Coordination and Antenna Domain Formation in Cloud-RAN systems

Hadi Ghauch*, Muhammad Mahboob Ur Rahman*, Sahar Imtiaz*, James Gross*,
*School of Electrical Engineering and the ACCESS Linnaeus Center, Royal Institute of Technology (KTH)

Abstract—We study here the problem of Antenna Domain Formation (ADF) in cloud RAN systems, whereby multiple remote radio-heads (RRHs) are each to be assigned to a set of antenna domains (ADs), such that the total interference between the ADs is minimized. We formulate the corresponding optimization problem, by introducing the concept of interference coupling coefficients among pairs of radio-heads. We then propose a low-overhead algorithm that allows the problem to be solved in a distributed fashion, among the aggregation nodes (ANs), and establish basic convergence results. Moreover, we also propose a simple relaxation to the problem, thus enabling us to characterize its maximum performance. We follow a layered coordination structure: after the ADs are formed, radio-heads are clustered to perform coordinated beamforming using the well known Weighted-MMSE algorithm. Finally, our simulations show that using the proposed ADF mechanism would significantly increase the sum-rate of the system (with respect to random assignment of radio-heads).

Index Terms—5G, Cloud RAN, radio head assignment, antenna domain formation, interference coupling, block coordination descent

I. INTRODUCTION

The Cloud-Radio Access Network (C-RAN) is identified as one of the promising architectures to address the challenges of 5G systems, namely, the requirement for high spectral efficiency within a particularly dense deployment (both users and access nodes) [1]. C-RAN systems are characterized as a centralized solution for interference coordination: remote radio-heads (RRH) act as access nodes, and their baseband processing capabilities vary from full digital signal processing capability (i.e., base stations), to ‘dumb antennas’ with no baseband capabilities (such as distributed MIMO systems). Such radio-heads are connected via high-capacity (possibly wireless) links to so-called aggregation nodes (ANs), each essentially acting as a large processing unit. Thus, such architectures are natural candidates for interference coordination.

In dense deployments, coordination among base stations was identified as the key to achieving high spectral efficiency: indeed the ideas of Coordinated Multi-point (CoMP) [2], [3] and Interference Alignment (IA) [4], [5] were central to achieve higher spectral efficiency. However, when applied to conventional cellular systems, such techniques have the stringent requirement that they need to be distributed, i.e., to only use local CSI at each node: the overhead associated with such techniques has been identified as a (potentially) limiting factor of the sum-rate gains brought about by techniques such as IA [6], [7], [8] and [9]. This essentially puts hard limits on the effectiveness of the latter techniques. However, this limitation is lifted in the C-RAN architecture, since ANs can be assumed to have perfect CSI of a large area, and can perform coordination in a centralized manner. Earlier related work has been reported in [10] where the authors investigate the beamforming design problem (for sum-rate maximization), in the context of cellular systems with limited backhaul capacity.

In [11], the authors study the problem of dynamic clustering in dense deployments (for joint transmission), by characterizing the statistics of the instantaneous signal-to-interference-and-noise ratio (SINR), via tools from stochastic geometry. In our earlier paper [12], we investigated radio-head coordination (namely coordinated beamforming), in a typical Cloud RAN setup, with a large number of radio-heads and users, served by one AN. In this work, however, we look higher into the coordination hierarchy, by investigating the so-called Antenna Domain Formation (ADF) problem: given a set of radio-heads (each serving a set of users), and a set of ANs, what is the best assignment of radio-heads to ANs, such that the total interference leakage between the ADs, is minimized. Studying the latter setup is the main contribution and novelty of this paper. In that sense, we formulate the ADF problem as integer programming problem, and devise an iterative algorithm for solving it. We also relax the latter problem to obtain a lower bound on the maximum performance of our algorithm. Moreover, we investigate the effect of using a layered coordination structure, whereby further coordination mechanism (coordinated beamforming) are put in place. We underline the fact that this work is currently being extended to a journal form[13].

In the following, we use bold upper-case letters to denote matrices, bold lower-case to denote vectors, and calligraphic letters to denote sets. Furthermore, for a given matrix $\mathbf{A}$, $\mathbf{A}_{i,j}$ denotes the matrix formed by taking columns $i$ to $j$, of $\mathbf{A}$. $||\mathbf{A}||_F$: its Frobenius norm, $|\mathbf{A}|$: its determinant, $\mathbf{A}^T$: its transpose, and $\mathbf{A}^\dagger$: its conjugate transpose. $[\mathbf{A}]_{i,j} = a_{i,j}$ denotes element $(i, j)$ in a matrix $\mathbf{A}$, and $[\mathbf{a}]_i$: element $i$ in a vector $\mathbf{a}$. While $\mathbf{I}_n$: the $n \times n$ identity matrix, $\mathbf{I}_n$: the $n \times 1$ vector of ones, $\mathbf{B}_N$: denotes space of $N$-dimensional binary vectors, and $\Pi_S[\mathbf{z}]$: is the Euclidean projection of a vector $\mathbf{z}$, into some (possibly non-convex) set $S$. 
II. SYSTEM MODEL

Given a large area of interest, i.e. a spatial area of a certain size, with $N$ of RRHs and $A$ ANs, we wish to address the issue of ADF, i.e. which radio heads should be assigned to which AN. In other words, given a set $\mathcal{R}$ of radio heads in an area, the problem is to assign them to a set of ADs (where each AD is controlled by one AN) $\mathcal{A}$, such that total interference coupling between the ADs is minimal. Let $\mathcal{A}$ denote a set of ANs (where $\mathcal{A} \bowtie |\mathcal{A}|$), and $\mathcal{R}$ the set of radio-heads where $N = |\mathcal{R}|$.

We assume that radio head $i \in \mathcal{R}$ is equipped with $M_i$ antennas, and serving users which are single-antenna receivers (we assume single antenna users for simplicity of exposition, though this can be extended to multi-antenna receivers). Let $\mathcal{U}_i$ be the set of users served by radio head $i \in \mathcal{R}$. Then, $H_{i,j} \in \mathbb{C}^{|\mathcal{U}_i| \times M_i}$ denotes the channel from the antennas of radio head $i \in \mathcal{R}$, to the users served by radio head $j \in \mathcal{R}$, i.e. $\mathcal{U}_j$, and $W_{i,j} \in \mathbb{C}^{M_i \times |\mathcal{R}|}$ be the precoder that RRH $i$ employs to serve its users $\mathcal{U}_i$. A small toy example is illustrated in Fig. 1.

This work essentially addresses the interference between ADs, as this type of interference limits significantly the performance of the entire system. It is clear that for any form of operation within the ADs, as well as for any type of interference mitigation technique applied between the ADs, there are more suitable and less suitable AD choices for the radio heads.

We assume that the load has been pre-allocated among all ANs, implying that the $k$th AN will have a load of $\gamma_k \geq 0$. A special case of this is when the load is equally distributed across all ANs, i.e., $\gamma_k$ is the same among all the ANs. Furthermore, each RRH is to be assigned to one AN only. Moreover, it is assumed that each AN has perfect CSI of the users to which it is associated.

Hence, the main idea in order to form ADs is to reduce the interference coupling between the different ADs. Denote by $\alpha_{i,j} \in \mathbb{R}_+$ the interference leakage between radio head $i \in \mathcal{R}$ and radio head $j \in \mathcal{R}$, and note that this denotes the leakage in the direction of $j$, i.e. $\alpha_{i,j}$ is a measure that represents the interference caused by RRH $i$ on terminals associated to RRH $j$, if $i$ and $j$ are assigned to be in different ADs. Due to asymmetric channel gains and terminal associations, notice in particular that $\alpha_{i,j} \neq \alpha_{j,i}$.

Thus, we define a coupling coefficient $c_{i,j}$ as the interference coupling coefficient between RRH $i$ and RRH $j$. The latter can be then viewed as the cost of having both radio heads in the same AD. Thus, one intuitive choice for the latter is to select $c_{i,j}$ as the total interference leakage between the two radio heads, i.e.,

$$c_{i,j} = \alpha_{i,j} + \alpha_{j,i} = \|H_{i,j}W_j\|_F^2 + \|H_{j,i}W_i\|_F^2 \quad \forall i \neq j$$

However, incorporating the precoders into the coupling coefficients makes the overall system quite challenging since the coupling coefficients have to be updated quite frequently. Thus, we formally define the latter quantity as,

$$c_{i,j} = \begin{cases} \|H_{i,j}W_j\|_F^2 + \|H_{j,i}W_i\|_F^2, & \forall i \neq j \\ 0, & \forall i = j \end{cases}$$

Let $\Psi$ be the matrix formed by gathering all those coefficients, i.e. $[\Psi]_{i,j} = c_{i,j}$, $\forall (i,j) \in \mathcal{R}^2$, where $\Psi \in \mathbb{R}^{N \times N}$ is such that $\Psi = \Psi^T$.

Furthermore, denote by $x_{i,k} \in \{0,1\}$ the binary decision variable if RRH $i$ is assigned to AD $k$ (or not), and denote by $x_k \in \{0,1\}^N$ the assignment vector of AN $k \in \mathcal{A}$. We formulate the ADF problem as the following integer programming problem:

$$\begin{align*}
\min f(x_{i,j}) &= \sum_{k=1}^A \sum_{l \neq k} \left( \sum_{i=1}^N \sum_{j \neq i} c_{i,j} x_{i,k} x_{j,l} \right) \\
\text{s. t.} & \sum_{i=1}^N \beta_{i,k} x_{i,k} = \gamma_k, \quad \forall k \in \mathcal{A} \\
& \sum_{k=1}^A x_{i,k} \leq 1, \quad \forall i \in \mathcal{R} \\
& x_{i,k} \in \{0,1\}, \quad \forall (i, k) \in \mathcal{A} \times \mathcal{R}
\end{align*}
$$

(P1)

Note that one can rewrite the above program in terms of $x_k$ as follows,

$$\begin{align*}
\min f(x_k) &= \sum_{k=1}^A \sum_{i \neq k} \beta_{i,k}^T x_k \Psi x_k \\
\text{s. t.} & \sum_{k=1}^A x_k \leq 1_N \\
& \beta_{i,k}^T x_k = \gamma_k, \quad x_k \in \mathcal{B}_N, \quad \forall k \in \mathcal{A}
\end{align*}
$$

(P2)

where the above inequality holds element-wise. In the above, $\sum_{k=1}^A x_k \leq 1_N$ is the assignment constraint forcing that each RRH is assigned to at most one AD. Moreover, denoting by $\beta_k \in \mathbb{R}_+^N$ the set of loading factors for AD $k$ (such that $1^T \beta_k = 1$, $\forall k \in \mathcal{A}$), $\beta_k^T x_k = \gamma_k$ is the loading constraint for the $k$th AD (i.e. the load profile for AD $k$). Using this equivalent form, it becomes evident that $f$ indeed is not jointly convex in all the variables, due to the coupling among $x_k$ and $x_l$. However, this does reveal an inherent bi-linear structure of $f$ (taken separately in each variable, $f$ is linear) that we exploit for the optimization. Intuitively, $f$ models the sum-total coupling that exists between the individual ADs.
III. PROPOSED ALGORITHM

A. Block Coordinate Descent

The coupling among the variables in \( f(x_1, \ldots, x_A) \), makes \((P2)\) a perfect candidate for a Block-Coordinate Descent (BCD) approach (also known as the Gauss-Siedel method). In a nutshell, BCD works iteratively, by fixing all variables but one block, i.e. \( f(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_A) \), and optimize for \( x_k \), iteratively. Let \( n \) denote the iteration number, i.e., \( x_k^{(n)} \) denotes the value of \( x_k \) at the \( n \)th iteration. Stated differently, in the above, \( x_k^{(n+1)} \) is obtained recursively, given the fixed \( x_k \), \( x^{(n)} \), \( x_k^{(n)} \) are fixed, \( x_k \) is optimized accordingly.

\[
\{x_k^{(n+1)}\} = \underset{x_k}{\arg\min} f(x_{k-1}^{(n+1)}, x_{k+1}^{(n+1)}, x_A^{(n+1)}),
\]

\[
= \underset{x_k}{\arg\min} f(x_k, z_k^{(n)}), \quad \forall k \in A
\]

where

\[
z_k^{(n)} = (x_1^{(n)}, \ldots, x_{k-1}^{(n)}, x_{k+1}^{(n)}, \ldots, x_A^{(n)})
\]

This denotes the block of fixed variables in the above BCD iteration. Stated differently, in the above, \( x_k^{(n+1)} \) indicates that block \( x_{k-1} \) (for instance) has been already updated, while block \( x_{k+1} \) hasn’t. Thus, as seen from (3), BCD generates a sequence of iterations \( \{x_k^{(n)}\} \) that converge to a limit point (this will be formalized later in this section). Moreover, we note that \( f(x_k, z_k^{(n)}) \) (which denotes the function \( f(x_k) \), when the variables in block \( z_k^{(n)} \) are fixed), can be rewritten as,

\[
f(x_k, z_k^{(n)}) = x_k^T \Psi \left( \sum_{l=1}^{k-1} x_l^{(n+1)} + \sum_{l=k+1}^{A} x_l^{(n)} \right) \triangleq x_k^T r_k^{(n)}
\]

where

\[
r_k^{(n)} \triangleq \Psi \left( \sum_{l=1}^{k-1} x_l^{(n+1)} + \sum_{l=k+1}^{A} x_l^{(n)} \right)
\]

is referred to as the residual of AD \( k \), at the \( n \)th iteration. Moreover, we clearly see that \( f(x_k, z_k^{(n)}) \) is linear, implying that when all but one block are fixed, \( f \) is linear. This indeed shows that \((P2)\) is equivalent to a series of problems that are solved iteratively, and block-by-block. Now that we have described the BCD framework, we focus on the solution of the optimization problem in (3), within each BCD iteration.

With that in mind, the update for \( x_k \), i.e. \( x_k^{(n+1)} \) in (3), is,

\[
x_k^{(n+1)} = \begin{cases} 
\text{arg\min}_{x_k} f(x_k, z_k^{(n)}) \\
\text{s. t. } \beta_k^T x_k = \gamma_k, \quad x_k \leq \omega_k^{(n)}, \quad x_k \in B_N 
\end{cases}
\]

(5)

where \( \omega_k^{(n)} \) is the vector of residual assignments dependent on the assignments of all other ADs.

\[
\omega_k^{(n)} \triangleq 1_N - \left( \sum_{l=1}^{k-1} x_l^{(n+1)} + \sum_{l=k+1}^{A} x_l^{(n)} \right)
\]

(6)

The above problem is known as a mixed integer linear program, and is NP-hard due to its inherent combinatorial nature. However, there exists many polynomial-time approximation algorithms that, for all practical purposes, solve it globally (such efficient solvers are found in MATLAB and CVX).

It becomes clear at this stage that BCD transforms \((P2)\) into a series of \( A \) parallel subproblems, where each can be solved in a distributed way, i.e. locally at each AN.

B. Algorithm Description

The use of BCD for solving \((P2)\) goes hand in hand with making the problem naturally decoupled: when \( \{x_k\}_{k \neq k} \) are fixed, the cost function decouples in \( x_k \) and can thus be solved separately by AD \( k \), without any loss in optimality. In that sense, the optimal update for \( x_k \) at AD \( k \), depends on the assignments at all the other ADs, that have to be shared. Given assignments from other ADs, \( x_1^{(n+1)}, \ldots, x_{k-1}^{(n+1)}, x_{k+1}^{(n+1)}, \ldots, x_A^{(n+1)} \), AD \( k \) forms the residual \( x_k^{(n+1)} \), and can proceed to solve its optimization problem locally, and update \( x_k^{(n+1)} \). The process is formalized in Algorithm 1.

Algorithm 1 ADF via BCD

Input: \( \Psi, N, A \)

\[ \text{for } n = 0, 1, \ldots, L - 1 \text{ do} \]

\[ \text{\quad \text{\quad \quad \text{\quad procedure at each AN}} } \]

\[ \text{\quad \text{\quad \quad \quad \text{\quad obtain } (x_1^{(n+1)}, \ldots, x_{k-1}^{(n+1)}, x_{k+1}^{(n+1)}, \ldots, x_A^{(n+1)} ) \text{ at AD } k } \]

\[ \text{\quad \text{\quad \quad \quad \text{\quad compute residual } r_k^{(n)} \text{, using (4)}} } \]

\[ \text{\quad \text{\quad \quad \quad \text{\quad compute residual assignment } \omega_k \text{, using (6)}} } \]

\[ \text{\quad \text{\quad \quad \quad \text{\quad compute } x_k^{(n+1)} \text{ as solution to (5)}} } \]

end for

Output: \( \{x_1^{(L)}, \ldots, x_A^{(L)}\} \)

C. Convergence

Let \( \{x_k^{(n)}\} \) be the sequence iterates produced by the BCD in (3), and \( \{x_k^{(n)}\} \triangleq \lim_{n \to \infty} \{x_k^{(n)}\} \). The monotonic nature of the BCD iterations is established below.

Lemma 1 (Monotonicity). With each update \( x_k^{(n)} \to x_k^{(n+1)} \), \( f \) is non-increasing. Moreover, the sequence of function iterates \( \{f(x_1^{(n)}, \ldots, x_A^{(n)})\} \) converges to a limit point \( f(\{x_k^{(n)}\}) \).

Proof: Refer to Appendix A

Although the above result establishes the convergence of the proposed BCD method, it only establishes convergence to a limit.

Remark 1. Although the updates generated by the BCD iteration are shown to converge monotonically to a limit point, it cannot be established that the latter corresponds to a stationary point of \( f \), namely due to

- the presence of the binary constraint, that prevents the use of advanced BCD convergence results such as [14]
- the coupling in the assignment constraint, i.e. \( x_1 + \ldots + x_A \leq 1_N \), cannot be handled by standard BCD convergence results.

D. Performance bounds

Here we attempt to shed light on the maximum performance that Algorithm 1 can deliver. This is achieved by relaxing the original problem in \((P2)\). One of the most well known relaxations for problems such as \((P2)\) is done by relaxing the
The result of combining the latter two ‘coordination mechanisms’ is a hierarchical cooperation model: at the highest level, RRHs are assigned into ADs (using Algorithm 1), then, RRHs are grouped to form coordination clusters (where each is performing coordinated beamforming). The resting network structure is illustrated in Fig. 2. Next, we mathematically formalized the operation and performance of the system.

1) Sum-rate performance: We assume for simplicity that each RRH is serving \( J \) users, and that the size of each cluster, \( |\mathcal{C}_m| \), is the same. Moreover, let \( C \) denote the number of clusters within each AD. Let \( C^l_m \) denote the \( m \)th cluster \((1 \leq m \leq C)\) in AD \( l \) \((1 \leq l \leq A)\): then \( C^l_m \subset \mathcal{R} \) is the set of cooperating radio heads within the \( l \)th AD (Fig 2).

For shorthand notation, we denote by \( i_k \), the \( i \)th user \((1 \leq i \leq J), \) served by RRH \( k \) \((1 \leq k \leq K)\). Then, \( \mathbf{h}^{(l,m)}_{i_k,j} \) is the (MISO) channel from RRH \( j \), to user \( i_k \), in \( C^l_m \). Similarly, we define \( \mathbf{v}^{(l,m)}_{i_k} \) at the transmit precoder, used to serve user \( i_k \), in \( C^l_m \). Letting \( P_l \) denote the transmit power of all radio-heads, the SINR of user \( i_k \), in \( C^l_m \), is given by (10). Thus, treating interference as noise at the users, the achievable sum-rate of the system is as follows,

\[
R_S = \sum_{(l,m)} \sum_{i_k} \sum_{i=1}^{J} \log_2(1 + \text{SINR}^{(l,m)}_{i_k})
\]

Under this setup we advocate, each user is subject to residual interference from users within its coordination cluster, interference emanating from users in other clusters (but still within the same AD), as well as interference coming from all users present in other ADs.

2) Practical Aspects: The matrix of coupling coefficients \( \mathbf{\Psi} \) should available at all the ANs, prior to the start of the algorithm: the latter can be “populated” sequentially, by having each AD estimate all channels (both to its served users, and to users from other ADs) via training, in an orthogonal manner.

Due to the fact that the assignment variables are binary, one only needs a low-rate control link between all the ANs. In the

\[\sum_{k=1}^{A} \sum_{l \notin \chi_k} w_k^T \mathbf{\Psi} \mathbf{w}_l \]

where \( \{w_k\} \), the optimization variables are no longer binary. The fact that the optimal solution of (P3) is a lower bound on the original problem (P2), follows immediately from the relaxation arguments. Despite its simple form, globally solving the above problem is not straightforward, namely due to coupling among the variables, \( \{w_k\} \), and that \( f \) is not convex in \( \{w_k\} \) (since \( \mathbf{\Psi} \) is not positive-definite). However, we recall that the same BCD procedure that was used to solve (P2), can be applied to the relaxed problem. Thus, the sequence of iterates generated by the BCD is given by,

\[
\begin{align*}
\mathbf{w}^{(n+1)}_k &= \left\{ \begin{array}{l}
\arg \min_{\mathbf{w}_k} f(\mathbf{w}_k, \mathbf{z}^{(n)}_k) \\
\text{s. t. } \mathbf{w}_k \leq \mathbf{\omega}^{(n)}_k, \mathbf{w}_k \in [0, 1]^N, \quad \forall k \in A
\end{array} \right.
\end{align*}
\]
among the two ADs, i.e. factors are identical, whereby the RRH load is split equally within each cluster. Moreover, we assume that the loading WMMSE is employed to iteratively optimize the precoders.

Global coordination (GC) cluster within each AD. Then, = 2 users). The RRHs are to be assigned to one of two ANs (i.e., deployed, each equipped with M = 4 transmit antennas and serving J = 2 single-antenna users (for a total of K = 32 users). The RRHs are to be assigned to one of two ANs (i.e., A = 2). In our simulations we assume that within each AD, radio heads form one cluster (i.e. C = 1), thereby forming global coordination (GC) cluster within each AD. Then, WMMSE is employed to iteratively optimize the precoders within each cluster. Moreover, we assume that the loading factors are identical, whereby the RRH load is split equally among the two ADs, i.e. \( \beta_k = 1 \), \( \forall k \) and \( \gamma_k = N/A \), \( \forall k \).

For the assignment of radio heads to ANs, we benchmark our proposed scheme, Algorithm 1, against a randomized assignment where radio heads are randomly assigned to each of the ANs. Despite the fact that the latter is not a good choice, it is however intended to be used as a lower bound on the performance of our algorithm (the latter assignment still takes into account the equal-loading constraint, and the fact that each RRH is assigned to one AD only). Moreover, for this particular (relatively small) case, we are able to find the globally optimal solution to the ADF problem, via exhaustive search of \((P1)\).

We also investigate two (extreme) special cases of ADF,

- ADF based on instantaneous CSI is the case detailed in (2) where matrix of coupling coefficients \( \Psi \) is based on instantaneous CSI: in this case, \( \Psi \) is updated every time the channel changes, and consequently, the ADs have to be recomputed at every channel realization.

- ADF based on statistical CSI, where the coupling coefficients are given by,

\[
\mathcal{S}_i,j = \begin{cases} 
|\mathbf{H}_{i,j}|^2 + |\mathbf{H}_{j,i}|^2, & \forall i \neq j \\
0, \forall i = j
\end{cases}
\]  

where, analogously to \( H_{i,j} \), we define \( \mathbf{H}_{i,j} \), as the matrix of pathloss factors from the antennas of radio head \( i \in \mathcal{R} \), to the users of radio head \( j \in \mathcal{R} \). Unlike the instantaneous CSI case, here \( \Psi \) is computed at the beginning of the simulation, the ADF performed, and remains static throughout the simulation run (since users are static).

**B. Results**

The sum-rate performance (refer to (11)) that results from the above setup is shown in Fig 3, where we compare the performance of our algorithm, against both the random assignment and the optimal exhaustive search. Instantaneous CSI is assumed for all schemes. We can clearly see a significant performance gap between the performance of our scheme, and that of the benchmark. Moreover, looking at the performance of the exhaustive search, reveals that there are rather large gains from finding the globally optimal solution to the ADF problem.

We next investigate the extent to which replacing Instantaneous CSI with long-term Statistical CSI, degrades the sum-rate performance of the system in question (we follow the exact same setup used for the instantaneous CSI case). As shown in Fig 4, we see the same trends discussed just above.
More importantly, as far as our algorithm is concerned, the degradation observed by going from instantaneous to statistical CSI, is extremely negligible. This can be exploited to massively reduce the overhead and complexity of the system (detailed in next subsection).

Following the same setup as above, we shed light on the behavior of the proposed ADF algorithm 1, as a function of the SNR. In Fig 5 we show the average value of $f$ that is achieved by Algorithm 1, as well as the lower bound discussed in III-D. Though our algorithm greatly outperforms the benchmark, it is relatively far from the lower-bound (derived from the relaxed problem $(P3)$): however, we reiterate the fact that solutions such as the latter, do not correspond to feasible solutions for the ADF problem (thus are potentially not achievable within our setup).

C. Discussions

The numerical results all point to a good performance brought about by the application of our proposed method (both in terms of achievable sum-rates for the system, and total interference leakage between the ADs). Moreover, as shown in Algorithm 1, the communication overhead requirement among the ANs is quite negligible since it consists of binary vectors only. With that in mind, the overhead would consist of obtaining all the CSI needed to compute $\Psi$, and propagate it to all the ANs (as often as needed). Though this might seem too high, our simulations also clearly indicated that for relatively static settings, ADF based on pathloss (i.e., location-based) has virtually the same performance as the one based on instantaneous CSI.

V. CONCLUSION AND FUTURE WORK

In this work we tackled the problem of ADF in C-RAN systems by formulating it as an integer programming problem. We employ the well-known BCD framework for solving the problem and devising an iterative algorithm for that purpose. We shed light on the convergence of the algorithm, as well as its ‘maximal performance’ via simple relaxations. Our simulations reveal that this approach promises to deliver good sum-rate performance, in typical C-RAN deployments.

As mentioned earlier, we started to investigate more analytical aspects of the ADF problem, namely, better relaxation techniques (e.g. Lagrange Relaxation), decomposition techniques (e.g. Dantzig-Wolfe Decomposition), and dual problem analysis. The latter will provide us with insights into fundamental lower bounds for the ADF problem. In addition, we extended the problem formulation used here, to include both the channel and precoder effect, rather than just the channel energy (as seen in (1)). All the above issues are investigated in great detail, in our subsequent work [13]. We also wish to investigate more practical scenarios, namely the so-called hybrid CSI case, consisting of a mix between instantaneous and statistical CSI.

VI. ACKNOWLEDGMENT

The authors gratefully acknowledge the funding and support of their counterparts at Huawei, Finland.
APPENDIX

A. Proof of Lemma 1

Note that the following is a direct consequence of (5)

\[ f(\{x_k^{(n)}\}) \geq f(\{x_2^{(n+1)}, z_2^{(n)}\}) \geq f(\{x_2^{(n+1)}, z_2^{(n)}\}) \geq f(\{x_A^{(n+1)}, z_A^{(n)}\}) \triangleq f(\{x_k^{(n+1)}\}) \]

where the last equality follows from the fact that \( f(\{x_A^{(n+1)}, z_A^{(n)}\}) \) corresponds to the case where all variables \( (x_1, \ldots, x_A) \), are updated. It follows that the sequence \( \{f(\{x_1^{(n)}, \ldots, x_A^{(n)}\})\}_n \) converges monotonically to a limit point \( f_0 \).

REFERENCES