RESOURCE ALLOCATION FOR OFDMA SYSTEMS WITH MULTI-CELL JOINT TRANSMISSION

Jingya Li†, Xin Chen‡, Carmen Botella*, Tommy Svensson† and Thomas Eriksson†

†Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
‡WTI lab, Beijing University of Posts and Telecommunications, Beijing, China
*IRTIC, Universitat de València, València, Spain

ABSTRACT
This paper considers the downlink resource allocation of a coordinated multi-cell cluster in OFDMA systems with universal frequency reuse. Multi-cell joint transmission is considered via zero-forcing precoding. Furthermore, joint optimization of the user selection and power allocation across multiple subchannels and multiple cells is studied. The objective is to maximize the weighted sum rate under per-base-station power constraints. Based on general duality theory, two iterative resource allocation algorithms are proposed and compared with the optimal solution, which requires an exhaustive search of all possible combinations of users over all subchannels. Simulation results show that the two proposed algorithms achieve a performance very close to the optimal, with much lower computational complexity. In addition, we show that joint user set selection across multiple subchannels significantly improves the system performance in terms of the weighted sum rate.

Index Terms—OFDMA, scheduling, power control, multi-cell joint transmission, base-station coordination

1. INTRODUCTION
Recently, multi-cell joint transmission has been considered as a promising technique to mitigate inter-cell interference (ICI) and improve the spectrum efficiency in wireless communication systems [1-2]. In this approach, multiple base stations (BSs) are inter-connected via a high-speed backhaul, so that both data and channel state information (CSI) of all users can be shared among the coordinated BSs. The ICI is then reduced by using the signals transmitted from other BSs to assist the transmission instead of acting as interference.

Many existing works have addressed the resource allocation problem with multi-cell joint transmission by assuming a narrow-band channel [3]-[5]. With the objective of maximizing the weighted sum rate, two power allocation schemes based on waterfilling distribution are proposed in [3] with block-diagonalization (BD) joint transmission subject to per-BS or per-antenna transmit power constraints. In [4], a closed-form expression for the optimal BD precoding matrix is derived, which specifies both the transmit beamforming vectors and power allocation for different beams. It has been shown in [3]-[4] that maximizing the weighted sum rate of a given user set under per-BS or per-antenna power constraints is a convex problem, if the total number of receive antennas is less than or equal to the total number of transmit antennas. In more practical cases, the total number of users in the system could be very large such that the above condition does not hold. In order to determine the set of users to be served simultaneously, different joint user selection and power allocation algorithms are proposed in [5].

LTE-based systems will adopt OFDMA as the downlink access technology. With a frequency-selective fading channel, OFDMA adds another dimension in the design of the resource allocation scheme, and the optimization problem is non-convex in general. Thus, the algorithms proposed in [3]-[5] may not be suitable for the multiple subchannels case. The optimal resource allocation solution for OFDMA systems with multi-cell joint transmission requires an exhaustive search over all possible combinations of users for each subchannel. In addition, the power allocation needs to be optimized over multiple subchannels. Hence, the computational complexity for obtaining the optimal solution would be prohibitively high. A recent study that considers resource allocation with BS coordination for OFDMA systems can be found in [6], wherein an iterative resource allocation algorithm is proposed by assuming that each user is served by only one BS, but the linear beam-vectors across a set of coordinated BSs are jointly designed. Note that this solution assumes coordinated beamforming, meaning that multi-cell joint transmission is not allowed. Other related works are [7] and [8], which consider multi-cell linear joint transmission under per-BS or per-subchannel power constraints. In [7] and [8], different power allocation schemes are presented by restricting an identical user set to be scheduled for all subchannels. Hence, the scheduling gain in the frequency domain cannot be fully utilized.

In this paper, the downlink resource allocation problem is addressed for a cluster of coordinated BSs in an OFDMA
system. Joint transmission is provided for the cell-edge users via zero-forcing precoding. With the objective of maximizing the weighted sum rate under per-BS power constraints, jointly optimizing the scheduled users and the transmit power across multiple subchannels and multiple BSs is studied. Motivated by the general duality theory proposed in [9], an iterative algorithm is proposed based on dual decomposition, which decomposes the optimization problem into several independent per-subchannel optimization subproblems. In order to further reduce the complexity of finding the optimal set of scheduled users for each subchannel, a low complexity algorithm is then proposed based on greedy user selection. Simulation results show that the two proposed algorithms achieve a performance very close to the optimal, with reduced computational complexity. Finally, compared with the algorithms proposed in [7] and [8], we show that joint optimization of the user scheduling across multiple subchannels provides a large weighted sum rate gain.

2. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the downlink of a cluster with $N$ coordinated BSs and $M$ users in a multi-cell OFDMA system. The BSs and the users are assumed to have one transmit antenna and one receive antenna, respectively. The system bandwidth is $B$, which is divided into $K$ subchannels. The frequency reuse factor is one, i.e., each BS uses all the $K$ subchannels during each time slot. The BSs in the cluster share both data and perfect CSI of all users, then they can act as a single antenna array and provide joint transmission to a subset of the $M$ users $S(k) \subseteq \{1, ..., M\}$ using the same subchannel $k$. Assuming that the interference from neighboring cells outside the cluster is effectively removed, the discrete-time received signal at the scheduled user $m \in S(k)$ on subchannel $k$ is

$$y_m^k = h_m^k x_k^k + n_m^k,$$

where $h_m^k = [h_{m1}^k, ..., h_{mK}^k]$ denotes the channel vector of the user $m$ on subchannel $k$. $x_k^k = [x_k^1, ..., x_k^K]^T$ is the signal vector transmitted from all the $N$ BSs on subchannel $k$, and $n_m^k$ is the additive white Gaussian noise (AWGN) at user $m$ on subchannel $k$ with covariance $\sigma_m^2$. By using linear precoding, the transmit signal vector $x_k^k$ can be expressed as

$$x_k^k = W_k^* b_k^k,$$

where $b_k^k \in \mathbb{C}^{[S(k)]}$ is the data symbols of the users in set $S(k)$ on subchannel $k$. $W_k^\epsilon \in \mathbb{C}^{N \times [S(k)]}$ is the beamforming matrix used to map the data symbols to the transmit signals. $|S(k)|$ denotes the cardinality of the set $S(k)$. Assume $|S(k)| \leq N$. Let $w_m^k = [w_{1m}^k, ..., w_{Nm}^k]^T$ denote the beamforming vector for user $m$ from the $N$ BSs on subchannel $k$, which is a column vector of $W_k^\epsilon$. Using zero-forcing beamforming among the coordinated $N$ BSs, the precoding matrix is designed such that

$$h_m^k w_j^k = \begin{cases} 0, & j \neq m \\ 1, & j = m \end{cases}, \quad m, j \in S(k).$$

Then, (1) becomes $y_m^k = h_m^k b_m^k + n_m^k$. Thus, the data rate of the user $m$ on subchannel $k$ is given by

$$r_m^k = \frac{B}{K} \log_2 (1 + p_m^k / \sigma_m^2),$$

where $B/K$ is the bandwidth of each subchannel. $p_m^k = \mathbb{E}[|b_m^k|^2]$ is the symbol power allocated to the user $m$ from all the $N$ BSs on subchannel $k$. According to (2), the transmit power of BS $n$ on subchannel $k$ can be derived as

$$p_n^k = \sum_{m \in S(k)} |w_{nm}^k|^2 p_m^k.$$ 

Assume each BS $n$ has a maximum transmit power constraint as $P_{max}$. Hence, $\sum_{k=1}^K p_n^k \leq P_{max}$ for all $n$. For any given time slot, the coordinated $N$ BSs need to jointly determine the set of selected users for each subchannel, and the symbol power allocated to each selected user, i.e., $S(k)$ and $p_m^k$ for all $k$ and $m$, so as to maximize the weighted sum rate under per-BS power constraints. The resource allocation optimization problem can be formulated as

$$\max \sum_{k=1}^K \sum_{m \in S(k)} \alpha_m r_m^k$$

s.t. 1) $\sum_{k=1}^K \sum_{m \in S(k)} |w_{nm}^k|^2 p_m^k \leq P_{max}, \forall n$,

2) $p_m^k \geq 0, \forall m$ and $\forall k$,

3) $S(k) \subseteq \{1, ..., M\}, \forall k$.

Constraints 3) and 4) guarantee that at most $N$ users are assigned to each subchannel. Note that there is no constraint on the number of subchannels to be allocated to a user set. $\alpha_m$ denotes the weight assigned to user $m$. The weights can be seen as the priorities of different users, which is related to the quality of the requested service. If $\alpha_m = 1$ for all $m$, the objective function in (6) becomes maximizing the system sum rate, which is measured in bps/Hz. If $\alpha_m$ is updated over time as a natural logarithm of the user $m$'s average data rate at the previous time slot, the objective becomes to maintain proportional fairness among users [10]. The user weights design is outside the scope of this paper. Given a set of channel vectors $\{h_m^k\}$ and user weights $\{\alpha_m\}$, for $k = 1, ..., K$ and $m = 1, ..., M$, the objective is to find the solution of (6). Although our analysis is general, the performance will be illustrated for equal user weights, i.e., the case $\alpha_m = 1$ for all $m$ will be considered in section 5, Simulation Results.

3. JOINT RESOURCE ALLOCATION

Assume that the scheduled user set on each subchannel is predetermined and feasible, i.e., for each subchannel $k$, $S(k)$ is given and satisfies the constraints 3) and 4). Then, problem (6) becomes a joint power allocation problem, which is convex since the objective function becomes concave and the remaining constraints 1) and 2) are linear. Therefore, it can be solved via standard convex optimization techniques. However, in general, problem (6) is not a convex optimization problem since it needs to find the optimal scheduled user set for each subchannel, which is a combinatorial problem. Specifically, let $S$ denote the set of
all feasible user sets per subchannel, i.e., $S = \{S(k) | S(k) \subseteq \{1, ..., M\}, |S(k)| \leq N\}$. Then, the number of user sets to consider for each subchannel is $|S| = \sum_{k=1}^{M} |S(k)| \leq O(M^N)$. With $K$ subchannels in total, we need to roughly search over $O(M^{NK})$ user set combinations, and the optimal power allocation of problem (6) needs to be solved for each combination. The complexity of this exhaustive search for the optimal solution is unacceptably high for a cluster with a large number of BSs $N$ and a large number of subchannels $K$. In this section, we first introduce a general duality theory [9] for non-convex optimization problems in multi-subchannel systems. Then, a joint resource allocation algorithm is developed via dual decomposition, which achieves a solution close to the optimal with lower complexity.

3.1. General duality theory

In general, with $N$ BSs, $K$ subchannels and $M$ users, the optimization problem in OFDMA systems can be written as
\[
\max \mathbf{x}_k \in \mathbb{R}^M \ s.t. \sum_{k=1}^{K} f_k(x_k) \leq p
\] (7)
where $x_k \in \mathbb{R}^M$ are vectors of optimization variables, $f_k(.)$ are $\mathbb{R}^M \rightarrow \mathbb{R}$ functions that are not necessarily concave, and $p \in \mathbb{R}$ is a vector of power constraints. The Lagrangian of (7) is $L(x_k, \lambda) = \sum_{k=1}^{K} f_k(x_k) + \lambda^T (p - \sum_{k=1}^{K} g_k(x_k))$, where $\lambda$ is the vector of dual variables. The dual objective of (7) is defined as an unconstrained maximization of the Lagrangian, denoted by $g(\lambda) = \max_{x_k} L(x_k, \lambda)$. Then, the dual optimization problem is
\[
\min g(\lambda) \ s.t. \lambda \geq 0.
\] (8)
The solution of the dual problem (8) provides an upper bound of the solution of the primal problem (7). The duality gap is defined as the difference between the upper bound and the optimal value of the primal problem. If (7) is a convex optimization problem, zero duality gap is guaranteed. When the convexity does not hold, the upper bound is not always tight. However, even if the primal problem is non-convex, the duality gap is still zero if the optimal $\sum_{k=1}^{K} g_k$ is a concave function in $p$, and this condition is always satisfied when the number of subchannels goes to infinity [9].

3.2. Joint resource allocation algorithm

Based on general duality theory, dual decomposition is used to solve (6) and to derive the joint resource allocation (JRA) algorithm. Let $p^k = [p^k_1, ..., p^k_M]^T$ denote the symbol power allocation vector on subchannel $k$. Recall that $\mathbb{S}$ denotes the set of all feasible user sets per subchannel satisfying constraints 3) and 4) of (6). Consider the feasible domain, where $S(k) \in \mathbb{S}$ for all $k$, the Lagrangian of (6) is
\[
L(p^k, \lambda) = \sum_{k=1}^{K} \alpha_{m} \sigma_{m} \ln 2 \sum_{n=1}^{N} \gamma_{m} \lambda_{n} |w_{nm}|^2 p_{m}^k - \sum_{n=1}^{N} \lambda_{n} |w_{nm}|^2 p_{m}^k.
\] (9)
Then, the dual objective function is given by
\[
g(\lambda) = \max_{S(k) \in \mathbb{S}} L(p^k, \lambda) = \sum_{k=1}^{K} g_k(\lambda) + \sum_{n=1}^{N} \lambda_{n} p_{max}.
\] (10)
where $g_k(\lambda) = \max_{S(k) \in \mathbb{S}} (\sum_{m \in S(k)} \alpha_{m} \sigma_{m} \ln 2 \sum_{n=1}^{N} \gamma_{m} \lambda_{n} |w_{nm}|^2 p_{m}^k - \sum_{n=1}^{N} \lambda_{n} |w_{nm}|^2 p_{m}^k)$.

(11)

Hence, the dual objective function is decomposed into $K$ independent per-subchannel optimization problems. Assume $\lambda$ is fixed and the user set $S(k)$ is selected on subchannel $k$.

Then, $g_k(\lambda)$ is a concave function of $p_m^k$, and the value of $p_m^k$ that maximizes $g_k(\lambda)$ can be derived by
\[
p_m^k = \frac{u_m - \sigma_m^2}{\alpha_m B / (K \ln 2 \cdot \sum_{n=1}^{N} \lambda_n |w_{nm}|^2)} \text{ and } [a]^+ = \max(a, 0).
\] (12)
Substituting (12) into (11), $g_k(\lambda)$ can be obtained by selecting the optimal user set $S(k)^*$ from $\mathbb{S}$. After solving (11) for all $k$, the dual objective function $g(\lambda)$ is derived by (10).

Hence, the complexity for finding the optimal user sets for all subchannels is reduced to $O(M^{NK} \cdot K)$, which is linear in $K$.

Based on the general duality theory, as the number of subchannels goes to infinity, the solution of the optimization problem (6) is the same as the minimal value of $g(\lambda)$. In this work, the optimal $\lambda^*$ that minimizes $g(\lambda)$ is found via the subgradient-based dual update method. Let $\lambda^{(l)}$ denote the updated dual vector on iteration $l$. A subgradient of $g(\lambda^{(l)})$, denoted by $d$, can be found as
\[
d_n = p_{max} - \sum_{k=1}^{K} \sum_{m \in S(k)} |w_{nm}|^2 p_m^k, \quad \forall n.
\] (13)
where $S(k)^*$ and $p_m^k$ are the optimizing variables obtained in the maximization problem (11) with respect to $\lambda^{(l)}$. Hence, an efficient dual-update method can be performed by simultaneously updating all dual variables along the subgradient direction [9],
\[
\lambda^{(l+1)} = [\lambda^{(l)} - t^{(l)} d]^+.
\] (14)
where $l$ is the iteration number, $t^{(l)}$ is the step size on iteration $l$, which can be chosen as $t^{(l)} = \beta / \sqrt{l}$ for some constant $\beta$ and is guaranteed to converge to the optimal dual solution. Finding the optimal dual solution has a polynomial complexity in the dimension of $\lambda$, which is the number of coordinated BSs.

Zero duality gap can be guaranteed only when the number of subchannels in the system goes to infinity. For a practical system with finite number of subchannels, the derived optimal dual solution may not be globally optimal. However, the simulation results show that the achieved dual solution is very close to the global solution of the primal problem, which requires an exhaustive search over all possible combinations of users with complexity of $O(M^{NK})$.

The JRA algorithm is summarized in Algorithm 1. Notice that the scheduled user set $S(k)^*$ can change when updating $\lambda$, leading to discontinuity of $p_m^k$. In this case, the subgradient-based search may converge to a solution, where the transmit power constraint is not satisfied for each BS. However, with the achieved user set selection $S(k)^*$ for each subchannel $k$, the problem (6) becomes convex. Therefore, in Step 2, similar to [2], the power allocation to these selected user sets can be redesigned based on waterfilling distribution so that per-BS power constraints are satisfied.
4. LOW COMPLEXITY ALGORITHM

As shown in Section 3, for each iteration $l$, solving per-subchannel optimization subproblem still requires an exhaustive search to find the optimal scheduled user set $S(k)^*$ from $\mathcal{S}$, i.e., line 4 in Step 1 of Algorithm 1. This exhaustive search for each subchannel has a complexity of $O(M^N)$. For a cluster with a large number of coordinated BSs and users, the proposed Algorithm 1, is still prohibitively complex.

In this section, similar to [5] a low-complexity greedy algorithm is proposed to select $S(k)^*$ and compute $g_k(\lambda^{(l)})$, $p_m^{k*}$ for each per-subchannel optimization subproblem. The low-complexity joint resource allocation (LC-JRA) algorithm is successively performed by selecting the best user with the largest $g_k(\lambda^{(l)})$ in (11), and then iteratively adds a new user from the remaining users until adding one more user reduces the value of $g_k(\lambda^{(l)})$ or the number of selected users equals to the number of BSs. Compared to [5], instead of using weighted sum rate, the user selection is based on $g_k(\lambda^{(l)})$, which also takes the power consumption of the selected users into account. Algorithm 2 lists the per-subchannel scheduling and power allocation procedure (line 4 and 5 of Step 1 shown in Algorithm 1) for the proposed LC-JRA algorithm. Note that $S(k)^* \cup \{m\}$ denotes the union of the user set $S(k)^*$ and the user $m$.

Algorithm 2 requires at most $N$ iterations, and the optimization problem in line 4 of Algorithm 2 is solved with complexity $O(M)$ in each iteration. Hence, the complexity of the per-subchannel optimization subproblem is further reduced to $O(MN)$. The remaining of the LC-JRA algorithm is the same as Algorithm 1.

5. SIMULATION RESULTS

As depicted in Figure 1, we consider the downlink of an OFDMA cellular system, where a cluster of two coordinated BSs simultaneously transmit on multiple subchannels. The system bandwidth is $B = 1$ MHz. The cell radius $R$ is 1 km. The path loss model is $PL(d) = 128.1 + 37.6 \log_{10}(d)$ in dB, with $d$ in km. Shadowing is log-norm distributed with zero mean and standard deviation 8. The fast fading for the subchannels is independently Rayleigh distributed. The AWGN power is -135dBW. Cell-edge SNR, denoted as $\gamma$, is defined to be the received SNR at the boundary of the cell, assuming full power transmission $P_{\text{max}}$ from the BS, accounting for only pathloss gain $PL(R)$ and ignoring shadowing and Raleigh fading. The users are randomly dropped in the cell-edge area with equal weights, i.e., $\alpha_m = 1$ for all $m$. The sum rate is averaged over 1000 independent locations of the users. The proposed two algorithms are named as joint resource allocation (JRA) and low complexity joint resource allocation (LC-JRA), respectively. Step 1 of these two algorithms are stopped when $|g(\lambda^{(l)}) - g(\lambda^{(l-1)})| \leq 10^{-5}$ or $l_{\max} = 100$.

In Figure 2, the convergence of the Step 1 of the JRA and LC-JRA algorithms is analyzed. Two users are randomly dropped into the overlapped cell-edge area. The dual objective values $g(\lambda^{(l)})$ of the JRA and LC-JRA algorithms are plotted versus the number of iterations $l$ for different cell-edge SNR. The Lagrangian dual vector is initialized as $\lambda^0 = [0.5, 0.5]^T$. The step size sequences chosen for $\gamma = 2, 10, 18$ dB are $\beta = 0.7, 0.1, 0.01$, respectively. For the sake of comparison, we also plot the optimal sum rate value provided by exhaustive search over all possible user set combinations, and by solving optimal power allocation for each combination (referred as Optimal-RA). As can be seen from Figure 2, the dual objective values of JRA and LC-JRA algorithms monotonically decrease at each iteration and converge to a value very close to the optimal solution. The duality gap decreases as the cell-edge SNR increases for both the two proposed algorithms. Notice that LC-JRA shows the same convergence behavior as JRA, however, with much lower complexity for user selection in each iteration.
Figure 3 shows the average sum rate of the JRA, LC-JRA and Optimal-RA algorithms. Considering the simulation feasibility for the Optimal-RA algorithm, only the cases with small number of users and small number of subchannels are simulated. It is shown that the proposed algorithms can achieve a performance very close to the optimal value.

6. CONCLUSION

In this paper, two iterative resource allocation algorithms are proposed for a cluster of coordinated BSs in an OFDMA system with multi-cell joint transmission. Jointly optimizing user scheduling and power allocation across multiple subchannels and multiple cells is studied. Simulation results show that the proposed two algorithms provide a solution very close to the optimal, however, with much lower complexity. Moreover, compared with the resource allocation algorithms restricting an identical user set to be allocated to all subchannels, the sum rate gain provided by jointly optimizing the user selection across multiple subchannels are roughly 45% and 20% when cell-edge SNR are 2 and 18 dB, respectively.

7. REFERENCES