi}_{j,i}$ among all SCs
12. Total interference cost for the i-th SC $\boldsymbol{\Omega}_{-i}^n = \sum_{j\neq i}^{N_{cells}} \boldsymbol{\Psi}_{j,i}$

#*Solving problem* $(\text{P1})_i \quad \forall i = 1 \ldots N_{cells}$

13. SC computes $\mathbf{P}_i^{aux} \;|\; \boldsymbol{\omega}_i, \boldsymbol{\Psi}_{i,i}, \boldsymbol{\Omega}_{-i}$ using (20) $\forall i = 1 \ldots N_{cells}$
14. At i-th SC set $\overline{\mathbf{P}}_i^0 = \mathbf{P}_i^{aux}$. t=0
15. Iterate
16. Update t=t+1
17. SC computes $\overline{\mathbf{R}}_{(k,i)}^t \,|\, \mathbf{N}_{(k,i)}^n, \overline{\mathbf{P}}_i^{t-1}$ using (22) $\forall k = 1 \ldots N_{UE}^i$
18. SC computes $\overline{\boldsymbol{\Psi}}_{i,i}, \overline{\boldsymbol{\omega}}_i \;|\; \overline{\mathbf{R}}_{(k,i)}^t$ using (18)
19. SC computes $\overline{\mathbf{P}}_i^t \;|\; \overline{\boldsymbol{\Psi}}_{i,i}, \overline{\boldsymbol{\omega}}_i, \boldsymbol{\Omega}_{-i}^n$ using (20)
20. Until convergence
21. Set $\mathbf{P}_i^n = \overline{\mathbf{P}}_i^t, \quad \forall i = 1 \ldots N_{cells}$
22. Until convergence

We would like to remark that the proposed algorithm also subsumes the training sequences design for point-to-multipoint (P2MP) systems, for which no optimal solution has been derived before in the literature. In such a case, the phase for exchanging the *interference cost* is not required.

V. Results

As a reference scenario we have assumed the one defined in [13] with a configuration of 4 SCs per cluster, each equipped with M=2 transmitting antennas serving multiple UEs with N=2 receive antennas. We consider four schemes for designing the training sequences:

- *Orthogonal allocation*: It is assumed that each SC has allocated M orthogonal symbols for *training*, imposing a total training phase equal to $L = M \times N_{cells}$ symbols. Nevertheless, different schemes are possible in this configuration:
  a. *Dumb scheme* ('Dumb Orth'): Each SC selects randomly an M×M training sequence, imposing orthogonality among antennas.
  b. *Wise scheme* ('P2P Orth' or 'P2MP Orth'). In P2P scenario it was shown in [2] that the optimal training sequence depends on the channel correlation matrix. This scheme will be referenced as 'P2P Orth'. On the other hand, for P2MP scenario we will use the algorithm proposed in this work to take into account information from all served UEs. This second scheme is referenced as 'P2MP Orth'.
- *Dumb Resource Reuse* ('Dumb Res. Reuse'): Every SC randomly obtains an M×L training sequence, keeping orthogonality among antennas, but not ensured over different SCs.
- *Coordinated training sequence optimization* ('Coordinated TSO'): Algorithm proposed in this work.
- *Uncoordinated training sequence optimization* ('Uncoordinated TSO'): The same algorithm considered in this work, but without exchanging the *interference cost.* Every SC designs its training sequences independently and it only takes into account the feedback from its associated users.

A. MSE minimization

First we have considered a scenario where each SC is only serving one UE ($N_{UE}^{i} = 1$) and the training phase consists of L=8 symbols. Figure 2 depicts the sum MSE attained by the different schemes. In this case, the 'Dumb Orth' and the optimal 'P2P Orth' attain almost the same result. Afterwards, the proposed 'Coordinated TSO', after being randomly initialized, monotonically converges to the same solution. It is also illustrated the benefits of using a better refinement of the training sequence by means of the local loop described in lines 14-20 of Table I, either with 1 or 20 iterations. Similarly, the 'Uncoordinated TSO' also minimizes the MSE, but presenting some fluctuations in terms of MSE. Finally, the 'Dumb Res. Reuse' is the one that gets the worst performance.

Figure 3 evaluates the performance of the previous schemes in case there are 4 SCs serving 3, 6, 6 and 1 UEs, respectively, with L=8 symbols. The 'Dumb Res. Reuse' continues getting the worst performance. Additionally, it can be observed that there are differences between the 'Dumb Orth' and the proposed 'P2MP Orth', because this latter one takes into account the statistics of all served UEs. Finally, we can see that the proposed 'Coordinated TSO' is the one that gets the best performance and the convergence of the proposed algorithm is shown to be very fast. The difference between the 'P2MP Orth' and the 'Coordinated TSO' is due to the presence of different local optima.

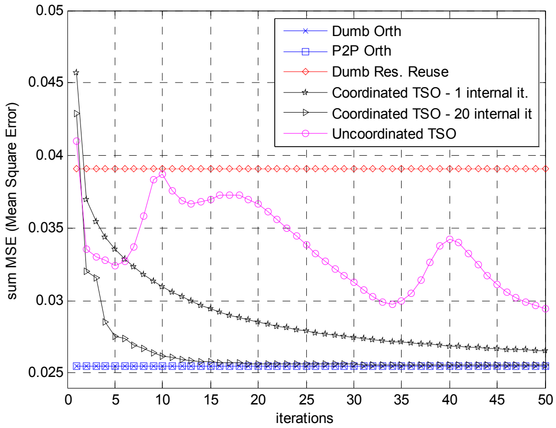

Fig. 2. Evolution of the sum MSE as a function of the number of iterations of the decentralized algorithm. I-P2P scenario: 4 SCs and 1 UE per SC. L=8.

Finally, Figure 4 illustrates the performance in terms of average MSE attained by the different schemes as a function of the number of training symbols $M \le L \le M \times N_{cells}$. Notice that the 'Dumb Orth' and 'P2MP Orth' schemes have the intrinsic constraint that they can only be designed for L=8. We can observe that the 'Coordinated TSO' achieves the minimum MSE. As it is expected, the MSE increases as the number of symbols for training decreases.

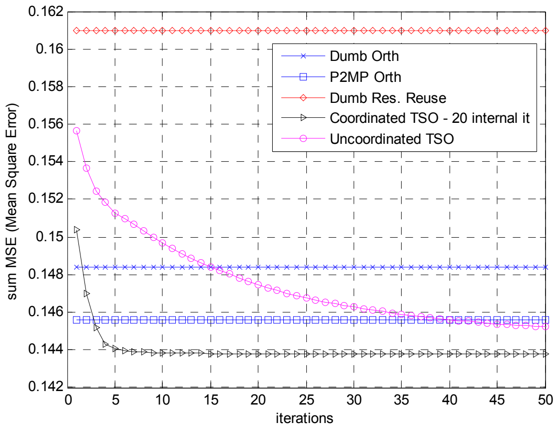

Fig. 3. Evolution of the sum MSE vs. the number of iterations of the decentralized algorithm. I-P2MP scenario: 4 SCs and {3,6,6,1} UEs per SC. L=8.

B. Benefits in terms of throughput

Previous results have elucidated that the proposed coordinated training sequence optimization is able to get the best performance in terms of MSE, and additionally it adapts to the number of available resources for training. Hereinafter, we explore how many symbols should be devoted for training or data transmission in a scenario with 4 SCs equipped with M=2 antennas and serving multiple UEs. The data transmission follows the signal model in (7).

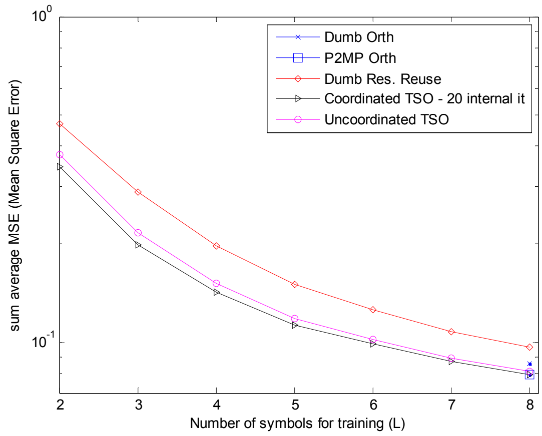

Fig. 4. Average sum MSE as a function of the number of symbols devoted for training (L).

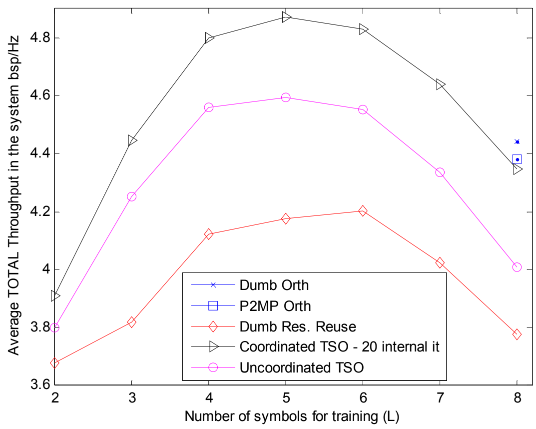

Fig. 5. Average total throughput (bits/s/Hz) vs. the number of symbols devoted for training (L). The frame consists of T=16 symbols.

Figure 5 shows the average total throughput in the system (in bits/s/Hz) as a function of the number of symbols devoted for training (L) where frames are composed of T=16 symbols, i.e. T-L symbols for data transmission. We can observe that the 'Dumb Res. Reuse' is a very inefficient scheme when it is compared with 'Uncoordinated TSO' and 'Coordinated TSO' schemes. These later schemes are able to adapt to the interference-dominated scenario. They can free more symbols for data transmission, hence increasing the average total throughput. However, as it is expected, reducing L so much leads to an average total throughput loss due to the high estimation errors. In the scenario considered, we can observe that just with L=4, 5 or 6 symbols devoted for training it is enough, getting throughput gains up to 10% and 16% with respect the 'Dumb Orth' (which can only be applied for L=8 when there are 4 SCs with M=2) and 'Dumb Res. Reuse' schemes, respectively.

VI. CONCLUSIONS

The strategy of allocating orthogonal resources for *training* at each small cell in a HetNet scenario is a simple way of getting a good channel estimation, but at the cost of requiring resources proportional to the number of antennas and number of SCs such that the system efficiency can be severely degraded. We can reduce the number symbols devoted to the *training* phase, but then interference comes up in the system (and the system can be modeled through the Interference Point-to-Multipoint channel). This work has shown that an adequate interference management technique for designing the training sequences is able to improve significantly the throughput of the system assuming a simple transmission scheme in the data phase. We have proposed a decentralized algorithm that exploits the coordination over neighboring SCs and additionally it takes into account statistical feedback from the served UEs. In this regard, the proposed algorithm is able to adapt to the current conditions of the scenario. Likewise, having a good design of training sequences permits having acceptable channel estimation at the receiver side, so that more resources can be devoted for data transmission. Further research will consider the use of advanced transmission schemes during the data phase.

REFERENCES

[1] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions", *IEEE Signal Proessing Magazine*, November 2004

[2] M. Biguesh, A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals", *IEEE Trans. on Signal Processing*, vol. 54, no. 3, March 2006.

[3] M. Biguesh, S. Gazor, M. H. Shariat, "Optimal training sequence for MIMO wireless sytems in colored environments", *IEEE Trans. on Signal Processing*, vol. 57, no. 8, Aug. 2009.

[4] D. Katselis, et al. "Training sequence design for MIMO channel: An applications-oriented approach", arXiv: 1301.3708v1, Jan. 2013.

[5] B. Hassibi, B. M. Hochwald, "How much training is needed in multiple antenna wireless links?", *IEEE Trans. on Information Theory*, vol. 49, no. 4, April 2003.

[6] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems" *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011

[7] G. Scutari, D. P. Palomar, F. Facchinei, J-S. Pang, "Distributed Dynamic Pricing for MIMO Interfering Multiuser Systems: A Unified Approach", *IEEE Int. Conf. on Network Games, Control and Optimization*, pp. 1-5, October 2011

[8] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, W. Utschick, "Distributed Interference Pricing for the MIMO Interference Channel," *IEEE Int. Conf. on Communications*, pp. 1796-1800, June 2009

[9] S. Lagen, A. Agustin, and J. Vidal, "Distributed inter-cluster interference management for CoMP-based cellular networks", in Proc. *IEEE Global Comm Conf.* (GLOBECOM), Atlanta, USA, Dec. 2013

[10] S. Christensen, R. Agarwal, E. Carvalho, J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO BC beamforming design", *IEEE Trans. on Wireless Comm.*, vol. 7, no. 12, Dec. 2008

[11] K. M. Abadir, and J.R. Magnus, *Matrix Algebra*, Cambridge University Press., 2005

[12] Q.Shi, M. Razaviyayn, Z. Luo, C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel", *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4331-4340, September 2011

[13] 3GPP-R1-130856, "Evaluation assumptions for SCE", Huawei, HiSilicon.